METABOLIC NETWORK CONSTRUCTION BASED ON THE GENOME OF THE MARINE DIATOM *THALASSIOSIRA PSEUDONANA* AND THE ANALYSIS OF GENOME-WIDE TRANSCRIPTOME DATA TO INVESTIGATE TRIACYLGLYCERIDE ACCUMULATION

A Thesis

Presented to

The Faculty of Moss Landing Marine Labs

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

Karen R. Parker

December 2013

The Designated Thesis Committee Approves the Thesis Titled

METABOLIC NETWORK CONSTRUCTION BASED ON THE GENOME OF THE MARINE DIATOM *THALASSIOSIRA PSEUDONANA* AND THE ANALYSIS OF GENOME-WIDE TRANSCRIPTOME DATA TO INVESTIGATE TRIACYLGLYCERIDE ACCUMULATION

By

Karen R. Parker

APPPROVED FOR MOSS LANDING MARINE LABS

SAN JOSÉ STATE UNIVERSITY

December 2013

Dr. Nick Welschmeyer	Moss Landing Marine Labs
Dr. Jason Smith	Moss Landing Marine Labs
Dr. Jon Geller	Moss Landing Marine Labs
Dr. Ginger Armbrust	University of Washington

ABSTRACT

METABOLIC NETWORK CONSTRUCTION BASED ON THE GENOME OF THE MARINE DIATOM *THALASSIOSIRA PSEUDONANA* AND THE ANALYSIS OF GENOME-WIDE TRANSCRIPTOME DATA TO INVESTIGATE TRIACYLGLYCERIDE ACCUMULATION

By Karen R. Parker

Algal lipids called triacylglycerols (TAGs) are essential compounds in marine ecology and are also used as the basis for commercial production of algal biofuels and omega-3 nutraceuticals. The goal of this thesis was to elucidate the biochemical processes associated with algal lipid accumulation in diatoms under different environmental conditions. The approach made use of the Mock et al. whole genome tiling microarray gene expression database for *Thalassiosira pseudonana* and gene annotations derived from the whole genome sequencing work of Armbrust et al. The analytical approach used bioinformatic tools, including Stanford Research Institute's (SRI) BioCyc Pathway tool and the Joint Genome Institute's (JGI) Integrated Microbial Genomes/Expert Review (IMG/ER) tool. This analysis has resulted in building the first whole genome BioCyc pathway model for *T. pseudonana* that includes over three hundred metabolic pathways. The analysis of the Mock transcriptome data in combination with the pathway model illustrated not only the activity of fatty acid and lipid pathways but also the interplay of other pathways that affect the accumulation of TAGs under different environmental conditions. This in silico analytical approach also revealed evidence for un-annotated gene functions and lack of the key regulatory protein PII pointing to the distinctive metabolic features of marine diatoms.

ACKNOWLEDGEMENTS

This thesis work could not have been done without the generous assistance of many people who buoyed my efforts. Jason Smith was a valuable sounding board who passed on interesting reading material and gave me detailed constructive criticism on the thesis manuscript. Nick Welschmeyer's insistence on high standards, especially for the oral defense, helped refine the thesis presentation and put it in perspective. Ginger Armbrust, the Director of the School of Oceanography at the University of Washington, led the effort of the original analysis of the genome of *Thalassiosira. pseudonana* that was the basis of the metabolic pathway model developed as part of this thesis work. She also provided valuable feedback, encouragement, and introductions to professional contacts. Jon Geller's attention to detail in the review of drafts contributed to the quality of the work. This work could not have been completed without the interaction with professionals in the field who took the time to help steer it in useful directions. I would specifically like to thank Robert Phair at Integrated Bioinformatics who helped guide me toward computational biology, Kate Dreher of the Carnegie Institute of Science and curator of the Plant Metabolic Network, Adrian Marchetti of the University of North Carolina, the BioCyc team at SRI, and, most significantly, Thomas Mock and the team from the University of Washington and the University of Wisconsin for making their transcriptome data available for my thesis analysis.

None of this work would have taken flight without the encouragement and loving support of my husband, Joe Fusco.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

There is a need to more efficiently and effectively leverage large-scale genome-wide databases to address contemporary issues such as improving algal lipid yields for algal biofuel and nutraceutical applications. This thesis work used bioinformatic technology to develop a metabolic pathway model from the genome of the marine diatom *Thalassiosira pseudonana*. The metabolic pathway model was used to analyze transcriptome data to address key questions with regard to the accumulation of triacylglycerol (TAG) for nutraceutical and biofuel applications.

Triacylglycerol for Algal Biofuel and Omega-3 Nutraceuticals

Concerns about climate change and the need for energy independence have driven the demand for more carbon-neutral algal biofuels. The triacylglycerols (TAGs) in algae can be used as jet fuel, but the cost for the fuel is currently not cost-competitive with fossil fuels. The omega-3 nutraceuticals DHA and EPA are high-value algal fatty acids contained in triacylglycerols. There is a need to improve TAG yields for both the nutraceutical and algal biofuel industries. The marine diatom *T. pseudonana* has been shown to be a good TAG producer.

T. pseudonana CCMP1335 is a single-cell marine diatom that has been studied extensively since the 1950s. Marine diatoms are unicellular eukaryotic algae that generate 20% of the organic carbon produced from photosynthesis on Earth (Field, 1998). Under certain conditions, up to 60% of diatom biomass can consist of neutral lipids stored

as TAGs (Hu, 2008). *T. pseudonana* has been shown to produce the long-chain polyunsaturated fatty acids (LC-PUFAs), eicosapentaenoic acid (EPA, C20:5n-3), and docosahexaenoic acid (DHA, C22:6n-3) during the stationary phase of growth (Tonon, 2002). These fatty acids are stored as neutral lipids called TAGs.

EPA and DHA are omega-3 fatty acids that have been shown to mitigate heart disease, cancer, and mental depression (Ginter, 2010). DHA is also used in infant formula to improve neurodevelopmental outcomes (Tonon, 2002). Currently, the principal biomass source of EPA and DHA is fish, but they are derived from their feedstock. As the demand for EPA and DHA continues to increase, it will accelerate the exploitation of already declining fish stocks to meet demand. Omega-6 fatty acids have been shown to lower blood pressure, cholesterol, and end the incidence of diabetes. Linoleic acid (LIN (C18:2 n-6) is an omega-6 fatty acid that is found in vegetable oils like corn, sunflower, and soy oil as well as in fish and algae. Vertebrates, including seals and humans, do not have the ability to produce the substrate A-linoleic acid (ALA (18:3n-3), which is required to produce EPA and DHA. Vertebrates cannot produce LIN and ALA because they lack delta 12 and delta 15 desaturase enzymes.

Lipid compounds are important energy currency in marine food webs, and lipid accumulation underlies successful maturation, egg development, and spawning. Deficiencies in essential fatty acids like DHA can negatively affect retina and nervous system development in fish to the extent that schooling behavior is affected (Masuda, 2007). Marine fish as well as most terrestrial plants cannot produce EPA or DHA. Fish

must consume EPA and DHA to obtain it. Although many invertebrates can produce EPA and DHA, algae are the primary source of EPA and DHA in aquatic ecosystems (Arts, 2009).

Concerns about energy independence led to the enactment of the 2007 Energy Independence and Security Act (EISA). EISA mandates that there will be an increase in renewable energy that decreases greenhouse gas emissions. Algal biofuel has several benefits compared to other biofuel sources like corn and soybean. Due to small cells and high biomass yields, algal-based production can yield more oil per acre and it does not compete with food cropland. The TAGs in algae are a good source of jet fuel because they have a composition similar to the lipids found in fossil fuels. As a result of these developments, large-scale investments are being made in algal biofuels (Scott, 2010).

The Aquatic Species Program (ASP) was funded by the Department of Energy from 1978 to 1996. The goal of the program was to develop renewable transportation fuels from algae. During this period, 3,000 strains of algae were tested in several environmental conditions to identify strains that were good TAG producers. Most algal species produce C14-C18 fatty acids in TAGs. Diatoms were shown to produce long-chain unsaturated fatty acids like EPA and DHA (Hu, 2008).

Of the factors used to screen for lipid accumulation, nitrogen limitation was shown to be the most consistent nutrient affecting lipid metabolism in algae of all taxa. Diatoms are exceptional in that silicon availability required for frustule development is equally important as a regulator of lipid accumulation (Hu, 2008). Diatoms were shown

to increase TAG content during starvation of either nitrogen or silica after day 7-9 of reaching the stationary growth phase but decreased in days 3-4 (Shifrin, 1981). In one experiment, diatom average lipid content was 23% Dry Cell Weight (DCW) in replete conditions and 45% DCW in low nitrogen and other stress conditions (Hu, 2006). Low temperature conditions increase unsaturated fatty acids, but a general lipid content trend has not been established (Hu, 2008). There are also experimental data that suggest that lipid accumulation increases in high pH conditions (> 9.5) (Guckert and Cooksey, 1990).

The yield and molecular species distribution of TAGs during starvation have been characterized in the marine diatoms *Thalassiosira pseudonana* (CCMP1335) and *Phaeodactylum tricornutum*. The maximal non-polar lipid yields were found to be 14% of the dry biomass in *T. pseudonana* in the low nitrogen condition and 18% in the low silicate condition . In the low nitrogen condition, neutral lipid content (as determined by Nile Red bulk assay) grew exponentially during the exponential growth phase and continued to grow exponentially until approximately 50 hours when its growth rate decreased to near zero (Yu, 2009). This result contradicts some previous low nitrogen studies but is in line with the Shifrin (1981) experiment that showed a decrease in lipid accumulation in the first four days of reaching the stationary phase. In the silicate starvation condition, *T. pseudonana* neutral lipid content grew exponentially as it transitioned into the stationary phase and continued to grow exponentially after entering the stationary phase but at a lower rate. The Yu (2009) data for low silicate are in line with previous studies indicating that lipid accumulation increases in low silicate conditions. The Yu experiment for low N suggests that lipid production decreases within

24 hours after reaching the stationary phase. A transcriptome study of *P. tricornutum* acclimated to light/dark cycles showed fatty acid synthesis enzymes upwardly expressed during light hours and beta-oxidation enzymes upwardly expressed during the dark period (Chauton, 2013). This suggests that fatty acid synthesis and degradation are under transcription control.

Fatty Acid and Lipid Chemistry

TAG synthesis starts with fatty acid initiation when acetyl CoA carboxylase (ACC) and malonyl CoA-acyl carrier protein transacylase (MCT) converts acetyl CoA to malonyl (Figure 1). Malonyl is used to create the fatty acid palmitate (C16:0), which is the most common fatty acid and the main building block of TAGs.

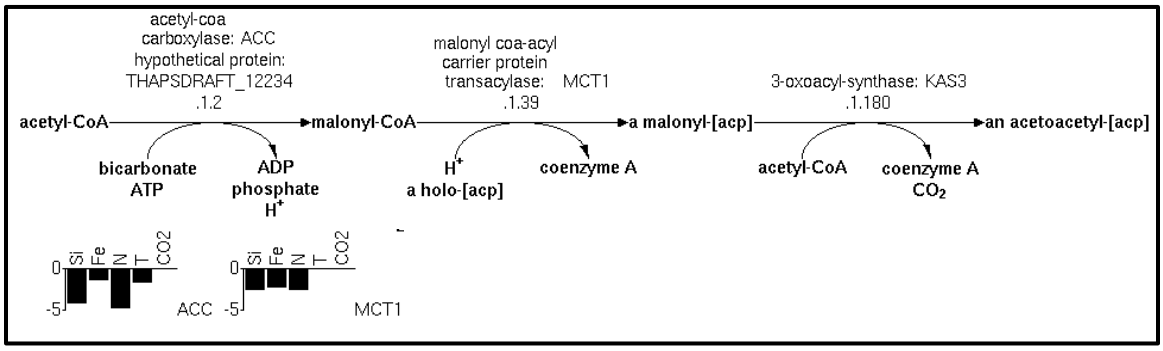

Figure 1. Fatty Acid Initiation Pathway with Expression Data

Acetyl-CoA carboxylase (ACC) and malonyl-CoA carrier protein transacylase (MCT1) were both downwardly expressed in the low nitrogen and low silicon conditions. The expression data are in (log_2) units and are relative to the nutrient replete condition.

Palmitate consists of 16 carbons and contains no double bonds between carbons. It has a methyl end (CH_3) and a carboxyl end (COOH). The hydrogen atom at the carboxyl end tends to be donated as a proton. This proton donor tendency is why it is called an acid. TAGs are created when three fatty acids are combined with glycerol. When fatty acids are combined with glycerol, they lose their tendency to donate a proton and are no longer referred to as acids. They are called lipids instead of fatty acids. Fatty acids can be elongated and double bonds can be added between carbon atoms to form unsaturated fatty acids like EPA and DHA .

Palmitate synthesis can be used as a proxy for TAG synthesis because it is the primary building block of TAGs. Palmitate synthesis requires acetoacetyl-(ACP) as a substrate and protons, NADPH, NADH, malonyl, and water as co-factors (Figure 2). The enzymes 3-oxoacyl acyl-carrier-protein (ACP) reductase, 3-hydroxyacyl -(ACP) dehydratase, enoyl-(ACP) reductase, and 3-oxoacyl-(ACP) synthase catalyze the reactions in palmitate synthesis. An increase or decrease in substrate, co-factors, and enzymes can affect the amount of palmitate produced.

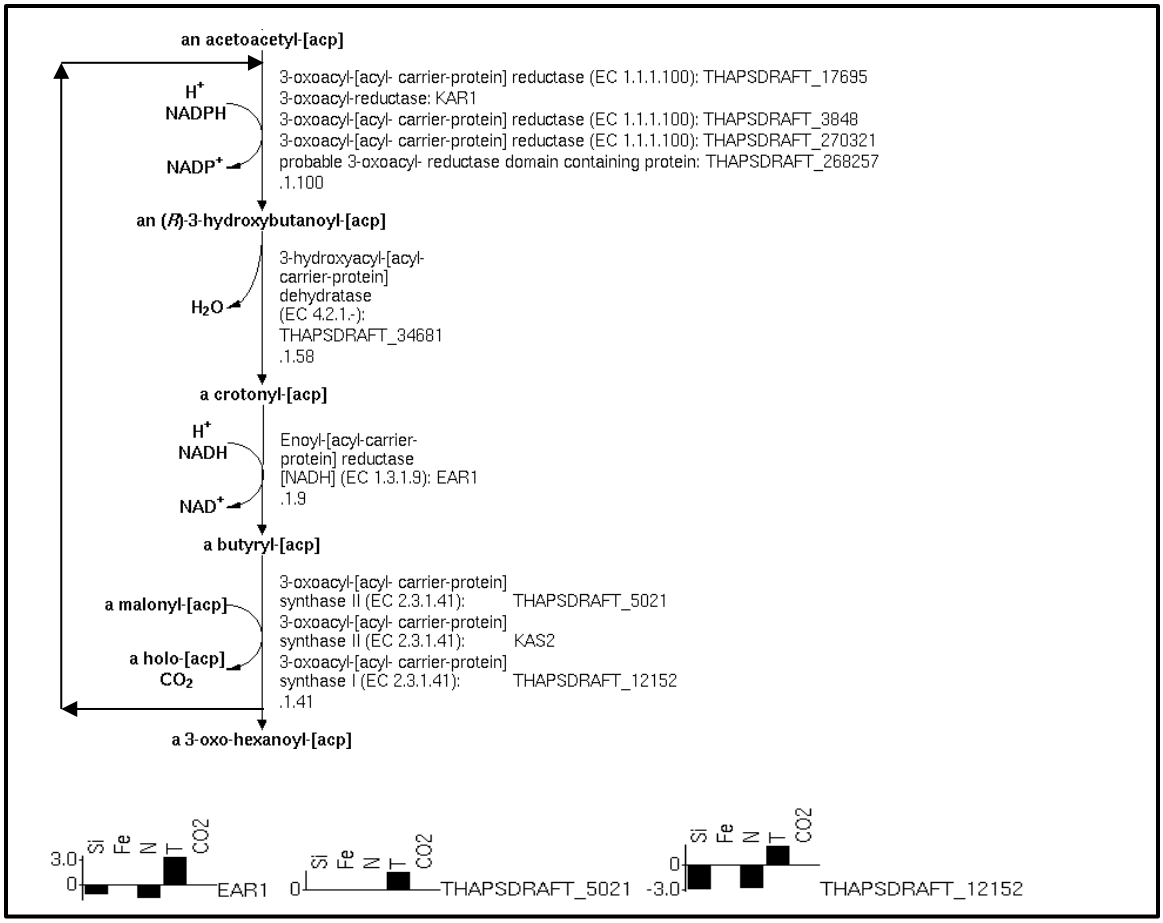

Figure 2. Fatty Acid Synthesis Cycle with Expression Data

The first four reactions of fatty acid synthesis are repeated until there are 16 carbons to form palmitate. In the low nitrogen and low silicon conditions, enzymes are downwardly expressed in this pathway. In the low temperature condition, they are upwardly expressed. Expression data are in units of ($\log_2$) relative to the nutrient replete condition.

Fatty acids are synthesized in the chloroplast with elongation of the fatty acid carbon chain and TAG synthesis takes place primarily in the endoplasmic reticulum that is attached to the chloroplast in the export path from the chloroplast (Lea, 1999). Fatty acid catabolism takes place in the mitochondria and in the peroxisomes. Medium chain fatty acids are transported to the mitochondria for degradation using a process called β-

oxidation. Peroxisome organelles catabolize very long chain fatty acids, branched chain fatty acids, and amino acids.

New Tools and Methods

An organism's genes determine its capability to respond to changes in environmental conditions. An organism's metabolic model is built using gene functional annotation. A metabolic pathway model maps the products of genes that are enzymes to biochemical reactions that make up the biochemical pathways. This mapping process can result in hundreds of pathways within an organism. The challenge is to understand and visualize the vast amount of data in a useful way. Kyoto Encyclopedia of Genes (KEGG) maps are one version of this attempt to organize the data. KEGG is an encyclopedia of enzymes in pathways of many organisms. The different versions of pathways are overlaid on top of each other and the genes of an organism can be highlighted over the complex and cryptic map. It is difficult to determine whether the organism has the pathway described in KEGG or merely some of the enzymes that are used in another pathway. There is a need to define individual pathways in an organism as compared to KEGG where the genes are overlaid on an encyclopedia of all known pathways. There is also a need to visualize experimental data better and to link databases with multiple genomes for comparison.

New bioinformatic tools have been developed to leverage genome and transcript data. BioCyc by Stanford Research Institute (SRI) is an example of a tool that uses semi-automated computer processes to build a metabolic model based on the functional gene

annotation of an organism's genome. Once the model is built, it has visualization tools to observe the entire metabolic pathway model in an organized way. Data are organized into three groups: biosynthesis, energy, and degradation. Within each section, there are subsections like glycolysis in the energy section or lipid synthesis in biosynthesis section. Within each subsection there are individual pathways; for example, in the glycolysis subsection, there are multiple versions of glycolysis: plastidic glycolysis and cytosolic glycolysis pathways. An organism's entire metabolic pathway model can be visualized effectively limited only by the user's screen size. This is extremely useful when one overlays the transcriptome data on the pathways. One can very quickly see which pathways are upwardly or downwardly expressed. Selecting the pathway of interest allows the user to zoom in on data to reveal which individual genes are differentially expressed and where they are in the pathway of interest. Discovery tools within BioCyc can be used to determine how an enzyme is used in other pathways and how pathway product compounds are connected to other pathways. Built-in analysis tools allow for systematic analysis of pathways and the transcript activity within functional categories. Without this type of tool, it may take months to examine each transcript individually, categorize its function, and determine the interconnection of expressed genes. New bioinformatic tools enable a wider investigative scope with fewer people.

In the case of this thesis, the BioCyc platform enabled the investigation into the accumulation of triacylglycerides to be expanded beyond the central pathways associated with the biosynthesis and degradation of triacylglycerides to include interactions with cellular energy systems and pathways that compete with TAG accumulation for metabolic resources.

Thesis Objectives and Strategy

The two main objectives of this thesis are to build a reusable metabolic pathway model for *T. pseudonana* and to analyze the Mock et al. 2008 expression data within the context of the metabolic model to elucidate the TAG accumulation response. The outcome of this work resulted in: 1) a reusable metabolic pathway model for *T. pseudonana,* 2) a cell-wide analysis that included not only triacylglycerol synthesis and degradation, but also energy systems and competing processes involving carbohydrates and proteins, 3) an analysis of the tools and workflow for pathway model development and transcript data analysis, and 4) a palmitate synthesis Flux Balance Analysis (FBA) model that illustrates how the metabolic pathway model can be used to build FBA models that can simulate differential fluxes based on differential gene expression data.

The overall strategy to build the metabolic pathway model was to refresh the automated pipeline annotation because it had been 10 years since it was last annotated, manually annotate probable enzymes, use BioCyc to semi-automatically build the metabolic model, review the model, fill holes in pathways, and add pathways based on literature on diatoms.

The overall approach to the analysis was to focus first on the fatty acid and lipid pathways in all five environmental conditions and then to broaden the analysis to evaluate competing pathways and energy systems. The Mock et al. 2008 transcript data were mapped onto the *T. pseudonana* BioCyc metabolic model and then analyzed to address key questions derived from the literature. Fatty acid/lipid synthesis and

degradation were evaluated to assess the accumulation trend. Differential upward expression of synthesis and downward differential expression of degradation pathways implies a trend toward fatty acid and lipid accumulation.

The palmitate co-factor NADPH is evaluated to look for indication that fatty acids may be limited by NADPH availability (Ratledge, 2002). High pH (10.2) has been shown to trigger lipid accumulation in algae (O'shea, 2009). The pH 9.4 data set was evaluated to look for lipid accumulation indicators.

Key lipid regulatory elements were assessed to determine whether they were present in the organism and, if so, evaluate their expression level. Key regulatory elements include ACC, which initiates fatty acid synthesis, PII signaling protein, PAT lipid droplet regulators, Carnitine acetyl-CoA transferase (CAT), and transcription factors.

METHOD

The method section is organized into two sections: building the model and analyzing the Mock transcription data within the context of the model. The pathway model was developed by mapping gene products to biochemical reactions that make up pathways. Once the model was developed, the transcript data were analyzed within the context of the metabolic functions of the organism.

Building the Metabolic Pathway Model

Building the model primarily consisted of updating the sequence annotation and then mapping gene products to biochemical pathways. Once the preliminary model was built, it was manually annotated to fill holes in pathways and to add pathways that were missing based on a review of diatom literature.

Genbank File Annotation Refinement

The *Thalassiosira. pseudonana* genome sequence data V 3.0 (Armbrust et al. in 2004) were used to build the metabolic pathway model. The sequenced genome consists of 34 million nucleic acid base pairs and almost 12,000 genes (Armbrust, 2004). The genome annotation for *Thalassiosira pseudonana* V 3.0 was updated in 2012 prior to pathway building. For comparison purposes, the annotation for the marine diatom *Phaeodactylum tricornutum* genome was also updated. Updated versions of databases like NCBI, KEGG, and Pfam allow the prediction of more genes and more precise annotations. The Genbank fasta files from NCBI were re-annotated using DOE-JGI

microbial annotation pipeline for eukaryotes (DOE-JGI MAP). The annotation includes both the identification of protein-coding and non-coding genes and repeats as well as the prediction of the function of each gene and the assignment of a product name (Mavromatis, 2009). Genes are identified using Hidden Markov Models and sequence similarity. The protein-coding genes are compared to COGs and Pfam protein families and the proteins of well-annotated genomes. Through this process, a product name is assigned. The protein sequences are compared to COG PSSMs using RPS BLAST at an e-value cutoff of $1e^{-2}$. The sequences searched against the KEGG gene database using BLASTp used an e-value cutoff of $1e^{-5}$. A KEGG Orthology rank of 5 or better is assigned with a greater than 70% alignment length on the query and KEGG gene sequences. The sequences are searched against the Pfam and TIGRfam databases. The newly annotated genome was output to JGI Integrated Microbial Genome Database and accessed using the JGI Integrated Microbial Genomes Expert Review (IMG/ER). The new version of the genome annotation is available to the public through the JGI-IMG website. The resulting Genbank files were exported from IMG/ER and used as the input files to build pathways in BioCyc. The newly updated Genbank files greatly improved the pathway annotation. The number of identified enzymes increased from 712 to 1,703. The additional enzyme information enabled over one hundred more pathways to be identified.

The JGI-IMG database includes over 9,000 datasets in its library that includes over 4,000 bacterial organisms and almost 200 eukaryotes. The IMG/ER database file for each gene in *Thalassiosira pseudonana* includes the sequence, name of the gene

product, KOG, Go terms, pfam, Enzyme Commission number (EC #), and a link to both the KEGG and MetaCyc pathway databases where the enzyme appears. During analysis, IMG/ER was used to examine sequences and annotation details and to make comparisons with other genomes including *Phaeodactylum tricornutum* CCAP1055/1, *Chlamydomonas reinhardtii* CC-503, *Arabidopsis thaliana Columbia*, *Synechococcus* sp. RCC 307, *Prochlorococcus marinus marinus* CCMP1375, *Escherichia coli* K12 DH, and *H. sapiens*.

Building Pathways from Genome Annotation Data

SRI's BioCyc Pathologic tool was used to build pathways from the genome database. The algorithm extracts gene product names and EC numbers from the Genbank files and maps them to pathways in the central MetaCyc pathway library that contains over 1,600 pathways with more than 600 organisms represented in the database (not necessarily whole genomes for all organisms). The MetaCyc pathway library is a curated library that contains pathways with significant experimental evidence to support each pathway description. Pathways that have a significant number of gene products mapped to them are included in the organism's pathway database. The phylogenetic pruning option was determined to be inaccurate for diatoms so this function was turned off. The resulting pathways are displayed with evidence glyphs, which identify pathway holes and whether the enzyme is unique to the pathway. The build resulted in over four hundred pathways including super pathways that combine smaller pathways into super pathways.

The set of pathways was evaluated to identify missing pathways. The pathways in *Thalassiosira pseudonana* CCMP 1335 were compared to pathways in other organisms

including *Phaeodactylum tricornutum* (DiatomCyc), *Chlamydomonas reinhardtii* (ChlamyCyc), *Arabidopsis thaliana* (AraCyc), *Synechococcus* sp. RCC 307, *Prochlorococcus marinus marinus* CCMP1375, *Escherichia coli* K12, *Plasmodium falciparum* 3D7, *Plasmodium yoelii yoelii* 17XNL, *Saccharomyces cerevisiae*, *Homo sapiens,* and a composite plant database called PlantCyc. Several reference papers on plants and diatoms were also used to edit pathways (Domergue, 2003; Fabris, 2012; Hockin, 2012; Kroth, 2008; Lee, 1999; Tonon, 2005; Yu, 2009). The pathways added included Palmitoleate biosynthesis II, TCA I, photorespiration, C4, and Chrysolaminaran. The pathways were imported from MetaCyc with the exception of Chrysolaminaran, which was imported from DiatomCyc (Fabris, 2012). Two reactions within the Chrysolaminaran pathway were not automatically assigned enzymes by BioCyc. The genes in DiatomCyc for P. tricornutum for the enzymes for the two reactions without genes assigned in *T. pseudonana* were evaluated in IMG/ER and blasted against the genomes listed above. The BLASTp resulted in the identification of three genes in *T. pseudonana,* which could be mapped to the Chrysolaminaran pathway. BLASTp for PHATRDRAFT_55324 resulted in a match to Thapsdraft_12695 with a good match to pfam02364 glucan synthase with a bit score of 1148 and e-value of o.o e-o.o (too small to register). For PHATRDRAFT_56509 with pfam 0393 blasted to THAPSDRAFT_ 262361 and THAPSDRAFT_3105 with bit scores over 400 and e-values of e-115 and e-143, respectively.

As part of the build process, BioCyc creates a list of probable enzymes. This is a list of gene products with enzyme-sounding names that it could not map to a biochemical

reaction because the data were too general or contained contradictory data. The list included 111 probable enzymes. Of the 111 probable enzymes, 25 were assigned to reactions manually.

The review process consisted of using IMG/ER to compare the enzyme to the six comparison genomes. If using BLASTp against the six genomes resulted in an e-value of 10^{-5} or less, a bit score of 100 or greater, and an alignment of approximately 70% or greater, the gene functional annotation, EC#, and pfams were compared to determine the function and reaction. Links to KEGG and MetaCyc were used to reference mapping to reactions in biochemical pathways. If the annotation was inconclusive or not specific enough to map to a reaction, Brenda, the enzyme information system, was consulted for clarification. If there were no good BLASTp hits in the six genomes, a BLASTp was done against the entire IMG/ER database and if that was inconclusive, a BLASTp was done on the NCBI database.

The Transport Inference Parser (TIP) built into SRIs pathologic identifies proteins in the current pathway database that are likely to catalyze transport reactions. Some of these transporters may be multi-subunit transporters. There are eight steps used by TIP to identify candidate transport proteins. The TIP steps are as follows: 1) find candidate transporter proteins using syntactic annotation, 2) filter candidates that are too generic or if they have a contradictory indicator word such as 'regulator', 3) search the annotation to identify transported substrate(s), and 4) assign an energy coupling classification to the transporter. Energy coupling classifications are defined as unknown, mechanically-

driven, light-driven, electron flow-driven, decarboxylation-driven, PEP-driven, or ATP-driven. 5) Identify the compartment of each substrate, 6) identify transporters that consist of subunits and group them into complexes. 7) Construct full reaction including substrates, compartment assignments and coupling, 8) construct enzymatic reaction linking each reaction with the transport protein. Transport proteins were not explicitly assigned to an organelle although some transport proteins have names that imply the type of membrane in which the transporter is located.

TargetP was used to aid in identifying the localization of enzyme activity. TargetP 1.1 predicts the subcellular location of eukaryotic proteins. The location assignment is based on the predicted presence of any of the N-terminal presequences: chloroplast transit peptide (**cTP**), mitochondrial targeting peptide (**mTP**) or secretory pathway signal peptide (**SP**). TargetP has been tested successfully on *Arabidopsis thaliana* and *Homo sapiens* genomes. There is evidence that it is less effective on diatoms because diatom chloroplasts have an additional membrane layer, derived from past host cells during diatom evolution (Kroth, 2008).

The refinement of the model increased the number of enzymes from 712 to 1,811 enzymes and the number of pathways from 131 to 466 pathways with 2,413 reactions and 1,508 compounds (Table 1). The total number of transport proteins was 166. Some of the pathways did not contain holes (reactions without an enzyme assigned), but a significant number did. In addition, many of the enzyme complexes were missing components of the complex. This result is not surprising, considering that most genomes

have 20-80% of the genes undefined. At this stage of genome knowledge, there are many discoveries to be made before a comprehensive model can be accurately built without holes. The data that we do have are enough to build out most of the pathways for central cell functions. The resulting model is very useful and reusable, even if it is imperfect.

Table 1. Pathway Model Improvements

Pathway Model	**Enzymes**	**Transporters**	**Pathways**	**Reactions**	**Compounds**
Without refinement	712	43	131	949	734
With refinement	1811	166	466	2413	1508
% Improvement	154%	286%	256%	154%	105%
Total added	1099	123	335	1464	774

Flux Balance Analysis Model

A flux balance analysis (FBA) model was developed for palmitate synthesis. The model was developed to illustrate the potential effect of substrate and co-factor limitation on palmitate synthesis flux (Figure 3). The model was used to evaluate limits implied by differential transcript data in the low nitrogen condition. Once the model was developed, fluxes were determined using the BioCyc FBA tool.

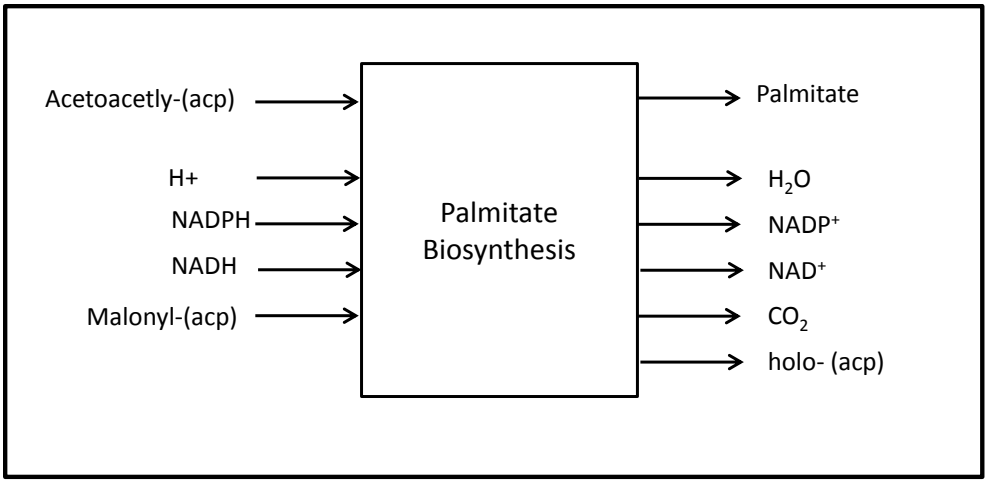

Figure 3. Palmitate Biosynthesis Inputs and Outputs

The model was developed from the metabolic pathway model for palmitate synthesis in the T. pseudonana BioCyc database. Reactions, substrates, co-factors, secretions, and products were defined in a FBA input file. The substrate limit was set to 1,000 units to simulate a nutrient-replete condition and the input file was used to generate flux of all inputs and outputs to the model using the Solving Constraints Integer Program (SCIP) in BioCyc. Transcript data were used to adjust input amounts of co-factors and predicted enzyme abundance, assuming a 1:1 correspondence between transcript abundance and new protein abundance relative to the replete condition.

This is a simple illustration of how FBA data can be used with differential transcript data. A FBA model of the whole cell or at least major components of interest within the cell would work in a similar way but nutrient inputs would be set based on actual growth medium concentrations. The model can also simulate gene knock out conditions by removing the reaction that the gene catalyzes.

Data Analysis

The Mock et al. 2008 genome-wide transcriptome analysis was performed on *T. pseudonana* strain CCMP 1335 under five different environmental conditions: low nitrate (low N), low silicic acid (Low S), low iron (Low Fe), low temperature (4°C), and high pH (9.4), with nutrient-replete cultures serving as reference conditions (Mock et al., 2008). As described in their publication, cultures were maintained from natural seawater, autoclaved, and supplemented with 2 x f/2 nutrients minus limited nutrient at 20°C and 100 μmol of photons $m^{-2} \bullet s^{-}1$ (24-h light cycle). All limitation experiments were conducted in parallel with nutrient-replete cultures. Cells were harvested for RNA when the growth rate began to decrease significantly relative to the control cultures. Over 2,000 genes were shown to be differentially expressed over all conditions. Tiling and gene specific array data were made publically available on NCBIs Gene Expression Omnibus (GEO) web page under the Accession number of GSE9697.

BioCyc Group Tools were used to analyze the data. First, all fatty acid and lipid synthesis pathways were selected, and then all the reactions for all the pathways in the category were determined using a transform function in BioCyc. Then the enzymes associated with each reaction were determined using the transform function. Once the list of all enzymes in the fatty acid and lipid synthesis category was determined, the list was compared with the Mock et al. 2008 differential expression database to find differentially expressed genes. Each resulting gene was evaluated to understand whether the enzyme was unique to the pathway, compared to any isozyme or complex member

assigned to the same reaction, and related to the overall expression pattern of the entire pathway. After fatty acid synthesis and degradation pathways were analyzed, the broader energy systems and competing pathways were analyzed using the same method (Figure 4). To form conclusions from a systems' perspective, the data for each pathway were summarized as an expression direction as represented by an arrow in a table. A data cutoff of ≥ 2 $\log_2$ and ≤ 2 $\log_2$ was used. This was a way to summarize the data concisely.

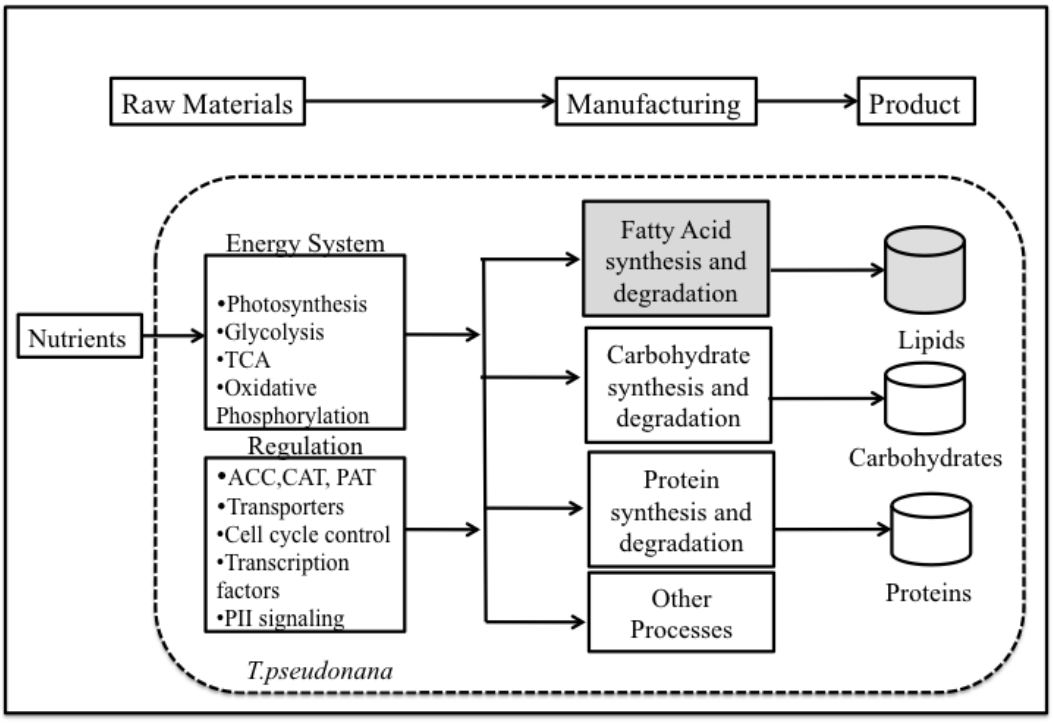

Figure 4. Cell Subsystems

Carbon fixed in photosynthesis can be converted into lipids, carbohydrates, and proteins. Lipid biosynthesis competes with carbohydrate and protein synthesis for resources. Regulatory elements can control the partitioning of carbon between competing pathways.

The pathways that were investigated beyond fatty acid and lipid synthesis and degradation included energy systems and competing pathways. Photosynthesis is important because this pathway fixes carbon that is used to produce fatty acids. Energy systems also produce co-factors like NADPH that are needed to produce fatty acids. If there are fewer hydrocarbons coming into the cell, there are fewer hydrocarbons to produce fatty acids. The photosynthesis and carbon fixation category included light reactions, Calvin Cycle, photorespiration, and C4 photosynthetic carbon assimilation pathways. The fucoxanthin chlorophyll a/c binding proteins were also assessed.

Carbohydrate synthesis and degradation pathways were included in the analysis because they compete with fatty acid and lipid synthesis for hydrocarbons. The synthesis pathways analyzed included Chrysolaminaran (the most common carbohydrate form in diatoms), gluconeogenesis, and sucrose biosynthesis. The degradation pathways included cytosolic and chloroplastic glycolysis, sucrose degradation, rubisco shunt, TCA, and pentose phosphate pathways. Many of the pathways in carbohydrate degradation share enzymes and many of the reactions are reversible. As a result, it is difficult to ascertain the activity based on transcripts alone.

Protein synthesis and degradation was analyzed because proteins also contain carbon and hydrogen and therefore compete with TAG synthesis. This category focused on the five classes of amino acids. The five classes of amino acid are categorized by their biochemical pathway precursor. The precursors are compounds of either glycolysis or the TCA cycle. Proteins are of particular interest in the low nitrogen condition because

they contain nitrogen (chrysolaminaran and TAGS do not). Since low nitrogen conditions have been shown to increase lipid accumulation, nitrogen assimilation is of interest to the investigation. Processes like ammonia assimilation to glutamate are of particular interest because it requires 2-oxoglutarte, which is a compound that is known to up regulate lipid synthesis in conjunction with the PII signaling protein (Bourrellier, 2010).

To address one of the investigated questions about the possibility that a lack of NADPH may inhibit fatty acid synthesis, pathways that consume or produce NADPH are of interest. In addition to the light reactions that produce NADPH, oxidative stress reactions were also analyzed because they consume NADPH. Oxidative stress is a destructive state for the cell. During oxidative stress, DNA and mRNA can be damaged and apoptosis (cell death) can be triggered. NADPH is used as a reducing agent to mitigate oxidative stress. A change in an environmental condition may upset the redox balance in a cell and result in oxidative stress. In addition, reduced photosynthesis efficiency may cause a shortage of NADPH.

Transporters were assessed because a lack of a transporter can reduce the productivity of a pathway because it lacks a substrate or co-factor that needs to be transported to the location where the pathway reactions take place. For example, carnitine acetyl-CoA (CAT) proteins transport fatty acids across the mitochondrial membrane for beta-oxidation.

Other fatty acid and lipid regulators that were investigated based on literature references included an analysis of CAT, PII, ACC, PAT proteins, and transcription factors. The analysis of the regulators first involved determining whether the genes for these gene products existed in the organism or one of the six comparison genomes. Once the existence and forms of the protein where established, the differential expression was investigated.

RESULTS AND DISCUSSION

The results of the transcript analysis section are organized into four sections: 1) fatty acid and lipid synthesis and degradation, 2) competing processes associated with carbohydrates and proteins, 3) energy systems, and, 4) regulatory elements. The initial fatty acid and lipid synthesis analysis in five environmental conditions revealed that only the low silicon, low nitrogen, and low temperature conditions showed significant differential expression. Therefore, the scope of the wider cell evaluation was focused only on these three environmental conditions.

Fatty Acid and Lipid Synthesis and Degradation

All five environmental conditions relative to nutrient-replete (Low Si, Low Fe, Low N, Low T [4°C], High pH [9.4]) were evaluated for gene transcript expression in the pathways involved in the synthesis and degradation of fatty acids and lipids. For TAGs in the test condition to accumulate at a rate higher than TAGs in the nutrient-replete conditions, fatty acid and TAG synthesis would be expected to increase relative to TAGs in the replete condition (assuming β-oxidation remained the same or decreased). The transcripts of the enzymes in the degradation of lipids and fatty acids pathways would be expected to remain the same or decrease.

Evaluation of the Mock 2008 transcriptome under all five environmental conditions showed that the low temperature condition was the only one in which the fatty acid/lipid synthesis transcripts increased and fatty acid/lipid degradation decreased

relative to the nutrient-replete condition (Figure 5). At low temperatures, gene transcripts are regulated to favor carbon storage in the form of lipids. It may not necessarily imply that the lipids are accumulating at a faster rate than under the replete condition because the rate of metabolic activities is lower at low temperatures as defined by the Q10 effect. A *T. pseudonana* experiment showed that nitrate reductase (NR) activity more than doubled between 8°C and 18°C (Berges, 2002).

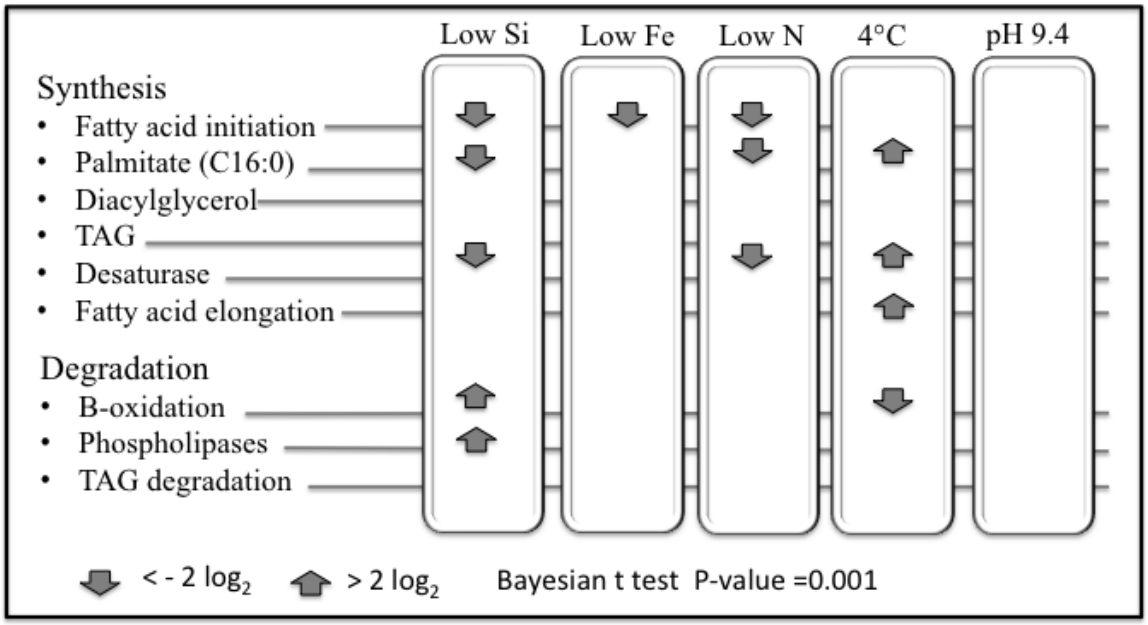

Figure 5. Fatty Acid and Lipid Synthesis and Degradation Differential Expression

The low temperature condition showed fatty acid synthesis and desaturase and elongation pathways upwardly expressed, while β-oxidation showed a downward expression. The low temperature differential expression profile suggests a condition that would favor lipid accumulation. The low silicon condition showed a downward differential expression for fatty acid synthesis and an upward differential expression for β-oxidation, which suggests a condition that would deplete fatty acids. The low nitrogen condition showed a downward differential expression for fatty acid synthesis that would suggest a lower rate of lipid accumulation.

The low Si condition showed the opposite trend as compared to the low temperature condition: the fatty acid/lipid transcripts were downwardly expressed in the synthesis pathways and upwardly expressed in the degradation (β-oxidation) pathways. This is not the expected result, since experimental data have shown an accumulation of lipids in low Si conditions relative to nutrient-replete conditions (Hu, 2008; Yu, 2009). However, there has been evidence that this result may suggest that accumulation does not increase until some time after reaching the stationary phase (Shifrin, 1981).

The low N condition showed transcripts of fatty acid and lipid synthesis decreased, and there were no significant transcript increases detected in the fatty acid and lipid degradation pathways. This result would be consistent with the Yu (2009) and Shifrin (1981) experiments that showed the lack of accumulation of lipids following the transition into the stationary phase.

Low Fe and high pH conditions showed little or no differentiation relative to the replete condition in fatty acid and lipid pathways. In the high pH environment, none of the 113 fatty acid and lipid gene transcripts showed any significant change at the 2 $\log_2$ cutoffs. Previous studies indicated that a pH greater than 10.2 showed an increase in lipid accumulation relative to nutrient-replete conditions (Guckert and Cooksey, 1990). The data imply that lipids continued to accumulate at the replete rate of accumulation. It may indicate that the pH was not high enough to trigger the increased accumulation response. Figure 5 summarizes the lipid accumulation in all five conditions.

Low Temperature

The low temperature condition was the only condition where the fatty acid/lipid synthesis transcripts increased and fatty acid/lipid degradation decreased relative to the nutrient-replete condition. Of the 113 enzymes in the fatty acid/ lipid synthesis pathways, only eight differentially expressed transcripts for those proteins at low temperature (Table 2). All differentially expressed genes were expressed upwardly for fatty acid synthesis.

Table 2. Fatty Acid and Lipid Synthesis Expression at Low Temperature

Locus Tag	Value	Gene product name
THAPSDRAFT_5021	2.02	3-oxoacyl-ACP synthase
THAPSDRAFT_34681	2.16	3-hydroxyacyl-[acyl-carrier-protein] dehydratase
THAPSDRAFT_12152	2.28	beta-ketoacyl-[acyl-carrier-protein] synthase I
THAPSDRAFT_3143	2.38	microsomal omega-6 fatty acid desaturase
THAPSDRAFT_264335	2.55	phosphatidylinositol-4-phosphate 5-kinase
THAPSDRAFT_40801	2.55	dimethylallyltranstransferase
THAPSDRAFT_23798	2.89	delta-12 desaturase
THAPSDRAFT_32860	3.26	enoyl-[acyl-carrier-protein] reductase

The fatty acid degradation pathway β-oxidation had one enzyme transcript that was downwardly expressed. The most differentially expressed gene was THAPSDRAFT_32860, which had a differential express level of 3.26 relative to the replete condition. THAPSDRAFT_32860 is a gene which codes for a protein called enoyl-[acyl-carrier protein] reductase I (EAR1). It is one of four transcripts that were differentially expressed upwardly in the palmitate (C16:0) fatty synthesis pathway (Figure 2). EAR1 is the enzyme of the third step of the palmitate fatty acid synthesis pathway. The reaction with EC 1.3.1.9 involves the reduction of NADH to NAD+. The

other three differentially expressed genes at low temperature in the palmitate pathway were THAPSDRAFT_12152, THAPSDRAFT_34681, and THAPSDRAFT_5021 that had expression levels between 2.28 and 2.02. THAPSDRAFT_334681 is a hydroxymyristoyl ACP dehydratase associated with the second step of fatty acid synthesis. THAPSDRAFT_12152 and THAPSDRAFT_5021 are isozymes for the reaction EC 2.3.1.41 in palmitate synthesis. They are 3-oxoacyl-[acyl-carrier-protein] synthase enzymes associated with EC 2.3.1.41, which plays a role in fatty acid synthesis and elongation. It is associated with very large chain fatty acids like mycolate as well as the pathway of the vitamin biotin that is a part of fatty acid initiation. There were no downwardly expressed genes within the cutoff range in fatty acid synthesis.

Two of the upwardly expressed genes in low temperature were desaturases that add a double bond between carbons. Fatty acids with double bonds are called unsaturated fatty acids. THAPSDRAFT_3143 and THAPSDRAFT_23798 are particularly interesting because they are both omega-6 fatty acid desaturases. THAPSDRAFT_264335 is associated with phospholipids and is active in the D-myo-inositol (1,4,5)-trisphosphate biosynthesis and 3-phosphoinositide biosynthesis pathways. Myo-inositol containing phospholipids are part of the intracellular signal transduction pathways. The THAPSDRAFT_40801 gene is associated with isoprenoid biosynthesis and diphosphate biosynthesis. It is noteworthy that THAPSDRAFT_263660, which codes for LAT1_1 in diacylglycerol and TAG biosynthesis pathways, was upwardly expressed at a value of 1.6, but was below the 2.0 cutoff.

The only downwardly expressed gene in the fatty acid and lipid synthesis and degradation pathways at low temperature was THAPSDRAFT_26365. Its expressed value was -3.32. This gene codes for a tri-functional enzyme in the β-oxidation and cholesterol degradation pathways. Interestingly, it also plays a role in pathways outside of the fatty acid lipid degradation pathway class. It is also part of the isoleucine degradation pathway and the valine degradation pathway. Isoleucine and valine are branch-chain amino acids.

The low temperature condition shows a response to optimize TAG accumulation because fatty acid synthesis and elongation are upwardly expressed in four genes. The fact that only these four genes in fatty acid synthesis showed differential change in any condition is significant. Interestingly, there was not a significant upward expression in the fatty acid synthesis initiation pathway that includes ACC. In fact, there was a -1.5 level of expression for ACC that was above the cutoff and trending in the opposite of the expected direction. There could be several explanations for this since ACC is regulated not only by enzyme availability, but also allosterically by citrate, fatty acids, and also by phosphorylation. Desaturases for omega-6 fatty acids are significant because omega-6 fatty acids have high value in the nutraceutical industry. It is also interesting that there is an upward expression for diphosphate biosynthesis that may increase both phospholipid and TAGs. There was only one gene downwardly expressed for the degradation of fatty acids and lipids, although the enzyme is tri-functional and plays a roll in three steps of β-oxidation.

Low Silicon

The low Si condition showed the opposite trend as compared to the low temperature condition: the fatty acid/lipid transcripts were downwardly expressed in the synthesis pathways and upwardly expressed in the degradation pathways. Five genes in fatty acid and lipid synthesis were downwardly expressed and no genes were detected to be upwardly expressed in these pathways (Table 3).

Table 3. Fatty Acid and Lipid Synthesis Expression at Low Silicon

Locus Tag	Value	Gene product name
THAPS_6770	-4.4	acetyl-coa carboxylase(ACC)
THAPSDRAFT_268480	-3.2	geranylgeranyl pyrophosphate synthetase(GGPS1)
THAPSDRAFT_5219	-2.7	malonyl coa-acyl carrier protein transacylase (MCT1)
THAPSDRAFT_12152	-2.9	3-oxoacyl-(acyl-carrier-protein) synthase
THAPSDRAFT_32145	-2.8	Fatty acid desaturase

THAPS_6770 and THAPSDRAFT_5219 are part of the fatty acid initiation pathway. THAPS_6770 is acetyl-CoA carboxylase (ACC) that had a differential expression value of -4.4 in the low Si condition. ACC is the first committed of fatty acid synthesis. ACC catalyzes the reaction with EC number 6.4.1.2 that converts acetyl Co-A to malonyl CoA. THAPSDRAFT_5219, which is part of the fatty acid initiation pathway (Figure 1), expressed at a value of -2.7 which was the same level as the low N condition. THAPSDRAFT_12152 codes for 3-oxoacyl-[acyl-carrier-protein] synthase enzyme associated with EC 2.3.1.41 which is a reaction in fatty acid synthesis and elongation. This gene was expressed upwardly in the low temperature condition. THAPSDRAFT_268480 was expressed at a value of -2.9. The transcript codes for a

protein called geranylgeranyl pyrophosphate synthetase (GGPS1) that is an enzyme in the isoprenoid biosynthesis pathway. Isoprenoids are part of the prenol lipid class. They are precursors of pigments (including chlorophyll), vitamins, and cholesterol. They have a five-hydrocarbon substructure. Isoprenoid biosynthesis uses three units of acetyl CoA. THAPSDRAFT_32145 codes for an enzyme that is a desaturase in the sphingolipid metabolism pathway. It expressed at a value of -2.8 relative to the replete condition. Sphingolipids contain nitrogen.

Three genes which code for fatty acid and lipid degradation enzymes were upwardly expressed: THAPSDRAFT_34809, THAPSDRAFT_26365, and THAPSDRAFT_263246 (Table 4). THAPSDRAFT_34809 was only expressed differentially in the low Si condition and expressed at 2.7. It codes for 3-ketoacyl-coa thiolase, mitochondrial (KCT1) that is an enzyme in the β-oxidation and cholesterol degradation pathways. THAPSDRAFT_26365 expressed at a value of 2.2 and is a putative 3-hydroxyacyl-CoA dehydrogenase that is an enzyme in the β-oxidation pathway. One of its isozymes of the reaction is also in the isoleucine and valine degradation pathways. THAPSDRAFT_263246 expressed at a value of 2.1 and codes for phospholipase C that is an enzyme of the phophatidate metabolism and phospholipases associated with cell signaling. It is also worth noting the expression of a THAPSDRAFT_262242 gene that codes for an enzyme in the β-oxidation pathway that was below the cut off at a value of 1.98. It is a putative long-chain-fatty-acid--CoA ligase associated with reaction EC 6.2.1.3.

Table 4. Fatty Acid and Lipid Degradation Expression at Low Silicon

Locus Tag	Value	Gene product name
THAPSDRAFT_34809	2.7	3-ketoacyl-coa thiolase, mitochondrial(KCT1)
THAPSDRAFT_26365	2.2	simular to 3-hydroxyacyl-CoA dehydrogenase
THAPS_263246	2.1	phospholipase C

Gene expression in the low Silicon condition showed an upward expression of four out of five of the reactions in beta-oxidation and a strong -4.4 downward expression of ACC which is the first committed step of fatty acid synthesis.

Low Nitrogen

The low N condition showed transcripts of fatty acid and lipid synthesis to decrease and there were no significant transcript increases detected in the fatty acid and lipid degradation pathways (Table 5). Four genes were differentially expressed in fatty acid and lipid synthesis: THAPS_6770, THAPSDRAFT_3143, THAPSDRAFT_5219 AND THAPSDRAFT_12152. They were all expressed in the downward direction (Table 3). There was no differential expression in any of the fatty acid and lipid degradation pathways at low N (no beta-oxidation, TAG degradation or phospholipases). The most differentially expressed gene was THAPS_6770 that is acetyl-CoA carboxylase (ACC) that had a differential expression value of -5.0. ACC is the first committed of fatty acid synthesis. ACC catalyzes the reaction with EC number 6.4.1.2. THAPSDRAFT 12234 is also ACC2 for this reaction and no differential expression was detected for it at low N or any of the other five conditions. ACC2 is thought to be associated with elongation in the endoplasmic reticulum. THAPSDRAFT_5219 is also part of the fatty acid initiation

pathway. It expressed at a value of -2.7. THAPSDRAFT_12152 is a gene that codes for 3-oxoacyl-(acyl-carrier-protein) synthase and expressed at a value of -2.9. This protein is part of fatty acid synthesis and elongation. THAPSDRAFT_3142 is an omega-6 fatty acid desaturase (DES9) that was differentially expressed at a value of -2.2. This is the same gene that was differentially expressed upwardly at low temperature. The expression pattern for fatty acid and lipid synthesis and degradation at low N shows a decrease in fatty acid and lipid synthesis and no change in degradation expression relative to the replete condition. This may suggest that fatty acids and lipids are still accumulating but at a slower rate relative to the replete condition.

Table 5. Fatty Acid and Lipid Synthesis Expression at Low Nitrogen

Locus Tag	Value	Gene product name
THAPS_6770	-4.97	acetyl-coa carboxylase(ACC)
THAPSDRAFT_3143	-2.16	omega-6 fatty acid desaturase(DES9)
THAPSDRAFT_5219	-2.69	malonyl coa-acyl carrier protein transacylase (MCT1)
THAPSDRAFT_12152	-2.9	3-oxoacyl-(acyl-carrier-protein) synthase

A Palmitate flux balance analysis (FBA) model was developed and used to evaluate the potential effect of the differential expression of ACC and ferredoxin NADPH reductase (Table 6). ACC catalyzes the reaction that generates Malonyl-ACP that is required for palmitate synthesis. Ferredoxin NADPH is an enzyme in the photosynthesis light reaction pathway that catalyzes a reaction that produces two NADPH that is also required for palmitate synthesis.

Table 6. Flux Balance Analysis for Palmitate

		Inputs					Output
Condition	**FBA Limit Setting**	**Acetoacetyl-ACP**	**Proton**	**NADPH**	**NADH**	**Malonyl-ACP**	**Palmitate**
Stoichiometry	Acetoacetyl-ACP=1	1	13	7	7	6	1
"Replete"	Acetoacetyl-ACP=1000	1000	13000	7000	7000	6000	1000
"Replete"	Malonyl-ACP=6000	1000	13000	7000	7000	6000	1000
LOW N: ACC -5 (log2)	Malonyl-ACP=188	31	407	219	219	188	31
"Replete"	NADPH= 7000	1000	13000	7000	7000	6000	1000
LOW N: ferridoxin reductase -3(log2)	NADPH =438	62	813	438	438	375	62
Ferridoxin nadp reductase -5(log2)	NADPH =109	15	202	109	109	93	15

The substrate limit was set to 1,000 units to simulate a nutrient-replete condition and the input file was used to generate flux of all inputs and outputs of the model. Transcript data were used to adjust input amounts of co-factors and predicted enzyme abundance, assuming a 1:1 correspondence between transcript abundance and new protein abundance relative to the replete condition.

In the case of ACC in the low nitrogen condition, the differential expression was -5 $\log_2$ relative to the nutrient-replete condition (1/32 of the nutrient-replete condition). In the case of a nutrient-replete condition generating 1,000 units of palmitate per time unit, 6,000 units of malonyl-ACP would be required per time unit. If ACC, which catalyzes the reaction that produces malonyl-ACP is reduced by 1/32, then 188 units of Malonyl would be the new input limit for Malonyl-ACP. The FBA results with the new limit setting indicated that for the reduction implied by the differential transcript expression the amount of palmitate produced in one time unit would be reduced from 1,000 to 31 units of palmitate. This implies that palmitate synthesis would be reduced by 96%.

In the case of ferredoxin NADPH reductase in the low nitrogen condition, the differential expression was -3 $\log_2$ that would imply that NADPH production in the light reaction pathway would be reduced by 1/8, which would be 438 NADPH units compared to the 7,000 units modeled in the nutrient-replete condition because the reaction produces two NADPH. The result of the FBA with the 438 NADPH limit produced 62 units of palmitate as compared to 1,000 units in the nutrient-replete case. This implies that if NADPH was the only limiting factor of palmitate synthesis, palmitate would be reduced by 92%. It is interesting to note that a differential expression of -5 $\log_2$ of ferredoxin NADPH would imply the production of 15 units of palmitate while -5 $\log_2$ of ACC would result in 31 one units of palmitate.

Competing Processes Involving Carbohydrates and Proteins

Carbohydrate and protein synthesis competes with lipid synthesis for metabolites. A shortage of a metabolite can reduce the production of fatty acids.

Carbohydrate Synthesis and Degradation Including Glycolysis

Carbon is stored in the form of carbohydrates and fatty acids. Carbohydrate synthesis competes with fatty synthesis for carbon storage. There has been interest in the algal biofuel community to understand the partitioning of carbon between the two storage pools better. Carbohydrate synthesis and degradation pathways were assessed in all five conditions. The low Si, low N, and low Fe conditions showed the only differential expression (Figure 6).

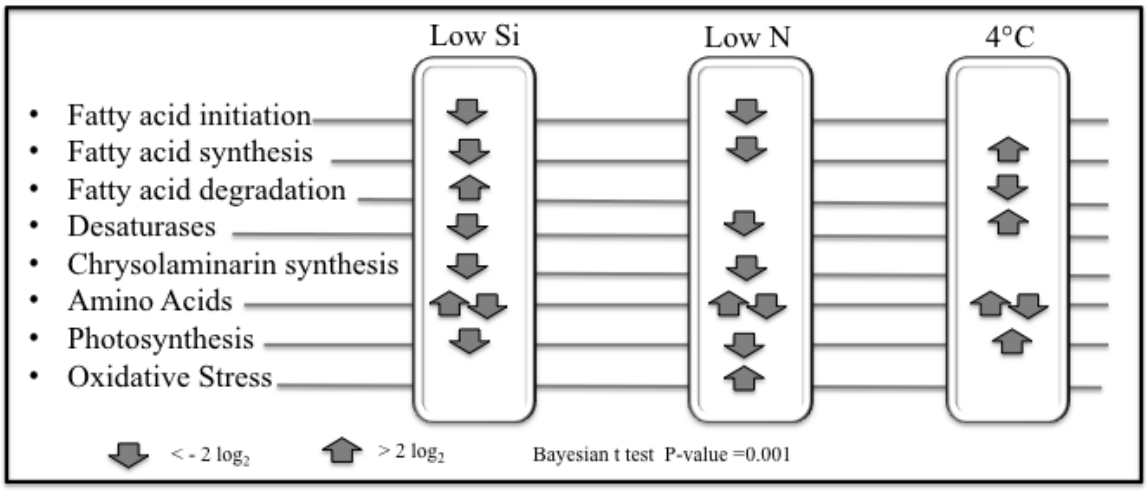

Figure 6. Cellular Overview of Differential Expression

Chrysolaminarin, the most common carbohydrate in diatoms, showed downward expression in both the low silicon and low nitrogen conditions. Photosynthesis enzymes downwardly expressed for both low silicon and low nitrogen conditions and upwardly expressed in the low temperature (4°C) condition. There was some indication of oxidative stress in the low nitrogen condition where glutathione reductase that consumes NADPH was upwardly expressed.

Carbohydrates are primarily stored as Chrysolaminaran in diatoms (Kroth, 2008). Two of the six reactions had catalytic enzyme expressed downwardly in both the low Si and low N condition (Table 7). Phosphoglucomutase (THAPSDRAFT_268621) had an expression level of -2.57 at low Si and -2.12 in the low N condition. The gene for the gene product 1,3 beta D-glucansynthase (THAPSDRAFT_12695) had an expression level of -2.2 for low Si and -4.1 for low N. This gene is unique to this pathway and is not used in any other pathway including sucrose synthesis. This is a significant finding because it implies that neither fatty acid synthesis nor Chrysolaminaran are upwardly expressed in either low Si nor low N condition, which suggests that they are not switching nor

competing between fatty acid synthesis and carbohydrate synthesis, but are reducing the synthesis of both types of carbon storage. This could be a response to a reduction in photosynthesis productivity or it may indicate that carbon is being diverted to another system. There was no significant up-regulation in any of the five conditions; however, there was a noteworthy expression level of +1.43 below the +2.0 cutoff for low temperature.

Table 7. Chrysolaminaran Biosynthesis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_31636		-2.2				aldose-1-epimerase
THAPSDRAFT_26862	-2.57		-2.12			phosphoglucomutase (PGM2)
THAPSDRAFT_12695	-2.2		-4.1			1,3, beta-D-glucan synthase

Gluconeogenesis is essentially the reverse of glycolysis. Many of the enzymes are the same bidirectional enzymes as in glycolysis and sucrose synthesis. In this pathway, the directionality is imagined to be from non-sugar to glucose. Non-sugar substrates like pyruvate, malate, glycerol, and glucogenic amino acids like L-alanine and L-glutamine are the inputs to this pathway that may be sourced from catabolic reactions like β-oxidation. In the beginning of this pathway, malate is converted to oxaloacetate or to pyruvate. NADH and NADPH are produced. Malic Enzyme (MAO1) catalyzes a reaction that converts malate to pyruvate. If NADP+ is the co-factor, NADPH and CO_2 are products of the reaction and its EC number is 1.1.1.40. If NAD+ is the co-factor that NADH and CO_2 are produced and the reaction EC number is 1.1.1.38. MAO1 is the catalytic enzyme for this reaction and it is only differentially expressed relative to the replete condition during the low Si (-2.64) and low N (-2.46) (Table 8). This is

significant to lipid accumulation because NADPH is required for fatty acid synthesis. The alternate path for Malate conversion is catalyzed by malate dehydrogenase (THAPS_41425) and has a reaction EC number of 1.1.1.37. This reaction also produces NADH (but not NADPH) and oxaloacetate. Interestingly, the gene for this enzyme is significantly upwardly expressed in the low N condition (2.12) but not in any other of the five conditions. Phosphoglycerate (PGAM2) (THAPSDRAFT_27850) was also upwardly expressed in the low N condition and none of the other conditions. It is also noteworthy that three other genes were expressed upwardly above 1.5 but below the 2.0 cutoff. This suggests an upward regulation of the sub-pathway from malate, to oxaloacetate to 3-phospho-D-glycerate. The part of the pathway below phosphor-D-glycerate shows downward expression for both low N (-3.43) and low Si (-3.61) in glyceraldehyde-3-posphate dehydrogenase (GAPD1) as well as fructose bisphosphate aldolase (ALDO2) (THAPSDRAFT_21748) with a value of -2.45 for low SI and -2.16 for low N. This may suggest a gene transcription regulation to increase 3-phospho-D-glycerate that is an intermediary of glycolysis as well as the Calvin cycle. These are bidirectional reactions so it could also suggest an increase in pyruvate generation in glycolysis from 3-phospho-D-glycerate.

Table 8. Gluconeogenesis Differential Expression

Locus Tag	Low Si	Low FE	Low N	Low T	CO2	Gene product name
THAPSDRAFT_34030	-2.64		-2.46			Malic Enzyme (MAO1)
THAPS_41425			2.12			cytosolic malate dehydrogenase
THAPSDRAFT_5186				-2.2		phosphoenopyruvate carboxy kinase (PCK1)
THAPSDRAFT_268546	-3.15	-2.59		-2.79		Phosphoenolpryruvate carboxylase (PPC1)
THAPSDRAFT_34543	-2.6		-2.79			Phosphoenolpryruvate carboxylase (PPC2)
THAPSDRAFT_27850			2.99			phosphoglycerate mutase (PGAM2)
THAPSDRAFT_28350				2.4		phosphoglycerate mutase (PGAM1)
THAPSDRAFT_269057	2.98					phosphoglycerate kinase (PGK4)
THAPSDRAFT_42577	-2.28			-2.04		phosphoglycerate kinase (PGK3)
THAPSDRAFT_25116	-2.01					phosphoglycerate kinase (PGK5)
THAPSDRAFT_31383	-3.61		-3.43			glyceraldehyde-3-phosphate dehydrogenas (GAPD1)
THAPSDRAFT_21748	-2.45		-2.16			fructose-biphosphate aldolase (ALDO2)

3-phospho-D-glycerate is also used to produce serine in the serine biosynthesis pathway. Serine is a precursor for protein synthesis as well as glycine, cysteine, tryptophan, and phospholipids**.** One gene showed noteworthy expression between 1.5 and 2.0 for low Si and low N in the serine biosynthesis pathway. Three genes in the serine pathway had gene expression between 2.0 and 2.5, which is consistent with studies showing upward expression of the serine pathway in response to low temperature.

Triose phosphate is the product of photosynthesis that is exported from the chloroplast to the cytoplasm where it is converted to hexose phosphate by the gluconeogenesis process. These hexoses are used by sucrose biosynthesis and chrysolaminaran synthesis in the cytoplasm.

Seven genes were differentially expressed in the sucrose pathway in the Mock experiment (Table 9). They were all downwardly expressed with the exception of one that was upwardly expressed in low Si. The high CO_2 condition showed no differential expression. The low Fe had one of the genes downwardly expressed at a value of -2.2 (THAPSDRAFT_31636). Low temperature had one differentially expressed downwardly at -2.04 for phosphoglycerate kinase (PGK3) (THAPSDRAFT_42577) that is also used in glycolysis. In the low Si condition, five genes were expressed downwardly and one (PGK4) (THAPSDRAFT_269057) was expressed upwardly. PGK4 is also part of the glycolysis pathway. Three of the downwardly expressed genes were also downwardly expressed in the low N condition: THAPSDRAFT_31383 glyceraldehyde-3-phosphate dehydrogenase (GAPD1) expressed at -3.61 for low Si and -3.43 for low N, THAPSDRAFT_21748 fructose bisphosphate aldolase (ALDO2) at a value of -2.45 for low Si and -2.16 for low N, and THAPSDRAFT_268621 phosphoglucomutase (PGM2) at a value of -2.57 for low Si and -2.12 for low N. Two other genes were also downwardly expressed in Low Si condition: THAPSDRAFT_ 42577 (PGK3) expressed at -2.28 that was one of three isozymes for phosphoglycerate kinase that was expressed. The other one that was expressed was PGK5 THAPSDRAFT_ 25116 that was expressed at a value of -2.01 in the low Si condition. Overall, the trend was downward expression for both the low Si and low N conditions.

Table 9. Sucrose Biosynthesis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_269057	2.98					phosphoglycerate kinase (PGK4)
THAPSDRAFT_42577	-2.28			-2.04		phosphoglycerate kinase (PGK3)
THAPSDRAFT_25116	-2.01					phosphoglycerate kinase (PGK5)
THAPSDRAFT_31383	-3.61		-3.43			glyceraldehyde-3-phosphate dehydrogenas(GAPD1)
THAPSDRAFT_21748	-2.45		-2.16			fructose-biphosphate aldolase (ALDO2)
THAPSDRAFT_31636		-2.2				aldose-1-epimerase
THAPSDRAFT_27850			2.99			phosphoglycerate mutase (PGAM2)

There is overlap with the chrysolaminaran pathway to UDP-D-glucose where it can be converted to sucrose or chrysolaminaran. The sucrose pathway expression for the reaction that consumes UDP-D-glucose did not express differentially as compared with the replete condition. The reaction in chrysolaminaran that consumes UDP-D-glucose was expressed downward.

Glycolysis converts glucose into pyruvate. Glycolysis takes place in both the cytoplasm and the chloroplast. ChloroP was used to identify chloroplast transit peptides in glycolysis genes. Four genes had chloroplast transit peptides and seven did not.

Seven genes in the cytosolic glycolysis pathway were differentially expressed (Table 10). The low Fe and low CO_2 conditions showed no differential expression in this pathway. There were five genes differentially expressed in the low Si condition and all but one were downwardly expressed. Three of the genes were isoforms for phosphoglycerate kinase (PGK4, PGK3 and PGK5). PGK4 (THAPSDRAFT_269057) was upwardly expressed at a value of 2.98. PGK catalyzes the reaction that converts

1,3,bisphospho-D-glycerate to 3-phospho-D-glycerate. The low Si condition was the only condition where the isozymes were expressed differentially in opposite direction.

Table 10. Cytosolic Glycolysis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_21748	-2.45		-2.16			fructose-biphosphate aldolase (ALDO2)
THAPSDRAFT_36462	-4.11		-4.06			triose-phosphate isomerase (TPI3)
THAPSDRAFT_269057	2.98					phosphoglycerate kinase (PGK4)
THAPSDRAFT_42577	-2.28			-2.04		phosphoglycerate kinase (PGK3)
THAPSDRAFT_25116	-2.01					phosphoglycerate kinase (PGK5)
THAPSDRAFT_28350				2.4		phosphoglycerate mutase (PGAM1)
THAPSDRAFT_22345			6.21	4.07		pyruvate kinase (PYK2)

In the low N condition, the fructose-bisphosphate aldolase (ALDO2) THAPSDRAFT_21748 and triose-phosphate isomerase (TPI3) THAPSDRAFT_36462 were downwardly expressed. ALDO2 converts fructose 1,6 bisphosphate to dihydroxyacetone phosphate (DHAP) and D-glyceraldehyde-3-phosphate. DHAP is one of 13 primary precursor metabolites that are used by sucrose biosynthesis and other processes. The other downwardly expressed gene in cytosolic glycolysis is triose-phosphate isomerase (TPI3) THAPSDRAFT_36462 that was expressed at a value of -4.06. This is a bidirectional reaction that converts between DHAP and D-glyceraldehyde-3-phosphate. Interestingly, it was strongly upwardly expressed in the chloroplast glycolysis (TPI1). The one gene which was upwardly expressed was pyruvate kinase (PYK2). THAPSDRAFT_22345 that was differentially expressed at a

value of 6.12 (it was also strongly upwardly expressed in the chloroplast). PYK converts phosphoenolypyruvate to pyruvate, which are both precursor metabolites. It is a one-way reaction, which is part of multiple pathways including Entner-doudoroff, rubisco shunt, and fermentation pathways. Pyruvate is a substrate for acetyl-CoA synthesis. Acetyl-CoA is a substrate for fatty acid synthesis, ketogenesis and ATP synthesis. Acetyl-CoA is also a product of fatty acid degradation, amino acid degradation, and ketone body degradation in the TCA cycle.

At low temperature there was one downwardly expressed gene THAPSDRAFT_42577 for phosphoglycerate kinase (PGK3) and two upwardly expressed THAPSDRAFT_28350 phosphoglycerate mutase (PGAM1) and pyruvate kinase (PYK2). PGAM catalyzes the reaction from 3-phospho-D-glycerate to 2-phospho-D-glycerate. In the low temperature, the upward expression of pyruvate generating enzyme is significant because fatty acid synthesis is upwardly expressed in this condition and pyruvate is a precursor to substrates required for fatty acid synthesis.

In total, four differentially expressed genes for glycolysis contained chloroplast transit peptides. In the low Si condition, two genes were differentially expressed (Table 11). Triose-phosphate isomerase (TPI) was expressed upwardly to a value of 3.22 (THAPSDRAFT_40958). Glyceraldehyde-3-phosphate (GAPD) THAPSDRAFT_31383 was downwardly expressed at a value of -3.61. This may suggest regulation to decrease 3-phospho-D-glycerate while increasing dihydroxyacetone phosphate (DHAP).

Table 11. Plastidic Glycolysis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_40958	3.22	3.49	3.3	4.92		triose-phosphate isomerase (EC:5.3.1.1)
THAPSDRAFT_31383	-3.61		-3.43			glyceraldehyde-3-phosphate dehydrogenas
THAPSDRAFT_27850			2.99			phosphoglycerate mutase (PGAM2)
THAPSDRAFT_40393			6.17	4.24		pyruvate Kinase (PYK1)

In the low N condition, four genes were differentially expressed in the chloroplast for glycolysis. Three of them were upwardly expressed (TPI, phosphoglycerate mutase PGAM2, and PYK1 pyruvate kinase. Glyceraldehyde-3-phosphate was significantly downwardly expressed to a value of -3.43. In the low temperature condition, there were two differentially expressed genes (TPI) and pyruvate kinase (PYK1) that were both expressed upwardly in the chloroplast.

The sucrose degradation pathway has only one gene out of eleven that was significantly differentially expressed in low Si, low N or low T conditions. This gene (THAPSDRAFT_268621) is phosphoglucomutase (PGM2) that is a bidirectional acting enzyme that is common to many pathways including Chrysolaminaran, sucrose biosynthesis, and glycogen degradation. The direction of a bidirectional reaction is difficult to predict based solely on transcriptional data. The gene product directionality can be affected by substrate concentration, localized pH level, temperature, and other factors. The differential expression of a small number of genes in degradation pathways that are also used in other pathways make it difficult to draw conclusions on the activity of degradation pathways. Since there was no significant expression of genes that were

unique to degradation pathways, it implies that there is not a significant difference in carbohydrate degradation as compared to nutrient-replete conditions.

The rubisco shunt as compared to glycolysis yields 20% more acetyl-CoA with 40% less carbon loss (BioCyc). Literature on diatoms does not refer to the rubisco shunt; however, enzymes for all the reactions in the rubisco shunt have been identified in *T. pseudonana* and none of the enzymes are unique to the pathway. The rubisco shunt has also been identified as a pathway in *P. tricornutum* (DiatomCyc) and *A. thalina* (BioCyc). The first stage of the rubisco shunt is the reverse of the non-oxidative pentose phosphate pathway. The second stage includes the carboxylase activity of Rubisco that fixes one CO_2 with one D-ribulose-1, 5 biphosphate yielding 3 molecules of 3 carbons 3-phospho-D-glycerate. Stage three is the same as the last three of reactions of glycolysis where 3-phospho-D-glycerate is converted to pyruvate. The third stage is upwardly express in the same way as glycolysis.

The first stage is downwardly expressed in the low Si and low N conditions. The middle stage is unchanged relative to the replete condition in all tested environmental conditions. There is a lot of overlap between glycolysis, the Calvin cycle, the pentose phosphate cycle, the Entner-Doudoroff pathway, and fermentation. In particular, the lower part of glycolysis that connects pyruvate to many precursor metabolites like phosphoenolypyruvate, 3-phospho-D-glycerrate, and dihydroxyacetone phosphate (DHAP). These pathways act as bidirectional metabolite 'switch' under gene control. The bidirectionality of the pathways makes it difficult to predict the directionality of the

catalytic activity because it is dependent on factors including local pH, temperature, and substrate concentration. A broader cell-wide perspective may provide insight into potential metabolite cell demand and suggest a reaction direction to meet the theoretical metabolite demand. For example, during photosynthesis when C fixation rate is high, there is a need to process the fixed carbon into energy and storage. This suggests a direction of the pathways from triosephosphate towards pyruvate generation. Pyruvate can then be transformed into acetyl CoA and processed for energy or fatty acid storage.

In the dark condition when energy stores are degraded and consumed, it is conceivable that the directionality of the pathway may be reversed. Many degradation pathways result in the generation of pyruvate or Acetyl-CoA. The pyruvate can be used to generate precursor metabolite if it reverses the directionality of the reactions. It can turn pyruvate in to precursor metabolites and back into carbohydrates (gluconeogenesis). Pyruvate from degradation can be used in the other direction to form Acetyl CoA that can be used to generate energy in the TCA and oxidative phosphorylation processes.

Protein Synthesis and Degradation

Protein synthesis competes with fatty acid synthesis for metabolic resources. An upward expression in these pathways may explain why acetyl CoA production is upwardly expressed in low Si and low N conditions, but fatty acid and carbohydrate synthesis are downwardly expressed. If it is not used for fatty acid synthesis or carbohydrate synthesis, it may be used for protein synthesis.

There are five classes of amino acids: Aspartate, glutamate, pyruvate, erythrose 4-phosphate, and ribose 5-phosphate families. They are classified by the precursor of their biochemical pathway. The aspartate family consists of asparagine, lysine, threonine, methionine, and isoleucine that are synthesized from the amino acid substrate aspartate. Aspartate is produced from oxaloacetate (from the TCA cycle) and also from ornithine and citrulline in the urea cycle. Leucine and valine do not use asparagine as a carbon source substrate (they use pyruvate), but they use the same enzymes that are associated with the isoleucine pathway (Lea, 1993). Amino acids formed from oxaloacetate compete with citrate production in the mitochondria. Citrate is formed from oxaloacetate and acetyl-CoA in the mitochondria. The citrate is exported from the mitochondria and the acetyl-CoA is removed from the citrate for fatty acid synthesis. The remaining oxaloacetate is converted into malate and recycled back into the mitochondria where it is converted back into oxaloacetate. This cycle is called the citrate shuttle. Although fatty acid synthesis does not consume oxaloacetate, a lack of oxaloacetate in the mitochondria could slow the transport of acetyl-CoA out of the mitochondria via the citrate shuttle and this in turn could slow the rate of fatty acid synthesis.

There was no differential expression in any of the five environmental conditions of any of the five isozymes for aspartate aminotransferase that catalyzes the reversible reaction of oxaloacetate and glutamate to/from aspartate and 2-oxaglutarate (EC 2.6.1.1) (Table 12). In addition, oxaloacetate, pyruvate, and 2-oxoglutarte are regulators of fatty acid synthesis (Bourrelier, 2010). PII inhibits ACC, 2-oxoglutarate inhibits PII. PII

senses C and N levels through allosteric binding which changes it structure (see PII regulation of ACC in regulation section).

Table 12. Amino Acid Biosynthesis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_28496	-2.89		-2.04	2.42		adenosylhomocysteinase (AHC3)
THAPSDRAFT_412	-4.68	-2.58	-3.32	-2.01		alanine aminotransferase (ALAT_1)
THAPSDRAFT_11411			2.09			precursor of synthase (CSN1)
THAPSDRAFT_34543	-2.6		-2.79			phosphoenolpyruvate carboxylase, type 1 (PPC2)
THAPSDRAFT_14147	-2.05	-2.19	-1.94	-2.23		L-threonine ammonia-lyase
THAPSDRAFT_20816			2.11			branched-chain amino acid aminotransferase
THAPSDRAFT_24248	-3.79	-2.97	-2.46	-2.59		multifunctional pyrimidine synthesis protein (CAD)
THAPSDRAFT_260934			3.08			branched chain aminotransferase (BCAT1)
THAPSDRAFT_268970				2.39		D-3-phosphoglycerate dehydrogenase
THAPSDRAFT_28544				4.03		dihydrodipicolinate reductase
THAPSDRAFT_31394			-2.24			LL-diaminopimelate aminotransferase apoenzyme
THAPSDRAFT_40323	-4.55	-2.94				CPSase
THAPSDRAFT_5954	-3.12		-2.12	2.55		S-adenosylhomocysteine hydrolase
THAPS_269942	-2.92		-2.57			glycine or serine hydroxymethyltransferase
THAPSDRAFT_21290	-2.27					N-acetyl-gamma-glutamyl-phosphate reductase
THAPSDRAFT_268546	-3.15	-2.59		-2.79		phosphoenolpyruvate carboxylase type 1 (PPC1)
THAPSDRAFT_262099	2.4	2.05				alanine racemase
THAPSDRAFT_5500		-3.03		-2.77		pyruvate phosphate dikinase (PDK1_2)
THAPSDRAFT_24816				2.32	-2.26	aspartate-ammonia ligase (ASA)
THAPSDRAFT_11320				2.53	2.42	3-deoxy-D-arabinoheptulosonate-7-phosphate
THAPSDRAFT_25130				2.45		D-3-phosphoglycerate dehydrogenase
THAPSDRAFT_26031				2.12		serine hydroxymethyltransferase
THAPSDRAFT_268335				-2.46		glutamate dehydrogenase
THAPSDRAFT_32137				2.22		phosphoserine phosphatase
THAPSDRAFT_32140				2.73		dihydrodipicolinate synthase

Isoleucine biosynthesis pathway expressed THAPSDRAFT_14147 downwardly less than -2.0 for low Si, low Fe and low temperature. Notably, low nitrogen expressed it at -1.94. This reaction is non-reversible and it converts threonine (another aspartate

amino acid) to 2-oxobutanoate. Interestingly, the last step of the pathway which is a reversible reaction expressed upwardly only in the low N condition where two isozymes expressed upwardly (THAPSDRAFT _20816 at 2.11 and THAPSDRAFT _260934 at 3.08). The same enzymes are also part of the leucine pathway. This suggests that aspartate amino acids are most likely being degraded in low N condition. Isoleucine degradation showed downward expression for low temperature for THAPSDRAFT _26365 and upward expression for low Si. This same enzyme is active in β-oxidation and cholesterol degradation. Lysine biosynthesis enzyme THAPSDRAFT _31394 is expressed downwardly in the low N condition. In the low temperature condition, where fatty acid synthesis generally expressed upwardly, lysine biosynthesis had two enzymes upwardly expressed (THAPSDRAFT _32140 and THAPSDRAFT _28544). Notably, several enzymes of isoleucine and lysine biosynthesis showed upward expression but were below the cutoff value. Asparagine biosynthesis enzyme THAPSDRAFT_24816 expressed upwardly in the low temperature condition suggesting that aspartate amino acid biosynthesis is increased in the low temperature condition. Methionine biosynthesis showed no significant differential expression but methionine degradation to homocysteine showed downward expression for low Si and low N (THAPSDRAFT _28496(AHC3)) and upward expression of two isozymes of the same reaction (EC 3.3.1.1.) THAPSDRAFT _28496(AHC3) and THAPSDRAFT _5954). It is the last reaction of the degradation pathway and it is not a reversible reaction.

The glutamate family consists of glutamine, arginine, and proline. Glutamate is a carboxylated anion of glutamic acid. Glutamic acid is an amino acid with a molecular

formula of $C_5H_9NO_4$. Glutamate biosynthesis showed differential downward expression in low temperature. Glutamate dehydrogenase (THAPSDRAFT _268335) that converts 2-oxoglutarate to/from glutamate was downwardly expressed at a value of -2.46 in low temperature. This is a reversible reaction. See regulation section below.

Arginine biosynthesis enzyme THAPSDRAFT _24248 differentially expressed downward in low Si, low Fe, low N, and low temperature conditions suggesting that arginine biosynthesis from bicarbonate decreased in those conditions relative to nutrient-replete condition. There was no differential expression for proline biosynthesis. There was no differential expression in any condition of the glutamate family of amino acids.

The pyruvate family of amino acids consists of alanine, serine, cysteine, and glycine. Pyruvate also donates carbon to lysine, isoleucine, leucine, valine, and phosphoenolpyruvate via the shikimate pathway (Lea, 1999). This family competes for pyruvate that would otherwise be converted to acetyl-CoA, which could be used as a carbon source for fatty acid synthesis.

Alanine biosynthesis and degradation consists of one bidirectional reaction. The reaction is catalyzed by alamine amino transferase (THAPSDRAFT _412) (ALAT-1). It was expressed downwardly in low Si (-4.68), low Fe, low N, and low T conditions. The reaction converts pyruvate and glutamate to/from alanine and 2-oxoglutarate.

Serine biosynthesis showed upward expression at low temperature. Three genes associated with this pathway expressed differentially upward (THAPSDRAFT _2689, THAPSDRAFT _25130, THAPSDRAFT _32137). The other conditions did not show

significant differential expression. There was no significant differential expression for cysteine biosynthesis or degradation.

Glycine biosynthesis was downwardly expressed for low N and low Si condition (THAPS_269942). An isozyme (THAPSDRAFT _26031) for the same reaction (EC: 2.1.2.1) showed upward expression for low temperature. Glycine degradation in the glycine cleavage complex showed downward expression in the low Si (THAPSDRAFT _36208, THAPSDRAFT _39799) and low T conditions (THAPSDRAFT _39799) (Table 13). Low N condition also showed downward expression but also had an upwardly expressed enzyme (THAPSDRAFT _36716-LPD) that also is an enzyme that is part of acetyl-CoA biosynthesis (pyruvate dehydrogenase complex).

Table 13. Amino Acid Degradation Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPS_41425			2.12			cytosolic malate dehydrogenase
THAPSDRAFT_20816			2.11			branched-chain amino acid aminotransferase
THAPSDRAFT_260934			3.08			branched chain aminotransferase (BCAT1)
THAPSDRAFT_260953	-6.13	-4.4	-5.91	-3.01		ornithine cyclodeaminase
THAPSDRAFT_26365	2.18			-3.32		long-chain 3-hydroxyacyl-CoA dehydrogenase
THAPSDRAFT_268335				-2.46		glutamate dehydrogenase
THAPSDRAFT_268594			3.63			tyrosine aminotransferase
THAPSDRAFT_269328				-2		beta subunit acetyl/propionyl-CoA carboxylase
THAPSDRAFT_28496	-2.89		-2.04	2.42		adenosylhomocysteinase (AHC3)
THAPSDRAFT_36208	-2.71					glycine decarboxylase t-protein GDCT
THAPSDRAFT_36291			2.12			dihydrolipoamide transacylase
THAPSDRAFT_36716			2.37			hydrogenase lipoamide dehydrogenase precursor
THAPSDRAFT_38807	-4.06	-2.32	-3.63			2-keto-3-deoxy-6-phosphogluconate aldolase
THAPSDRAFT_39799	-4.64	-2.1	-4.38	-2.54		glycine decarboxylase p-protein (GDCP)
THAPSDRAFT_412	-4.68	-2.58	-3.32	-2.01		alanine aminotransferase (ALAT_1)
THAPSDRAFT_5954	-3.12		-2.12	2.55		S-adenosylhomocysteine hydrolase
THAPSDRAFT_795	2.64		4.83	-2.61		branched chain alpha-keto acid dehydrogenase

The Erythrose 4-phosphate family of aromatic amino acids consists of phenylalanine, tyrosine, and tryptophan via the shikimate pathway. There was no significant differential expression for biosynthesis or degradation of phenylalanine or tryptophan in any of the environmental conditions.

Tyrosine biosynthesis showed no differential expression in any of the conditions. Tyrosine degradation was strongly upwardly expressed (THAPSDRAFT _268594) at a value of 3.63 for low N condition. This reaction is unidirectional and not part of biosynthesis. There was also upward expression at low T of the gene for HDPI (THAPSDRAFT _32153).

There was no significant differential expression for biosynthesis or degradation of tryptophan in any of the environmental conditions. There was however a significant upward expression of the gene that codes for DPHA1 (THAPSDRAFT _11320) for both low temperature and high pH conditions. DPHA1 is the unidirectional first step of the shikimate pathway that produces the substrate that feeds into tryptophan biosynthesis. The input to the shikimate pathway (also called chorismate biosynthesis) is the substrate D-erythrose-4-phosphate from the pentose phosphate pathway (non-oxidative branch). The ribose 5-phosphate amino acid family consists of histidine. Histidine biosynthesis and degradation showed no differential expression in any of the conditions.

Proteins compete against fatty acid thesis for carbon, acetyl-CoA, and NADPH. Amino acids are formed from different substrates and co-factors. For example, the first step of arginine biosynthesis uses a L-glutamate substrate and acetyl-CoA as a co-factor. The third step of the arginine synthesis pathway utilizes NADPH in a dehydrogenase

reaction. Aspartate biosynthesis uses oxaloacetate that is a TCA intermediary as the substrate and L-glutamate as a co-factor to form L-aspartate using just one enzyme called aspartate aminotransferase. L-glutamate can be formed from the amino acid Glutamine. Many of the amino acids are formed from smaller amino acids.

The last four steps of isoleucine, valine, and leucine are the same for all three of these amino acids. The last step in these pathways is a branched-chain amino acid aminotransferase reversible reaction with the EC number of 2.6.1.42. During amino acid biosynthesis, this reaction uses L-glutamate as a co-factor to produce isoleucine, valine or leucine depending on the substrate. In the degradation mode, this same reaction degrades isoleucine, valine or leucine and produces L-glutamate.

Interestingly, two of the five isozymes for EC 2.6.1.42 expressed differentially in only one of the five conditions and that was for the low N condition. THAPSDRAFT_20816 and THAPSDRAFT_260934 expressed upwardly at values of 2.11 and 3.08, respectively. In a low N condition, glutamate concentrations are presumably low, suggesting that the reaction supports the degradation of leucine, isoleucine, and valine to scavenge nitrogen in the form of glutamate. The branched-chain alpha-ketoacid dehydrogenase complex (BCDHC) pathway is one of a family of pathways that convert 2-oxo acids to acyl-CoA derivatives and produce NADH and CO_2. BCDHC plays a key role in the degradation of L-valine, L-leucine, and L-isoleucine. The BCDHC pathway consists of three reactions. Enzymes for all three reactions were differentially expressed upwardly in the low N conditions. THAPSDRAFT_795 was

expressed at a value of 4.83, THAPSDRAFT_36291 was expressed at 2.12 and THAPSDRAFT_36716 expressed at a value of 2.37. This upward expression in the low N condition also supports the interpretation that L-valine, L-leucine, and L-isoleucine are degraded in low N condition. One enzyme of the BCDHC pathway was upwardly expressed in the low Si condition and one was downwardly expressed in the low temperature condition.

Energy Systems

Energy systems consist of the photosynthesis and carbon fixation pathways, the TCA and pentose phosphate pathways, and the oxidative stress mitigation reactions that consume the reducing agent NADPH that is a product of the photosynthesis light pathway.

Photosynthesis and Carbon Fixation

The rate of photosynthesis affects carbon availability for fatty acid synthesis as well as the availability of NADPH that is also required for fatty acid synthesis. Photosynthesis efficiency is reduced in low Si and low N conditions (Berges and Falkowski, 1998; Hockin, 2012). It is expected that gene expression in photosynthetic pathways would be downwardly expressed. The photosynthesis class consists of four pathways: Light reaction pathway, Calvin Benson cycle, photorespiration, and C4 photosynthetic carbon assimilation cycle pathway. In these pathways, eighteen genes were expressed significantly in low Si, low N or low temperature conditions.

In the light reaction pathway, photons split water and releases oxygen to harvest electrons so that NAP+ can be reduced to NADPH. The light reactions also create a proton motive force across the thylakoid membrane to generate ATP. NADPH is required for fatty acid synthesis (14 units for palmitate). Three genes were differentially expressed in low Si, low N or low T conditions (THAPSDRAFT_34592, THAPSDRAFT_4914, THAPSDRAFT_25892) that are isozymes for ferredoxin-NADP+ reductase (EC: 1.18.1.2) (Table 14). These enzymes reduce ferredoxin to create an oxidized ferredoxin and NADPH. THAPSDRAFT_4914 and THAPSDRAFT_25892 were both downwardly expressed in both low Si and low N conditions as anticipated. THAPSDRAFT_34592 was upwardly expressed in the low temperature and low N conditions. A comparison of the three isoforms which were differentially expressed for reaction EC 1.18.1.2 in IMG/ER showed that all three had good BLASTp homology to each other and similar genes in *P. tricornutum* and *A. thaliana*. Interestingly, THAPSDRAFT_4914 and THAPSDRAFT_25892 that were the downwardly expressed isoforms only contained one pfam: pfam00175 (NAD_binding_1). THAPSDRAFT_34592 that was upwardly expressed in the low temperature and low N conditions contained two pfams: pfam00175 and pfam00970 (FAD_binding_6).

Table 14. Photosynthesis Light Reactions Expression

Locus Tag	Low Si	Low N	Low T	Gene product name
THAPSDRAFT_34592		2.18	2.28	ferredoxin--NADP(+) reductase
THAPSDRAFT_4914	-3.23	-3.32		ferredoxin--NADP+ reductase
THAPSDRAFT_25892	-4.33	-3.64		ferredoxin-nadp reductase

Carbon is fixed in the Calvin Benson cycle when the enzyme ribulose 1,5 bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the reductive carboxylation of 2-ribulose 1.5 bisphosphate. One carbon is added for each turn of the cycle. After three turns, 3-phospho-D-glycerate is formed. After six cycles, hexose sugar is formed. Of the eight differentially expressed genes in the Calvin Benson cycle, six of the genes were downwardly expressed and five of these genes were also enzymes of the glycolysis pathway (Table 15). Of the eight differentially expressed genes in the Calvin Benson cycle, only one (THAPSDRAFT_40958) was upwardly expressed in low Si, low N and low T conditions. It is triose-phosphate isomerase (TPI1) (EC: 5.3.1.1) that is also used in the sucrose degradation and glycolysis pathways. Interestingly, an isozyme of this enzyme called TPI3 was downwardly expressed as TPI1 was upwardly expressed. The only other upwardly expressed gene was phosphoglycerate kinase (PGK4) (EC: 2.7.2.3) (THAPSDRAFT_269057) that was only upwardly expressed in the low Si condition. In the same low Si condition, isozymes for PGK4 called PGK3 and PGK5 were both downwardly expressed. In summary, all the enzymes were downwardly expressed except in those associated with reactions that had at least one isozyme that was downwardly expressed. The general expression is downward for low N and low Si conditions as expected since photosynthesis efficiency has been measured to decrease in those conditions.

Table 15. Calvin Cycle Differential Expression

Locus Tag	Low Si	Low N	Low T	Gene product name
THAPSDRAFT_269057	2.98			phosphoglycerate kinase (PGK4)
THAPSDRAFT_40958	3.22	3.3	4.92	triose-phosphate isomerase (TPI1)
THAPSDRAFT_21748	-2.45	-2.16		fructose-bisphosphate aldolase (ALDO2)
THAPSDRAFT_31383	-3.61	-3.43		glyceraldehyde-3-phosphate dehydrogenase (
THAPSDRAFT_42577	-2.28		-2.04	phosphoglycerate kinase precursor (PGK3)
THAPSDRAFT_25116	-2.01			phosphoglycerate kinase (PGK5)
THAPSDRAFT_21175	-4.53	-5.62		transketolase (TKT2)
THAPSDRAFT_36462	-4.11	-4.06		triose-phosphate isomerase (TPI3)

Fucoxanthin chlorophyll a/c binding proteins (FCP) are associated with photosynthesis pigments. Nine of the 30 identified FCPs were differentially expressed (Table 16). They were all downwardly expressed in all conditions except high pH where there was no differential expression of FCPs. The downward regulation of light harvesting capacity may enable growth-limited cells to reduce potential impacts of oxidative stress.

Table 16. Fucoxanthin Chlorophyll a/c Binding Protein Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_2601	-4.6	-1.7	-4.57	-2.1		fucoxanthin chl a/c protein, lhca clade (Lhca2)
THAPSDRAFT_3815	-5.93	-3.43	-3.8	-5.97		fucoxanthin chlorophyll a/c family (Lhca6)
THAPSDRAFT_38583	-4.98	-1.28	-4.31	-2.28		fucoxanthin chlorophyll a/c protein 1 (Lhcf1)
THAPSDRAFT_38667	-5.92	-1.49	-4.18	-3.48		fucoxanthin chlorophyll a/c protein 4 (Lhcf4)
THAPSDRAFT_33018	-6.19	-1.67	-6.41	-3.45		fucoxanthin chlorophyll a/c protein 6 (Lhcf6)
THAPSDRAFT_5174	-3.9		-5.07	-4.63		fucoxanthin chlorophyll a/c protein 8 (Lhcf8)
THAPSDRAFT_268127	-5.69	-1.11	-7.68	-5.58		fucoxanthin chlorophyll a/c protein 8 (Lhcf9)
THAPSDRAFT_264921	-4.12	-2.46	-1.86	2.68		fucoxanthin chlorophyll a/c protein (lhcx1)
THAPSDRAFT_38879	-4.6	-2.6	-1.84	2.65		fucoxanthin chlorophyll a/c protein (Lhcx2)
THAPSDRAFT_31128	-4.22	-4.03	-4.35			fucoxanthin chlorophyll a/c protein (lhcx5)
THAPS_30385	-4.26	-1.45	-5.35	-3.14		fucoxanthin chlorophyll a/c protein (Lhcx6_1)

Chlorophyll synthesis showed differential expression in both the low Si and the low N conditions (Table 17). There was no other differential expression in any of the other conditions. This result is also consistent with the reduced efficiency of photosynthesis in the low N and low Si conditions.

Table 17. Chlorophyll Synthesis Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_3974	-2.85		-1.7			uroporphyrinogen III decarboxylase (UroD2)
THAPSDRAFT_31012	-3.51		-2.97			coproporphyrinogen III oxidase(Cpx1)
THAPSDRAFT_26573	-3.65		-3.28			chelatase of mg-protoporphyrin IX (Ch1H1)

Photorespiration reduces the effect of the oxygenation reaction catalyzed by Rubisco. When oxygen is utilized instead of CO_2, the enzyme ribulose 1,5 bisphosphate carboxylase/oxygenase (Rubisco) catalyzes the oxidation of 2-ribulose 1,5 bisphosphate and creates 3-phosphoglycerate and 2-phosphoglycolate which cannot be used by the Calvin cycle. In addition, 2-phosphoglycolate inhibits TPI, which is part of glycolysis as well as the Calvin cycle (Kroth, 2008). Interestingly, triose-phosphate isomerase TPI1 of the chloroplast was the only upwardly expressed gene in the Calvin cycle for low Si, low N and low T; its upward expression may counteract the effects of 2-phosphoglycolate (Table 18). O_2 and CO_2 compete for Rubisco activity; therefore, if there is a higher concentration of CO_2, there is less of a need for photorespiration. Photorespiration takes place in the chloroplast, the peroxisome, and the mitochondria (Kroth, 2008). The only differentially expressed gene in low Si, low N or low T conditions that was unique to this pathway was l-lactate dehydrogenase (EC: 1.1.3.15) (THAPSDRAFT_3353) and it was downwardly expressed. The only other two differentially expressed genes were isozymes

for glycine or serine hydroxymethyltransferase, serine methylase (EC: 2.1.2.1) (THAPS_269942, THAPSDRAFT_26031) that are also used in the glycine betaine degradation, purine nucleotides degradation, and glycine biosynthesis. THAPS_269942 was downwardly expressed in low Si and low N conditions. THAPSDRAFT_26031 was upwardly expressed in the low T condition.

Table 18. Photorespiration Differential Expression

Locus Tag	Low Si	Low N	Low T	Gene product name
THAPSDRAFT_3353	-4.96	-4.51		l-lactate dehydrogenase
THAPS_269942	-2.92	-2.57		glycine or serine hydroxymethyltransferase
THAPSDRAFT_26031			2.12	serine hydroxymethyltransferase

The C4 photosynthetic carbon assimilation cycle pathway is a carbon concentration mechanism that suppresses photorespiration (Lea and Leegood, 1993). By increasing the CO_2 level in the cell, it increases the efficiency of photosynthesis by increasing the carboxylation of 2-ribulose 1,5 bisphosphate and decreasing the competitive oxidation of 2-ribulose 1,5 bisphosphate. There is some controversy as to the working presence of the C4 pathway in diatoms (Kroth, 2008). There were enzymes identified in *T. pseudonana* for each of the reactions in this pathway and none of the genes in this pathway are unique to this pathway. Only downward expression was significantly detected in any of the conditions (Table 19). In both the low Si and low N conditions, PPC2 and the MAO1 genes transcripts were downwardly expressed. The proteins that are derived from these genes are also used in gluconeogenesis.

Table 19. C4 Differential Expression

Locus Tag	Low Si	Low N	Low T	Gene product name
THAPSDRAFT_34030	-2.64	-2.46		NAD dependent malic enzyme (MAO1)
THAPSDRAFT_268546	-3.15		-2.79	phosphoenolpyruvate carboxylase (PPC1)
THAPSDRAFT_34543	-2.6	-2.79		phospoenolpyruvate carboxylase (PPC2)
THAPSDRAFT_5500			-2.77	pyruvate phosphate dikinase (PDK1_2)

Although Rubisco was not differentially expressed, most genes expressed in the low N and low Si conditions were downwardly expressed and in the case where there was an upward expression, there was a downward expression of an isozyme of the same reaction. This is the expected result since photosynthesis efficiency is reduced in low Si and low N conditions (Berges and Falkowski, 1998). This is an indicator that carbon fixation and NADPH production is reduced. This may inhibit fatty acid synthesis, but not necessarily. If carbohydrate synthesis were decreased, it would reduce the overall usage of carbon and make more available for fatty acid synthesis. If the reduction of carbon usage by carbohydrate synthesis is more than the reduction of carbon fixation, fatty acid synthesis could increase despite the decrease in carbon fixation.

TCA and Pentose Phosphate Pathways

The TCA cycle is a catabolic process of aerobic respiration that generates energy and the reducing equivalents NADP and NADPH, which are required for palmitate fatty acid synthesis. The TCA cycle occurs in the mitochondria in eukaryotes. Isocitrate dehydrogenase (IDH1) catalyzes a reaction that generates NADPH. The gene for this gene product is THAPSDRAFT_1456 that was not significantly differentially expressed in any of the five test conditions. Malate dehydrogenase is an enzyme of the TCA cycle

that catalyzes a reaction that produces NADH. The gene for Malate dehydrogenase (THAPSDRAFT_41425) was upwardly expressed in the low N condition to a value of 2.12. It is interesting to note that IDH1 was not differentially expressed upward in low T condition that shows an indication of an increase in FA synthesis that requires NADPH.

The pentose phosphate pathway diverts glucose from glycolysis to produce pentose (ribose-5-phosphate) to make RNA, DNA, and several coenzymes (Tomanek, 2010). In addition, the pentose can be recycled into glucose-6-phosphate via the non-oxidative branch of the pathway. The oxidative branch converts glucose-6-phosphate into ribulose-5-phosphate and produces NADPH. Glucose-6-phosphate-1-dehydrogenase (EC 1.1.1.49) is the gene product for THAPSDRAFT_34514 that generates NADPH that was not differentially expressed in any condition. The non-oxidative branch is dependent on a thiamine pyrophosphate transketolase (THAPSDRAFT_21175) is highly expressed in a downward condition for both low Si and low N conditions (-4.53 and -5.62, respectively). It is not differentially expressed in any of the three other conditions. The output of the non-oxidative branch is D-erythrose-4-phosphate that is used by the shikimate pathway to produce chorismate. Chorismate is an important intermediary in the biosynthesis of the amino acids phenylalanine, tyrosine, and tryptophan as well as vitamin E and K.

In the low Si, low N, and low T conditions, the pentose phosphate pathway is not producing more NADPH relative to the replete condition. This may be most significant in the low T condition where fatty acid synthesis pathway enzymes are upwardly

expressed suggesting that it is increasing lipid accumulation (demand for NADPH), but not increasing NADPH production in the oxidative pentose phosphate pathway. The downward expression of thiamine pyrophosphate transketolase may inhibit the biosynthesis of amino acids and or other co-factors like vitamin E.

Oxidative Stress and NADPH Availability

Oxidative stress can inhibit lipid accumulation by reducing the availability of NADPH for fatty acid synthesis and by increasing fatty acid degradation. It can also cause mRNA destruction and interrupt the protein synthesis of enzymes required for lipid synthesis. Oxidation is part of the normal cellular reduction-oxidation (Redox) activity in the cell. ATP synthase produces reactive oxidation species (ROS) as it produces ATP. ROS and reactive nitrogen species (RNS) include nitric oxide (NO), super oxide (O_2^-), peroxynitrite ($ONOO^-$), nitrite (NO_2^-), nitrate (NO_3^-), and hydrogen peroxide (H_2O_2). ROS-RNS can damage DNA, degrade proteins, and trigger apoptosis (cell death). Hydroxide ions can breakdown unsaturated fatty acids into aldehydes through a process called lipid peroxidation (Tomanek, 2010). The cell has mechanisms to inhibit the accumulation of reactive species by reducing them to non-reactive species. The cell system works to keep the system in a balanced state where the rate of oxidation is balanced to the rate of reduction of the ROS. During some conditions, including changes in environmental condition or high temperature, the cell may be in a state of oxidative stress, which triggers a response to ameliorate the condition of oxidative stress to reach homeostasis. Since the diatoms were sampled during a transition from exponential

growth to stationary phase, they may have been experiencing oxidative stress at the time of sampling.

During oxidative stress period, unsaturated lipids may be degraded and there may be a shortage of NADPH that is used as a reducing equivalent in reactions that reduce ROS-RNS to non-destructive molecular forms. When a cell reduces photosynthesis efficiency, less NADPH is produced in the light reaction pathway which may further increase oxidative stress.

Oxidative stress is primarily caused by peroxynitrite ($ONOO^-$) that is created when super oxide ($O2^-$) reacts with Nitric oxide (NO^-). Peroxynitrite and its damaging effects can be limited if superoxide is consumed by super oxide dismutase (SOD). SOD is an enzyme that catalyzes a reaction that converts $O2^-$ into hydrogen peroxide (H_2O_2). Hydrogen peroxide is also reactive; therefore it needs to be acted on, too. There are two main ways that H_2O_2 are converted to H_2O.

Glutathione is the substrate for a reaction that is catalyzed by glutathione peroxidases (GPx) that converts H_2O_2 into H_2O. When glutathione is oxidized, it becomes (GSSG) glutathione disulfide. It then needs to be reduced back to glutathione so that it can be reused as a substrate to convert H_2O_2 to H_2O. The enzyme to catalyze the reduction of GSSG to GSH is called glutathione reductase. This reaction requires NADPH. The other path to degrading H_2O_2 depends on peroxiredoxins (PRx) that depends on Thioredoxin (TRx) that is reduced using NADPH and catalyzed by thioredoxin reductase (Murphy, 2009).

Several enzymes associated with redox homoeostasis were found in *T. pseudonana.* Most of the redox-related genes that were significantly differentially expressed were in low N and low Si conditions. Low Fe, low T, and high pH showed no significant differential expression.

Glutathione reductase (THAPSDRAFT_41433) was upwardly expressed at a value of 3.04 in the low N condition. Glutathione reductase (GR) is a gene product that catalyzes a reaction that uses NADPH as a reducing equivalent to reduce glutathione disulfide to glutathione (a major cellular antioxidant) to enable it to scavenge ROS to reduce oxidative stress. GR is unique to this process and is not used in other pathways. This may be an indication of oxidative stress in the low N condition.

Another indicator of oxidative stress is superoxide dismutase (SOD) when it is upwardly expressed. Mitochondrial manganese superoxide dismutase (SOD) has multiple isoforms in *T. pseudonana*. SOD gene THAPSDRAFT _32874 was expressed downwardly for both the low Si and low N conditions (-2.36 and -2.02 respectively) and was not significantly differentially expressed in the low Fe and high pH conditions. Notably, at low temperature the gene was expressed above the -2.0 cutoff at -1.7. This suggests that there is not oxidative stress in the low N and low Si conditions.

Cytochrome C peroxidase (THAPSDRAFT_35409) gene was also downwardly expressed at a value of -2.51 in the low N condition and not expressed differentially relative to the replete condition in the other four environmental conditions. Peroxidase catalyzes the reaction, which transforms H_2O_2 to H_20.

There is not conclusive evidence of oxidative stress in the low N condition because there are contradictory indicators. In support of oxidative stress in the low N condition, the heat shock transcription factor was upwardly expressed in the low N condition. See regulation section. The other conditions do not show evidence of oxidative stress.

Regulatory Elements

In addition to catalytic activity of enzymes and substrate/co-factor availability, there are several other regulatory factors that may effect lipid accumulation. Post translational modification, phosphorylation/dephosphorylation of an enzyme, allosteric binding and transportation also contribute to the control of lipid accumulation. Transcription and translation of a gene can take several seconds to minutes to accomplish. Allosteric, phosphorylation, and cAMP signaling regulation can respond to environmental conditions within seconds.

Regulation of Acetyl CoA Carboxylase

Acetyl CoA carboxylase (ACC) is a key regulatory enzyme for fatty acid synthesis. There are multiple posttranslational events that can modify the activity of ACC in time frames shorter than the time to promote, transcribe and translate a gene. In addition there are three states to ACC activity: inactive, active, and highly active.

There are two distinct forms of ACC (Table 20). The heteromeric form which is primarily found in prokaryotes consists of four subunit proteins: Biotin carboxyl carrier

protein (BCCP), Biotin carboxylase (BC), alpha –carboxyltransferase (α-CT), and beta-carboxyltransferase (β-CT). The homomeric form that is found in eukaryotes is a large polypeptide with six pfam domains. Most plants have both forms with the heteromeric form in the plastid and the homomeric form in the cytosol although some plants like wheat and rice which have the only the homomeric form in both the cytosol and the plastid (Sasaki, 2004). There are two forms of homomeric ACC: ACC1and ACC2 that is thought to block β-oxidation by preventing the transport of fatty acids into the mitochondria (Sasaki, 2004).

Table 20. Homomeric and Heteromeric Forms of ACC

	HETEROMERIC			
Organism	**BCCP**	**BC**	**α-CT**	**β-CT**
A. thaliana	AT5G16390,AT5G15530	AT5G35360	AT2G38040	ArthCp031
T. pseudonana				
P. tricornutum				
C. reinhardtii	CHLREDRAFT_185478, _183660	CHLREDRAFT_122970	CHLREDRAFT_184945	CHLREDRAFT_133238
H. sapiens				
E. coli	EcDH1_0451	EcDH1_0450	EcDH1_3418	EcDH1_1340
Synechococcus RCC 307	SynRCC307_0032	SynRCC307_0341	SynRCC307_1584	SynRCC307_0993
P. marinus	Pro0027	Pro0074	Pro0534	Pro0859
	HOMOMERIC ACC			
A. thaliana	AT1G36160 (1)			
T. pseudonana	THAPS_6770, _12234(2)			
P. tricornutum	PHATRDRAFT_54926,_55209			
C. reinhardtii				
H. sapiens	ACACA, ACACB			
E. coli				
Synechococcus RCC 307				
P. marinus				

There are two homomeric forms in *T. pseudonana*: ACC1 (THAPS_6770) and ACC2 (THAPSDRAFT_12234). THAPS_6770 has a product name of ACC and BLASTp resulted in the best homology to PHATRDRAFT_54926(ACC1) with a high bit

score of 2939. The homology to Arabidopsis ACC1 had a very low 0.0 e-00 e-value (too low to register) and a bit score of 712. THAPSDRAFT_12234 has an assigned product name of hypothetical protein although the reaction EC numbers, pfams, and KOG terms identify it as ACC. A blast of THAPSDRAFT_12234 identified it as ACC2 as compared to its best homology to PHATRDRAFT_55209 which had a gene name ACC2 with a bit score of 2939.

There are heteromeric forms of ACC in plants and bacteria (Sasaki, 2004). Heteromeric forms of ACC were found in the IMG database in Arabidopsis thaliana, *Chlamydomonas reinhardtii, Escherichia coli*, *Synechococcus,* and *Prochlorococcus marinus marinus*. No heteromeric forms were found in *T. pseudonana* or *P. tricornutum*. The lack of heteromeric ACC forms in diatoms has not been previously reported. This is a significant finding because it may in part explain why diatoms accumulate lipids differently than *Chlamydomonas reinhardtii*.

ACC1 (THAPS_6770) was the only ACC gene differential expressed in any of the conditions of the Mock experiment as compared to the nutrient-replete condition. ACC1 is thought to be active in the chloroplast for short-chain fatty acid synthesis. THAPSDRAFT_12234 (ACC2) was not differentially expressed in any of the Mock experimental conditions. ACC2 is thought to be active in the endoplasmic reticulum in the elongation process.

ACC has three activity levels: inactive, active, and highly active. In the Mock experiment ACC (THAPS_6770) was downwardly expressed relative to the replete

condition in low Si (-4.39) and low N (-4.97) conditions where there was also downward expression of fatty acid synthesis. Notably, it was also downwardly expressed at a value of -1.7 in the low temperature condition where there is evidence of fatty acid biosynthesis upward expression. Due to the state variability of the ACC enzyme the gene expression of ACC may be secondary to the activation level effect on fatty acid initiation. Factors affecting the activity level of ACC may suggest the ACC activity state.

The monomeric form of ACC is considered a low activity state. Phosphorylation can transition the monomeric form of ACC it to an inactive state. The phosphorylation is thought to be triggered by signal cascade resulting from a low glucose condition. The activity can transition back to the low activity by the dephosphorylation of ACC by protein phosphatase. In H. sapiens, high blood sugar levels cause the pancreas to secrete insulin that in turn promotes the dephosphorylation of ACC. There has yet to be identification of kinases and phosphates involved in these activities. Plants do not produce insulin (Sasaki, 2004). The *T. pseudonana* genome does not contain genes associated with insulin production.

ACC in the monomeric low activity form can be transitioned to the active forms allosterically through the binding of citrate to ACC. Conversely, ACC activity can be reduced from its active state to its low activity monomeric state by the presence of palmitoyl-CoA that is the product of fatty acid synthesis. This action provides a negative feedback loop to prevent the over-accumulation of free fatty acids. In the low N

condition, citrate production is up in the TCA cycle. This could be increasing the activity of ACC in the low N condition.

Carnitine Acetyl-CoA Transferase (CAT)

Carnitine acetyl-CoA transferase (CAT) is present in both mitochondria and peroxisomes. It transfers activated acetyl groups to carnitine to form acetylcarnitine, which can be shuttled across membranes. CAT transports fatty acids to mitochondria for β-oxidation. Lack of CAT can inhibit β-oxidation. Malonyl-CoA blocks CAT activity to prevent β-oxidation when fatty acid synthesis is up-regulated. There are two forms of CAT in *T. pseudonana*: CAT1-2 and CAT2. CAT1-2 in *T.* pseudonana has a locus tag of THAPSDRAFT_13065 and is homologous to PHATRDRAFT_859 in *P. tricornutum*. In *T. pseudonana* CAT2 is locus tag THAPSDRAFT_31762, which is homologous to PHATRDRAFT_48078 in *P. tricornutum*. There were homologs for CAT 1-2 in H. sapiens (CPT2) and S. cerevisiae (Yaro35w). There were no homologs for either CAT1-2 or CAT2 in *Arabidopsis thaliana*, *Chlamydomonas reinhardtii, Escherichia coli*, *Synechococcus* or *Prochlorococcus marinus marinus*. This is a significant finding because it indicates that diatoms may regulate β-oxidation differently than C. reinhardtii, A. thaliana and bacteria. CAT1-2 and CAT2 showed no differential expression in any of the five environmental conditions in the Mock experiment. Presumably in the replete condition CAT for β-oxidation is not needed. The fact that it was not differentially expressed upwardly may indicate that transportation of fatty acids is inhibiting β-oxidation.

PII Signaling Protein

Plants have a signaling protein called PII that has been shown to inhibit chloroplastic ACC activity in *A. thaliana*. The inhibiting effect of PII is reversed in the presence of 2-oxoglutarate (2-OX) (Bourrellier, 2010). This demonstrates a reversible control of the metabolic effects of PII. This implies that when central metabolism increases the level of 2-oxoglutarate from TCA or other sources there is a fast-acting increase in ACC activity as the activity level transitions from low activity state to the active state. PII acts like an inhibit switch for ACC activity until 2-OX levels increase to switch ACC activity on.

PII is a regulatory protein that is highly conserved in bacteria, archea, and plants (Mizuno, 2007). It is thought to sense carbon energy availability and is involved in the mediating of carbon and nitrogen assimilation (Forchhammer, 1995). PII interacts with other enzymes like ACC as well as transcription factors and transporters like nitrate transporters to regulate cellular metabolism. PII knockout experiments have shown altered levels of carbon and nitrogen metabolites like starch and glutamine (Bourrellier, 2010).

PII has been identified in *Arabidopsis, E. coli,* and *Synechococcus* (Table 21). In *Synechococcus* it has been shown to be a nitrogen regulatory protein (Forchhammer, 1995). The PII protein signals nitrogen status by being phosphorylated or dephosphorylated at a serine residue. E. coli contains a PII protein GlnK that regulates ammonium uptake as it interacts with an ammonium transporter (Bourrellier, 2010).

Table 21. PII Regulatory Proteins Not Found in *T. Pseudonana*

Genome Name	**Locus Tag**	**Gene Name**	**Length**	**E-value**	**pfam**
Synechococcus sp. RCC 307	SynRCC307_1895	PII (glnB)	112		pfam00543
Prochlorococcus marinus	Pro1616	PII (glnK)	112	2.00E-58	pfam00543
Escherichia coli K12	EcDH1_1115	PII (glnB)	112	1.00E-44	pfam00543
Escherichia coli K12	EcDH1_3159	PII (glnK)	112	2.00E-40	pfam00543
Arabidopsis thaliana	AT4G01900	PII (GLB1)	196	6.00E-36	pfam00543
Chlamydomonas reinhardtii	CHLREDRAFT_206116	PII (GLB1)	205	1.00E-24	pfam00543
Phaeodactylum tricornutum					
Thalassiosira pseudonana					
Homo sapiens					

An evaluation of the *T. pseudonana* and *P. tricornutum* genomes revealed that there is not a PII gene in either of these diatoms. The gene was found in *E. coli*, *A. thaliana*, *C. reinhardtii*, *Prochlorococcus*, and *Synechococcus*. Table 21 shows the result of a BLASTp against the PII protein (glnB) in *Synechococcus*. The gene in Synechococcus was the topic of the Forchhammer (1995) paper. All of the PII genes contain pfam00543. IMG/ER was used to search for this pfam in *T. pseudonana* and *P. tricornutum* genomes and no genes were found with this pfam. This is a significant finding because it implies that diatoms have a different mechanism not only for controlling the activation level of ACC but also a broader carbon and nitrogen assimilation mechanism. If diatoms do not have PII functionality, it may help explain why diatoms have been shown to accumulate lipid differently than Chlamydomonas reinhardtii. Further research may uncover a new metabolism control mechanism in diatoms.

There is evidence that ACC activity is modulated through several mechanisms that lay outside the scope of this analysis. For example, the pH level in the stroma can affect the activity level of ACC. When photosynthesis occurs, the pH in the stroma can increase from 7 to 8. Mg ion concentration can also increase from 1mM to 3mM and the activity level of ACC can triple in A. thaliana leaves (Sasaki, 2004). This is an example of how light can increase the level of ACC activity in several ways. It also illustrates how there are many non-transcription factors that effect fatty acid synthesis. It underscores the limitations of this transcription analysis on the regulation of fatty acid synthesis.

PAT Proteins in Lipid Droplets

PAT proteins bind to intracellular lipid droplets and regulate cellular lipid metabolism. Members of this family of proteins are found in H. sapiens as well as insects, slime molds, and fungi (Bickel, 2009).

There are three species of PAT proteins: Perlipin, Adipose differentiation-related protein (ADRP) also called adipophilin, and the Tail interacting protein 4t (TIP47). PAT is an abbreviation based on the terms Perlipin, Adipose, and Tail. There are three forms of Perilipin: A, B, and C that are different splicing conditions of one gene. A is the most abundant one. In one form, Perilipin acts as a protective coating on the lipid droplet to prevent lipases from breaking down TAGs. In another form, Perilipin promotes TAG degradation and may have a role in the activation of the first step in the degradation of triglycerides. The step is catalyzed by adipose triglyceride lipase (ATGL) and its co-activator CGI-58 perilipin interacting protein. It suggests that the phosphorylation of

perilipin A results in its disassociation from CGI-58 which then associates with ATGL on lipid droplets to allow lipolysis (Ehnholm, 2009).

ADRP plays a role in the assembly of lipid droplets. Upward expression results in the increased formation of lipid droplets. Accumulation of intracellular TAGS decreases the degradation of ADRP. TIP47 plays a role similar to ADRP but tends to favor the biosynthesis of phospholipids verses TAGS. Other PAT protein members are S3-12 and OXPAT (MLDP, LSDP5) (Simons, 2011).

Five PAT proteins were found in H. sapiens (PLIN, ADRP, TIP47 (M6PRBP), S3-12 (KIAA1881) and OXPAT (LSDP5). They all had pfam03036 in them. I was unable to find any of the genes for these proteins in Arabidopsis thaliana, Chlamydomonas reinhardtii, Escherichia coli, Synechococcus or Prochlorococcus marinus marinus. This is significant because it implies that there is no protective cover on lipid droplets or the protective cover on the lipid droplets has yet to be discovered. If diatoms do have a regulated protective lipid droplet in a form other than PLIN, control of it allosterically could lead to the ability to enable diatoms to release their TAGs.

Transcription Factors

Numerous transcription factors genes have been identified in *T. pseudonana*, however only two differentially expressed in any of the five conditions. In the low N condition, heat shock factor35 (HSF35) (THAPS_269238) expressed upwardly at a value of 2.57. HSFs regulate the expression of heat shock proteins. Stress conditions like high temperate can cause the misfolding of proteins. Heat shock proteins (HSP) are involved

in the folding and unfolding of other proteins. Production of high levels of heat shock proteins can also be triggered by exposure to different kinds of environmental stress conditions, such as infection, exposure of the cell to toxins, starvation, hypoxia (oxygen deprivation) nitrogen deficiency (in plants), or water deprivation. Heat shock proteins (or stress proteins) are known to be upwardly expressed in plants with nitrogen deficiency so it is not surprising that HSP35 is upwardly expressed in the low nitrogen condition and may indicate that the low N condition is causing misfolding of proteins and potentially may inhibit enzyme efficiency.

The other expressed transcription factor that was downwardly differentially expressed in the low temperature condition was a Sir2-type regulatory transcription factor silent information regulator (SIR) protein (THAPS_264494) which was expressed downwardly at a value of -2.13. SIR protein complex has been implicated in transcription silencing and suppression of recombination. It is also thought to promote longevity and genome stability in response to calorie restriction (Blander, 2004).

Transporters

Nine out of 82 genes identified as transporter activity proteins (GO:0005215) showed differential expression in any of the five environmental conditions (Table 22). As expected, silicic acid transporters SIT1 and SIT2 (THAPSDRAFT_268895, THAPS_41392) were strongly upwardly expressed in the low Si condition at values of 7.15 and 3.15, respectively. Interestingly, nitrate transporters NRT1 and NRT2 were not differentially expressed at low N although xanthine/uracil permease was upwardly

expressed at a value of 3.07. Uracil is part of the pyrimidine family that occurs as a component of ribonucleic acid (RNA). Xanthine is a precursor of uric acid, which contains for molecules of nitrogen. Nitrate transporters, and xanthine uracil permease were downwardly expressed for both low T and low Si conditions.

Table 22. Transporter Activity Differential Expression

Locus Tag	Low Si	Low Fe	Low N	Low T	CO2	Gene product name
THAPSDRAFT_27414	-3.05			-2.03		nitrate transporter (NRT1)
THAPSDRAFT_269274	-3.27			-2		nitrate transporter (NRT2)
THAPSDRAFT_268895	7.15					silicic acid transporter (SIT1)
THAPS_41392	3.15					silicic acid transporter (SIT2)
THAPSDRAFT_13485	-2.79	-1.78	3.07	-4.51		xanthine uracil permease
THAPSDRAFT_262849	-3.61	-2	-4.3	-2.39		succinate/fumarate mitochondrial transporter
THAPSDRAFT_264353	2.34					chromate transporter
THAPSDRAFT_31085	2.74	3.15				amino acid/polyamine transporter
THAPSDRAFT_681	3.06					amino acid/polyamine transporter

Succinate/fumarate mitochondrial transporter was downwardly expressed in all conditions except the high pH (CO_2) condition. Succinate is converted to fumarate by succinate dehydrogenase in the TCA cycle. Fumarate is converted and then converted to malate in TCA or can be exported from the mitochondrial through the mitochondrial transporter. The low Si condition showed upward differential expression of chromate and amino acid/polyamine transporters. No differential expression of transporters was detected in the high pH condition.

CONCLUSION

The availability of the *Thalassiosira pseudonana* genome database allowed the systematic building of a BioCyc metabolic pathway model, which reflects the capability of the diatom to respond to its environment. The availability of a whole genome expression profiling dataset for *Thalassiosira pseudonana* (Mock, 2008) provides a unique opportunity to analyze the potential for lipid accumulation in *T. pseudonana* within the context of the whole cell metabolic processes. This enabled the investigation to assess not only lipid accumulation processes but also carbon and protein accumulation pathways that compete with lipid synthesis. In most cases, the enzymes of a pathway had the same directionality indicating a coordinated gene regulation and directed control. Results were both anticipated and surprising. The photosystem responded as expected in a downward direction in stressed conditions (Table 23). The low temperature condition response was in line with previous experimental results and showed a response favoring lipid accumulation: fatty acid producing transcripts were upwardly expressed and fatty acid degradation pathways were downwardly expressed. In the low nitrogen and low silicon conditions, lipid accumulation and carbohydrate accumulation processes were downwardly regulated. Silicate transporters were strongly upwardly expressed in the low silicon condition and amino acids were degraded in the low nitrogen condition. Surprisingly, the nitrogen transporters were not upwardly expressed in the low nitrogen conditions. It is difficult to draw conclusions from one sample point in a dynamic metabolic system undergoing a transition into a stressful environmental condition. In the future, time series transcriptome experiments may further illuminate this complex and dynamic metabolic system.

Table 23. Summary of the Number of Differentially Expressed Genes by Category

	Low Si		Low Fe		Low N		Low T		High pH	
	Down	Up	Down	Up	Down	Up	Down	Up	Down	Up
Fatty Acid										
Synthesis	5				4			8		
Degradation		3								
Carbohydrates										
Synthesis										
Chrysolaminarin	2		1		2					
Sucrose	4	1	1		2	1	1			
Glucosneogenesis	7	1	1		4	2	3	1		
Degradation										
Proteins										
Synthesis	10	1	6	1	8	3	6	10	1	1
Degradation	7	2	4		6	7	7	2		
Energy										
Light reactions	2				2	1		1		
Calvin	6	2			4	1	1	1		
Photorespiration	2		1		2			1		
C4	3		1		2		2			
FCP	11		10		11		8	2		
Glycolysis-plastidic	4	1	1		1	3		2		
Glycolysis-Cyto	1	1			2	1	1	2		
TCA						4				
Pentose Phos non OX	2				2					
Pentose Phos OX										
Rubisco Shunt					2	3		3		
Regulation										
ACC	1				1					
CAT										
Transcription factors						1	1			

The pathway model building process yielded several valuable insights into the genome and inner workings of diatoms. Unlike *A. thaliana* and *C. reinhardtii*, Acetyl-CoA carboxylase exists only in its homomeric forms. Unlike *E. coli*, *A. thaliana*, *C. reinhardtii*, *Prochlorococcus*, and *Synechococcus, T. pseudonana* and *P. tricornutum* do not have the signaling protein PII that has been shown to play an important role in carbon and nitrogen assimilation as well as the regulation of lipid synthesis. Unlike *Arabidopsis thaliana*, *Chlamydomonas reinhardtii, Escherichia coli*, *Synechococcus* or *Prochlorococcus marinus marinus*, both *T. pseudonana and P. tricornutum* have genes for Carnitine acetyl-CoA transferase (CAT). *T. pseudonana* like *A. thaliana* and *C. reinhardtii* do not have PAT lipid droplet regulatory proteins that are found in H. sapiens. These findings further refine our understanding of diatoms and underscore our future challenges.

As the cost of sequencing continues to drop and as bioinformatic tools continue to be adopted, the library of reusable metabolic models will quickly grow. Analysis tools will evolve to be more automated and standardized. Indicators of lipid, carbohydrate and protein synthesis, oxidative stress, transcription factors, transporters, cell cycle control, and photosynthesis activity would be easy to automate and would be a useful starting point for any analysis. Technology to create flux balance analysis models will continue to evolve and eventually there will be a large library of flux balance analysis models available. Flux balance analysis models can be used to simulate nutrient limitation, gene knockouts and the effects of differential gene expression. These metabolite flux simulations can be used to refine hypothesizes and design new experiments. For example, FBA can be

used to define optimal nutrient formulas and gene based strain selection for maximum algal lipid yields.

Kinetic modeling tools like Virtual Cell developed at the University of Connecticut or ProcessDB by Integrative Bioinformatics will be more widely used in the future to simulate the dynamics of the cell based on the physics of diffusion and biochemical reactions. Kinetic models can solve differential equations to model complex signaling networks and biological switches (Tyson, 2003) as well as protein movement through cell sub compartments (Phair, 2000). The adoption of these computational biology tools will widen our breath of understanding, quicken our learning cycles, and reduce experimental costs.

REFERENCES

Journal Articles

Aguilar, Pablo S., and Diego de Mendoza. "Control of Fatty Acid Desaturation: A Mechanism Conserved from Bacteria to Humans." *Molecular Microbiology* 62.6 (2006): 1507-14. Print.

Alonso, Diego López, et al. "Acyl Lipid Composition Variation Related to Culture Age and Nitrogen Concentration in Continuous Culture of the Microalga Phaeodactylum Tricornutum." *Phytochemistry* 54.5 (2000): 461-71. Print.

Arao, Tomohito, Akihiko Kawaguchi, and Mitsuhiro Yamada. "Positional Distribution of Fatty Acids in Lipids of the Marine Diatom Phaeodactylum Tricornutum." *Phytochemistry* 26.9 (1987): 2573-76. Print.

Armbrust, E. Virginia. "The Genome of the Diatom Thalassiosira Pseudonana: Ecology, Evolution and Metabolism." *Science* 306 (2004): 79-86. Print.

Bazan, Nicolas G., and Geoffey Allan. "Signal Transduction and Gene Expression in the Eye: A Contemporary View of the Pro-Inflammatory, Anti-Inflammatory and Modulatory Roles of Prostaglandins and Other Bioactive Lipids." *Survey of ophthalmology* 41 (1997): S23-S34. Print.

Beard, Daniel A., Shou-dan Liang, and Hong Qian. "Energy Balance for Analysis of Complex Metabolic Networks." *Biophysical journal* 83.1 (2002): 79-86. Print.

Behrens, Paul. "Photobioreactors and Fermentors: The Light and Dark Sides of Growing Algae, Algal Culturing Techniques." *Academic Press* (2005): 189-99. Print.

Beisson, Frederic, et al. "Arabidopsis Genes Involved in Acyl Lipid Metabolism. A 2003 Census of the Candidates, a Study of the Distribution of Expressed Sequence Tags in Organs, and a Web-Based Database." *Plant Physiol.* 132.2 (2003): 681-97. Print.

Berges, J. "Effects of Temperature on Growth Rate, Cell Composition and Nitrogen Metabolism in the Marine Diatom Thalassiosira Pseudonana (Bacillariophycease)." *Marine Ecology Progress* 225 (2002): 139-46. Print.

Berges, J. A., et al. "Differential Effects of Nitrogen Limitation on Photosynthetic Efficiency of Photosystems I and Ii in Microalgae." *Plant Physiol.* 110.2 (1996): 689-96. Print.

Berges, John A., and Paul G. Falkowski. "Physiological Stress and Cell Death in Marine Phytoplankton: Induction of Proteases in Response to Nitrogen or Light Limitation." *Limnology and Oceanography* 43.1 (1998): 129-35. Print.

Berges, John A., and Paul J. Harrison. "Relationships between Nitrate Reductase Activity and Rates of Growth and Nitrate Incorporation under Steady-State Light or Nitrate Limitation in the Marine Diatom *Thalassiosira Pseudonana* (Bacillariophyceae)." *Journal of Phycology* 31.1 (1995): 85-95. Print.

Bickel, P. "Pat Proteins, an Ancient Family of Lipid Droplet Proteins That Regulate Cellular Lipid Stores." *Biochim Biophys Acta.* 1791(6) (2009): 419–40. Print.

Blander, Gil, and Leonard Guarente. "The Sir2 Family of Protein Deacetylases." *Annual Review of Biochemistry* 73.1 (2004): 417-35. Print.

Borowitzka, Michael. "Algal Biotechnology Products and Processes - Matching Science and Economics." *Journal of Applied Phycology* 4 (1992): 267-79. Print.

Bourrellier, Ana Belen Feria. "Chloroplast Acetyl-Coa Carboxylase Activity Is 2-Oxoglutarate–Regulated by Interaction of Pii with the Biotin Carboxyl Carrier Subunit." *PNAS* 107.1 (2010): 502-07. Print.

Bowler, Chris, et al. "The Phaeodactylum Genome Reveals the Evolutionary History of Diatom Genomes." *Nature* 456.7219 (2008): 239-44. Print.

Brownsey, R.W. "Regulation of Acetyl-Coa Carboxylase." *Biochemical Society* 34.part 2 (2006). Print.

Brutlag, Douglas L. "Genomics and Computational Molecular Biology." *Current Opinion in Microbiology* 1 (1998): 340-45. Print.

Burman, Chloe, and Nicholas T. Ktistakis. "Regulation of Autophagy by Phosphatidylinositol 3-Phosphate." *FEBS Letters* 584.7 (2010): 1302-12. Print.

Caviglia, J. Matias, et al. "Phosphatidylcholine Deficiency Upregulates Enzymes of Triacylglycerol Metabolism in Cho Cells." *J. Lipid Res.* 45.8 (2004): 1500-09. Print.

Chauton, Matilde Skogen, et al. "Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom Phaeodactylum Tricornutum Acclimated to Light/Dark Cycles." *Plant Physiology* 161.2 (2013): 1034-48. Print.

Chen, Wei. " A High Throughput Nile Red Method for Quantitative Measurement of Neutral Lipids in Microalgae." *Journal of Microbiological Methods* 77 (2009): 41-47. Print.

Clarke, H. T., and Abraham Mazur. "The Lipids of Diatoms." *Journal of Biological Chemistry* 141.1 (1941): 283-89. Print.

Coe, Natalie Ribarik, et al. "The Fatty Acid Transport Protein (Fatp1) Is a Very Long Chain Acyl-Coa Synthetase." *Journal of Biological Chemistry* 274.51 (1999): 36300-04. Print.

Covert, Markus W., Christophe H. Schilling, and Bernhard Palsson. "Regulation of Gene Expression in Flux Balance Models of Metabolism." *Journal of Theoretical Biology* 213.1 (2001): 73-88. Print.

Davis, Aubrey K., Mark Hildebrand, and Brian Palenik. "Gene Expression Induced by Copper Stress in the Diatom Thalassiosira Pseudonana." *Eukaryotic Cell* 5.7 (2006): 1157-68. Print.

Dehesh, Katayoon, et al. "Production of High Levels of 8:0 and 10:0 Fatty Acids in Transgenic Canola by Overexpression of *Ch Fatb2*, a Thioesterase Cdna from *Cuphea Hookeriana*." *The Plant Journal* 9.2 (1996): 167-72. Print.

Domergue, Frederic. "New Insight into Phaeodactylum Tricornutum Fatty Acid Metabolism. Cloning and Functional Characterization of Plastidial and Microsomal 12-Fatty Acid Desaturases." *Plant Physiology,* Vol. 131 (2003): pp. 1648–60,. Print.

Dunahay, Terri. "Manipulation of Microalgal Lipid Production Using Genetic Engineering." *Applied Biochemistry and Biotechnology* 57.58 (1996): 223-31. Print.

Dundr, Miroslav, et al. "A Kinetic Framework for a Mammalian Rna Polymerase in Vivo." *Science* 298.5598 (2002): 1623-26. Print.

E Ginter. "Polyunsaturated Fatty Acid N-3: New Data on Heart Disease, Cancer, Immune Resistance and Mental Depression." *Bratisl Lek Listy* 111.12 (2010): 680-85. Print.

Elsey, Danielle. "Fluorescent Measurement of Microalgal Neutral Lipids." *Journal of Microbiological Methods* 68 (2007): 639-42. Print.

Fabris, Michele, et al. "The Metabolic Blueprint of Phaeodactylum Tricornutum Reveals a Eukaryotic Entner-Doudoroff Glycolytic Pathway." *The Plant Journal* (2012): no-no. Print.

Feria Bourrellier, Ana Belén, et al. "Metabolite Regulation of the Interaction between Arabidopsis Thaliana Pii and N-Acetyl-L-Glutamate Kinase." *Biochemical and Biophysical Research Communications* 387.4 (2009): 700-04. Print.

Fernie, Alisdair R., Fernando Carrari, and Lee J. Sweetlove. "Respiratory Metabolism: Glycolysis, the Tca Cycle and Mitochondrial Electron Transport." *Current Opinion in Plant Biology* 7.3 (2004): 254-61. Print.

Field, Christopher B., et al. "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." *Science* 281.5374 (1998): 237-40. Print.

Fokina, Oleksandra, et al. "A Novel Signal Transduction Protein Pii Variant from Synechococcus Elongatus Pcc 7942 Indicates a Two-Step Process for Nagk-Pii Complex Formation." *Journal of Molecular Biology* 399.3 (2010): 410-21. Print.

Forchhammer, K, and N Tandeau de Marsac. "Functional Analysis of the Phosphoprotein Pii (Glnb Gene Product) in the Cyanobacterium Synechococcus Sp. Strain Pcc 7942." *J. Bacteriol.* 177.8 (1995): 2033-40. Print.

Franklin, Scott. "Prospects for Molecular Farming in the Green Alga Chlamydomonas Reinhardtii." *Current Opinion in Plant Biology* 7 (2004): 159-66. Print.

Ginter, E. "Polyunsaturated Fatty Acids N-3: New Data on Heart Disease, Cancer, Immune Resistance and Mental Depression." *Bratislavske Lekarske Listy* 111(12 (2010): 680-85. Print.

Goltsov, Alexey, et al. "Kinetic Modelling of Nsaid Action on Cox-1: Focus on in Vitro/in Vivo Aspects and Drug Combinations." *European Journal of Pharmaceutical Sciences* 36.1 (2009): 122-36. Print.

Gostincar, C., et al. "Expression of Fatty-Acid-Modifying Enzymes in the Halotolerant Black Yeast Aureobasidium Pullulans (De Bary) G. Arnaud under Salt Stress." *Stud Mycol* 61.1 (2008): 51-59. Print.

Guckert, James B., and Keith E. Cooksey. "Triglyceride Accumulation and Fatty Acid Profile Changes in Chlorella (Chlorophyta) During High Ph-Induced Cell Cycle Inhibition1." *Journal of Phycology* 26.1 (1990): 72-79. Print.

Guschina, Irina A., and John L. Harwood. "Lipids and Lipid Metabolism in Eukaryotic Algae." *Progress in Lipid Research* 45.2 (2006): 160-86. Print.

Harwood, John L., and Irina A. Guschina. "The Versatility of Algae and Their Lipid Metabolism." *Biochimie* 91.6 (2009): 679-84. Print.

Hay, Jordan, and Jörg Schwender. "Metabolic Network Reconstruction and Flux Variability Analysis of Storage Synthesis in Developing Oilseed Rape (Brassica Napus L.) Embryos." *The Plant Journal* 67.3 (2011): 526-41. Print.

Henry, Christopher S., et al. "Connecting Genotype to Phenotype in the Era of High-Throughput Sequencing." *Biochimica et Biophysica Acta (BBA) - General Subjects* In Press, Corrected Proof (2011). Print.

Hildebrand, Mark, and Katherine Dahlin. "Nitrate Transporter Genes from the Diatom *Cylindrotheca Fusiformis* (Bacillariophyceae): Mrna Levels Controlled by

Nitrogen Source and by the Cell Cycle." *Journal of Phycology* 36.4 (2000): 702-13. Print.

Hockin, Nicola L., et al. "The Response of Diatom Central Carbon Metabolism to Nitrogen Starvation Is Different to That of Green Algae and Higher Plants." *Plant Physiology* (2011). Print.

Holland, Ross, and D. Grahame Hardie. "Both Insulin and Epidermal Growth Factor Stimulate Fatty Acid Synthesis and Increase Phosphorylation of Acetyl-Coa Carboxylase and Atp-Citrate Lyase in Isolated Hepatocytes." *FEBS Letters* 181.2 (1985): 308-12. Print.

Hu, Qiang. "Microalgal Triacylglycerols as Feedstocks for Biofuel Production: Perspectives and Advances." *The Plant Journal* 54 (2008): 621-39. Print.

Huang, Guan-Hua, Gu Chen, and Feng Chen. "Rapid Screening Method for Lipid Production in Alga Based on Nile Red Fluorescence." *Biomass and Bioenergy* 33.10 (2009): 1386-92. Print.

Iniative, Arabidopsis Genome. "Analysis of the Genome Sequence of the Flowering Plant *Arabidopsis Thaliana*." *Nature* 408 (2000). Print.

Irigoien, Xabier. "Some Ideas About the Role of Lipids in the Life Cycle of Calanus Finmarchicus." *Journal of Plankton Research* 26 (2004): 259-63. Print.

Jako, Colette, et al. "Seed-Specific over-Expression of an Arabidopsis Cdna Encoding a Diacylglycerol Acyltransferase Enhances Seed Oil Content and Seed Weight." *Plant Physiol.* 126.2 (2001): 861-74. Print.

Jiang, Shan, et al. "Activated Ampk and Prostaglandins Are Involved in the Response to Conjugated Linoleic Acid and Are Sufficient to Cause Lipid Reductions in Adipocytes." *The Journal of Nutritional Biochemistry* In Press, Corrected Proof. Print.

John Ohlrogge, Allen D., Berguson B., DellaPenna D., Shachar-Hill Y., Stymne S. "Driving on Biomass." *Science* 324 (2009). Print.

Kalogeropoulos, Nick, et al. "Nutritional Evaluation and Bioactive Microconstituents (Carotenoids, Tocopherols, Sterols and Squalene) of Raw and Roasted Chicken Fed on Dha-Rich Microalgae." *Food Research International* 43.8: (2006):13. Print.

Kates, M., and B. E. Volcani. "Lipid Components of Diatoms." *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 116.2 (1966): 264-78. Print.

Kelder, Thomas. "Finding the Right Questions: Exploratory Pathway Analysis to Enhance Biological Discovery in Large Datasets." *PLOS Biology* 8.8 (2010). Print.

Keseler, Ingrid M., et al. "Ecocyc: Fusing Model Organism Databases with Systems Biology." *Nucleic Acids Research* 41.D1 (2013): D605-D12. Print.

Khalil, Maroun Bou, et al. "The Level and Compartmentalization of Phosphatidate Phosphatase-1 (Lipin-1) Control the Assembly and Secretion of Hepatic Vldl." *J. Lipid Res.* 50.1 (2009): 47-58. Print.

Kim, Minjung, et al. "Gene Expression Profiling of *Dunaliella* Sp. Acclimated to Different Salinities." *Phycological Research* 58.1 (2010): 17-28. Print.

Kliewer, Steven. "Fatty Acids and Eicosanoids Regulate Gene Expression through Direct Interactions with Peroxisome Proliferator-Activated Receptors " *biochemistry* 94 (1997): 4318-23. Print.

Kroth, Peter G., et al. "A Model for Carbohydrate Metabolism in the Diatom *Phaeodactylum Tricornutum* Deduced from Comparative Whole Genome Analysis." *PLoS ONE* 3.1 (2008): e1426. Print.

Lars Tomanek, Eric Sanford. "Heat Shock Protein 70 (Hsp) as a Biochemical Stress Indicator: An Experimental Field Test in Two Congeneric Intertidal Gastropods (Genus: Tegula)." *Biology Bulletin* 205 (2003): 276-84. Print.

Latendresse, Mario, et al. "Construction and Completion of Flux Balance Models from Pathway Databases." *Bioinformatics* (2012). Print.

Leversen, Nils Anders, et al. "Evaluation of Signal Peptide Prediction Algorithms for Identification of Mycobacterial Signal Peptides Using Sequence Data from Proteomic Methods." *Microbiology* 155.7 (2009): 2375-83. Print.

Liang, Yanna. "Biomass and Lipid Productivities of Chlorella Vulgaris under Autotrophic, Heterotrophic and Mixotrophic Growth Conditions." *Biotechnol Lett* 31 (2009): 1043-49. Print.

Liu, Liming, et al. "Use of Genome-Scale Metabolic Models for Understanding Microbial Physiology." *FEBS Letters* 584.12 (2010): 2556-64. Print.

Los, Dmitry A., and Norio Murata. "Structure and Expression of Fatty Acid Desaturases." *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* 1394.1 (1998): 3-15. Print.

M. Amin Hejazi, Wifffels, Jene. "Milking of Microalgae." *Trends in biotechnology* 22 (2004): 4. Print.

Ma, Zhengping, Ching-Hsuen Chu, and Dong Cheng. "A Novel Direct Homogeneous Assay for Atp Citrate Lyase." *J. Lipid Res.* 50.10 (2009): 2131-35. Print.

Maheswari, Uma. "Update of the Diatom Est Database: A New Tool for Digital Transcriptomics." *Nucleic Acids Research* 37.D1001-D1005 (2009). Print.

Maheswari, Uma, et al. "Digital Expression Profiling of Novel Diatom Transcripts Provides Insight into Their Biological Functions." *Genome Biology* 11.8 (2009): R85. Print.

Maheswari, Uma, et al. "The Diatom Est Database." *Nucleic Acids Research* 33.suppl 1 (2005): D344-D47. Print.

Mandal, Shovon. "Microalga Scenedesmus Obliquus as a Potential Source for Biodiesel Production." *Appl Microbiol Biotechnol* 84 (2009): 281-91. Print.

Mansour, Maged Pl. "Lipid and Fatty Acid Yield of Nine Stationary-Phase Microalgae: Applications and Unusual C_{24}-C_{28} Polyunsaturated Fatty Acids." *Journal of Applied Phycology* 17 (2005): 287-300. Print.

Marchetti, Adrian, et al. "Phytoplankton Processes During a Mesoscale Iron Enrichment in the Ne Subarctic Pacific: Part I—Biomass and Assemblage." *Deep Sea Research Part II: Topical Studies in Oceanography* 53.20–22 (2006): 2095-113. Print.

Markowitz, Victor M., et al. "The Integrated Microbial Genomes System: An Expanding Comparative Analysis Resource." *Nucleic Acids Research* 38.suppl 1 (2010): D382-D90. Print.

Markowitz, Victor M., et al. "Img Er: A System for Microbial Genome Annotation Expert Review and Curation." *Bioinformatics* 25.17 (2009): 2271-78. Print.

Masuda, Reiji, David A. Ziemann, and Anthony C. Ostrowski. "Patchiness Formation and Development of Schooling Behavior in Pacific Threadfin Polydactylus Sexfilis Reared with Different Dietary Highly Unsaturated Fatty Acid Contents." *Journal of the World Aquaculture Society* 32.3 (2001): 309-16. Print.

Mayfield, Stephen. "Chlamydomonas Reinhardtii Chloroplasts as Protein Factories " *Current Opinion in Biotechnology* 18 (2007): 126-33. Print.

McGinnis, k. Ml. "Characterization of the Growth and Lipid Content of the Diatom Chaetoceros Muelleri." *Journal of Applied Phycology* 9 (1997): 19-24. Print.

Melnick, Michael, et al. "Salivary Gland Branching Morphogenesis: A Quantitative Systems Analysis of the Eda/Edar/Nfkappab Paradigm." *BMC Developmental Biology* 9.1 (2009): 32. Print.

Mizuno, Yutaka, Greg B. G. Moorhead, and Kenneth K.-S. Ng. "Structural Basis for the Regulation of N-Acetylglutamate Kinase by Pii in Arabidopsis Thaliana." *Journal of Biological Chemistry* 282.49 (2007): 35733-40. Print.

Mock, Thomas. "Whole-Genome Expression Profiling of the Marine Diatom *Thalassiosira Pseudonana* Identifies Genes Involved in Silicon Bioprocesses." *PNAS* 105.5 (2008): 1579-84. Print.

Mock, Thomas, and Bernd M. A. Kroon. "Photosynthetic Energy Conversion under Extreme Conditions--I: Important Role of Lipids as Structural Modulators and Energy Sink under N-Limited Growth in Antarctic Sea Ice Diatoms." *Phytochemistry* 61.1 (2002): 41-51. Print.

Montsant, Anton, et al. "Identification and Comparative Genomic Analysis of Signaling and Regulatory Components in the Diatom *Thalassiosira Pseudonana*." *Journal of Phycology* 43.3 (2007): 585-604. Print.

Montsant, Anton, et al. "Comparative Genomics of the Pennate Diatom Phaeodactylum Tricornutum." *Plant Physiol.* 137.2 (2005): 500-13. Print.

Morais, Sofia, et al. "Expression of Fatty Acyl Desaturase and Elongase Genes, and Evolution of Dha:Epa Ratio During Development of Unfed Larvae of Atlantic Bluefin Tuna (Thunnus Thynnus L.)." *Aquaculture* 313.1-4 (2011): 129-39. Print.

Nagle, Cynthia A., et al. "Identification of a Novel Sn-Glycerol-3-Phosphate Acyltransferase Isoform, Gpat4, as the Enzyme Deficient in Agpat6-/- Mice." *J. Lipid Res.* 49.4 (2008): 823-31. Print.

Orlova, Irina V., et al. "Transformation of Tobacco with a Gene for the Thermophilic Acyl-Lipid Desaturase Enhances the Chilling Tolerance of Plants." *Plant Cell Physiol.* 44.4 (2003): 447-50. Print.

Papin, Jason A., et al. "Comparison of Network-Based Pathway Analysis Methods." *Trends in Biotechnology* 22.8 (2004): 400-05. Print.

Park, Jong Myoung, Tae Yong Kim, and Sang Yup Lee. "Prediction of Metabolic Fluxes by Incorporating Genomic Context and Flux-Converging Pattern Analyses." *Proceedings of the National Academy of Sciences* 107.33 (2010): 14931-36. Print.

Phair, Robert D, and Tom Misteli. "High Mobility of Proteins in the Mammalian Cell Nucleus." *Nature* 404.6778 (2000): 604-08. Print.

Pienkos, Philip T., and Al Darzins. "The Promise and Challenges of Microalgal-Derived Biofuels." *Biofuels, Bioproducts and Biorefining* 3.4 (2009): 431-40. Print.

Piorreck, Margret, Klaus-Hinnerk Baasch, and Peter Pohl. "Biomass Production, Total Protein, Chlorophylls, Lipids and Fatty Acids of Freshwater Green and Blue-

Green Algae under Different Nitrogen Regimes." *Phytochemistry* 23.2 (1984): 207-16. Print.

Qiu, Xiao. "Biosynthesis of Docosahexaenoic Acid (Dha, 22:6-4, 7,10,13,16,19): Two Distinct Pathways." *Prostaglandins, Leukotrienes and Essential Fatty Acids* 68.2 (2003): 181-86. Print.

Raclot, Thierry, Cecilia Holm, and Dominique Langin. "Fatty Acid Specificity of Hormone-Sensitive Lipase: Implication in the Selective Hydrolysis of Triacylglycerols." *J. Lipid Res.* 42.12 (2001): 2049-57. Print.

Ratledge, C. "Regulation of Lipid Accumulation in Oleaginous Micro-Organisms." *Biochemical Society Transactions* 30, part 6 (2002): 1047-50. Print.

Robertson, D. "Unraveling the Regulation of Nitrogen Assimilation in the Marine Diatom Thalassiosira Pseudonana(Bacillariophyceae):Diurnal Variations in the Transcript Levels for Five Genes Involved in Nitrogen Assimilation." *J. Pycol.* 45 (2009): 413-26. Print.

Roduit, Raphael. "A Role for the Malonyl-Coa/Long-Chain Acyl-Coa Pathway of Lipid Signaling in the Regulation of Insulin Secretion in Response to Both Fuel and Nonfuel Stimuli." *Diabetes* 53 (2004): 1007-18. Print.

Ruppin, Eytan, et al. "Metabolic Reconstruction, Constraint-Based Analysis and Game Theory to Probe Genome-Scale Metabolic Networks." *Current Opinion in Biotechnology* 21.4 (2010): 502-10. Print.

Sasaki, Yukiko, and Yukio Nagano. "Plant Acetyl-Coa Carboxylase: Structure, Biosynthesis, Regulation, and Gene Manipulation for Plant Breeding." *Bioscience, Biotechnology, and Biochemistry* 68.6 (2004): 1175-84. Print.

Scott, Stuart A., et al. "Biodiesel from Algae: Challenges and Prospects." *Current Opinion in Biotechnology* 21.3 (2010): 277-86. Print.

Secq, Marie-Pierre Oudot-Le, and Beverley R. Green. "Complex Repeat Structures and Novel Features in the Mitochondrial Genomes of the Diatoms Phaeodactylum Tricornutum and Thalassiosira Pseudonana." *Gene* 476.1-2 (2011): 20-26. Print.

Sherman, David. "Enzyme's Black Box Cracked Open." *Nature* 461 (2009): 1068-69. Print.

Shifrin, Neil S., and Sallie W. Chisholm. "Phytoplankton Lipids: Interspecific Differences and Effects of Nitrate, Silicate and Light-Dark Cycles." *Journal of Phycology* 17.4 (1981): 374-84. Print.

Shrager, Jeff, et al. "Chlamydomonas Reinhardtii Genome Project. A Guide to the Generation and Use of the Cdna Information." *Plant Physiol.* 131.2 (2003): 401-08. Print.

Sialve, Bruno. "Anaerobic Digestion of Microalgae as a Necessary Step to Make Microalgal Biodiesel Sustainable." *Biotechnology Advances* 27 (2009): 409-16. Print.

Smith, G. Jason. "Amino Acid Profiles in Species Adn Strains of Pseudo-Nitzschia from Monterey Bay California: Insights into the Metabolic Roles of Domoic Acid." *Harmful Algal Blooms 200* (2001). Print.

Sriharan, Shobha. "Effects of Nutrients and Temperature on Lipid and Fatty Acid Production in the Diatom Hantzhia Di-60." *Applied Biochemistry and Biotechnology* 24.25 (1990): 309-16. Print.

Sundaram, Meenakshi, et al. "Expression of Apolipoprotein C-Iii in Mca-Rh7777 Cells Enhances Vldl Assembly and Secretion under Lipid-Rich Conditions." *J. Lipid Res.* 51.1 (2010): 150-61. Print.

Sweetlove, Lee J., and Alisdair R. Fernie. "Regulation of Metabolic Networks: Understanding Metabolic Complexity in the Systems Biology Era." *New Phytologist* 168.1 (2005): 9-24. Print.

T.V. Ramachandra, Mahapatra D.M., Karthick B., Gordon R. "Milking Diatoms for Sustainable Energy: Biochemical Engineering Versus Gasoline-Secreting Diatom Solar Panels." *Industrial Engineering Chemical Research*.July 25th (2009). Print.

Tonon, Thierry, et al. "Fatty Acid Desaturases from the Microalga *Thalassiosira Pseudonana*." *FEBS Journal* 272.13 (2005): 3401-12. Print.

Truan, Daphne, et al. "A New Pii Protein Structure Identifies the 2-Oxoglutarate Binding Site." *Journal of Molecular Biology* In Press (2010), Corrected Proof. Print.

Tyson, John J., Katherine C. Chen, and Bela Novak. "Sniffers, Buzzers, Toggles and Blinkers: Dynamics of Regulatory and Signaling Pathways in the Cell." *Current Opinion in Cell Biology* 15.2 (2003): 221-31. Print.

Vanhaesebroeck, Bart, et al. "Synthesis and Function of 3-Phosphorylated Inositol Lipids." *Annual Review of Biochemistry* 70.1 (2001): 535-602. Print.

Vaulot, D., et al. "Cell-Cycle Response to Nutrient Starvation in Two Phytoplankton Species, Thalassiosira Weissflogii and Hymenomonas Carterae." *Marine Biology* 95.4 (1987): 625-30. Print.

Viswanathan, Ganesh A., et al. "Getting Started in Biological Pathway Construction and Analysis." *PLoS Comput Biol* 4.2 (2008): e16. Print.

Vizcaíno-Ochoa, Verónica, et al. "The Effect of Dietary Docosahexaenoic Acid (Dha) on Growth, Survival and Pigmentation of California Halibut Paralichthys Californicus Larvae (Ayres, 1810)." *Aquaculture* 302.3-4 (2001): 228-34. Print.

Waltz, Emily. "Algae Trailblazer Shuts." *Nat Biotech* 27.7 (2009): 586-86. Print.

Weldy, Chad Share. "Lipid Production by Dunaliella Salina in Batch Culture: Efficts of Nitrogen Limitation and Light Intensity." *US Deptment of Energy Journal of Undergraduate Research* (2007): 115-22. Print.

Wheeler, Mariel C. Gerrard. "A Comprehensive Analysis of the Nadp-Malic Enzyme Gene Family of Arabidopsis." *Plant Physiology* 139 (2005): 39-51. Print.

Work, Victoria H., et al. "Increased Lipid Accumulation in the Chlamydomonas Reinhardtii Sta7-10 Starchless Isoamylase Mutant and Increased Carbohydrate Synthesis in Complemented Strains." *Eukaryotic Cell*: EC.00075-10 (2010). Print.

Yu, Eizadora. "Triacylglycerol Accumulation and Profiling in the Model Diatoms Thalassiosira Pseudonana and Phaeodactylum Tricornutum (Baccilariophyceae) During Starvation." *J Aappl Phycol* DOI 10 (2009). Print.

Zaslavskaia, Lioudmila A., et al. "Transformation of the Diatom *Phaeodactylum Tricornutum* (Bacillariophyceae) with a Variety of Selectable Marker and Reporter Genes." *Journal of Phycology* 36.2 (2000): 379-86. Print.

Zemke, Peter E., Byard D. Wood, and Daniel J. Dye. "Considerations for the Maximum Production Rates of Triacylglycerol from Microalgae." *Biomass and Bioenergy* 34.1 (2010): 145-51. Print.

Zhanf, Hailong. "Crystal Structure of the Carboxyltransferase Domain of Acetyl-Coenzyme a Carboxylase." *Science* 299 (2003): 2064-67. Print.

Books

Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Boca Raton, Fl, Chapman & Hall/CRC, 2007. Print.

Andersen, R. A., Ed. *Algal Culturing Techniques*. Burlington, Ma, Elsevier, 2005. Print.

Arts, Michael T., Michael T. Brett, and Martin J. Kainz. Eds. *Lipids in Aquatic Ecosystems*. New York, Springer, 2009. Print.

Baclawski, Kenneth. *Ontologies for Bioinformatics*. Cambridge, Massachusetts Institute of Technology, 2006. Print.

Baxevanis, Andreas D., Ed. *Bioinformatics: A practical guide to the analysis of genes and proteins*. Hoboken, NJ, John Wiley & Sons, 2005. Print.

Bessant, Conrad I. S., and Darren Oakley *Building Bioinformatics Solutions with Perl, R and MySQL*. Oxford, Oxford University Press, 2009. Print.

Bhakuni, D. S. R. *Bioactive Marine Natural Products*. New Delhi, Springer, 2010. Print.

Blankenship, R. *Molecular Mechanisms of Photosynthesis*. Malden, Ma, Blackwell Science, 2008. Print.

Bohnert, Hans J., and Norman G. Lewis, Ed. *Bioengineering and Molecular Biology of Plant Pathways*. Advances in Plant Biochemistry and Molecular Biology. San Diego, Elsevier, 2011. Print.

Bolouri, H. *Computational Modeling of Gene Regulatory Networks*. London, Imperial College Press, 2008. Print.

Choi, S., Ed. *Systems Biology for Signaling Networks*. New York, Springer, 2010. Print.

Ehnholm, C., Ed. *Cellular Lipid Metabolism.* Berlin, Springer, 2009. Print.

Fall, Christopher P., John M. Wagner, and John J. Tyson, Ed. *Computational Cell Biology*. Interdisciplinary Applied Mathematics. New York, Springer, 1999. Print.

Falkowski, Paul G. *Aquatic Photosynthesis*. Princeton, Princeton University Press, 2007. Print.

Gunstone, F. *Fatty Acid and Lipid Chemistry*. Gaithersburg, Maryland, Aspen, 1999. Print.

Gurr, M. I., and K.N. Frayn. *Lipid Biochemistry*. Oxford, Blackwell, 2002. Print.

Helms, V. *Principles of Computational Cell Biology*. Weinheim, Wiley-VCH, 2008. Print.

Lea, Peter J., and R. C. Leegood Eds. *Plant Biochemistry and Molecular Biology*. West Sussex, England, John Wiley & Sons Ltd, 1999. Print.

Michal, G., Ed. *Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology*. New York, Wiley, 1999. Print.

Nagasaki, M. *Foundations of Systems Biology: Using Cell Illustrator and Pathway Databases*. London, Springer, 2009. Print.

Pevsner, J. Bioinformatics and Functional Genomics. Hoboken, John Wiley & Sons, 2009. Print.

Simons, K., Ed. *Biology of Lipids: Trafficking, Regulation, and Function*. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory Press, 2011. Print.

Tisdall, J. D. *Beginning Perl for Bioinformatics*. Sebastopol, Ca, O'Reilly, 2001. Print.

Tisdall, J. D. *Mastering Perl for Bioinformatics*. Sebastopol, Ca, O'Reilly, 2003. Print.