UNIVERSITY OF CALIFORNIA, SAN DIEGO

Molecular Mechanisms of Urea Uptake in Marine Diatoms

A Thesis submitted in partial satisfaction of requirements
for the degree of Master of Science

in

Biology

by

Katherine Elizabeth Oakley

Committee in Charge:

Professor Milton H. Saier, Chair
Professor Andrew E. Allen
Professor James Golden
Professor William Loomis

2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

This Thesis of Katherine Elizabeth Oakley is approved and acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2010

DEDICATION

This thesis is dedicated to my mother, Nancy Oakley, and to my father, John L. Oakley. I am very grateful for their exceptional support and encouragement. I would like to share all of my love with my parents.

TABLE OF CONTENTS

LISTS OF FIGURES AND TABLES

ACKNOWLEDGMENTS

I would like to acknowledge the contributions of Dr. Milton H. Saier, Dr. Andrew E. Allen, Dr. Christopher Dupont and members of the Environmental Genomics Department at the J. Craig Venter Institute. I would like to thank Dr. Saier, Dr. Allen and Dr. Dupont for their dedication and support throughout this project. This work would not be possible without their efforts, and I would like to express my deep appreciation for the opportunity to work with such remarkable scholars and mentors.

I would also like to acknowledge the members of the Environmental Genomics department at the J. Craig Venter Institute, particularly Lisa Zeigler, James McCarthy, Hong Zheng and Ahmed Moustafa. Their assistance was extremely helpful throughout the duration of this project.

ABSTRACT OF THE THESIS

Molecular Mechanisms of Urea Uptake in Marine Diatoms

by

Katherine Elizabeth Oakley

Master of Science in Biology

University of California, San Diego, 2010

Professor Milton H. Saier, Chair

The genome sequence and established expressed sequence tags for the pennate marine diatom *Phaeodactylum tricornutum* reveal a full complement of genes encoding proteins for a complete metazoan-type urea cycle, a urease enzyme and two urea transporters that could function to take up urea as an organic nutrient. Amidst this set, two genes involved in the uptake of urea as a nutrient source have been identified: JGI_Pt_20424, homologous to high-affinity plant urea transport proteins, and

JGI_Pt_39772, with homology to bacterial and metazoan urea transporter proteins, including the well characterized vertebrate renal and erythrocyte transporters. Both genes were cloned into expression vectors containing a YFP fluorophore tag by Gateway technology (Invitrogen), and the subcellular localizations of the two urea transporter proteins were determined following overexpression studies and *in vivo* epifluorescence and confocal microscopy in transfected diatoms. By observing yellow fluorescence of the transporter-C-terminal YFP fusion protein *in vivo*, the urea transporter JGI_Pt_20424 localized to *P. tricornutum*'s plasma membrane. The JGI_Pt_39772 urea transporter was localized to the mitochondrion using the same method.

Urea uptake rates of ammonium grown cultures of both wild-type and transgenic *P. tricornutum*, containing the JGI_Pt_20424 urea transporter overexpression construct, were determined using ^{15}N urea. The overexpression of JGI_Pt_20424 increased urea uptake approximately ten-fold compared to the wild-type; therefore, the protein efficiently transports urea. Work has begun on the molecular mechanisms and kinetics of urea transport by JGI_20424 and JGI_39772 using a heterologous expression system in *Xenopus laevis* oocytes.

Introduction

1.1 Background

Marine phytoplankton are responsible for roughly one half of all photosynthesis on Earth (Falkowski and Raven, 1997). Diatoms, which are unicellular, eukaryotic algae including an estimated 100,000 species, represent a major phytoplankton group, and they contribute 20% of all marine primary production of organic compounds and generate 40% of the forty-five to fifty billion metric tons of organic carbon fixed annually in the oceans (Armbrust et al, 2004). Marine diatoms are responsible for the assimilation of inorganic carbon into organic molecules that are the basis of the entire marine food web, and they provide a major mechanism for the removal of atmospheric carbon dioxide. Of great concern to the human population are the threatening effects of global warming; therefore, understanding the mechanisms controlling primary production in the ocean is of critical importance.

Nitrogen assimilation in marine phytoplankton is a tightly regulated process that fuels autotrophic and heterotrophic carbon metabolism. Marine phytoplankton provide aquatic ecosystems with remarkable ecological flexibility, and they make a significant contribution to the biogeochemical cycling of carbon, nitrogen, phosphorus, silica, iron and various trace metals. The physiological and molecular mechanisms that govern nitrogen sensing and utilization in phytoplankton communities are pivotal to an understanding of their ecological success.

Nitrogen utilization affects primary productivity as well as the

ecological success of phytoplankton groups. Phytoplankton play an important role in the well-characterized marine nitrogen cycle. The nitrogen sources utilized by marine phytoplankton include urea $(NH_2)_2CO$, nitrate (NO_3^-), nitrite (NO_2^-) and ammonium (NH_4^+). These usable forms of nitrogen are available after fixation of atmospheric nitrogen by marine microbes. Utilization of nitrogenous molecules leads to the synthesis of amino acids, nucleotides and other molecules, and ultimately the regeneration of nitrogen substrates, in the form of ammonium. Ammonium is oxidized to nitrite and then nitrate through nitrification, a chemoautotrophic process catalyzed by restricted lineages of archea and bacteria. Phytoplankton utilize regenerated nitrate in the photic zone which is primarily supplied through episodic upwelling events in the ocean. Sinking and degradation of organic debris to aphotic waters leads to accumulation of inorganic nitrogen. This creates a nutrient separation in the ocean; however, upwelling and vertical mixing in the oceans provide a flux of nitrate to the photic zone which stimulates primary production. Fluctuating nutrient conditions and light availablity between the aphotic and photic regions create competition among phytoplankton for nitrogen resources that result from episodic, physically forced vertical mixing events in the ocean. Nitrate uptake and utilization stimulate phytoplankton growth and primary production in marine ecosystems. Episodic upwelling events in the oceans deliver a high concentration of nitrate to the photic zones, which induces competition among phytoplankton for nitrate and a large increase in phytoplankton biomass. Marine diatoms are among the most successful competitors for nitrate, and this activity has been characterized as a “shift-up” in diatom metabolism. Various experiments have examined the changing stoichiometic

relationships of nutrients to which phytoplanktonic groups are exposed. Nutrient stoichiometry has been manipulated both *in situ* and in pure cultures of marine diatoms to demonstrate that the enzymatic activities of nitrate reductase and urease increase in response to higher concentrations of nitrogen substrates in photic zones after oceanic mixing events. The "shift-up" in diatom metabolism underscores the nitrogen assimilation capabilities of marine diatoms (Zimmerman 1987, Berges et al. 2004).

Recent whole genome analysis has revealed that diatoms contain genes and metabolic pathways that are novel in photosynthetic eukaryotes. Enzyme localization, physiological studies, gene transcript profiling and bioinformatics analyses have revealed differences in nitrogen assimilation and turnover pathways in marine diatoms as compared to vascular plants, green algae and cyanobacteria. Diatoms have a unique evolutionary history that has fostered their distinctive combination of biochemical pathways. Bloom-forming diatoms are members of the chlorophyll c (Chl c) lineage, whose common ancestor probably developed from a host heterotrophic eukaryote enveloping a red algae containing a previously enveloped an aerobic bacterium (Yoon et al., 2004b, Bachvaroff et al. 2005). The initial endosymbiotic event and subsequent diversification produced photosynthetic eukaryotes with diverse biochemical pathways. Genomic sequencing and biochemical and gene expression studies have revealed in diatoms a variety of biochemical components, previously observed in bacteria and metazoans, which probably contribute to their ecological success. Combinations of genes collectively encode a complete metazoan-like urea cycle that was most likely acquired from the original heterotrophic host (Armbrust et al. 2004). The presence of a urease

enzyme in the genome, however, suggests the alternative function of nitrogen storage for urea in diatoms (Peers et al. 2000, Armbrust et al. 2004). Diatom genomes also encode a GS-GOGAT cycle, typically found only in plastids of green lineage eukaryotes (Allen et al., 2006, Bowler et al. 2010). However, in contrast, the diatom GS-GOGAT components are also located in the mitochondrion (Siaut et al. 2007a). Diatoms also contain other proteins, such as glutamine synthase, hydroxylamine reductase-like hybrid cluster protein and an agmatinase family protein, targeted to the mitochondria. The urea cycle, urease enzyme and glutamine and glutamate synthase (GS-GOGAT) pathways that are localized to the mitochondria in diatoms are novel molecular mechanisms for nitrogen assimilation in phytoplankton that provide a cellular basis for the competitive advantage of diatoms during seasonal upwelling events. Therefore, the physiological and molecular understanding of intracellular nitrogen metabolism in marine diatoms is critical for determining the mechanisms responsible for their competitive advantage for nitrogen substrates among marine phytoplankton groups.

A molecular understanding of nitrogen assimilation and turnover pathways in the marine pennate diatom *Phaeodactylum tricornutum* has not yet been obtained. The Joint Genome Institute (http://genome.jgi-psf.org/Phatr2/Phatr2.home.html) has sequenced the complete genomes of both the centric diatom *Thalassiosira pseudonana* (Ambrust et al. 2004) and the pennate diatom *P. tricornutum* (Bowler et al, 2008). The complete *P. tricornutum* genome comprises 27.4 Mega base pairs (Mb), consisting of about 11,000 predicted genes, while the *T. pseudonana* genome is 32Mb containing between 10,000 and 14,000 genes. Both genomes contain unique combinations of genes that encode

pathways of nutrient assimilation and metabolite management. *P. tricornutum* is an ideal organism for both sequence analysis and experimental procedures, because of the ease with which the organisms can be cultured as well as its remarkable ability to genetically transform. With the complete genome and a wealth of expressed sequence tags (ESTs), now over 100,000 in number, *P. tricornutum* is an excellent representative diatom for comparative genomics studies. The diatom genomes bear traces of both red and green algal origin and serial secondary endosymbiosis, in which a eukaryotic organism hypothetically engulfed an ancient green or red algal endosymbiont to generate the diatom plastids (Sanchez-Puerta and Delwiche 2008, Moustafa et al. 2009, Worden et al, 2009). The *P. tricornutum* genome contains 784 genes of clear bacterial origin, comprising about 7.5% of the total gene content. More than half of the bacterial genes are shared with *T. pseudonana*, signifying that the gene transfer probably occurred prior to diatom diversification (Bowler et al. 2008).

A full complement of urea cycle genes has been identified in the genomes of *T. pseudonana* and *P. tricornutum*, and genomic analysis suggests the set of genes and pathways that connect to the urea cycle are most likely of bacterial origin, resulting from horizontal gene transfer, viral infection, or phagotrophy (Parker et al 2008, Allen et al 2006). Previous studies of EST databases had identified an unusual combination of sequences encoding a complete urea cycle and a urease enzyme (Ambrust et al. 2004). Despite the gene sequence information, the mechanisms by which diatoms accumulate and store urea have yet to be determined. In metazoans, the urea cycle has been fully described. Carbamoyl phosphate synthetase III (CPSIII) is of Opisthokont origin and is

absent in plants and green algae. CPSIII initiates the urea cycle combining carbon and nitrogen substrates (ammonium) in the mitochondrion, and it was found in both diatoms to localize to the mitochondrion (Bowler et al, 2004). The CPSIII in the mitochondrion of diatoms is unique in that it utilizes ammonium as its substrate instead of glutamine, incorporating ammonium with a bicarbonate ion to produce carbamoyl phosphate. The diatom's CPSIII utilization of ammonium as its substrate may indicate that the urea cycle could act as a pathway of ammonium assimilation (Allen et al. unpublished).

Urease is localized to the cytoplasm in plants (Faye et al. 1986); however, *in-silico* analysis in marine diatoms has revealed that the urease enzyme (JGI_Pt_29702) contains a putative mitochondrial targeting presequence, with a *TargetP* score of 0.96. *In-vivo* analysis of the constructed urease-YFP fusion protein, followed by immunogold labeling, has determined that urease is indeed localized within the mitochondria (Dupont et al, 2010). Urease's localization in the mitochondrion implies that exogenous urea must be transported not only across the plasma membrane of the diatom but also through the mitochondrial membranes through active urea transporters. In bacteria, plants, fungi and most protists urea is broken down by the Ni-containing urease enzyme into carbon dioxide and ammonia in the mitochondrion. As urease is a strictly mitochondrial protein, a mitochondrial urea transporter must be present to provide its substrate. It had been previously assumed that urea transporters operated solely to transport urea into the cytosol. Urease, an enzyme present in all algae, was also found in genomic sequencing results from red algal lineage phytoplankton species, such as *T. pseudonana, P.*

tricornutum and *A. anophagefferens*, as well as green algal lineage species, such as *Ostreococcus*.

The genomic sequences of both *T. pseudonana* and *P. tricornutum* encode multiple transporters for nitrate, nitrite, ammonium and urea. Diatom genomes encode copies of NO_2^- transporters and a ferredoxin NO_2^- reductase located in the plastid, both of cyanobacterial origin. They also encode genes of endosymbiotic origin, specifically nitrate reductase (Allen et al. 2005), which is presumably located in the cytosol where it functions to deliver NO_2^- to the plastid. Ammonium is assimilated into amino acids in the diatom's chloroplast via the GS-GOGAT cycle. The diatom mitochondrion is a vital organelle that has several functions in nitrogen metabolism including maintenance of the overall cellular balance of nitrogen within the microorganism, which is unique to photosynthetic eukaryotes (Allen et al. 2006, Bowler et al. 2010). The diatom's CPSIII and urease enzymes are localized to the mitochondria as well as the glutamine and glutamate synthase (GS-GOGAT) pathways that contribute to nitrogen balance. Diatoms also assimilate ammonium into glutamate, glutamine and other amino acids in the GS-GOGAT cycle components in the mitochondrion. A glutamine synthase III (Siaut et al. 2007a), ferredoxin and NADPH dependent glutamate synthase, and putative transaminases are most likely localized to the mitochondria based on targeting presequences and *in silico* analyses (Allen et al., unpublished). The mitochondria of marine diatoms therefore potentially manage overall cellular nitrogen nutrient status.

In the marine environment, urea is the most abundant form of organic nitrogen, supporting approximately 50% of the total nitrogen utilization by phytoplankton

communities. Urea exists as an available nitrogen source in concentrations ranging from 0.05μM-N to 50μM-N (McCarthy, 1972). The concentration of urea in the marine environment is less than that of nitrate an ammonium. As a regenerated nitrogen substrate, urea is essential to the growth and development of marine phytoplankton. Anthropogenic activities and natural processes, such as regeneration by heterotrophic bacteria, excretion by marine animals and zooplankton as a waste product of purine catabolism and a by-product of the urea cycle, release by phytoplankton and atmospheric deposition are responsible for the release of urea into the marine environment (Wafar et al. 1995, Berman and Bronk 2003). To encourage phytoplankton growth and thus remove carbon dioxide from the atmosphere, proposals have been made to fertilize the ocean with nitrogen-rich urea fertilizers (Gilbert et al. 2008); however, the proposals are controversial due to the potential lethal side effects on the marine ecosystem.

The urea cycle and associated enzymes in marine diatoms most likely function in carbon and nitrogen recycling, storage and redistribution catabolically derived ammonia and carbon dioxide especially in conditions of nutrient starvation or perhaps more importantly, during recovery from nutrient starvation. It is believed that the urea cycle in marine diatoms serves to detoxify, assimilate and redistribute ammonia in the mitochondria. In periods of nutrient stress, the stored nitrogen and carbon organic compounds are therefore accessible for nutrient metabolism. The urea cycle produces urea endogenously, which serves as a non-toxic organic nitrogen storage base (Allen et al. 2006, Parker et al. 2008). On the other hand, the GS-GOGAT cycle serves to assimilate urea-derived nitrogen into amino acids. The prevailing hypothesis for the role

of the urea cycle in phytoplankton is that the produced urea regulates other mechanisms of cellular metabolism, regulates amino acid catabolism and serves as an osmolyte for the cell. For example, ornithine is used to make spermidine, which helps to precipitate silica for cell wall formation (Armbrust et al.2004, Allen et al 2006). Also, arginine could be used for protein synthesis or to generate polyamine substrates for frustule biogenesis (Kroger et al. 2000).

Exogenous urea must be transported across the outer plasma membrane of diatom species and into the mitochondrion through active urea transporters in order to activate urease and the complete set of urea cycle proteins targeted to the mitochondria. The enzymes of the urea transporter family, TC# 1.A.28 (TCDB, http://www.tcdb.org/tcdb/subfamily.php?tc=1.A.28), are typically found in vertebrate animals, plants, fungi and bacteria, but not in other eukaryotes, or archaea (Minocha et al, 2003). Many mammalian urea transporters are homologous to those found in bacteria. Typically, urea transporters are about 380 to 400 amino acid residues, and contain 10 putative transmembrane domains. The urea transporters operate a channel mechanism, because urea is a highly polar molecule that does not freely diffuse across cell membranes (Levin et al. 2009).

Analysis of the predicted coding sequences of *P. tricornutum's* genome revealed proteins with homology to transporter proteins. Two genes encode the urea transporters- the first, JGI_Pt_20424 (TC# 1.A.28.1.3 (TCDB)), is orthologous to the yeast and plant transporters, and its expressed sequence tags are highly expressed in condition of nitrogenous nutrient stress, while the second, JGI_Pt_39772 is orthologous to both

bacterial and vertebrate renal and erythrocyte urea transporters. JGI_Pt_20424 is the second most highly expressed EST under nitrate stress conditions indicating the importance of urea metabolism, regeneration and storage in the diatom.

1.2 Bioinformatics Analysis of Urea Transporters

The complete genome of the pennate diatom *Phaeodactylum tricornutum* has been sequenced and compared to other diatom sequences to construct an evolutionary tree based upon its functional domains and gene homologies. Diatoms genomes contain a unique phylogenomic profile as well as a remarkable combination of genes that encode the components of nutrient assimilation and metabolic pathways. There are hundreds of genes of bacterial origin present in and shared between *P. tricornutum* and *T. pseudonana*, which indicates that gene transfer occurred prior to diatom diversification. The genes encoding specific metabolic processes, such as the urea cycle, are orthologous with multicellular organisms. Genomic analysis supports the proposed evolutionary hypothesis that marine diatoms and dinoflagellates, belonging to supergroup Chromalveolata, are closely related to the green algae pico-eukaryotes prasinophytes, probably as a result of gene transfer from an ancient prasinophyte endosymbiont as it was suggested recently in Moustafa et al. 2009. Each urea transporter appears to be transcriptionally upregulated under nitrogen limitation conditions, which has not been previously observed in diatoms and marine phytoplankton. *P. tricornutum's* expressed sequence tags encode JGI_Pt_20424 and JGI_Pt_39772 with high homology to other urea transporters.

1.2.a. JGI_Pt_20424

JGI_Pt_20424 is found within chromosome 9 at position 102,984-106,378 base pairs, and its transcript is 2,571 base pairs and coding sequence of 2,115 base pairs. It contains fifteen transmembrane domains. Gene annotation results suggest that the molecular function is transport activity, the biological process is transport and the cellular component in which it is found is the membrane. Phylogenetically, this membrane-targeted urea transporter is homologous to *DUR3* urea transporters, a plant-like high-affinity transporter, and it belongs to the sodium solute symporter family. JGI_Pt_20424 is the second most highly expressed gene in the nitrate starved EST library. The WoLF PSort (http://wolfpsort.org/) program determined a significant prediction score for the target and subcellular localization of JGI_20424 of 10.0 to the plasma membrane.

1.2.b. JGI_Pt_39772

JGI_Pt_39772 is found within chromosome 20 at position 670,083-671,381 base pairs, and it contains five transmembrane domains. Its 1,299 base pair genome sequence reveals that it is similar to members of solute carrier family 14, a large family of integral membrane transport protein. Its suggested molecular function is urea transport activity. The *TargetP* program produced a significant prediction score for the subcellular localization of JGI_39772 of 0.815 to the mitochondria.

1.2.c. Phylogenetic Tree Analysis

The phylogenomic pipeline "iTree" (Moustafa et al., 2008) was used to determine the phylogenomic profiles of both urea transporters JGI_20424 and JGI_39772. This pipeline was designed to assimilate all of the hits from BLAST, as opposed to only a certain number of the top hits; however, the pipeline filters all BLAST hits to include the

top hits from each taxonomic group by setting a maximum number of hits at each taxonomic rank to be included in the alignment for subsequent analyses. This approach provides a uniform distribution among taxonomic groups, reducing the effect of the biases in the taxonomic groups chosen for sequencing projects.

To maximize taxon coverage, iTree runs against a genome database that is assembled from NCBI RefSeq, Joint Genome Institute, and EST libraries of taxa with no complete genome data. The generated BLAST results are aligned using MUSCLE, and the produced alignment is refined and then the FASTA alignment is converted into the PHYLIP format. Following this, a maximum likelihood phylogeny is inferred using PhyML.

The maximum likelihood bootstrap phylogenetic tree constructed for the plant-like high affinity urea transporter (JGI_20424, "Chromalveolata Heterokontophyta *Phaeodactylum_tricornutum* CCAP_10551_gi219119") indicated that JGI_20424 belongs to the "supergroup" Chromalveolata, phylum Heterkontophyta (Figure 1). It specifically clusters most closely with pennate polar diatom *Fragilariopsis cylindrus,* centric diatom *Thalassiosira pseudonana*, and the pelagophyte *Aureococcus anophagefferens* and the dinoflagellate *Heterocapsa triquetra,* respectively. The closest relative sister group to Chromalveolata is the Plantae kingdom, specifically the green algae prasinophytes *Ostreococcus lucimarinus* and *Micromonas*, and *Populus trichocarpa* and *Vitis vinifera* of phylum Streptophyta. The next closest relative belongs to Chromalveolata, phylum Heterkontophyta, specifically *Ectocarpus siliculosus*, which is followed by the kingdom Plantae phylum Streptophyta sister group, consisting of

Arabidopsis thaliana and *Oryza sativa Japonica* organisms. Within this internal node are organisms belonging to the supergroup Opisthokonta (animals and fungi) phyla Fungi such as *Neurospora crassa* and *Penicillium marneffei*. The eukaryotic supergroups consist of the following: Archaeplastida or Plantae (consisting of red algae, green algae and plants), Excavata (anaerobes), Chromalveolates (single celled eukaryotes such as dinoflagellates and ciliates), Rhizaria (amoebae, flagellates), Opisthokonta (Metazoa, Fungi and close protist relatives) and Amoebozoa (protozoan eukaryotes, specifically amoebae (Burki et al. 2007). JGI_20424 has recent ancestors from four of the six eukaryotic supergroups. The ancient ancestral node, is represented by Archaea (phyla Thaumarchaeota *Cenarchaeum symbiosum*), and kingdom Bacteria phyla Proteobacteria (*Idiomarina baltica*) and phyla Bacteroidetes (*Rhodothermus marinus*). The extant ancestor is kingdom Bacteria phyla Cyanobacteria (*Synechococcus*), phyla Proteobacteria (*Nitrosococcus oceani*) and phyla Actinobacteria (*Rhodococcus erythropolis*). This phylogenetic tree reveals that urea transporter JGI_20424 encoded by genome sequences of *P. tricornutum* and other lineages of the supergroup Chromalveolata (e.g., other diatoms and dinoflagellates) is the closely related to the green algae pico-eukaryotes prasinophytes, probably as a result of gene transfer from an ancient prasinophyte endosymbiont as it was suggested recently (Moustafa et al. 2009). The Plantae taxa supergroup is the most recent ancestor of JGI_20424 and other plant-like descendents belonging to Chromalveolata. The Opisthokonta taxa, specifically phyla fungi, preceded Plantae, and thus, it is the next most recent common ancestor of JGI_20424 and its plant-like descendents. The phylogenetic tree also indicates that the urea transporter in

eukaryotes has a bacterial origin, represented in the tree by Proteobacteria and Actinobacteria species. The Bacteria taxa common ancestor has the most distant evolutionary relationship from JGI_20424 in the tree, indicating that the urea transporter originated in bacteria, its extant ancestor. The phylogenetic history of JGI_20424 infers its evolutionary relationship to the ancestral taxonomic kingdoms Plantae and Opisthokonta; therefore, this urea transporter is indeed a plant-like high affinity transporter that is homologous with members of the sodium solute symporter family in these two kingdoms.

The JGI_39772 urea transporter has a unique evolutionary footprint, as it is a descendent of both metazoans and bacteria (Figure 2). In the maximum likelihood bootstrap phylogenetic tree, JGI_39772 (Chromalveolata.Heterokontophyta *Phaeodactylum tricornutum* CCAP_10551_gi219126831) is a member of Chromalveolata taxa, phylum Heterkontophyta. It specifically clusters most closely with bloom-forming haptophyte *Emiliania huxleyi* and the diatoms *Thalassiosira pseudonana,* and *Fragilariopsis cylindrus*. The closest neighbors of JGI_39772 are Metazoa. For example, the sister group contains ancestors such as *Takifugu rubripes, Danio rerio, Mus musculus* and *Homo sapiens*. As noted in the tree, there is considerable homology with bacterial genes, suggesting a common ancient ancestor for diatoms and metazoans. Independent horizontal gene transfer from the bacterial source to both the common ancestors of diatoms and animals could explain this intriguing phylogeny. Alternatively, JGI_39772 was transferred from the bacterial source to the ancestor of eukaryotes but subsequently, it was lost from all the eukaryotic lineages except diatoms and animals,

which managed to retain it in their nuclear genomes. As a descendent of bacteria, the metazoan-like urea transporter confirms that serial secondary endosymbiosis did occur after a eukaryote engulfed a red algal endosymbiont, which had engulfed a photosynthetic aerobic bacterium.

1.2.d. Qualitative Analysis of the EST Library as an Expression Profiling System

The bioinformatics analysis of the EST libraries constructed for the *Phaeodactylum tricornutum* genome provide significant insight into the functional significance of various genes based upon expression profiling analysis. Bowler et al, 2010 added fifteen non-normalized cDNA libraries to the previous cDNA library generated by Maheswari et al. (2005). The newly constructed cDNA libraries were made to provide quantitative data to demonstrate the responses of *P. tricornutum* to different growth conditions. The gene expression patterns for each of the sixteen culture conditions are compared to one another after EST numbers were converted to frequencies in each library. The frequencies of the ESTs in the various conditions were examined for redundancy patterns and diversity (Bowler et al. 2010).

Following the completion of generated EST libraries and gene clustering patterns based on gene expression from the sequenced *Phaeodactylum tricornutum* genome (Maheswari et al. 2005, Bowler et al. 2010), the functional significance of the library clusters was performed using sequence and domain conservation analysis. Histograms were generated for *P. tricornutum's* urea transporters, JGI_20424 (Figure 3) and JGI_39772 (Figure 4), by determining the frequency of each gene in each EST library for a particular culture condition. The data was normalized by dividing the number of ESTs

in a particular condition for the gene of interest by the sum of all ESTs in the libraries for that particular growth condition from all genes in *P. tricornutum*. The value of R for a specific gene represents the amount of differential gene expression across the sixteen growth conditions. The R-value of JGI_20424 is 506.75719, while that of JGI_39772 is 19.75272868.

The frequency of ESTs were determined and normalized for each of the sixteen conditions for JGI_20424 and JGI_39772. The x-axis of the plot represents the normalized EST library and the y-axis represents the frequency of the ESTs for a specific gene. Bar charts were made to illustrate the changes in the frequencies of ESTs across all sixteen conditions for JGI_20424 (Figure 3) and JGI_39772 (Figure 4). The EST histogram demonstrates that JGI_20424 urea transporter is highly upregulated in conditions of nitrate stress. These results are consistent with observations that JGI_20424 is the second most highly expressed gene in the nitrate-starved library. The greatest frequencies of ESTs for JGI_20424 are in conditions of nitrate stress (0.01962533) and high CO_2 for one day (0.01804378). The greatest frequencies of ESTs for JGI_39772 are in the tropical accession (0.0002094241) and ammonium adapted (0.0001122586) conditions, and it has the same EST frequency as JGI_20424 in the silica plus condition (0.0001352082). By comparison, the ESTs for JGI_39772 show much less change with nutrient conditions. Its EST frequency is far less in nitrate stress conditions (0.0001115076); therefore, the expression level of this gene does not change in nitrate-starved conditions.

1.3 Summary of the Experimental Purpose

The subcellular localization and function of the diatom's two urea transporters were the goals of the work described herein. The urea transporters supply exogenous urea, both into the cell and into the mitochondrion. The two urea transporters were examined by overexpression studies in *Phaeodactylum tricornutum* conducted with expression vectors that are generated based on Gateway technology (Invitrogen) for high-throughput protein fusion. The expression vectors were created using diatom destination vectors fused to fluorescent proteins that allow one to determine the subcellular localization of each urea transporter. The coding sequences of the genes of interest (JGI_20424 and JGI_39772) in *P.tricornutum* were cloned into a commercially available entry vector, pENTR, using a topoisomerase. The entry vector recombined with a destination vector (pDEST) that contains a coding sequence for a yellow fluorescent protein tag, in order to fuse the coding sequence with a YFP protein at its C-terminus. The localization of both transporters as overexpression constructs was determined *in vivo* following gene transformation by high-pressured particle delivery system. The transgenic diatoms, expressing native-protein-fluorophore fusions (YFP) were examined by fluorescence microscopy and epifluorescence confocal microscopy.

Urea uptake rates in the generated overexpression constructs were determined by examining the uptake rate of nitrogen-15 isotope ^{15}N-labeled urea. Wild type *P. tricoruntum* and JGI_20424 transgenic *P. tricornutum* samples were grown in seawater media containing ammonia as the nitrogen source. Following growth to mid-exponential phase, both cultures had an inactive native copy of the urea transporter gene; however, the transgenic, overexpressing JGI_20424 line had additional copies of the urea

transporter gene. The two cultures were exposed to ^{15}N-labeled urea through a filtration system to measure the urea uptake. The concentration and isotopic ratio of ^{15}N labeled urea was determined using gas chromatography/mass spectrometry at the UC Davis Metabolomics Facility.

The functionality, molecular mechanisms and kinetics of urea transport by JGI_20424 and JGI_39772 will be determined using the heterologous expression of optimized expression vectors in *Xenopus laevis* oocytes. The cDNAs encoding the two transporters were cloned into an optimized expression vector, pGEM, for transcription of the respective cRNAs in oocytes. The *Xenopus* oocytes will translate, process and target the urea transporter proteins to allow structural and functional studies of the proteins. This robust expression system will provide an ideal system for ion channel properties of the two urea transport proteins. Voltage-clamp studies will be applied to the system to measure the intracellular pH and membrane potential during active urea transport in oocytes, as well as the degree to which the urea transporters depend on voltage across membranes to drive active transport.

Materials and Methods

2.1. JGI_20424 and JGI_39772 C'YFP Fusion Constructs and Subcellular Localization within *Phaeodactylum tricornutum*

2.1.a. cDNA Synthesis

Phaeodactylum tricornutum RNA was isolated and harvested from *P. tricornutum* batch cultures grown in seawater media containing 75uM nitrate, a low nitrate condition, 7.5uM phosphate and 100nM silica. The *P. tricornutum* batch cultures were grown in an 18 degree Celcius incubator under a light-dark cycle, or diel cycle. The batch cultures were exposed to the light cycle in white light for a fourteen hour period, followed by ten hours in dark light conditions. The concentration of the RNA isolated from the low Nitrogen cultures (or Nitrogen starved (-N)) was 108 ng/μl. 5μg of –N RNA was mixed with primers 1μl dNTP, 1μl oligo DT, 1μl hexamer and DEPC treated H_2O, followed by incubation for 5 minutes in 65°C water bath. The cDNA synthesis mix, including 10X Reverse Trancriptase buffer, 25nM $MgCl_2$, 0.1M DTT, 40U/ml RNase OUT recombinase RNase inhibitor, 200U/ml Superscript III Reverse Transcriptase, was prepared, and 10μl was added to the RNA/primer mixture and centrifuged. Random hexamer primer was incubated with the RNA/cDNA mixture prior to amplification by Polymerase Chain Reaction (PCR) program. 1μl RNase H was added and then incubated in 85°C water bath, allowing complete cDNA synthesis.

2.1.b. Primer Design for cDNA Amplification

The PCR primers were designed to amplify the *P. tricornutum* urea transporter genes of interest, JGI_20424 (coding sequence 2,115bp) and JGI_39772 (coding sequence 1,299bp), for protein expression. These primers were designed to fuse JGI_20424 and JGI_39772 with the pENTR/D-TOPO entry vector and to fuse the genes in frame with a C-terminal fusion protein tag supplied by the destination vector after Gateway recombination reactions. In order to achieve expression of native proteins, that will contain C-terminal fusion proteins after recombination of the entry clone with a Gateway destination vector, the amplified genes contained the translational start site (ATG), and the stop codon was removed. The primers were designed in order to ensure that the gene insert would be directionally cloned into the entry vector (pENTR) in frame. The forward primers for both genes contained the four base pair sequence CACC at the 5' end of the primer, before the first nucleotide of the start codon. The 5'-CACC-3' sequence base pairs with the complementary overhang sequence 5'-GTGG-3' present in the pENTR/D-TOPO vector. To express the PCR products, the sequence contains an ATG initiation codon to ensure proper translation activity. Due to the subsequent in frame fusion of the PCR amplified genes to a C-terminal tag following Gateway recombination, the reverse PCR primers were designed to start amplification with the first codon upstream of the stop codon.

The following gene-specific primers were constructed to amplify the synthesized –N cDNA: for 2.1 kilo base pair amplicon JGI_Pt_20424, the forward primer 5'-CACCATGAGCACGAACACCGTCAG-3' and reverse primer 5'-GGCTTCCATCTCGGCAGCGA-3', and for 1.3 kilo base pair amplicon JGI_Pt_39772

the forward primer 5'- CACCATGCCGTCCTTGATTTTGGC-3' and reverse primer 5'-CTGGTTTTGTTCCGGCGAAT-3' were used. There were four reactions that were created to amplify the genes of interest. Two of the reactions served as controls, because the wild-type *P. tricornutum* genomic cDNA was amplified with the gene specific primers for JGI_20424 and JGI_39772. In the first control reaction, 0.5μl wild-type genomic cDNA, 1.5μl 20424_F, 1.5μl 20424_R and 46.5μl Accuprime Supermix, and 0.5μl wild-type genomic cDNA, 1.5μl 39772_F, 1.5μl 39772_R and 46.5μl Accuprime Supermix were prepared for PCR amplification. The two experimental amplification reactions consisted of the following reactants. For the JGI_20424 specific amplification, 0.5μl Nitrogen starved (-N) cDNA, 1.5μl 20424_F, 1.5μl 20424_R and 46.5μl Accuprime Supermix were mixed to amplify the JGI_20424 gene. For the JGI_39772 specific amplification, 0.5μl Nitrogen starved (-N) cDNA, 1.5μl 39772_F, 1.5μl 39772_R and 46.5μl Accuprime Supermix were mixed to amplify the JGI_39772 gene.

The PCR thermocycler program consisted of an initialization step of 96°C for five minutes, a denaturation step of 96°C for thirty seconds, and an annealing step of 55°C for thirty seconds. The extension/elongation steps differ for the cDNA samples of the two genes due to their base pair length. The extension time depends on the Taq polymerase enzyme and on the length of the DNA fragment, and the enzyme synthesizes at a speed of 1kbp per minute. The extension step for the JGI_20424 cDNA sample is 68°C for 2:15 minutes, which corresponds to the 2.1kb length of JGI_20424 cDNA, and 68°C for 1:35 min JGI_39772 cDNA sample (1.3kb). This cycle was repeated thirty-five times before completion, and the final hold step was held at 4°C.

2.1.c. Purification of cDNA Amplicons

The PCR amplified JGI_39772 and JGI_20424 –N cDNA DNA fragments were purified from the PCR reactions with a QIAquick PCR Purification Kit that utilizes a silica-membrane based binding, washing and elution procedure.

2.1.d. pENTR Directional Cloning

Following PCR amplification by gene-specific primers and purification of JGI_20424 and JGI_39772 cDNA, the two genes were directionally cloned into pENTR/D-TOPO vectors (Invitrogen). The recombinant vectors containing the urea transporter genes were then transformed into One Shot TOP10 chemically competent *E. coli*. The molar ratio of the PCR product to the TOPO vector used in the reaction had the highest efficiency when the molar ratio was about 2:1. The pENTR vector contains a directional TOPO cloning site for efficient directional cloning of the blunt-end PCR product, a Kanamycin resistance gene for selection in *E. coli*, a pUC origin for high-copy replication and plasmid maintenance in *E. coli, rrnB* transcription termination sequence to prevent expression of PCR product in *E. coli*, and *att*L1 and *att*L2 sites for site-specific recombination of the entry clone. The double-stranded PCR product gene of interest, containing the 5' CACC was directionally cloned by the annealing of this four base pair sequence to the 5'-GTGG-3' overhang in the pENTR vector. The vector's GTGG overhang recognizes the 5' end of the PCR product, anneals to the CACC sequence of the PCR product, allowing the PCR insert to stabilize in the correct orientation in the pENTR vector. The Topoisomerase I from *Vaccinia* virus, whose function is to bind and supply charge to a 5'-CCCTT-3' site in the pENTR vector, cleaved the phosphodiester backbone

of the incoming PCR insert. The broken phosphodiester bond provides energy to form a covalent bond between the 3' phosphate of the cleaved cDNA strand and the tyrosyl residue, Tyr-274, of Topoisomerase I.

The purified PCR products JGI_20424 and JGI_39772 were mixed in separate reactions containing the following reagents: PCR product (3μl), salt solution (1μl), sterile water (1μl) and TOPO vector (1μl). Following this reaction, the two pENTR D/TOPO constructs were transformed into chemically competent *E. coli* One Shot TOP10 cells, and the transformations were plated and incubated overnight on Luria Broth and kanamycin plates. Colonies were picked and cultured overnight in LB medium containing 50μg/ml kanamycin; these cultures contained 3ml of LB media and 1.5μl of kanamycin. The cultures were subjected to miniprep procedure, using the Qiagen Miniprep Kit, to isolate and purify the plasmid DNA from the *E. coli* bacterial cells. The NanoDrop spectrophotometer determined the concentrations of the mini-prepped plasmids.

The purified plasmids of JGI_20424 and JGI_39772 were screened by PCR to determine whether the PCR purified gene inserts had been cloned into the pENTR vector successfully. PCR supermix (26.5μl), gene-specific primers (1.5μl of forward and 1.5μl reverse gene-specific primer) and the purified plasmids (0.5μl) were mixed, subjected to PCR amplification and applied to 1% agarose gel electrophoresis to determine the size of the amplified band.

2.1.e. Sequencing

All of the transformants containing the recombinant pENTR plasmids with the successfully cloned inserts were analyzed by sequencing using gene-specific primers

JGI_20424_F, JGI_20424_R, JGI_39772_F, and JGI_39772_R. These samples were submitted to Genewiz, Inc. to perform Sanger sequencing on the samples. Each of the sequence results from Sanger sequencing with gene-specific primers JGI_20424_F, JGI_20424_R, JGI_39772_F, JGI_39772_R were analyzed by aligning and comparing each sequence to the JGI_20424 coding sequence (CDS) or the JGI_39772 CDS, respectively, with the SerialCloner program. The base pairs that were not covered by sequencing using the forward and reverse gene-specific primers on the purified pENTR vectors were sequenced using an internal primer. The forward internal primers were designed to begin 150bp upstream of the last base pair sequenced in the JGI20424_F and JGI39772_F sequences, and both primers were 18bp in length.

2.1.f. LR Clonase Recombination

The LR clonase recombination reactions, which allow the recombination of the diatom gene of interest from the pENTR vector into the Gateway pDEST diatom vectors, were carried out using the Gateway LR Clonase II enzyme mix (Invitrogen) according to the manufacturer's instructions (Invitrogen). The two purified pENTR vectors contain the JGI_20424 gene insert and the JGI_39772 insert, respectively, between its *att*L1 and *att*L2 sites. The destination vectors, pDEST, also contain an *att*L and *att*R site, between which the JGI_20424 and JGI_39772 gene inserts were inserted by recombination with the two pENTR vectors. The pDEST_C'YFP_GUS vector contains a suicidal *ccdB* gene, which is removed from the system when replaced by the gene of interest after recombination. The vector also contains an ampicillin resistance gene, and an *FcpB* promoter (Fucoxanthin chlorophyll a/c-binding protein B for *P. tricornutum*), located

upstream of its *att*L site, and a fused YFP protein immediately downstream of its *att*R site. The LR reaction operates as follows:

*att*L1-gene-*att*R2 ×	*att*L-*ccd*B-*att*R →	*att*L1- *ccd*B -*att*L2 ×	*att*R1-*gene*-*att*R2
(pENTR entry clone)	(pDEST C'YFP destination vector)	(pENTR)	(Recombined pDESTcontaining gene of interest, now an expression vector)

The LR reaction required the following mixture: 100ng pENTR vector, 150ng pDEST_C'YFP_GUS (concentration of 278.5 ng/μl), LR Clonase II enzyme mix and TE buffer. The concentration of the purified plasmid JGI_20424_4 was 212.5 ng/μl; therefore, 0.471μl of pENTR containing JGI_20424 amplicon was mixed with 0.54 μl pDEST_C'YFP_GUS. The concentration of purified plasmid JGI_39772_7 was 239.8 ng/μl; therefore 0.417μl were mixed with 0.54 μl pDEST_C'YFP_GUS. The two resulting LR products, the pEXPR vectors, transformed chemically competent TOP10 *E. coli* cells. Following plating on selectable antibiotic 100μg/ml ampicillin and LB agar plates, culturing of selected colonies (LB media and amplicillin), and mini-prep procedures to purify the pEXP plasmids, restriction digests were performed to determine whether the JGI_20424 and JGI_39772 inserts were recombined into the pEXP vectors.

2.1.g. Purification of Expression Vectors

Following retransformation of the mini-prepped pEXP_20424 and pEXP_39772 vectors onto LB and ampicillin plates, a colony of transformed *E.coli* cells was selected from each of the two transformations and grown in an overnight culture containing 75mL LB media and 75μl ampicillin and placed in a 37 Celcius shaker overnight. The pEXP plasmid vectors were purified to yield a high-quality plasmid DNA. The midiprep

procedure was performed with PureLink HiPure Plasmid Filter Midiprep Kit to ensure efficient purification of high-quality pEXP plasmid DNA isolated from high copy number plasmids.

The two expression clones, each expressing one of the urea transporter gene sequences JGI_20424 or JGI_39772, were transfected into the diatom by delivering the plasmid DNA via *in vitro* gene transfer, through the Bolistic PDS-1000/He instrument particle bombardment system. As opposed to using a microinjections process, the particle delivery system used high-pressured Helium and a partial vacuum system to force delivery of microparticles coated with DNA into *P. tricornutum* wild-type cells grown on plates without phleomycin. This system was initially used for *Cyclotella cryptic* (Roessler et al. 1994), and then it was used for *P. tricornutum* (Apt et al. 1996) and *T. pseudonana* (Poulsen et al. 2007). The transgenes introduced to the diatoms were both under the control of and transcriptionally upregulated by the diatom's light sensitive *FcpB* promoter. This promoter regulates gene expression, and the subcellular localization of the overexpressing urea transporter genes can be determined *in vivo* by visualizing the gene product yellow fluorescent reporter protein.

The Biolistic PDS-1000/He system particle bombardment system used a partial vacuum and high-pressured helium, released by a rupture disk, to forcibly accelerate a macrocarrier sheet containing microcarrier tungsten beads coated with both the pEXP DNA and pAF6 vector, which contains a phleomycin resistance gene. This forced the plasmid DNA onto the seawater agar plate of wild-type *P. tricornutum* cells. The

stopping screen stopped the macrocarriers after a short distance in the system, while the microcarriers penetrated and transformed the *P. tricornutum* cells.

After sterilizing the stopping screens, microcarriers, assembly parts and microprojectiles in the autoclave, and sterilizing the chamber, rupture disk, microcarrier and holders in 70% ethanol, the microcarrier and DNA were prepared. Tungsten beads were prepared by placing 60mg of dry tungsten in 1mL 100% ethanol. The liquefied tungsten beads were subjected to vortex and microcentrifuge. This mixture was washed with 1ml sterile distilled water and aliquotted into microfuge tubes. The 5μl pAF6 plasmid DNA, 5μl pEXP purified plasmid DNA for the gene of interest, 50μl of tungsten suspended in dH20, 50μl of 2.5M CaCl2 (aids in penetration of DNA through cell wall and plasma membrane of diatom cells) and 20μl of 0.1M spermidine were added together, vortexed, washed with 100% ethanol and resuspended in 100% ethanol. The microcarriers were placed onto the macrocarrier and set to air-dry prior to placement in the delivery system. The rupture disk was tightly placed on the end of the acceleration tube, and the sterilized stopping screen was assembled with the microcarriers placed in the holder, dried DNA-coated microcarriers facing down along the acceleration pathway. The assembly was placed in the shelf and the plate of wild-type *P. tricornutum* cells was placed in the dish holder at the closest position to the system. With helium gas pressure of 200psi above the pressure of the rupture disk, the system fires to release the pressure and accelerate the DNA from the microcarriers into the plate of cells. The pAF6 and pEXP vectors recombine with the *P. tricornutum* chromosome randomly.

Seawater agar plates were prepared as follows: 500ml seawater, 490ml miliQ H_2O, 1ml 880μM Nitrate, 1ml 55μM Phosphate, 1ml 100nM Silica, 5ml Tris base stock, and 1% Bacto agar. This media was sterilized in the autoclave, and the trace metals and vitamins were added to the seawater agar media before plating procedures. After forced transformation of the *P. tricornutum* cells on seawater agar plates without phleomycin, the plates were placed in a cycling light incubator for many days. After detectable colonies grew on the plates, 5ml of seawater media was added to each of the two plates to resuspend the transformed cells. 0.5ml of each of the two sets of resuspended transformed cells were then placed on each of three seawater agar plates containing 100μg/ml phleomycin antibiotic for each of the two cultures of transfected *P. tricornutum* cells, JGI_20424_C'YFP and JGI_39772_C'YFP. These plates were placed in the cycling light incubator to allow growth of colonies.

2.1.h. Localization of Overexpression Fusion Constructs JGI_20424_C'YFP and JGI_39772 C'YFP In *P. tricornutum*

Following transformation of the *P. tricornutum* cells and selection on seawater agar plates containing the phleomycin selectable marker (provided by pAF6 vector), the colonies were selected and resuspended in seawater liquid media and zeocin antibiotic. The seawater liquid media contained the following: 880μM Nitrate, 55μM Phosphate, 100nM Silica, 2nM Tris base and miliQ water. Following growth of the cells to a biomass with which the diatoms were expected to be in early to mid exponential growth phase, the cells were prepared on microscope slides to view under the microscope. The *P. tricornutum* cell lines were screened for the overexpressing JGI_20424 and JGI_39772

constructs.

2.2.a. Growth on Ammonium Seawater Media and ^{15}N-Labeled Urea Uptake Assay

To examine the rate of absolute urea uptake by the transgenic *P. tricornutum* line with the overexpressing JGI_20424 C'YFP urea transporter gene compared to wild-type *P. tricornutum* cells, an uptake assay was conducted using the stable nitrogen isotope nitrogen 15 (^{15}N) labeled urea. Two 100ml cultures, one of wild-type and one of the JGI_20424 C'YFP transgenic line, were grown in 100 μM ammonia based seawater media; therefore, ammonia was present rather than nitrate or urea. The cells in both cultures must be in mid-exponential growth phase, with the same biomass. The cultures were both pelleted at 5000rpm for ten minutes, and each pellet was resuspended in 50mL of artificial seawater without any added nitrogen (NF Aquil). After the cultures were pelleted once more, they were resuspended in 40ml of artificial seawater, Aquil containing 100μM NH_4, trace metals and vitamins. Both cultures were spiked with 10μM ^{15}N urea tracer at the same time, and the spiked bottles sat for ten minutes before applying the samples to the filtration system. The ^{15}N urea spiked samples were filtered on pre-combusted Whatman GF/F filters. The filters were rinsed with artificial seawater lacking urea, and dried at 55°C in an oven before being sent to the University of California, Davis Metabolomics Facility for analysis. The concentrations of the urea metabolite and the isotope ratio were determined by gas chromatography/ mass spectrometry at the UC Davis Metabolomics Facility. A comparison was made between the percent of ^{15}N-labeled urea uptake rate by ammonia-grown wild-type and JGI_20424_C'YFP transgenic *P. tricornutum* cultures.

2.3.a. Dual Transgenic JGI_39772 C'CFP and JGI_20424 C'YFP *P. Tricornutum* Cell Line

A dual transgenic construct was created with the two overexpressing urea transporters, the mitochondria localized JGI_39772 fused at its C-terminus to a CFP protein, while the JGI_20424 outer membrane transporter remained fused at its C-terminus to a YFP protein. To create the JGI_39772_C'CFP construct, the JGI_39772 pENTR entry vector recombined with a destination vector pDEST_C'CFP_GUS (concentration of 278.5 ng/μl) through the LR clonase reaction previously described. After transformation into TOP10 *E. coli* cells, plating on LB and ampicillin plates, and colony selection, the recombined plasmid was purified from *E. coli* cells by the midiprep procedure.

The two expression clones, JGI_20424_C'YFP or JGI_39772_C'CFP, were transfected into the diatom by delivering both plasmid DNAs and pAF6 plasmid DNA into the *P. tricornutum* wild-type cells through the Bolistic PDS-1000/He instrument particle bombardment system. The transgenes were under the control of and transcriptionally upregulated by the diatom's light sensitive *FcpB* promoter. The *FcpB* promoter localized the overexpressing urea transporter genes *in vivo* by regulating expression of the transporter genes JGI_20424_C'YFP and JGI_39772_C'CFP, fused to yellow and cyan fluorescent reporter proteins, respectively.

2.4.a. Heterologous Expression System in *Xenopus laevis* Oocytes Functionally Characterize Cloned Urea Transporters

The two urea transporters will be functionally characterized by heterologous expression in *Xenopus laevis* oocytes. In this *Xenopus* assay system, the JGI_20424 and JGI_39772 cDNA will be overexpressed in pGEM plasmid vector, followed by linearization of the plasmid vector containing the cDNA transcript. The *Eco*R1 linearized pGEM vectors, one containing JGI_20424 and one containing JGI_39772 cloned into the gGEM vector, will be transcribed to cRNA prior to injection into defolliculated oocytes.

To construct the vectors, the following steps were performed on a pGEM vector and the two pENTR vectors constructed for JGI_20424 and JGI_39772. The pGEM vector was transformed into TOP10 *E. coli* competent cells, cultured in LB media and ampicillin, and midi-prepped to purify a great amount of plasmid from the bacterial cells. Primers were designed to amplify the original JGI_20424_pENTR vector and JGI_39772_pENTR vector for the two transporters. The forward primers for JGI_20424 and JGI_39772 contained the 5'-CCG-3' non-specific sequence, immediately followed by the *Eco*R1 restriction digest, 5'-GAATTC-3', followed by the first twenty-one base pairs of each of the two JGI sequences after the start codon ATG. The reverse primers contained the non-specific sequence 5'-GGC-3', followed by the *Xba*1 restriction digest site 5'-TCTAGA-3', followed by the reverse complement of the last 21 base pairs, including the stop codon, of JGI_20424 and JGI-39772.

The following gene-specific primers were constructed to amplify the synthesized two pENTR constructs: for the 2.1 kb JGI_20424 amplicon, forward primer 5'-CCGCAATTCATGAGCACGAACACCGTCAGC-3' and reverse primer 5'-CTATCTAGACTAGGCTTCCATCTCGGCAGC-3', and for the 1.3kb JGI_39772

amplicon, forward primer 5'-CCGCAATTCATGCCGTCCTTGATTTTGGCA-3' and reverse primer 5'- CTATCTAGATTACTGGTTTTGTTCCGGCGA-3'. The PCR conditions were as follows: Initialization step of 96°C for 5 minutes, denaturation step of 96°C for 30 seconds, annealing step of 53°C for 30 seconds, extension step of 68°C for 2:15 minutes for JGI_20424 and 1:35 minutes for JGI_39772, repeated for 37 cycles. The PCR products were applied to 1% agarose gel electrophoresis to determine the size of the products, and they were purified from the gel with a GeneClean kit. A double digest of the pGEM vector (3μg) and the gel-purified PCR products (20μl total) was performed using *Eco*RI and *Xba*I, 1X BSA and 1X CA buffer. Following digestion of the pGEM vector and the two PCR products, the samples were applied to an agarose gel to determine the respective sizes and purified with GeneClean kit. 2μl of each purified product was applied to an agarose gel with a 1kb plus DNA ladder that was loaded in a dilution series of 500ng, 200ng, 100ng and 50ng to determine the concentration of each product.

After determining the relative concentrations of the doubly digested pGEM vector, JGI_20424 and JGI_39772 inserts, the two inserts were ligated into the digested pGEM vector with T4 DNA ligase. The JGI_20424 was ligated into the pGEM vector in a 1:2 insert to vector molar ratio, whereas the JGI_39772 was ligated into pGEM in a 1:5 molar ratio. For the JGI_20424 ligation reaction, 0.5μl digested pGEM, 1μl of digested JGI_20424, 1μl T4 DNA ligase, 1.5μl 10X T4 Ligase Buffer, and 11μl of dH_2O. For the JGI_39772 ligation reaction, 0.5μl digested pGEM, 12.1875μl of digested 39772, 1μl

10X T4 DNA ligase, 1.5μl T4 Ligase Buffer. These ligation reactions were left overnight in 16°C.

The two ligations reaction (5μl of each) were separately transformed into TOPO ONE SHOT *E. coli* cells, followed by plating all of each transformation reaction onto Luria Broth agar and ampicillin. Selection of ampicillin plates identified the successfully ligated vectors containing pGEM ligated to JGI_20424 insert and pGEM ligated to JGI_39772 insert, which occurred due to the ligation of the respective sticky ends generated from the double restriction digestion reactions. The plasmids were mini-prepped, and the gene-specific primers created to amplify the JGI_20424 pENTR and JGI_39772 pENTR vectors were used to screen the colonies to ensure that the ligations were successful. Successfully ligated JGI_20424_pGEM and JGI_39772_pGEM vectors were retransformed onto LB and ampicillin plates, and a selected colony from each transformation was cultured in 75ml LB and 75μl of ampicillin. The cultures were midi-prepped to yield a highly purified plasmid with a high concentration for each of the completed JGI_20424_pGEM and JGI_39772_pGEM vectors.

Results

3.1. JGI_20424 and JGI_39772 C'YFP Fusion Constructs and Subcellular Localization within *Phaeodactylum tricornutum*

3.1.a cDNA Synthesis

The two cDNA samples that were amplified by PCR using gene-specific primers JGI_Pt_20424_Forward, JGI_20424_Reverse, JGI_39772_Forward and JGI_39772_Reverse were subjected to 1% agarose gel electrophoresis to determine the size of the amplified product. Experimental samples containing amplified Nitrogen starved (-N) cDNA by JGI_20424 specific forward and reverse primers yielded a band at 2.1kb. Experimental samples containing amplified –N cDNA by JGI_39772 specific forward and reverse primers yielded a band at 1.3kb. Control samples containing amplified wild-type *P. tricornutum* genomic cDNA by JGI_20424 forward and reverse primers yielded a band at 2.1kb. Control samples containing amplified wild-type *P. tricornutum* genomic cDNA by JGI_39772 forward and reverse primers yielded a band at 1.3kb. The control lanes reveal that the forward and reverse primers for both JGI_20424 and JGI_39772 are gene-specific, as they amplified the wild-type *P. tricornutum* genomic cDNA to the kilo-basepair size of the urea transporter genes, 2.1kb and 1.3kb, respectively (Figure 5). The controls also revealed that the -N cDNA was properly synthesized.

3.1.b. Purification of cDNA Amplicons

The PCR amplified JGI_39772 and JGI_20424 –N cDNA DNA fragments that were purified from the PCR reactions were quantified with the NanoDrop 1000

Spectrophotometer to reveal that the concentration of JGI_20424 was 97.4ng/uL and JGI_29772 was 89.5 ng/uL. The optical spectrophotometer measurements of $A_{260/280}$, measuring DNA/RNA concentration, were 1.89 for JGI_20424 and 1.91 for JGI_39772; the $A_{260/230}$ values, measuring protein concentration, were 2.12 for JGI_20424 and 2.03 for JGI_39772. These spectrophotometric measurements indicate highly purified cDNA fragments.

3.1.c. pENTR Directional Cloning

The NanoDrop Spectrophotometer was used to determine the concentration and the absorbance ratios from the mini-prepped, purified pENTR plasmids, one containing the JGI_20424 transcript and the other containing JGI_39772 transcript. The NanoDrop Spectrophotometer result for the selected pENTR JGI_20424 mini-prepped, purified plasmid was 212.5ng/μl. This plasmid was selected for the subsequent LR recombination reaction, because its concentration was high and its absorbance values indicate that it is a very pure sample ($A_{260/280}$ of 1.89 and $A_{260/230}$ of 2.00). The selected pENTR JGI_39772 mini-prepped, purified plasmid was 239.8ng/μl. This plasmid was selected for the subsequent LR recombination reaction, because its concentration was high and its absorbance values indicate that it is a very pure sample ($A_{260/280}$ of 2.19 and $A_{260/230}$ of 1.90).

The PCR screened JGI_20424 purified plasmids were correctly cloned into the pENTR vector, because 1% agarose gel electrophoresis revealed that the inserts were 2.1kb, the size of the gene (Figure 6). The PCR screened JGI_39772 purified plasmids

were also correctly cloned into the pENTR vector, as the 1% agarose gel revealed that the inserts were 1.3kb, the size of JGI_39772 gene (Figure 7).

3.1.d. Sequencing

The forward sequence of the mini-prepped JGI_ 20424_4 pENTR clone was locally aligned and compared to JGI_ 20424 CDS sequence. This local alignment was followed by amino acid translations for both sequences, allowing the start codon, ATG encoding Methionine, for both transcripts to align and translation of the codons for both transcripts into amino acids. The local align tool on Serial Cloner uses a block-based method as well as a Smith-Waterman algorithm to align two sequences. The two aligned sequences had 100% similarity. The JGI_20424_4_F sequence was translated to Frame 3 to align to the CDS translated to Frame 1. The forward sequence contained the desired CACC sequence before the start codon ATG, and it was 841bp in length. The two sequences were compared through this alignment by identifying any amino acid changes between the two sequences at a particular location in the transcript. The 20424_4 sequence and the 20424 CDS did not differ at any position in aligned amino acids. The reverse sequence of mini-prepped 20424_4 pENTR clone was also aligned and compared to 20424 CDS, after converting the reverse sequence to the anti-parallel composition. The 20424_4_R was translated to Frame 1, and that of the CDS was translated to Frame 1. The reverse sequence was 853bp in length. There were not any amino acid changes present between the two sequences, indicating that the reverse sequence and CDS were aligned.

With a 2.115kbp CDS sequence for the 20424 transcript, the base pairs that were not covered by sequencing using the forward and reverse gene-specific primers on the mini-prepped 20424 pENTR vector were sequenced using an internal primer. The forward internal primer that was designed was an 18bp primer that sequenced the 20424 pENTR clone at a position 150bp upstream from the last base pair sequenced from the 20424_4_F sequence; therefore, as the end of the forward sequenced ended at 865bp in the forward sequence, the internal primer's sequence begins at position 715 in the forward sequence. The internal primer's sequence is the following: Forward: 5'-TACATCAAGTTCTACTCG-3'. The internal primer sequenced the 20424 clone transcript from 760bp to 1017bp. There were not any amino acid changes in the sequence when compared to the CDS.

The forward sequence of mini-prepped JGI_39772_7 was aligned in the Serial Cloner program to the JGI_ 39772 CDS sequence, followed by amino acid translations for both sequences and frame alignment. Both the 720bp 39772_7_F sequence and 1356bp JGI_ 39772 CDS were aligned in Frame 2. At base pair position in the 39772_7_F forward sequence, there was an amino acid change from Valine in 39772_7_F to Methionine in the 39772 CDS. There was also an amino acid change at base pair position 197 in the 39772_7_F sequence from Histidine to Arginine in 39772 CDS. There were not any amino acid changes present in the locally aligned and translated reverse sequence of 757bp 39772_7_R compared to 1356bp 39772 CDS, indicating that the PCR and cloning reactions proceeded without any mistakes. The internal primer designed, whose sequence is 5'-ACAGCGCCAACACAGTGG-3', sequenced the

39772_7 transcript to reveal that there were no amino acid changes in the sequence compared to the CDS.

3.1.e. LR Clonase Recombination

Both JGI_20424 and JGI_39772 pENTR amplicons successfully recombined into the destination vector pDEST_C'YFP_GUS, to generate the pEXP_20424 and pEXP_39772 vectors. The mini-prepped pEXP 20424_4 digested by restriction enzyme *Kpn*1 revealed on the 1% agarose gel electrophoresis unit the expected band sizes 4.325kb and 2.172kb (Figure 8). The mini-prepped pEXP 39772_7 digested with *Eco*RI provided the expected band sizes on the 1% agarose gel of 4.0kb and 1.7kb (Figure 9).

3.1.f. Purification of Expression Vectors

The concentrations of the midi-prepped (with PureLink HiPure Plasmid Filter Midiprep Kit) and purified pEXP_20424 and pEXP_39772 are noted in Figure 5. The midi-prepped plasmid pEXP_20424_C'YFP had a concentration of 1154.0 ng/μl, $A_{260/280}$ of 1.90 and an $A_{260/230}$ of 2.20. The midi-prepped plasmid pEXP_39772_C'YFP had a concentration of 2150.8 ng/μl, $A_{260/280}$ of 1.89 and an $A_{260/230}$ of 2.22. The pAF6 vector, containing the phleomycin resistance gene, had a concentration of 2054.3ng/μl, $A_{260/280}$ of 1.88 and an $A_{260/230}$ of 2.25.

3.1.g. Localization of Overexpressing Transgenic JGI_20424 and JGI_39772 C'YFP Fusion Constructs In *P. tricornutum*

The JGI_20424 C'YFP overexpressing construct localized to the outer plasma membrane in *P. tricornutum*. This urea transporter is an orthologue of a plant-type high affinity urea transporter. The construct was strongly overexpressed, and the intense YFP

signal was visible around the outer membrane of *P. tricornutum* (Figure 10). The JGI_20424 localization to the outer plasma membrane confirmed the WoLF PSort program prediction score of 10.0 to the plasma membrane. The differential interference contrast microscopy (DIC) image reveals the bright image of the JGI_20424 C'YFP overexpressing construct, and the size of the diatom from its central region to outer membrane is about 20μm. The red color represents the chlorophyll auto-fluorescence that localizes to the plastid, while the green color represents the YFP fluorescence.

The JGI_39772 overexpressing transgenic C'YFP fusion construct was localized to the mitochondria in *P. tricornutum*. JGI_39772 is a homologue to metazoan-type renal and erythrocyte transporters (Figure 11). The differential interference contrast microscopy (DIC) image is the bright field image of JGI_39772 C'YFP fusion overexpressing construct. This localization to the mitochondria confirms the *TargetP* prediction score of 0.815 to the mitochondria. The red color represents the chlorophyll auto-fluorescence that localizes the plastid, while the green color represents YFP fluorescence. The scale bar represents 20μm. Confocal fluorescence microscopy provided a three-dimensional image of the mitochondrial localization of the overexpressed urea transporter in JGI_39772 in *P. tricornutum* (Supplementary figure 1).

3.2 Growth on Ammonium Seawater Media and ^{15}N-Labeled Urea Uptake Assay

The results of the percent of ^{15}N urea uptake by the wild-type *P. tricornutum* and JGI_20424 transgenic *P. tricornutum* cultures grown on ammonium indicate that the overexpression of *P. tricornutum* JGI_20424 urea transporter leads to an approximate

ten-fold increase in urea uptake; therefore, the protein efficiently transports urea. Urea uptake is relatively low in wild-type *P. tricornutum* cells grown on ammonia.

The natural abundance of ^{15}N is 0.366%; therefore, the ammonia in the media in which the cultures were grown contained 0.366% ^{15}N, while the added urea was 99% ^{15}N. The percent of ^{15}N of the particulate nitrogen reflects uptake of the ^{15}N-labeled urea. The percent of ^{15}N was measured in each of the four *P. tricornutum* cultures grown on ammonium (NH_4) seawater. The first sample of wild-type *P. tricornutum* culture with a biomass of 40.1μg contained 0.41 percent ^{15}N, while the second wild-type sample of biomass 40.3μg contained 0.42 percent ^{15}N. The first sample of the *P. tricornutum* JGI_20424 overexpression culture of biomass 11.5μg contained 0.59 percent ^{15}N, while the second overexpression culture of biomass 12.5μg contained 0.57 percent ^{15}N. The particulate nitrogen of the wild-type ammonium-grown *P. tricornutum* cultures was 1005.078; therefore, the percent of ^{15}N-labeled urea uptake per hour was 0.286 percent. The particulate nitrogen of JGI_20424 overexpression *P. tricornutum* culture was 300.02; therefore, the percent of ^{15}N-labeled urea per hour was 1.26 percent (Table 1 and Figure 12).

3.3 Dual Transgenic JGI_39772 C'CFP and JGI_20424 C'YFP *P. tricornutum* Cell Line

3.3.a. LR Clonase Recombination

The mitochondria-localized JGI_39772 was fused at its C-terminus to a CFP protein by recombining the 39772_pENTR vector with the pDEST_C'CFP_GUS vector in the LR recombination reaction. To create the JGI_39772_C'CFP construct, 100ng of JGI_39772 pENTR entry vector (concentration of 265.1ng/μl) recombined with 150ng

pDEST_C'CFP_GUS (concentration of 203.1 ng/μl) through the LR clonase reaction previously described; therefore, 0.377μl 39772_pENTR and 0.739μl pDEST_C'CFP_GUS were added to the LR clonase reaction. After transformation into TOP10 *E. coli* cells, plating on LB and ampicillin plates, and colony selection, pEXP_39772_C'CFP was purified from *E. coli* cells by the miniprep procedure. The mini-prepped pEXP_39772_C'CFP plasmids were digested with restriction enzyme *Nae1* to ensure that the JGI_39772 inserted correctly into the pDEST_C'CFP_GUS vector. The expected band sizes of 3.818kbp and 1.923kbp were determined on a 1% agarose gel electrophoresis unit (Figure 13).

3.3.b. Purification of Expression Vectors

The recombined plasmid pEXP_39772_C'CFP was retransformed onto LB and amplicillin plates, followed by purification from TOP10 *E. coli* cells by the midiprep procedure (with PureLink HiPure Plasmid Filter Midiprep Kit). The purified plasmid yielded a concentration of 2150.8 ng/μl, an $A_{260/280}$ of 1.87 and an $A_{260/230}$ of 2.22. The midi-prepped Plasmid pEXP_20424_C'YFP, which had a concentration of 1154.0 ng/μl, $A_{260/280}$ of 1.90 and an $A_{260/230}$ of 2.20 was used in the particle gene delivery system. The pAF6 vector, containing the phleomycin resistance gene, had a concentration of 3.7μg/μl, $A_{260/280}$ of 1.89 and an $A_{260/230}$ of 2.19.

3.3.c. Localization of Dual Transgenic *P. tricornutum* Construct Overexpressing JGI_39772 C'CFP and JGI_20424 C'YFP

A cell line overexpressing both JGI_39772 C'CFP and JGI_20424 C'YFP urea transporters has not been found to date. However, many cell lines that have been screened

for the dual transgenic construct have one channel overexpressing its respective urea transporter. For example, many cell lines from this transformation strongly overexpress in the YFP wavelength along the plasma membrane, which is where the urea transporter JGI_20424 C'YFP localizes. Screening will continue to attain a cell line overexpressing both urea transporters in their respective YFP and CFP wavelengths.

3.4 Heterologous Expression System in *Xenopus laevis* Oocytes to Functionally Characterize Cloned Urea Transporters

In the vector construction, the pGEM vector (1μg) was retransformed on Luria broth plates containing ampicillin. The selected colonies were cultured and midiprepped, and the midiprep yielded a high concentration of 789.6ng/μl. The 20424_F_EcoI and 20424_R_Xba1 primers amplified the original JGI_20424_4_pENTR vector, and the 39772_F_EcoI and 39772_R_Xba1 primers amplified the original JGI_39772_7_pENTR vector. The PCR products were applied to 1% agarose gel electrophoresis to determine the size of the products. The JGI_20424_4_pENTR vector was amplified with gene-specific primers including restriction enzyme sites to yield a band at 2.1kb, and JGI_39772_7_pENTR vector was amplified with gene-specific primers including restriction enzyme sites to yield a band at 1.3kb (Figure 14).

The PCR products that were applied to 1% agarose gel electrophoresis were purified from the gel with a GeneClean kit. The gel-purified PCR products that were run on a 1% agarose gel electrophoresis to revealed that the purified JGI_20424_4_pENTR vector amplified with gene-specific primers was 2.1kb, and the purified JGI_39772_7_pENTR vector amplified with gene-specific primers was 1.3kb.

A double digest of the pGEM vector (3ug) and the gel-purified JGI_20424 and JGI_39772 PCR products (20ul total) was performed using *Eco*RI and *Xba*I restriction enzymes, 1X BSA and 1X CA buffer. The bands that resulted from the *Eco*RI and *Xba*I digested JGI_20424, JGI_39772 and pGEM vector were the correct size based on *in silico* analysis of the restriction digest reaction. The double-digested JGI_20424 band appeared at 2.1 kb on the 1% agarose gel electrophoresis. The double-digested JGI_39772 bands appeared at 949 bp and 419 bp, whose sum is the expected 1.4 kb, and there was a very faint band at 1.4 kb. The double-digested pGEM vector band appeared at 2.945 kb (Figure 15).

Following double digestion of the pGEM vector, PCR amplified JGI_20424 and PCR amplified JGI_39772, the products were applied to a 1% agarose gel to determine the respective sizes of the digested products. 2μl of each purified product was run on the agarose gel alongside a DNA ladder that contained a dilution series of 500ng, 200ng and 50ng, respectively. Comparison of the intensity of the products' bands to that of the ladder's bands in the dilution series determined the relative concentration of each product. The concentration of the digested pGEM vector was approximately 600ng/μl, the concentration of double digested JGI_20424 was approximately 300ng/μl, and the concentration of double-digested JGI_39772 was approximately 40ng/μl (Figure 16).

Each of the two digested inserts, JGI_20424 and JGI_39772, were ligated into the digested pGEM vector with T4 DNA ligase. The JGI_20424 insert ligated into the pGEM vector in a 1:2 vector to insert molar ratio, and the JGI_39772 insert ligated into the pGEM vector in a 1:10 vector to insert molar ratio. Following transformation of the two

ligated reactions into TOP 10 chemically competent *E. coli* cells, colonies were selected, cultured overnight and mini-prepped to purify the plasmids. To ensure that the inserts were successfully ligated into the pGEM vector, a PCR screen was performed on the purified, mini-prepped plasmids with the gene-specific primers designed to amplify the two pENTR vectors with the restriction digestion sites. The PCR screen revealed a band at 2.1kbp for the JGI_20424_pGEM vector (Figure 17), and a band at 1.3kbp for JGI_39772_pGEM vector (Figure 18) on a 1% agarose gel electrophoresis unit. These results indicated that the ligations were successful and that the JGI_20424_pGEM vector and JGI_39772_pGEM vector were constructed properly.

Discussion

The experiments described above shed light on the role of urea as a nitrogen source in diatoms. Urea is a molecule that releases atmospheric carbon dioxide when it is catabolized, and it can act as a source and storage depot for nitrogen and a growth stimulant for phytoplankton, a family of organisms responsible for 25-40% of the 45-50 billion tons of organic carbon fixed annually by the sea (Nelson et al. 1995, Falkowski & Raven 1997). Use of *Phaeodactylum tricornutum* lines constructed to overexpress fluorescence-tagged urea transporters has shown that marine diatoms would have the ability to assimilate exogenous urea as a nitrogen source in order to fuel processes that store organic nitrogen.

The analysis of the predicted coding sequences of *Phaeodactylum tricornutum*'s genome and subcellular localization of two urea transporter proteins demonstrated that exogenous urea can be transported both across the outer membrane and into the mitochondria of the diatom. The plant-like high-affinity urea transporter JGI_20424 C'YFP fusion construct localized to the plasma membrane, while the JGI_39772 C'YFP construct, homologous to bacteria and metazoan urea transporters, localized to the mitochondria in *P. tricornutum*. The JGI_20424 urea transporter serves to transport exogenous urea, while the JGI_39772 urea transporter presumably transports urea of extra- and intracellular nitrogen to the mitochondria where it may be used as a substrate for urease. The identification of targeting presequences in both urea transporters supported the results of the overexpression and fluorescence imaging studies. The subcellular localization prediction scores for JGI_20424 and JGI_39772 of 10.0 to the

plasma membrane (WoLF PSort) and 0.815 to the mitochondria (*TargetP*), respectively, were confirmed by *in vivo* analysis.

The cellular localization of *P. tricornutum*'s urease enzyme to the mitochondria indicates that urea must be transported exogenously both across the outer plasma membrane and into the mitochondria. Urea produced exogenously and transported into the diatom through the plasma membrane urea transporter supports global carbon fixation, while the mitochondrial urea transporter provides nitrogen to the urea cycle for anabolic activity. The localization of JGI_20424 to the plasma membrane and of JGI_39772 to the mitochondria confirmed that the two transporters could indeed supply urease with urea. The localization studies support the theory that the urea cycle and CPSIII facilitate repackaging and reallocation of nitrogen products, generated from catabolic activities, to compounds other than amino acids during periods of nitrogen-limited growth. Ammonium and carbon dioxide are detoxified and recovered through the CPSIII and urea cycle activity (Allen 2005, Allen et al. 2006, Parker et al. 2008). The marine diatom contains the GS-GOGAT (glutamine synthetase/glutamate synthase) cycle in the mitochondria that assimilates ammonia into amino acids (Allen et al. 2006, Bowler et al. 2010). Therefore, both the mitochondrial GS-GOGAT cycle and the urea cycle pathways in marine diatoms most likely contribute to their ecological success.

The localization of a mitochondrial JGI_39772 urea transporter supports the theory that the mitochondria of marine diatoms are utilized in a unique fashion for nitrogen assimilation. The mitochondrion of marine diatoms potentially serves as the central organelle that maintains the overall cellular nitrogen status by linking nitrogen

turnover, storage and regeneration pathways, which is unique to photosynthetic eukaryotes. The evolutionary development of marine diatoms provided a unique combination of genes that collectively encode biochemical pathways of nutrient assimilation, which are processes found separately in metazoans, plants or bacteria. The specific gene products localized to the mitochondria of marine diatoms and Chlorophyll c algae confirm the central role of the organelle for nitrogen assimilation and regeneration. The metazoan-like urea transporter (JGI_39772) that localized to the mitochondria has an intriguing phylogenetic history, which suggests that among all other eukaryotes JGI_39772 is shared only between diatoms and animals with a potential bacterial origin.

The urea uptake assay demonstrated that urea uptake is relatively low in wild-type *P. tricornutum* cells grown on ammonium. The overexpression of *P. tricornutum* JGI_20424 urea transporter leads to an approximate five-fold increase in urea uptake, indicating that the protein efficiently transports urea. This increase in uptake rates would be even more striking if the results are normalized by biomass; however, the wild-type culture's biomass was four times greater than that of the overexpressing cultures. When cultured on ammonium, both the wild-type and the transgenic cultures native urea transporters downregulated urea uptake and utilization. During growth on ammonium, the ^{15}N-labeled urea uptake rates were expected to be low as the native urea transporters are downregulated; however, the overexpressing JGI_20424 transgenic line had additional copies of the native urea transporter that continued to be active.

Future experiments to validate the results of the *in-vivo* localization of JGI_20424 and JGI_39772 determined by epifluorescence microscopy, antibodies specific to the

JGI_20424 and JGI_39772 proteins will be incubated in with a wild-type *P. tricornutum* culture and used for *in vivo* localization of the native proteins.

Future experiments will determine the kinetics of urea transport by these urea transporters, including the limiting velocity (V_{max}) and the concentration of substrate giving half-maximal velocity (K_m). Additional experiments will be performed with the ^{15}N-labeled urea uptake assay to compare the urea uptake rate of the mitochondrial urea transporter JGI_39772 overexpressing *P. tricornutum* culture with wild-type and JGI_20424 overexpressing transgenic *P. tricornutum* cultures. Experiments with ^{15}N labeled nitrate and ammonium would help to determine the cellular assimilation pathways in wild-type and transgenic *P. tricornutum* cultures. A dual transgenic *P. tricornutum* culture containing two overexpressing constructs, JGI_39772 C'CFP in the mitochondria and JGI_20424 C'YFP in the outer membrane, will also be cultured on ammonia. This will directly compare the influence of JGI_20424 on JGI_39772 urea transporters and the effect on their respective (^{15}N-labeled) urea uptake rates.

The functional characterization, molecular mechanisms and kinetics of urea transport by the two urea transporters of interest will be examined by the heterologous expression system in *Xenopus laevis* oocytes. Expression vectors containing cloned JGI_20424 and JGI_39772 cDNA will be expressed by the oocytes (Shroeder et al. 1994). As described above, the respective cDNAs encoding the two transporters have been successfully cloned into an optimized expression vector, pGEM, to analyze the physiological mechanisms and structures of urea uptake in the two transporters. The constructed JGI_20424_pGEM vector and JGI_39772_pGEM vector will be linearized

by restriction enzyme *Eco*R1 for *in vitro* cRNA transcription and microinjection into *Xenopus* oocytes. The defollicuated *Xenopus* oocytes injected with the respective, linearized pGEM vector cRNA will translate, process and target the urea transporter proteins to allow structure and function studies of the proteins. This is a robust expression system for ion channel studies of the two urea transport proteins. Oocytes will be held under a voltage-clamp to obtain electrophysiological recordings that will measure the membrane potential changes associated with urea transport. JGI_20424 localized to the plasma membrane and JGI_39772 localized to the mitochondria are membrane transporter proteins that are expected to operate a channel mechanism of transport that is regulated by ionic currents. Voltage clamp studies in the transfected *Xenopus* oocytes will demonstrate the molecular mechanisms of urea transport. Mutational studies of the cloned urea transporters can be performed as well to pinpoint crucial amino acid residues.

Additional expression profiling experiments will be performed for the urea transporters, including both quantitative real-time PCR (qRT-PCR) and growth curve experiments (Siaut et al. 2007). qRT-PCR will analyze the transcript levels and gene expression of the two urea transporters. qRT-PCR will determine the genes' transcript levels in different growth conditions, and it will also demonstrate the influence of factors such as light availability, temperature and trace metal conditions on gene expression. Specifically, qRT-PCR experiments examining urea transporter expression will be performed on high or low urea, ammonia and nitrate conditions using housekeeping genes to normalize the urea transporter gene expression data.

Growth experiments will be conducted to compare the growth rates of the two

JGI_20424 and JGI_39772 overexpressing transgenic *P. tricornutum* cultures to the wild-type culture following initial growth on nitrate prior to inoculation with urea and nitrogen limitation, respectively. Growth rates will be determined using in-vivo fluorescence and flow cytometry cell counts (Dupont et al. 2010). The wild-type and JGI_20424 and JGI_39772 overexpressing transgenic cultures will be grown on nitrate seawater media initially. Following growth on nitrate, the three cultures will be inoculated with urea. It is expected that the two overexpressing transgenic cultures will have a higher growth recovery rate in the presence of urea. Batch cultures of wild-type and the two overexpressing transgenic *P. tricornutum* will be grown under nitrogen limited conditions until growth stops and the cultures have reached a stationary growth phase. The three cultures will be spiked with phosphate, silica and urea to examine the rapid responses of the urea transporters from conditions of nutrient limitation to the presence of urea. Recovery rates following nitrogen starvation are expected to be significantly different for the wild-type compared to the transgenic cultures. The recovery rate of the two overexpressing transgenic cultures is expected to be higher than that of the wild-type culture.

The electrophysiology and molecular mechanisms of urea uptake through the two urea transporters are unknown. The JGI_20424 transporter is member of the sodium-solute symporter superfamily; however, the mechanism of urea transport across plasma membrane, specifically its dependency on either a proton gradient or a Na^+ gradient to drive symport, has not yet been elucidated. The sodium-solute symporter is directed by energy stored by an intracellular Na^+/K^+ ATPase pump. This pump creates an

electrochemical sodium or proton gradient using ATP hydrolysis. The energy stored in the sodium gradient drives urea solute against its concentration gradient and into the diatom. To determine whether a Na^+ or an H^+ gradient allows symporter activity and subsequence urea uptake, the rate of uptake of ^{15}N-labeled urea would be measured in a balanced medium containing 300nM Na^+ compared to a control in the absence of Na^+ using both wild-type and overexpressing JGI_20424 transgenic *P. tricornutum* cultures. It is expected that ^{15}N-labeled urea uptake rates will be higher in 300nM Na^+ conditions.

Further clues to the nature of the transport mechanism could be obtained by the use of a Na^+ channel blocker, such as the anti-epileptic drug Phenytoin sodium, or a protonophore, such as carbonyl cyanide m-chlorophenylhydrazone.

The urea transporter localization studies and urea uptake activity results are consistent with the hypothesis that diatoms assimilate urea as a nitrogen source in the absence of ammonia or other forms of nitrogen, and that they most likely use urea for organic nitrogen storage (Figure 19). The urea cycle most likely functions in an anabolic capacity, converting ammonia to urea in the mitochondria to retain organic nitrogen and carbon compounds during periods of nutrient stress.

Figures and Tables

Figure 1. Phylogenetic Tree of JGI_20424 Urea Transporter.
The maximum likelihood "bootstrap" phylogenetic tree constructed for JGI_20424 (Chromalveolata Heterokontophyta *Phaeodactylum tricornutum* CCAP_10551_gi219119531) shows eighty-three protein sequences from BLAST search using the JGI_20424 sequence. The clusters with the highest homology with JGI_20424 protein sequence are grouped near to the identified JGI_20424 *P. tricornutum* sequence.

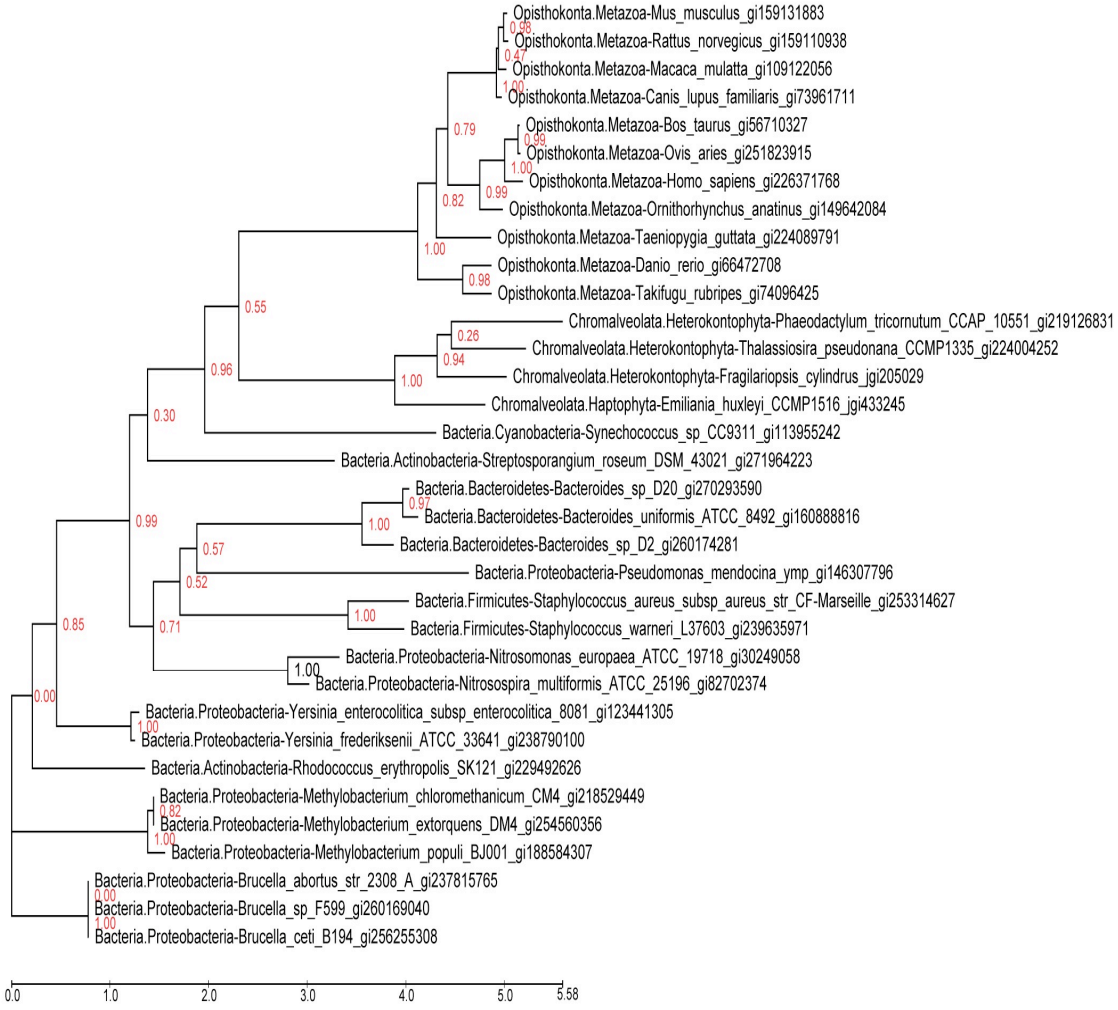


Figure 2. Phylogenetic Tree of JGI_39772 Urea Transporter.
The maximum likelihood "bootstrap" phylogenetic tree constructed for JGI_39772 (Chromalveolata Heterokontophyta Phaeodactylum tricornutum CCAP_10551_gi219126831) shows thirty-four protein sequences from BLAST search using the JGI_39772 sequence. The clusters with the highest homology with JGI_39772 protein sequence are grouped near to the identified JGI_39772 *P. tricornutum* sequence.

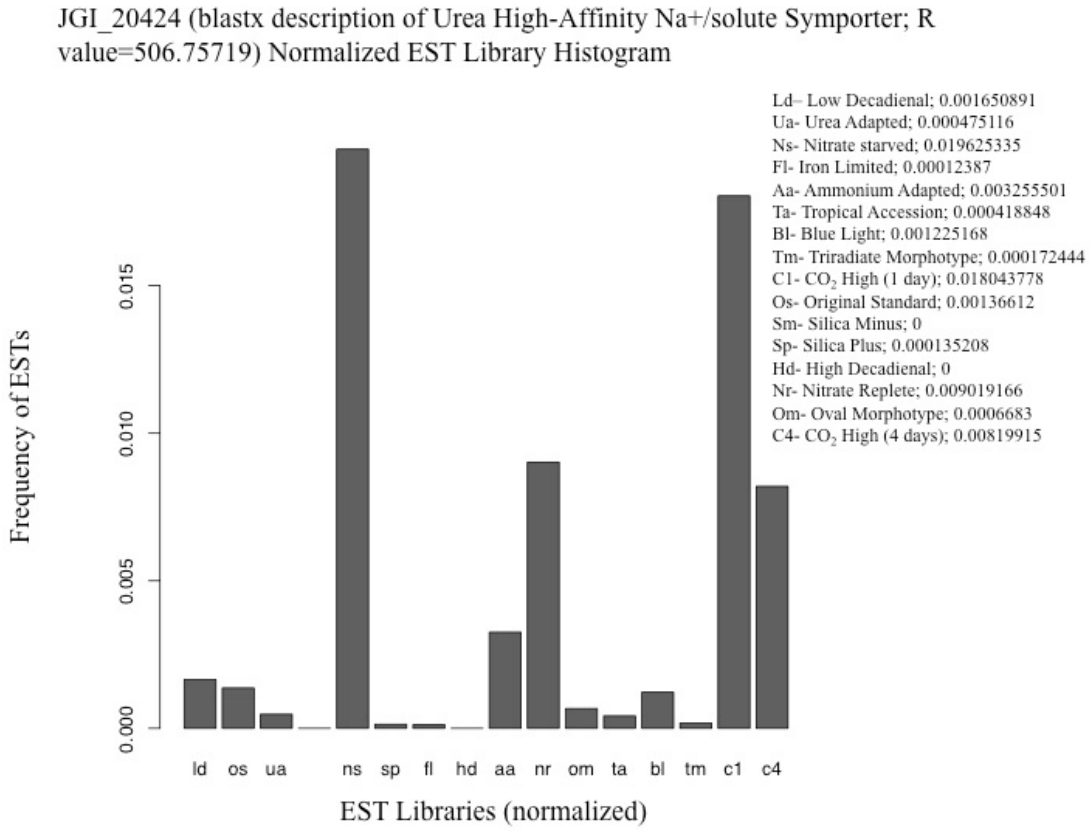


Figure 3. Qualitative Analysis of the EST Library as an Expression Profiling System for JGI_20424 Urea Transporter.

This histogram was generated for JGI_20424, by determining the frequency JGI_20424 in each EST library for the sixteen different culture conditions (data from Maheswari et al, 2010). The data was normalized by dividing the number of ESTs in a particular condition for JGI_20424 by the sum of all ESTs in the libraries for a particular growth condition from all genes in *P. tricornutum*. The value of R for a specific gene represents the amount of differential gene expression across the sixteen growth conditions. The R-value of JGI_20424 is 506.75719. The x-axis of the plot represents the normalized EST library and the y-axis represents the frequency of the ESTs for a specific gene. Bar charts were made to illustrate the changes in the frequencies of ESTs across all sixteen conditions.

The EST histogram demonstrates that JGI_20424 urea transporter is highly upregulated in conditions of nitrate stress. These results are consistent with observations that JGI_20424 is the second most highly expressed gene in the nitrate-starved library. The greatest frequencies of ESTs for JGI_20424 are in conditions of nitrate stress (0.01962533), high CO_2 for one day (0.01804378).

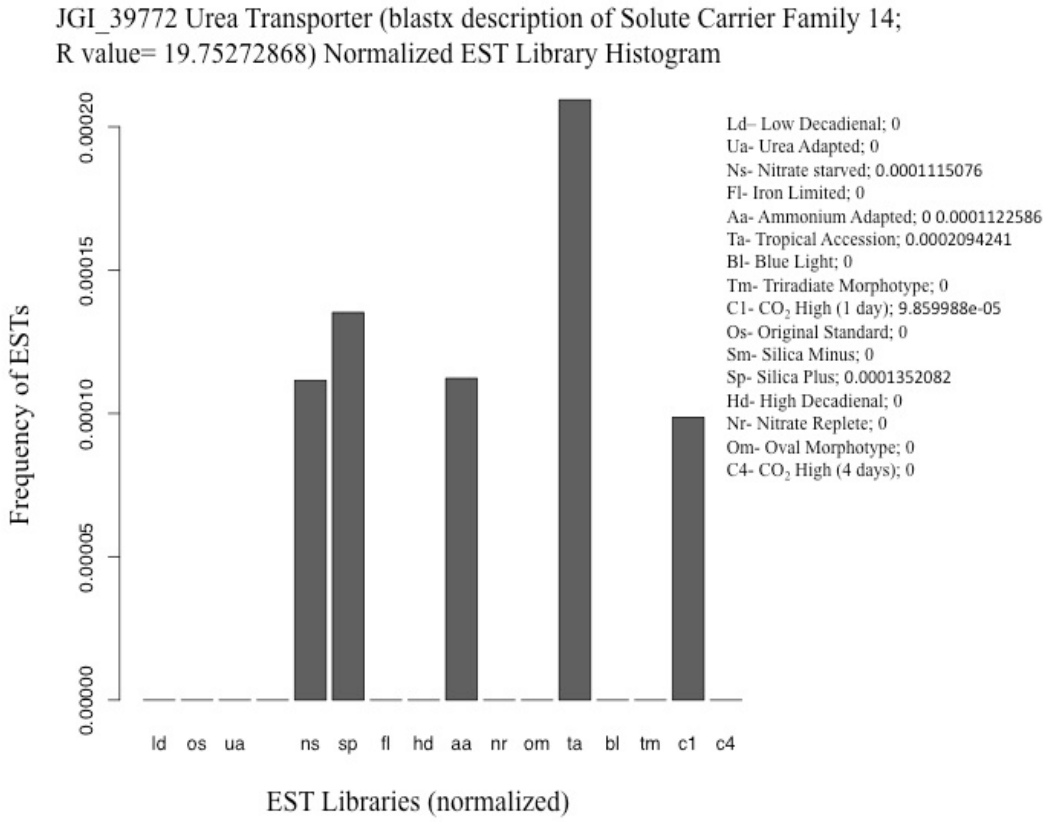


Figure 4. Qualitative Analysis of the EST Library as an Expression Profiling System for JGI_39772 Urea Transporter.

This histogram was generated for JGI_39772, by determining the frequency JGI_39772 in each EST library for the sixteen different culture conditions (data from Maheswari et al, 2010). The data was normalized by dividing the number of ESTs in a particular condition for JGI_39772 by the sum of all ESTs in the libraries for a particular growth condition from all genes in *P. tricornutum*. The R-value, representing differential gene expression in 16 culture conditions, of JGI_39772 is 19.75272868. The x-axis of the plot represents the normalized EST library and the y-axis represents the frequency of the ESTs for a specific gene.

The greatest frequencies of ESTs for JGI_39772 are in the tropical accession (0.0002094241) and ammonium adapted (0.0001122586) conditions, and it has the same EST frequency as JGI_20424 in the silica plus condition (0.0001352082).

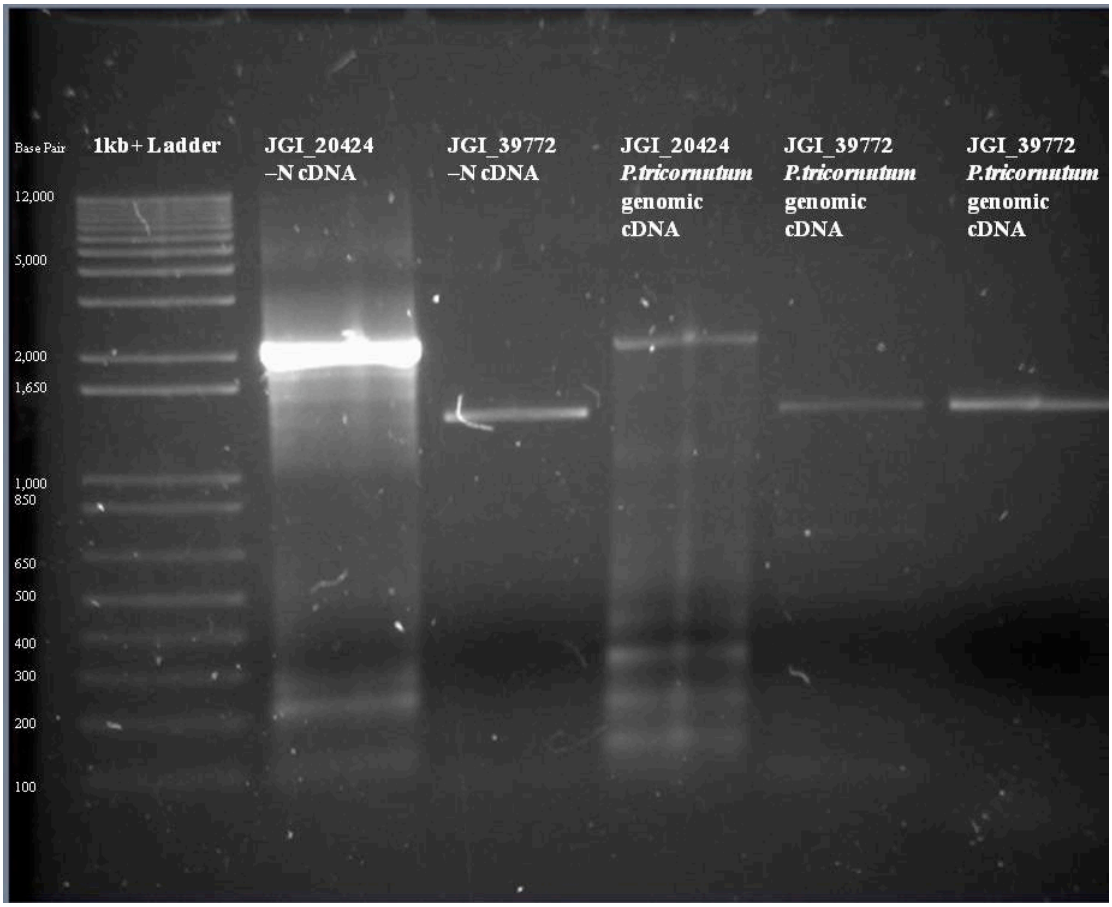


Figure 5. 1% Agarose Gel Electrophoresis to Examine Amplification of Nitrogen-Starved *P.tricornutum* cDNA with Gene-specific Primers by PCR.

Gene-specific primers, JGI_Pt_20424_Forward, JGI_20424_Reverse, JGI_39772_Forward and JGI_39772_Reverse, amplified *P.tricornutum* genomic or nitrogen starved (-N). The amplicons were subjected to 1% agarose gel electrophoresis to determine the size of the respective amplicons. Experimental samples containing amplified –N cDNA by JGI_20424 specific forward and reverse primers yielded a band at 2.1kb. Experimental samples containing amplified –N cDNA by JGI_39772 specific forward and reverse primers yielded a band at 1.3kb. Control samples containing amplified wild-type *P.tricornutum* genomic cDNA by JGI_20424 forward and reverse primers yielded a band at 2.1kb. Control samples containing amplified wild-type *P.tricornutum* genomic cDNA by JGI_39772 forward and reverse primers yielded a band at 1.3kb.

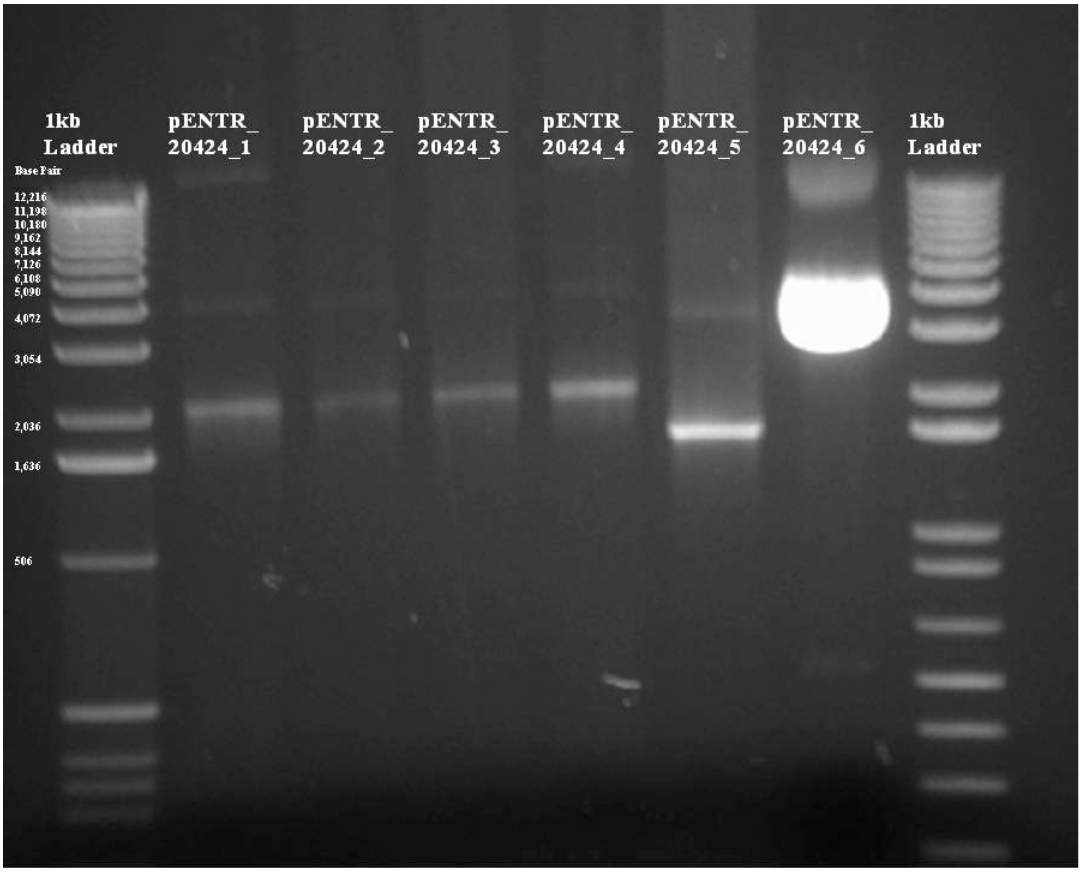


Figure 6. 1% Agarose Gel Electrophoresis to Examine PCR Screen Verification that JGI_20424 Insert Cloned into the pENTR Vector.

The purified, mini-prepped JGI_20424 pENTR plasmids were screened by PCR to ensure that the insert was properly cloned into the pENTR vector. The purified plasmids were amplified with JGI_20424_F and JGI_20424_R gene-specific primers, and they were then separated by 1% agarose gel electrophoresis to determine the insert band sizes. The purified plasmids 20424_1, 20424_2, 20424_3, 20424_4 revealed a band size of 2.1kbp. The JGI_20424 insert was properly cloned into the pENTR vector.

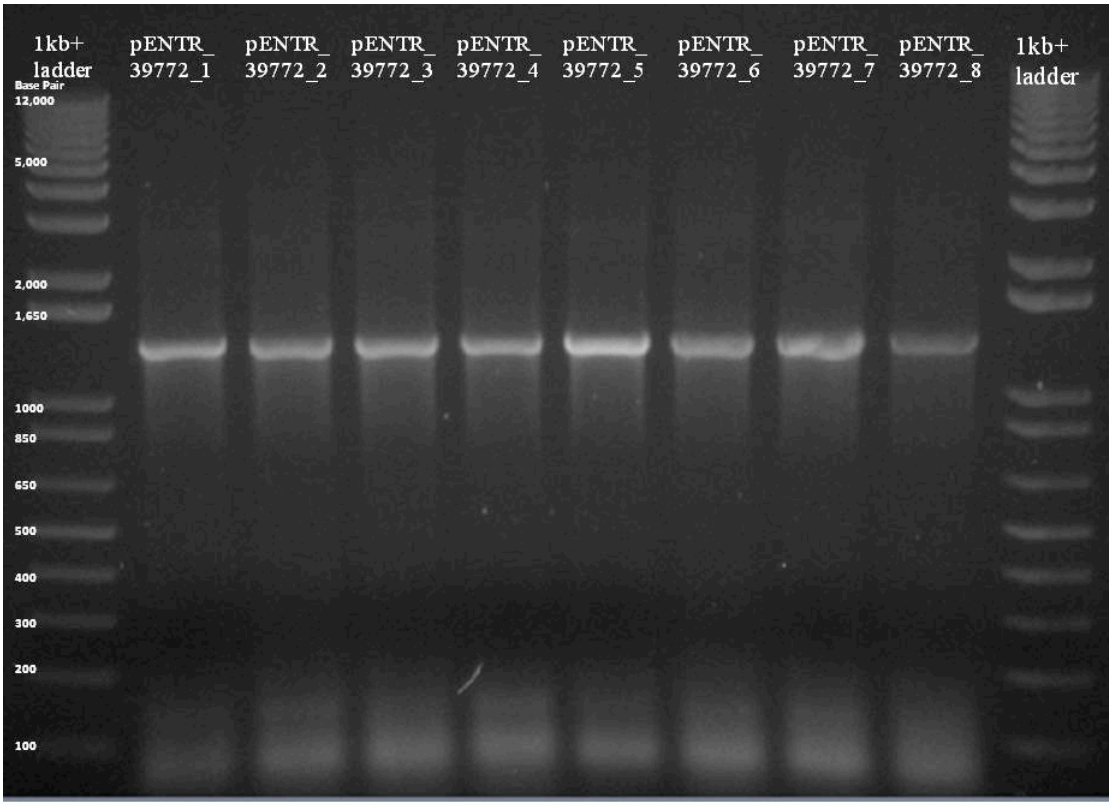


Figure 7. 1% Agarose Gel Electrophoresis to Examine PCR Screen Verification that JGI_39772 Insert Cloned into the pENTR Vector.

The purified, mini-prepped JGI_39772 pENTR plasmids were screened by PCR to ensure that the insert was properly cloned into the pENTR vector. The purified plasmids were amplified with JGI39772_F and JGI39772_R gene-specific primers, and they were then separated by 1% agarose gel electrophoresis to determine the insert band sizes. The purified plasmids 39772_1, 39772_2, 39772_ 3, 39772_ 4, 39772_ 5, 39772_ 6, 39772_7, 39772_8 revealed a band size of 1.3kbp. The JGI_39772 insert was properly cloned into the pENTR vector.

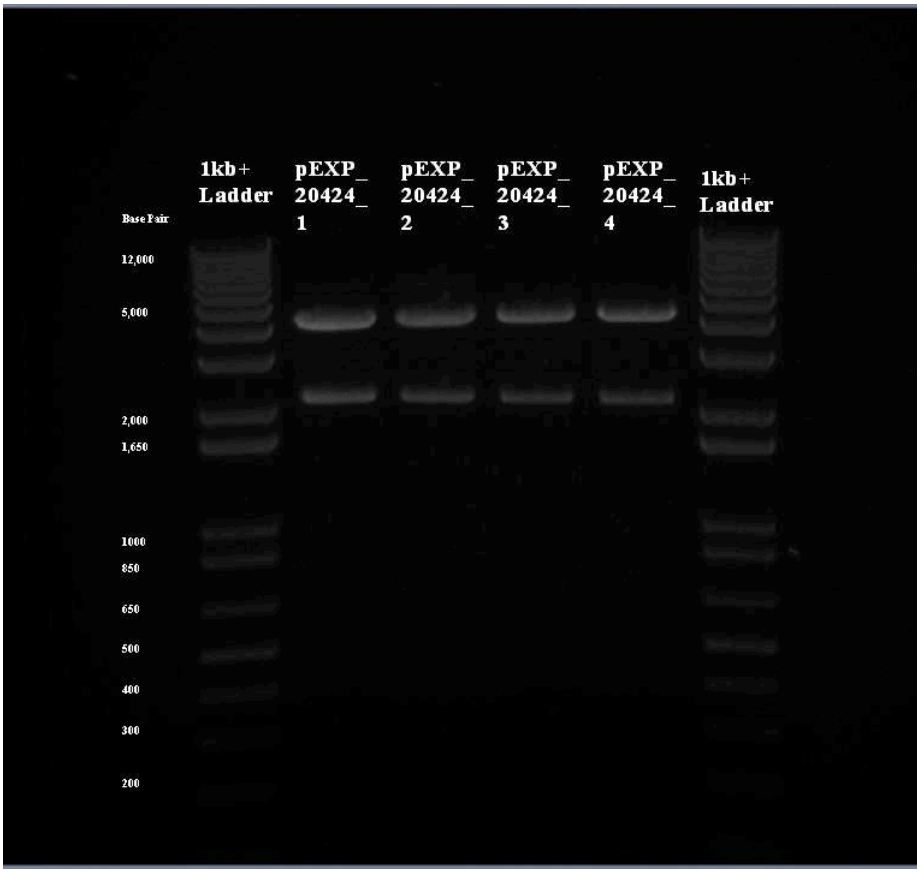


Figure 8. 1% Agarose Gel Electrophoresis to Examine Restriction Digest of Purified JGI_20424 pEXP Vectors (Determine that JGI_20424 pENTR recombined with pDEST_C'CFP_GUS Vector).

The purified, mini-prepped pEXP 20424_4 was digested with restriction enzyme *Kpn1*. The expected band sizes resulted on the 1% agarose gel electrophoresis unit. The bands were at 4.325kbp and 2.172kbp, as expected from *in silico* restriction digest analysis. These results indicate that the LR recombination reaction inserted the JGI_20424 into the pDEST_C'CFP_GUS vector efficiently, to create a pEXP vector.

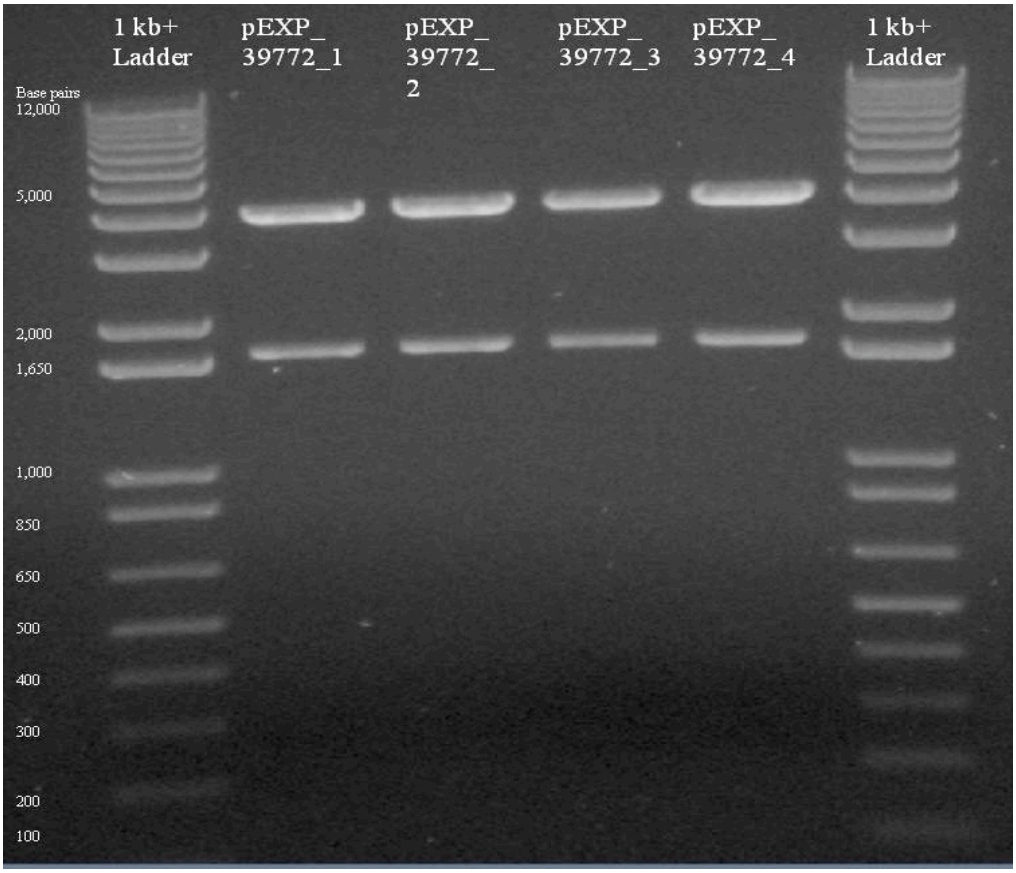


Figure 9. 1% Agarose Gel Electrophoresis to Examine Restriction Digest Purified JGI_39772 pEXP Vectors (Determine that JGI_39772 pENTR Recombined with pDEST_C'CFP_GUS vector).

The purified, mini-prepped pEXP 39772_7 digested with restriction enzyme *EcoR1*. The expected band sizes resulted on the 1% agarose gel electrophoresis unit. The bands were at 4.030kbp and 1.711kbp, as expected from *in silico* restriction digest analysis. These results indicate that the LR recombination reaction inserted the JGI_39772 into the pDEST_C'CFP_GUS vector efficiently, to create a pEXP vector.

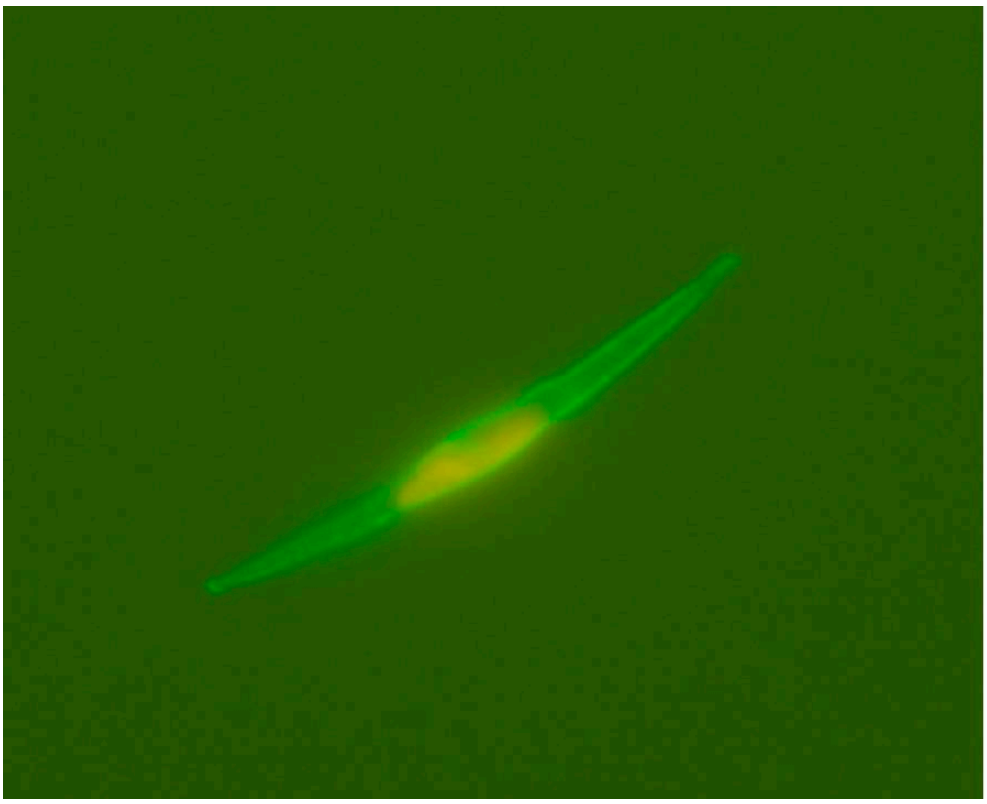


Figure 10. Localization of JGI_20424 C'YFP Overexpression *P. tricornutum* Construct to the Plasma Membrane.

The JGI_20424 C'YFP fusion overexpressing construct, an orthologue of plant-type high affinity urea transporter localized to the outer plasma membrane in *P. tricornutum* (above). The differential interference contrast microscopy (DIC) image represents the bright field image of JGI_20424 C'YFP fusion overexpressing construct (below). The red color represents the chlorophyll auto-fluorescence that localizes to the plastid, while the green color represents the YFP fluorescence. Scale bar represents 20μm.

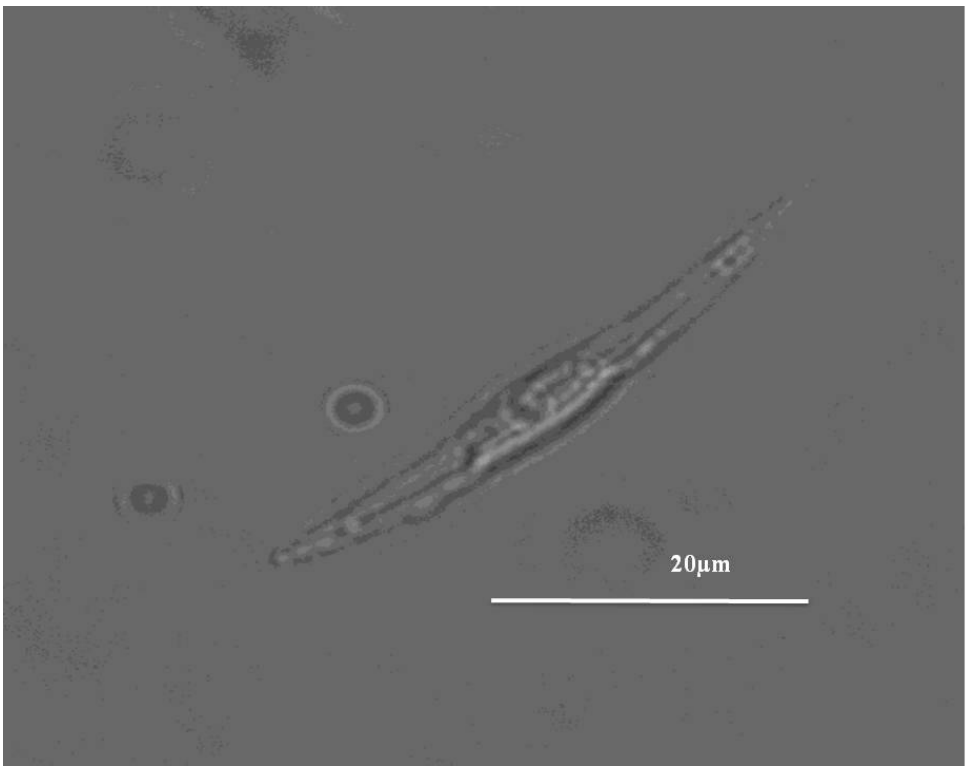


Figure 10 (continued)

Figure 11. Subcellular Localization of JGI_39772 C'YFP Overexpression *P. tricornutum* Construct to the Mitochondria.

The JGI_39772 C'YFP fusion overexpressor construct, a homologue to metazoan-type renal and erythrocyte transporters, localized to the mitochondria in *P. tricornutum* (above). The differential interference contrast microscopy (DIC) image is the bright field image of JGI_39772 C'YFP fusion overexpressing construct (below). This localization to the mitochondria confirms the *TargetP* prediction score of 0.815 for the subcellular localization of JGI_39772 to the mitochondria. The red color represents the chlorophyll auto-fluorescence that localizes the plastid, while the green color represents YFP fluorescence.

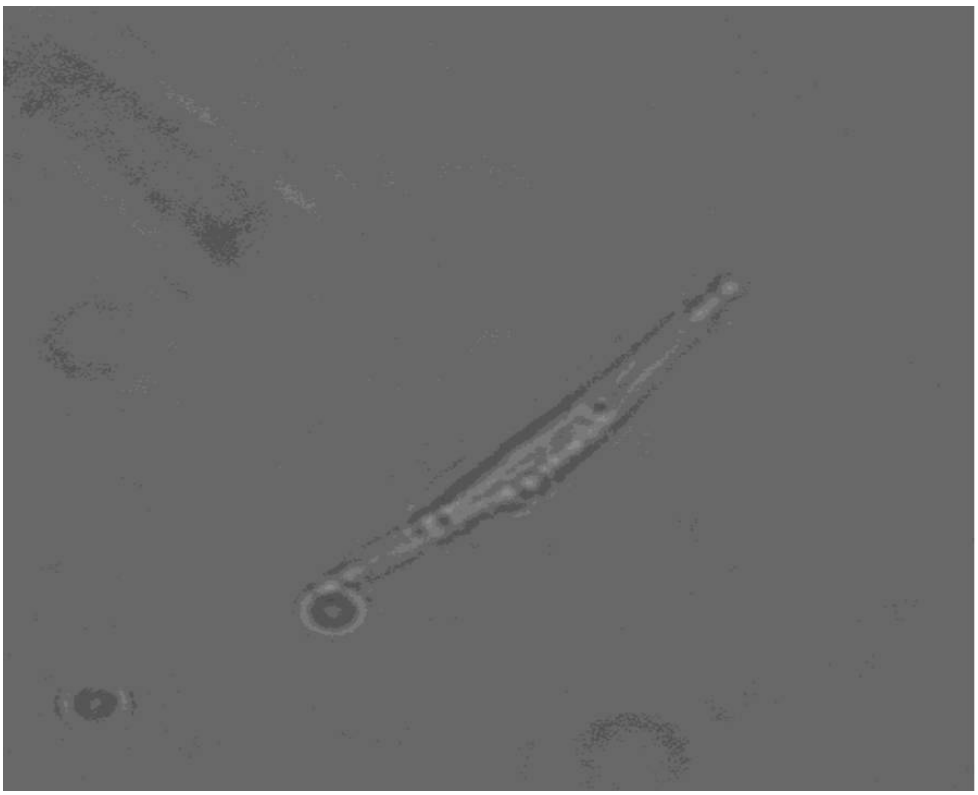


Figure 11 (continued)

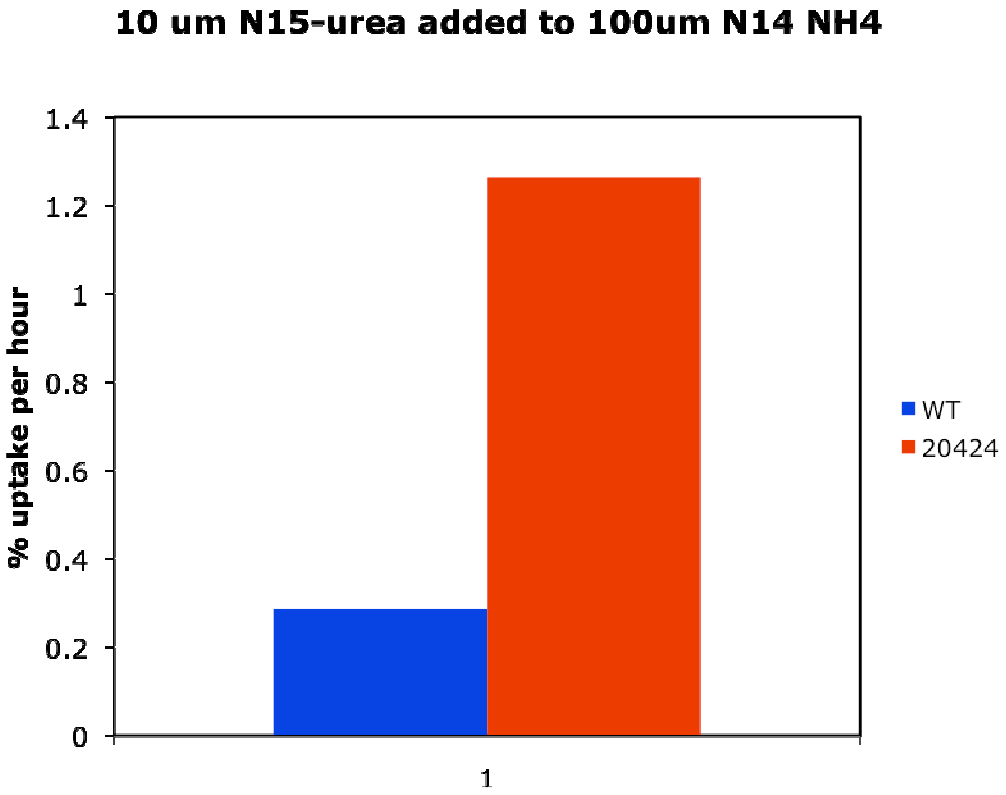


Figure 12. Graph of Percent ^{15}N-labeled Urea Uptake by the Wild-type and JGI_20424 overexpressing *P. tricornutum* cultures.

The percent of ^{15}N urea uptake by the wild-type *P.tricornutum* and JGI_20424 transgenic *P. tricornutum* cultures grown on ammonium indicate that the overexpression of *P. tricornutum* JGI_20424 urea transporter leads to an approximate ten-fold increase in urea uptake; therefore, the protein efficiently transports urea. Urea uptake is relatively low in wild-type *P. tricornutum* cells grown on ammonia.

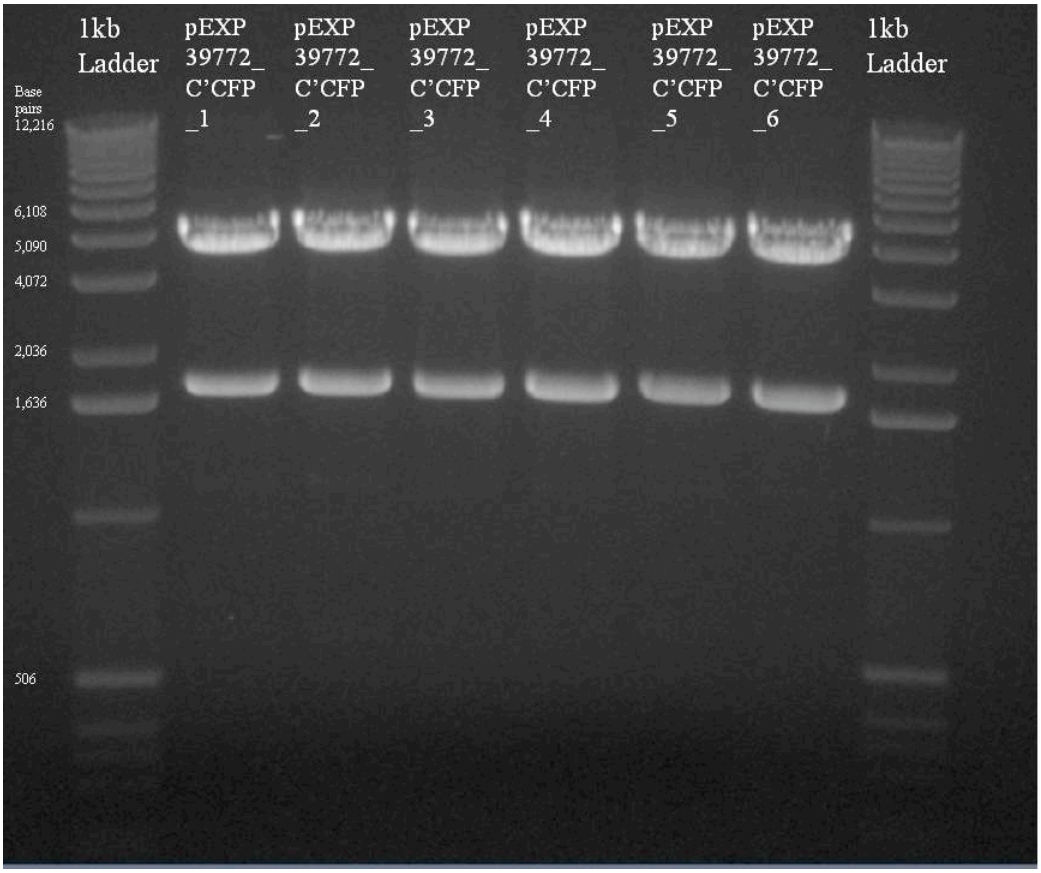


Figure 13. 1% Agarose Gel Electrophoresis to Examine Restriction Digest of Purified pEXP_39772_C'CFP (Determine that JGI_39772 Inserted Into the pDEST_C'CFP_GUS Vector).

The purified, mini-prepped pEXP_39772_C'CFP plasmids were digested with restriction enzyme *Nae1* to ensure that the JGI_39772 inserted correctly into the pDEST_C'CFP_GUS vector. The expected band sizes, generated from *in silico* analysis, resulted on the 1% agarose gel electrophoresis unit. The band sizes were 3.818kbp and 1.923kbp, as expected. Purified pEXP_39772_C'CFP digested with restriction enzyme *Nae1* demonstrated that JGI_39772 inserted correctly into the pDEST_C'CFP_GUS vector.

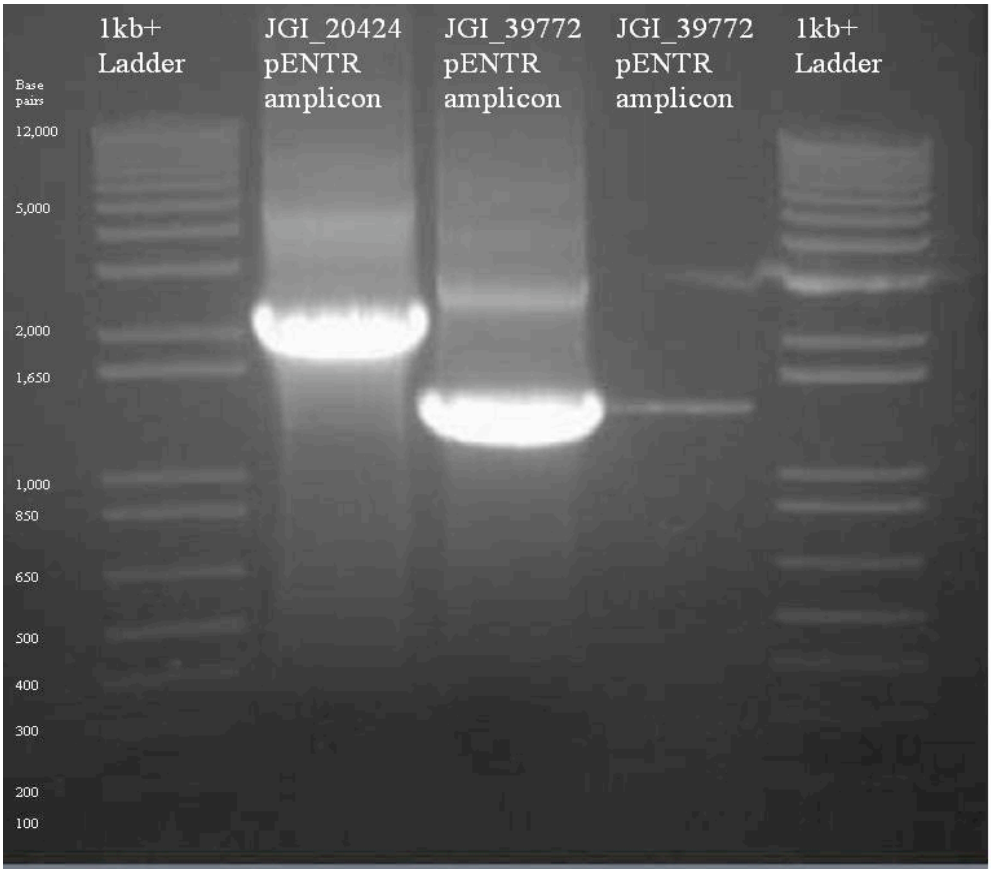


Figure 14. 1% Agarose Gel Electrophoresis to Examine JGI_20424 pENTR and JGI_39772 pENTR Plasmids Amplified by Gene-specific Primers Containing Restriction Enzyme Sites.

JGI_20424 pENTR and JGI_39772 pENTR plasmids amplified to correct sizes of 2.1kbp and 1.3kbp, respectively, following amplification with gene-specific primers containing restriction enzyme sequences. The 20424_F_EcoI and 20424_R_Xba1 primers amplified the original JGI_20424_4_pENTR vector, and the 39772_F_EcoI and 39772_R_Xba1 primers amplified the original JGI_39772_7_pENTR vector. The two PCR amplicons were applied to 1% agarose gel electrophoresis to determine the size of the products. The JGI_20424_4_pENTR vector was amplified with gene-specific primers including restriction enzyme sites to yield a band at 2.1kbp, and JGI_39772_7_pENTR vector was amplified with gene-specific primers including restriction enzyme sites to yield a band at 1.3kbp.

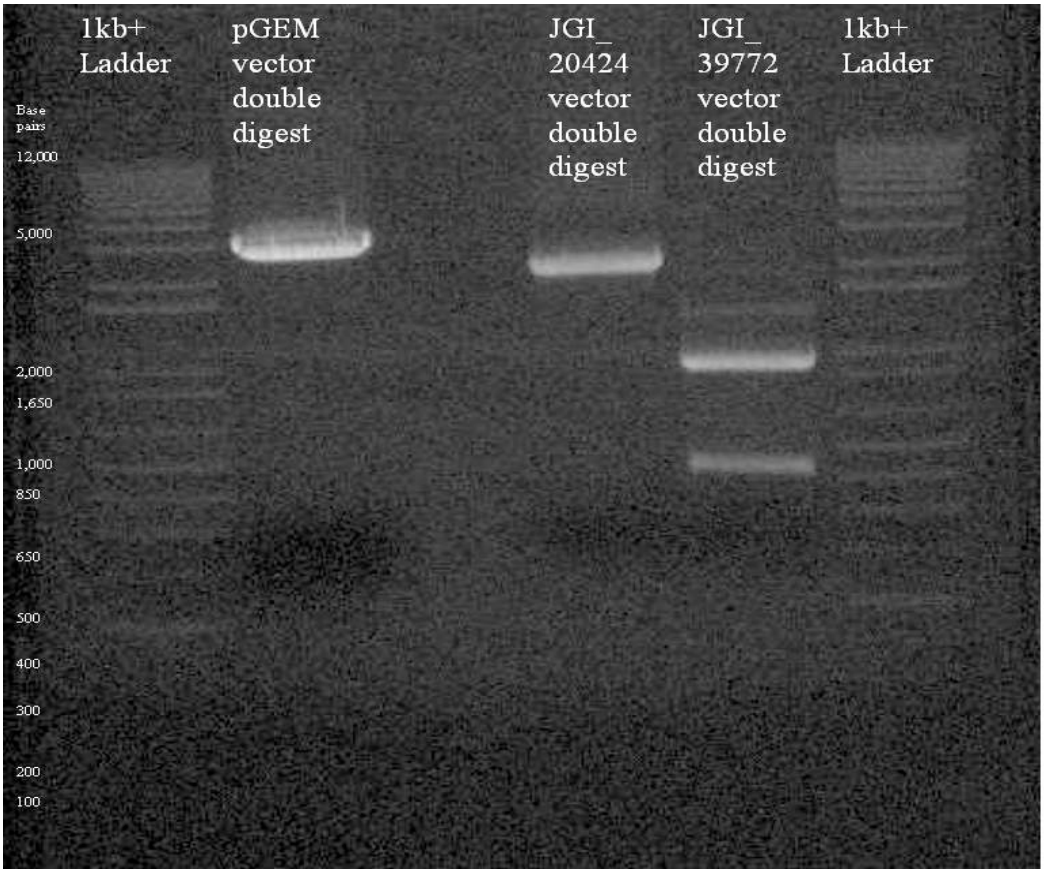


Figure 15. 1% Agarose Gel Electrophoresis to Examine Restriction Double-Digest of pGEM Vector and Gel-purified JGI_20424 and JGI_39772 Amplicons with *EcoRI* and *XbaI*.

A restriction double digest of the pGEM vector (3μg) and the gel-purified JGI_20424 and JGI_39772 PCR products (20μL total) was performed using *EcoRI* and *XbaI* restriction enzymes, 1X BSA and 1X CA buffer. The bands that resulted from the *EcoRI* and *XbaI* digested JGI_20424, JGI_39772 and pGEM vector were the correct size based on *in silico* analysis of the restriction digest reaction. The double-digested JGI_20424 band size was 2.121kbp on the 1% agarose gel electrophoresis. The double-digested JGI_39772 band sizes were at 949bp and 419bp, whose sum is the expected 1.4kbp, and there was a very faint band at 1.4kbp. The double-digested pGEM vector band size was 2.945kbp.

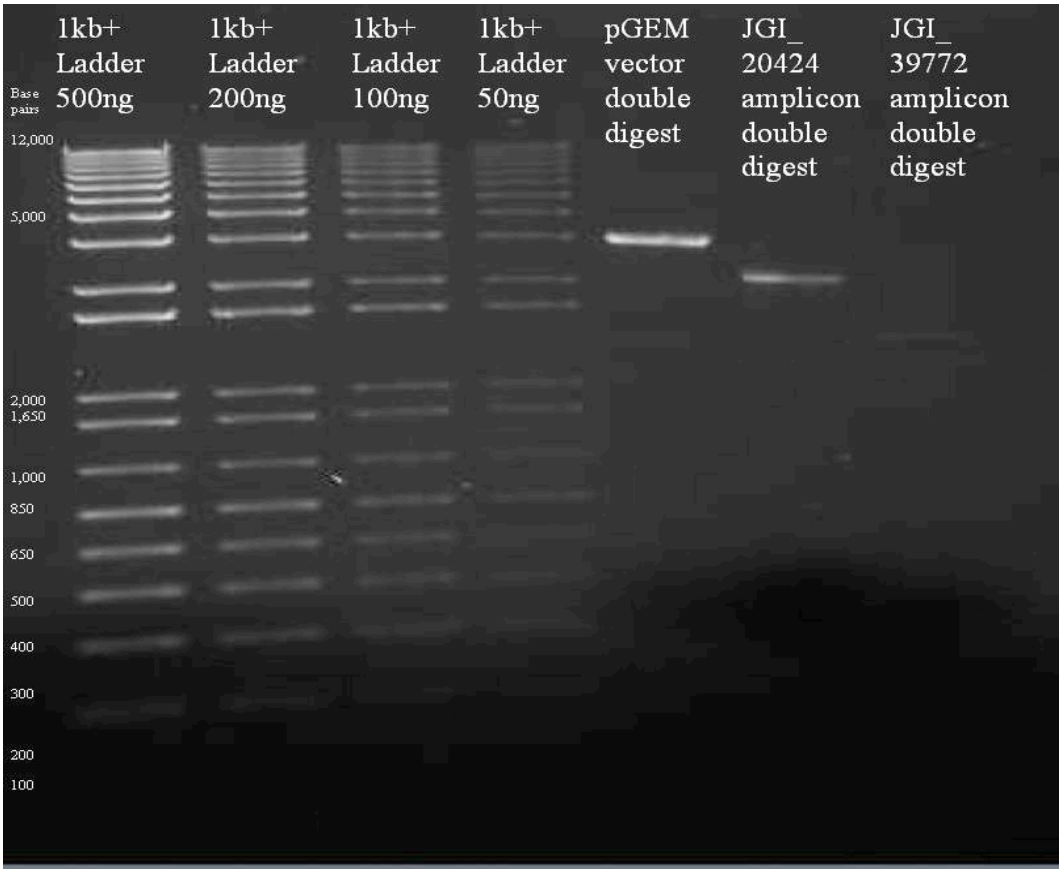


Figure 16. 1% Agarose Gel Electrophoresis to Determine Relative concentrations of Double-Digested pGEM vector and JGI_20424 and JGI_39772 Amplicons With a Dilution Series of the 1kb+ Ladder.

The PCR amplified JGI_20424 and PCR amplified JGI_39772 that had been digested with *EcoRI* and *XbaI* restriction enzymes were applied to a 1% agarose gel to determine the respective sizes of the digested products. 2μL of each purified product was run on the 1% agarose gel alongside a 1kb+ DNA ladder that contained a dilution series of 500ng, 200ng, 100ng and 50ng, respectively. Comparison of the intensity of the products' bands to that of the ladder's bands in the dilution series determined the relative concentration of each product. The concentration of the digested pGEM vector was approximately 600ng/μL, the concentration of double digested JGI_20424 was approximately 300ng/μL, and the concentration of double-digested JGI_39772 was approximately 40ng/μL.

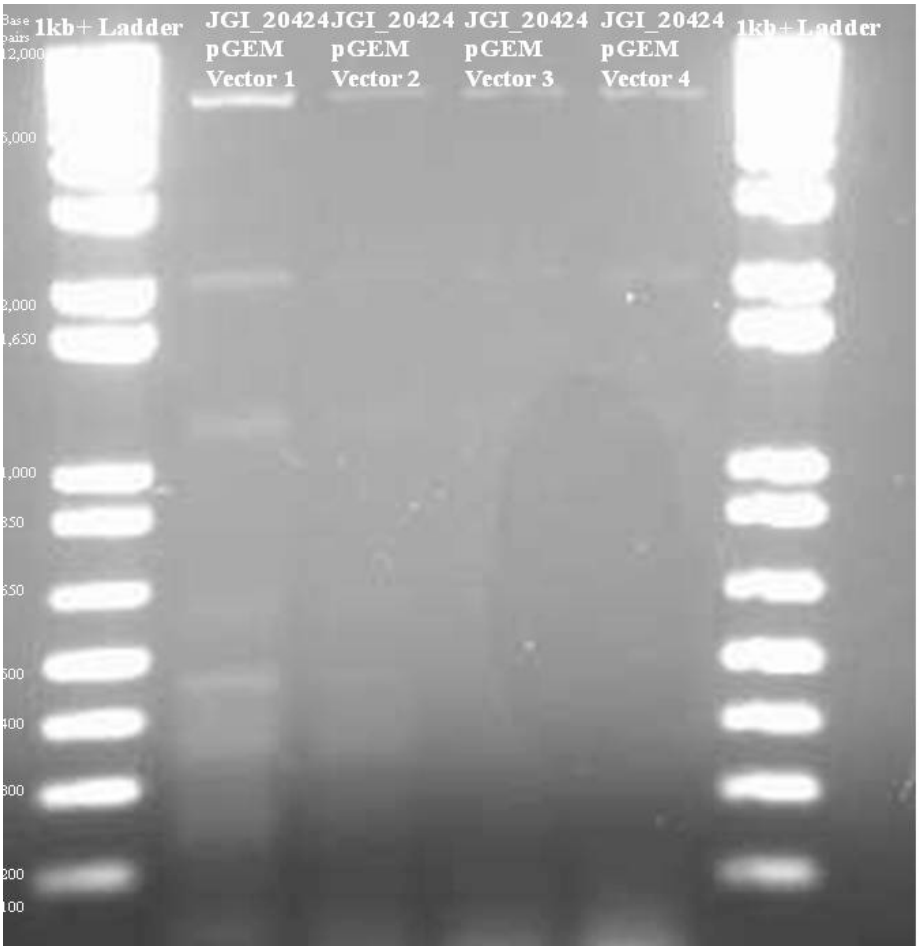


Figure 17. 1% Agarose Gel Electrophoresis to Determine that JGI_20424 Insert Ligated to the pGEM Expression Vector.

The JGI_20424 insert was properly ligated to the pGEM expression vector. The purified, mini-prepped JGI_20424 pGEM expression vectors were subjected to a PCR screen to ensure that JGI_20424 properly ligated into the pGEM vector following double restriction digestion. The purified plasmids were amplified with JGI_20424_Eco1 and JGI_20424_Xba1 gene-specific primers. They were then separated by 1% agarose gel electrophoresis to determine the insert band sizes. The purified plasmids 20424_pGEM_1 band size was 2.1kbp. 20424_pGEM_2, 20424_pGEM_3 and 20424_pGEM_4 were all 2.1kbp bands as well, although their bands are much less intense. The JGI_20424 insert was properly ligated to the pGEM expression vector.

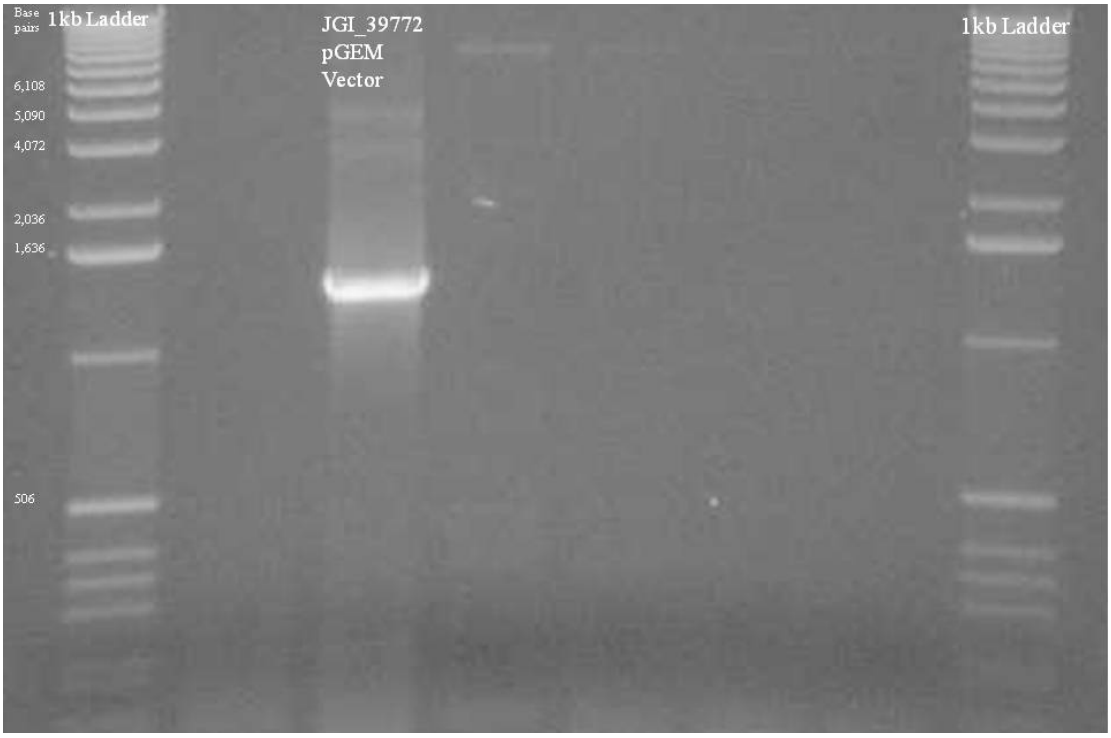


Figure 18. 1% Agarose Gel Electrophoresis to Determine that JGI_39772 Insert Ligated to the pGEM Expression Vector.

The JGI_39772 insert was properly ligated to the pGEM expression vector. The purified, mini-prepped JGI_39772 pGEM expression vectors were subjected to a PCR screen to ensure that JGI_39772 properly ligated into the pGEM vector following double restriction digestion. The purified plasmids were amplified with JGI_39772_Eco1 and JGI_39772_Xba1 gene-specific primers. They were then separated by 1% agarose gel electrophoresis to determine the insert band sizes. The purified plasmids 39772_pGEM_2 band size was 1.3kbp. The JGI_39772 insert was properly ligated to the pGEM expression vector.

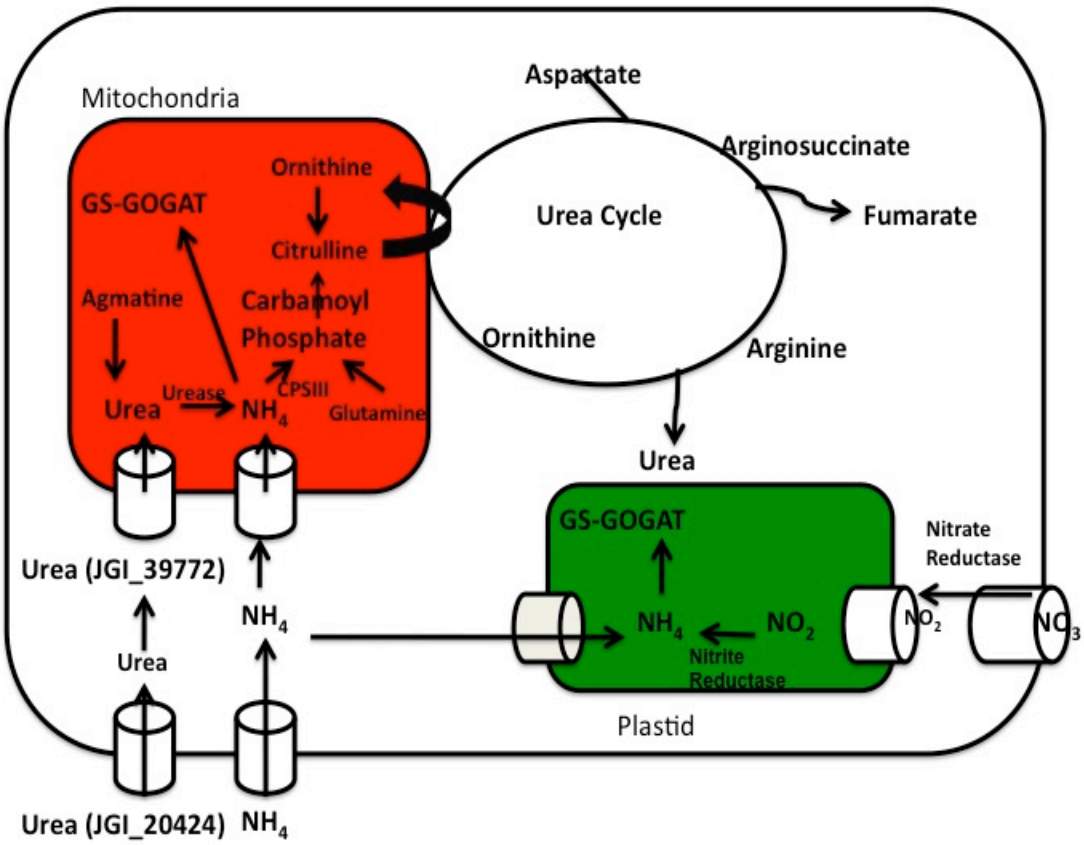


Figure 19. A Hypothetical Model of Nitrogen Assimilation Pathways in Marine Diatoms.

The simplified hypothetical model of the nitrogen assimilation pathways in marine diatoms includes the subcellular localization of transporters, various enzymes and substrates. JGI_Pt_20424 urea transporter is localized to the plasma membrane to transport exogenous urea into *P. tricornutum*, and the JGI_Pt_39772 urea transporter transports urea into the mitochondria. Nitrogen assimilation pathways, such as the GS-GOGAT cycle and the urea cycle, are noted in the model to indicate the manner with which urea is assimilated, redistributed and stored.

Table 1 A. Tables for Growth on 100μM Ammonium Seawater Media and 10μM ^{15}N-Labeled Urea Uptake Assay.

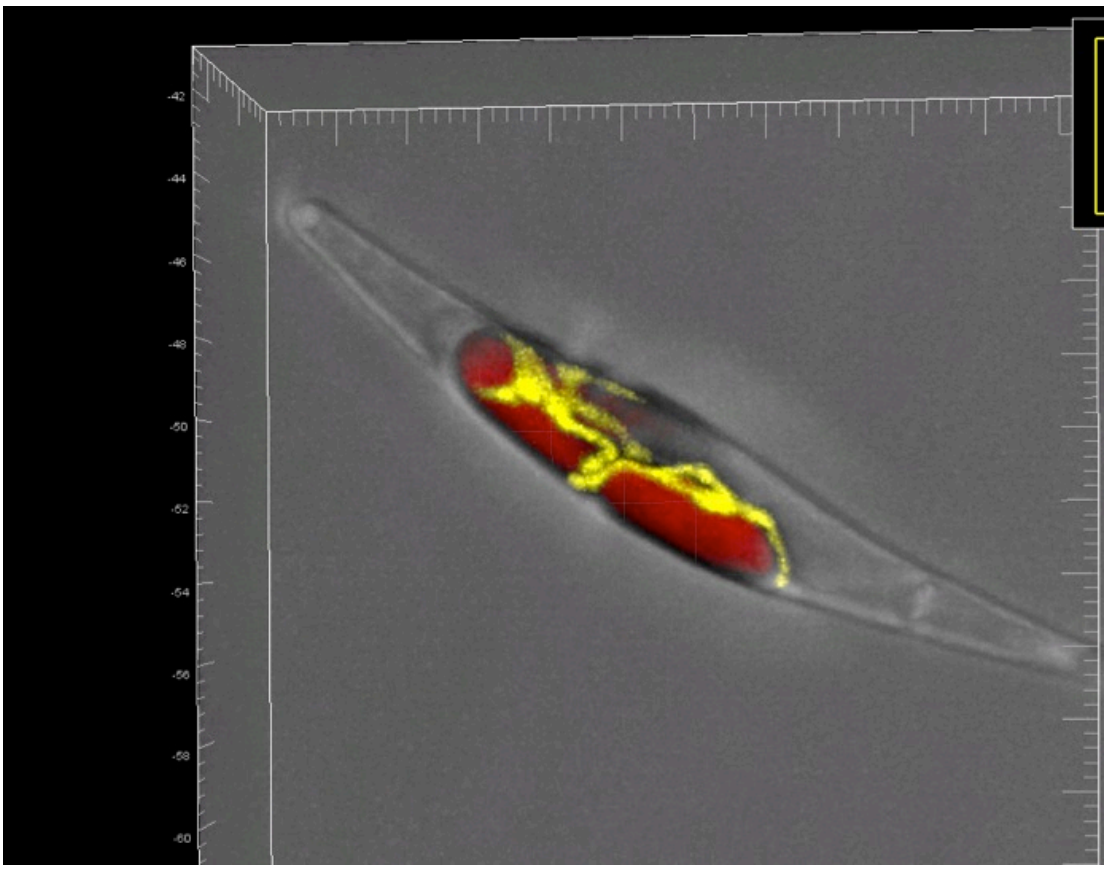
The natural abundance of ^{15}N is 0.366%; therefore, the ammonia in the media in which the cultures were grown contained 0.366% ^{15}N, while the added urea was 99% ^{15}N. The percent of ^{15}N of the particulate nitrogen reflects uptake of the ^{15}N-labeled urea. The percent of ^{15}N was measured in each of the four *P. tricornutum* cultures grown on 100 μM ammonium (NH_4) seawater. The first sample of wild-type *P. tricornutum* culture with a biomass of 40.1μg contained 0.41217 percent ^{15}N, while the second wild-type sample of biomass 40.3μg contained 0.41513 percent ^{15}N. The first sample of the *P. tricornutum* JGI_20424 overexpression culture of biomass 11.5μg contained 0.58888 percent ^{15}N, while the second overexpression culture of biomass 12.5μg contained 0.56451 percent ^{15}N.

P. tricornutum culture grown on 100 μM Ammonia (NH_4) Seawater	Biomass	Percent ^{15}N of the Sample
Wild-Type *P. tricornutum* Sample 1	40.1	0.41217
Wild-Type *P. tricornutum* Sample 2	40.3	0.41513
P. tricornutum JGI_20424 Overexpression Sample 1	11.5	0.58888
P. tricornutum JGI_20424 Overexpression Sample 2	12.5	0.56451

Table 1B. Table for Growth on 100μM Ammonium Seawater Media and 10μM ^{15}N-Labeled Urea Uptake Assay.

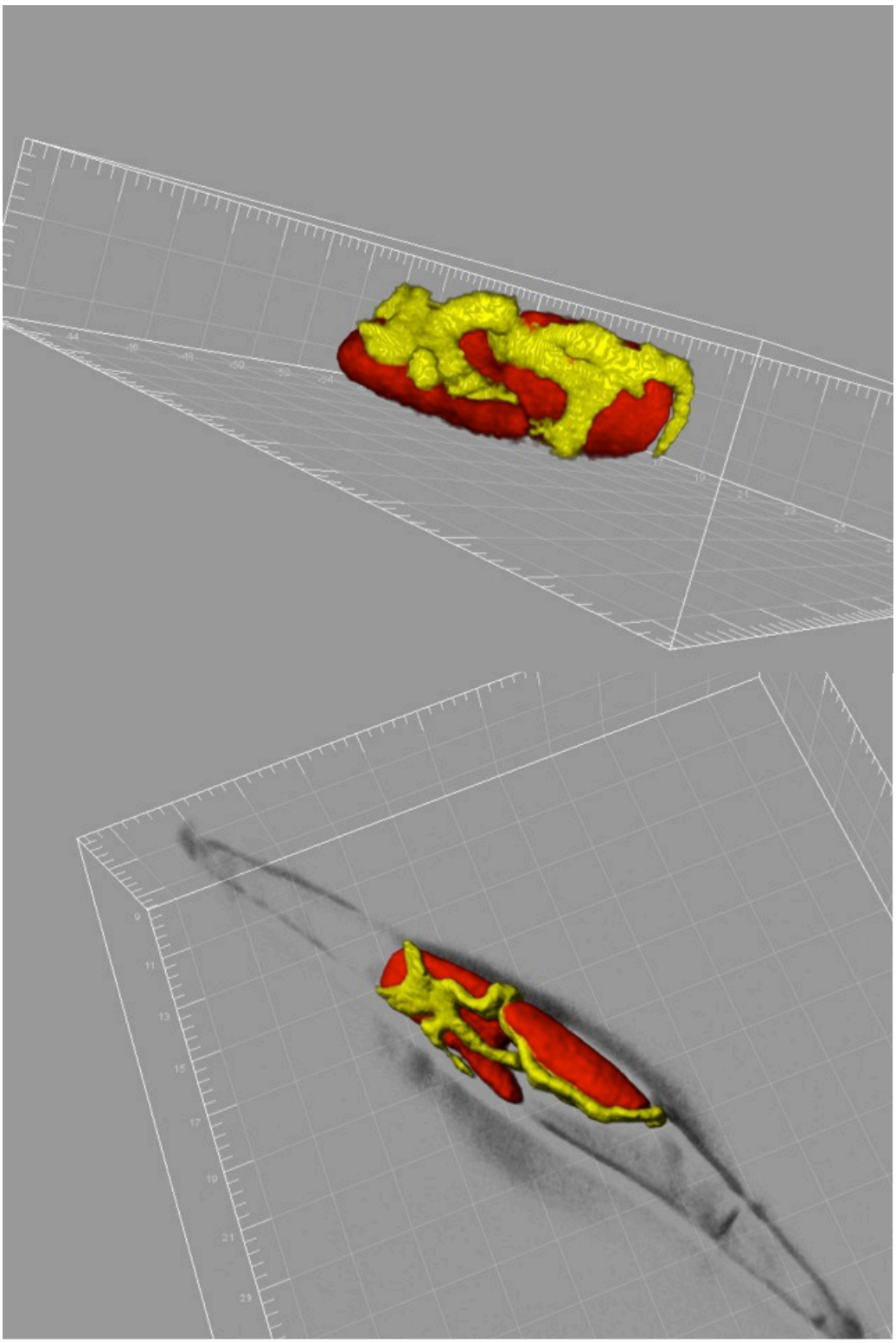
Particulate nitrogen of the wild-type ammonium-grown *P. tricornutum* cultures was 1005.077822; therefore, the percent of ^{15}N-labeled urea uptake per hour was 0.286 percent. The particulate nitrogen of JGI_20424 overexpression *P. tricornutum* culture was 300.0228453; therefore, the percent of ^{15}N-labeled urea per hour was 1.26 percent.

P. tricornutum culture grown on 100 μM Ammonia (NH_4) Seawater	Particulate nitrogen	Percent ^{15}N uptake (per hour)
Wild-Type *P. tricornutum*	1005.077822	0.28588343
P. tricornutum JGI_20424 Overexpression	300.0228453	1.264183826



Supplementary Figure 1. Three-Dimensional Image of Overexpressed JGI_39772 Localized to the Mitochondria from Confocal Fluorescence Microscopy.

Images were generated by Dr. Christopher Dupont. The three-dimensional images of JGI_39772 were obtained from confocal fluorescence microscopy. These pictures were manipulated in Microsoft Powerpoint to create the images in the supplemental figure. Strong mitochondrial localization is confirmed in all three images. The first is a top view, while the second image in the figure is a side view. The third image is a comprehensive and high-resolution three-dimensional image of the mitochondrial localization of the overexpressed JGI_39772 urea transporter in *P. tricornutum*.



Supplementary Figure 1 (continued)

References

Allen, A.E., Song, B., Ward, B.B. (2005). "Characterization of Diatom (Bacillariophyceae) Nitrate Reductase Genes and Their Detection in Marine Phytoplankton Communities." *Journal of Phycology* 41(1):95-104.

Allen, A.E., Vardi, A., Bowler, C. (2006). "An Ecological and Evolutionary Context for Integrated Nitrogen Metabolism and Related Signaling Pathways in Marine Diatoms." *Current Opinion in Plant Biology* 9: 264-273.

Apt, K.E., Korth-Pancic, P.G., Grossman, A.R. (1996). "Stable Transformation of the Diatom *Phaeodactylum tricornutum." Molecular Genetics and Genomics* 252:572-579.

Armbrust, E.V., Berges, J.A., Bowler, C., Green, B.R. (2004). "The genome of the Diatom *Thalassiosira pseudonana*: Ecology, Evolution and Metabolism." *Science*, Vol 306:79-86.

Bachvaroff, T.R., Sanchez Puerta, M.V., Delwiche, C.F. (2005). "Chlorophyll C-Containing Plastid Relationships Based on Analyses of a Multigene Data Set with All Four Chromalveolate Lineages." *Molecular Biology and Evolution* (9):172-82.

Berges, J.A., Gibson, C.E., Stewart. B.M. (2004). "Physiological Responses of Phytoplankton Communities in the Irish Sea to Simulated Upwelling." *Hydrobiologia* 517:121-132.

Berman, T., Bronk D. A. (2003). "Dissolved Organic Nitrogen: a Dynamic Participant in Aquatic Ecosystems." *Aquatic Microbial Ecology* 31: 273-305.

Bowler, C., Allen, A.E., Badger, J.H., Grimwood, J. (2008). "The *Phaeodactylum* genome reveals the Evolutionary History of Diatom Genes." *Nature* 456:239-244.

DOE Joint Genome Institute (1997). "Joint Genome Institute *Phaeodactylum tricornutum* Genome Sequence." The Regents of the University of California. <http://genome.jgi-psf.org/Phatr2/Phatr2.home.html>.

Dupont, C., Allen, A.E., et al (2010), "Nitrogen Assimilation in Marine Diatoms." Unpublished, National Science Foundation Grant.

Faye, L, Greenwood, J.S., Chrispeels, M.J. (1986) "Urease in Jack Bean *(Canavalia ensiformis* DC) Aeeds is a Cytosolic Protein." *Planta* 168:579-585.

Falkowski P.G., Katz M.E., Knoll A.H., Quigg A., Raven J.A., Schofield O., Taylor, F.J.R. (2004). "The Evolution of Modern Eukaryotic Phytoplankton." *Science* 305:354–360.

Gilbert, P.M. (2008). "Ocean Urea Fertilizer for Carbon Credits Poses High Ecological Risks." Marine Pollution Bulletin 56(5):1049-1056.

Kröger, N., Deutzmann, R., Bergsdorf, C., Sumper, M. (2000). "Species-Specific Polyamines from Diatoms Control Silica Morphology." *Proceedings of the National Academy of Sciences* 97:14133–14138.

Levin, E.J., Quick, M., Zhou, M. (2009). "Crystal Structure of a Bacterial Homologue of the Kidney Urea Transporter." *Nature* 462:757-761.

Maheswari, U., Jabbari, K., Petit, J.L., Porcel, B.M., Allen, A.E., Cadoret, J.P., Martino, A.D., Heijde, M., La Roche, J., Lopez, P.J., Martin-Jezequel, V., Mock, T., Parker, M.S., Vardi, A., Armbrust, E.V., Weissenbach, J., Katinka, M., Bowler, C. (2010). "Digital Expression Profiling of Novel Diatom Transcripts Provides Insight Into Their Biological Functions."

Maheswari, U., Mock, T.E., Armbrust, V., Bowler, C. (2008). "Update of the Diatom EST Database: a New Tool for Digital Transcriptomics." *Nucleic Acids Research* 37:D1001-D1005.

Maheswari, U., Monstant, A., Jabbari, K., Bowler, C. (2005). "Comparative Genomics of Pennate Diatom *Phaeodactylum tricornutum*." *Plant Physiology* 137:500-513.

McCarthy, J.J. (1972). "Uptake of Urea by Natural Populations of Marine Phytoplankton." *Limnology and Oceanography* 17:738-748.

Minocha, R., Studley, K., Saier, M.H. (2003). "The Urea Transporter (UT) Family: Bioinformatic Analyses Leading to Structural, Functional, and Evolutionary Predictions." *Receptors & Channels* 9: 345-352.

Moustafa, A., Beszteri, B., Maier, U.G., Bowler, C., Valentin, K., Bhattacharya, D. (2009). "Genomic Footprints of a Cryptic Plastid Endosymbiosis in Diatoms." *Science* 324:1724-1726.

Moustafa, A., Chan, C.X., Danforth, M., Zear, D., Ahmed, H., Jadhav, N., Savage, T., Bhattacharya, D. (2008). "A Phylogenomic Approach for Studying Plastid Endosymbiosis." *Genome Informatics*. 21:165-176.

Nelson, D.M., Treguer, P., Brzezinski, M.A., Leynaert, A., Queguiner, B., (1995). "Production and Dissolution of Biogenic Silica in the Ocean-Revised Global Estimates, Comparison with Regional Data and Relationship to Biogenic Sedimentation." *Global Biogeochemical Cycle* 9, 359-372.

Parker, M.S., Mock, T., Armbrust, E.V. (2008). "Genomic Insights into Marine Microalgae." *Annual Review of Genetics*, Vol 42: 619-645.

Peers, G.S, Milligan A.J., Harrison, P.J. (2000). “Assay Optimization and Regulation of Urease Activity in Two Marine Diatoms.” *Journal of Phycology* 36:523–528.

Poulsen, N., Chesley, P.M., Kroger, N. (2006). “Molecular Genetic Manipulation of the Diatom *Thalassiosira pseudonana* (Bacillariophyceae).” *Journal of Phycology* 42:1059–1065.

Roessler, P.G., Brown, L.M., Dunahay, T.G., Heacox, D.A., Jarvis, E.E, Schneider, J.C., Talbot, S.G., Zeiler, K.G. (1994). "Genetic Engineering Approaches for Enhanced Production of Biodiesel Fuel from Microalgae." *ACS Symposium Series*. 566:255-267.

Saier, M.H. "Transporter Classification Database". University of California, San Diego. <http://www.tcdb.org/>.

Sanchez-Puerta M.V., Delwiche CF (2008). “A hypothesis for plastid evolution in chromalveolates.” *Journal of Phycology* 44:1097-1107.

Schroeder, J.I. (1994). “Heterologous Expression and Functional Analysis of High Plant Transport Proteins in Xenopus Oocytes.” *Methods* 6:70-81.

Siaut, M., Heijde, M., Mangogna, M., Montsant, A., Coesel, S., Allen, A.E., Manfredonia, A., Falciatore, A., Bowler, C. (2007). “Molecular Toolbox for Studying Diatom Biology in *Phaeodactylum tricornutum*.” *Gene* 406:23-35.

Vardi, A. (2008). “Cell Signaling in Marine Diatoms.” *Communicative and Integrative Biology* 1(2):134-136.

Wafar, M.V.M., Le Corre, P., L'Helguen, S. (1995) “f-Ratios Calculated With and Without Urea Uptake in Total Nitrogen Uptake by Phytoplankton.” *Deep-Sea Research Part I*, 42:1669-1674.

Worden, A.Z. et al (2009). “Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes Micromonas.” *Science* 324:268-272.

Yoon, H.S., Hackett, Y.D., Ciniglia, C., Pinto, G., Bhattacharya, D. (2004b). “A Molecular Timeline for the Origin of Photosynthetic Eukaryotes.” *Molecular Biology and Evolution* 21:809-818.

Zimmerman, R., Kremer, J.N., Dugdale, R.C. (1987). “Acceleration of Nutrient Uptake by Phytoplankton in a Coastal Upwelling Ecosystem: A Modeling Analysis.” *Limnology and Oceanography* 32(2):359-367.