

AN ABSTRACT OF THE THESIS OF

Michelle Pombrol for the degree of Master of Science in Microbiology presented on May 28, 2019.

Title: Strategies for Altering Energy Flux in *Thalassiosira pseudonana* Elucidated by Transcriptomics

Abstract approved: _____

Kimberly Halsey

Diatoms play a major role in ocean biogeochemical cycles and are important tools in bioengineering for natural products and nanotechnology. Diatoms and other algae growing at varying resource-limited growth rates allocate carbon to different metabolic pathways to optimize growth; however, the molecular mechanisms controlling these pathway gating strategies are not well understood. We used RNA-Seq to investigate how the model diatom *Thalassiosira pseudonana* balances photosynthetic energy flux in cells grown in continuous culture under slow and fast light-limited growth rates. We explored fold-change thresholds for differential expression in cells grown under low light (5 μE) with a steady-state growth rate of 0.20 d^{-1} and cells grown under high light (200 μE) with a steady-state growth rate of 1.54 d^{-1} . Under the conservative threshold ($|\text{fold change}| > 4$, $p < 0.05$), only approximately 5% of genes were differentially expressed between low and high light conditions. Under the less conservative threshold ($|\text{fold change}| > 2$, $p < 0.05$), approximately 25% of genes were differentially expressed. Under both thresholds, the majority of differentially expressed genes were not annotated in the KEGG database, highlighting the need for further efforts in functional annotation of diatom genomes.

Several genes involved in the TCA and glyoxylate cycles, photorespiratory pathway, and peroxisomal functioning were differentially expressed. Slow-growing cells upregulated genes involved in carbon conservation (*i.e.*, gluconeogenesis and the glyoxylate cycle) and downregulated genes involved in carbon catabolism (*i.e.*, glycolysis). This research identified patterns of gene expression that help explain fundamental differences in metabolic flux in response to growth rate limitation in *T. pseudonana*. The genes identified in this study are well-conserved across phytoplankton groups, suggesting that they function as master regulators of carbon flux and macromolecular composition in algae. Knowledge of the relationship between growth rate and expression of master regulatory genes may allow for rapid assessment of phytoplankton growth rate *in situ*, a major goal of oceanographers interested in measuring primary production. This knowledge will also facilitate prediction of cellular responses to a changing climate and increase the feasibility of manipulating metabolic flux for bioengineering and production of important natural products.

© Copyright by Michelle Pombrol
May 28, 2019
All Rights Reserved

Strategies for Altering Energy Flux in *Thalassiosira pseudonana* Elucidated by
Transcriptomics

by
Michelle Pombrol

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented May 28, 2019
Commencement June 2019

Master of Science thesis of Michelle Pombrol presented on May 28, 2019

APPROVED:

Major Professor, representing Microbiology

Head of the Department of Microbiology

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Michelle Pombrol, Author

ACKNOWLEDGEMENTS

I am grateful to Dr. Kimberly Halsey, Dr. James Fox, Eric Moore, Bryce Penta, Lindsay Collart, Mary Fulton, and my committee members. All others who offered support, encouragement, or honest thoughts along the way—thank you.

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction.....	1
2. Literature Review.....	2
2.1 Diatom genomes and metabolism.....	4
2.2 Transcriptomic studies in diatoms.....	7
2.3 Steady-state light limitation.....	10
3. Methods	17
3.1 Culture conditions.....	17
3.2 RNA extraction and sequencing.....	18
3.3 Quality trimming and sequence alignment.....	18
3.4 Quality control and differential expression analysis.....	20
3.5 Functional annotation.....	22
4. Results & Discussion	23
4.1 Quality control analyses.....	23
4.2 Overview	25
4.3 Patterns of differential gene expression	28
4.4 Magnitude of differential expression.....	32
4.5 The TCA and glyoxylate cycles	33
4.6 Increasing flux through gluconeogenesis.....	39
5. Conclusion.....	41
5.1 Conclusions.....	41
5.2 Impact.....	43

TABLE OF CONTENTS (Continued)

	<u>Page</u>
5.3 Future Directions.....	44
6. Figures & Tables.....	48
7. References.....	65

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Generalized relationship between growth irradiance (μE) and specific growth rate (d^{-1}) in phytoplankton.	48
2. The percentage of total carbon allocated to carbohydrate, protein, and lipid in fast growing cells (1.5 d^{-1}) and slow growing cells (0.2 d^{-1}).....	49
3. Visualizations of data transformed via A) the shifted logarithm transformation, B) the variance stabilizing transformation, and C) the regularized logarithm transformation.....	50
4. Sample distance matrix calculated using Euclidean distances, showing similarity between samples.....	52
5. PCA plot showing relationships between the six samples chosen for differential expression analysis.....	53
6. Fraction of small p-values binned by normalized counts.....	54
7. Summary of annotations for genes called differentially expressed using the thresholds $ \text{FC} > 4, p < 0.05$	56
8. Summary of annotations for genes called differentially expressed using the thresholds $ \text{FC} > 2, p < 0.05$	57
9. Heatmap showing relative degrees of differential expression for the 1,000 genes with highest variance between all samples.....	58
10. Results from Phyre 2 program for <i>THAPSDRAFT_8336</i>	59
11. Genes involved in catabolic processes that produce acetyl-CoA are upregulated in slow-growing cells.....	62
12. Slow-growing cells differentially expressed genes involved in the glyoxylate cycle, TCA cycle, and photorespiratory pathway in order to conserve carbon.....	63
13. Slow-growing <i>T. pseudonana</i> upregulated genes involved in reactions that form PEP in order to increase flux through gluconeogenesis.....	64

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Alignment statistics for nine original samples.....	51
2. Comparison of the number of genes that are called significantly differentially expressed using two fold-change cutoffs	55
3. Differentially expressed genes involved in the glyoxylate cycle, TCA cycle, and related pathways.....	60-61

Chapter 1: Introduction

Diatoms—unicellular, eukaryotic algae with unique silica cells walls—are a major component of marine phytoplankton communities. They are highly abundant and incredibly diverse, with an estimated 200,000 species distributed throughout all aquatic environments with sufficient light and nutrients (Armbrust 2009a). Diatoms are the dominant primary producers in many parts of the ocean (Uitz et al. 2010). They play important roles in biogeochemical cycling of nitrogen, phosphorous, and silicon (Sarhou et al. 2005), contribute to the export of carbon from ocean surface layers (Scharek et al. 1999; Guidi et al. 2016), and produce organic matter that serves as the base of the marine food web (Armbrust 2009a). Diatoms are dominant in coastal upwelling zones (Margalef 1978), grow in sea ice and polar oceans (Thomas and Dieckmann 2002), and thrive in environments with fluctuating light (Mitrovic et al. 2003; Huisman et al. 2004). Their success in a variety of environments suggests that diatoms employ unique adaptive strategies (Crombet et al. 2011; Thamatrakoln et al. 2012; Nunn et al. 2013). Indeed, diatoms have been shown to possess a novel combination of metabolic pathways which may contribute to their dominance (Bowler et al. 2008; Allen et al. 2011). However, the fine-scale genomic and physiological mechanisms that control these adaptive strategies are not well understood.

Diatoms' unique ability to tolerate fluctuating and extreme light levels may contribute to their success. The marine light environment is highly dynamic, so photosynthetic organisms must be able to adapt to changes in light availability across a range of timescales. Photosynthetically active radiation (PAR) fluctuates on short timescales due to changes in incident solar radiation (modulated by local atmospheric

conditions), time of day, and the presence of dissolved organic matter or suspended particles (Marra 1980; Falkowski and Wirick 1981; Depauw et al. 2012). Phytoplankton may also experience changes in PAR depending on their vertical position in the water column, which is modulated by their sinking rate and water turbulence. When surface waters are well-mixed, phytoplankton may be exposed to full sunlight at the surface and complete darkness below the photic zone within a relatively short time period (Geider et al. 1986). Diatoms, which have heavy silica frustules and tend to sink when they are not actively growing (Malone et al. 1983; Acuña et al. 2010), are able to control their buoyancy by altering the contents of their vacuoles (Raven and Waite 2004). Diatoms likely rely on a combination of this vacuolar control and wind-driven mixing for distribution throughout the water column.

Over longer timescales, climate change will be an important driver of widespread changes in light availability in marine environments (Sarmiento et al. 2004; Marinov et al. 2010). However, the effects of climate warming will vary across ecosystems depending on a range of local factors (Hoegh-Guldberg and Bruno 2010). Generally, warming ocean temperatures will lead to stabilization of the temperature-driven density stratification of the water column (Behrenfeld et al. 2006), decreasing the effects of wind-driven mixing. The resultant decrease in the depth of the mixed layer will keep phytoplankton closer to the surface where light is stronger. In some areas, however, climate warming will lead to more intense storms (Knutson et al. 2010) that may increase the effects of wind-driven mixing on the water column. In coastal areas, where diatoms are common, higher rates of erosion on land due will lead increased sediment inputs (Nearing et al. 2004) and lower PAR. The polar ocean will also experience changes in

PAR as sea ice melts and allows more light into the underlying waters (Nicolaus et al. 2012). The input of fresh water from melting sea ice will also lead to stronger density stratification, decreasing vertical mixing and keeping phytoplankton closer to the surface.

Frequent and extreme fluctuations in light availability are known to decrease productivity and fitness of photosynthetic organisms (Long et al. 1994), although the effects are dependent on the specific organism as well as on the frequency and amplitude of the fluctuations (Litchman 2000). Diatoms, however, have been shown to be particularly tolerant of fluctuations in the light environment and extremely low light availability (Geider et al. 1986; Mitrovic et al. 2003; Wagner et al. 2005). In fact, Fisher and Halsey (2016) showed that the model diatom *Thalassiosira pseudonana* is extremely efficient at converting harvested light energy into biomass. Across a wide range of growth irradiances (5–200 $\mu\text{mol photons m}^{-2} \text{s}^{-2}$), *T. pseudonana* converted 57% of harvested light energy into biomass. That work explored the distribution of energy through interconnected metabolic pathways and provided evidence that *T. pseudonana* shifts the flow of energy through more efficient pathways under low light. However, the specific molecular mechanisms underlying these metabolic shifts have thus far not been investigated. The relatively recent progress in -omics technologies and methods has provided new tools that can now be used to explore these molecular mechanisms. This study uses whole-transcriptome sequencing to explore differential gene expression in *T. pseudonana* grown at two levels of continuous light; its goal is to connect transcriptomic data with existing physiological data to build a more complete understanding of how *T. pseudonana* responds to differences in light-mediated growth rate.

Chapter 2: Literature Review

Diatom genomes and metabolism

T. pseudonana's complete genome sequence was published in 2004 (Armbrust et al. 2004). Four years later, the genome sequence of the pennate diatom *Phaeodactylum tricorutum* was published (Bowler et al. 2008). Genome assemblies are now available for five other diatom species: *Thalassiosira oceanica*, *Fistulifera solaris*, *Cyclotella cryptica*, *Fragilariopsis cylindrus*, and *Pseudo-nitzschia multistriata* (Lommer et al. 2012; Tanaka et al. 2015; Traller et al. 2016; Mock et al. 2017; Basu et al. 2017). Study of these genomes has revealed that diatoms have a complex evolutionary history involving rapid evolution, pervasive horizontal gene transfer, and two endosymbiotic events. This evolutionary complexity is reflected in their genomes, which contain many genes that are diatom-specific, and in their metabolisms, which combine novel pathways (e.g., silicon metabolism) with pathways that do not coexist in other organisms (e.g., the Calvin-Benson cycle and the urea cycle).

Molecular clock analyses have suggested that diatoms originated in the Mesozoic era, approximately 250 million years ago (Medlin 2011). They are thought to be the result of two separate endosymbiotic events. In the first endosymbiotic event, which occurred about 1.5 billion years ago (Yoon et al. 2004), a non-photosynthetic eukaryote acquired chloroplasts from a prokaryotic cyanobacterium. Approximately 500 million years later, the second endosymbiotic event—wherein a non-photosynthetic eukaryote engulfed a photosynthetic red alga (Bhattacharya et al. 2017)—occurred. The red algal symbiont eventually became the plastids of the Stramenopiles, the group which includes diatoms,

brown macroalgae, and plant parasites (Armbrust 2009a). Diatoms eventually diverged into the centric diatoms (cells with radial symmetry) and pennate diatoms (cells with bilateral symmetry). *T. pseudonana* and *P. tricornutum* are now considered the model species for centric and pennate diatoms, respectively.

Studies of the available diatom genomes also revealed that genomes differ greatly among species; for example, the genomes of *T. pseudonana* and *P. tricornutum* are about as different as those of mammals and fish, despite the fact that mammals and fish diverged around 550 million years ago and the two diatom species diverged only 90 million years ago (Armbrust 2009b). The high degree of divergence between these two organisms is highlighted by comparative genomic studies. These studies have found that *P. tricornutum* shares only 57% of its genes with *T. pseudonana* (Bowler et al. 2008). Additionally, 13% of *P. tricornutum* genes and 12% *T. pseudonana* genes are diatom-specific. 42% of genes in the *P. tricornutum* genome are unique genes—genes that are not found in other diatom species—while 33% of *T. pseudonana* genes are unique genes. These highly divergent genomes can complicate -omics-based analyses of these organisms. In fact, only about 50% of diatom genes can be assigned putative functions based on homology-based methods (Maheswari et al. 2010). In transcriptomic studies, it is not uncommon for the majority of genes of interest to be unannotated (Alexander et al. 2015; Diner et al. 2016). Previous studies and reviews on the topic have highlighted the need for further functional annotation of protist genes (Caron et al. 2016); in particular, bench-based work is needed to identify the functions of genes with no known orthologs.

The high degree of divergence and large number of non-orthologous genes in diatom genomes may be explained by the fact that these organisms are among the most

rapidly evolving taxa on Earth (Oliver et al. 2007). This rapid evolution has been attributed to the relatively high number of transposable elements and insertion/deletion mutations in diatom genomes (Vardi et al. 2009). Diatom-specific genes have also been shown to evolve faster than other genes in diatom genomes, possibly contributing to the high rate of divergence between diatoms (Bowler et al. 2008). Another source of variability in diatom genomes is horizontal gene transfer. Analysis of the *P. tricornutum* genome has revealed that about 5% of its genes are derived from bacteria, and about half of these bacterial-derived genes are shared with *T. pseudonana* (Bowler et al. 2008). This suggests that horizontal gene transfer is much more pervasive in diatoms than in other eukaryotes (Bowler et al. 2008; Martens et al. 2008; Keeling and Palmer 2008). Cyanobacteria and heterotrophic bacteria have been found in close association with diatoms (Zehr et al. 2000; Carpenter and Janson 2000); these close associations may explain the high rate of horizontal gene transfer between bacteria and diatoms.

Diatoms' mosaic genomes encode novel combinations of metabolic pathways. For example, diatoms have the capacity for photosynthetic carbon fixation (via the Calvin-Benson cycle) and organic nitrogen production (via the urea cycle). Photosynthetic carbon fixation is often seen as a hallmark of plant metabolism, while the urea cycle is generally considered to be part of animal metabolism. In metazoans, the urea cycle functions in the removal of fixed nitrogen; in diatoms, it is involved in the turnover and reallocation of inorganic nitrogen and carbon and is important to the adaptive response to nitrogen fluctuation (Allen et al. 2011). Diatoms are also unique in their ability to precipitate silicic acid in their environment to build silica cell walls. Although this unique metabolic function means that diatoms contribute significantly to marine biogeochemical

cycling of silica (Tréguer and De La Rocha 2013), this pathway has remained largely unexplored (Montsant et al. 2005). The first investigations into the components of diatom cell wall synthesis revealed that these components are unique to diatoms (Hildebrand et al. 1998; Poulsen et al. 2003; Montsant et al. 2005). Other unique cellular components encoded in diatom genomes include highly abundant heat shock transcription factors (Rayko et al. 2010), diatom-specific cyclins (Huysman et al. 2010), and far-red light sensors (Fortunato et al. 2016; Tirichine et al. 2017).

Transcriptomic studies in diatoms

A cell's transcriptome is the complete set of RNA transcripts it contains at a given time or under a given condition. Transcriptome profiling can provide insights into cellular growth, regulation, reproduction, and responses to environmental variables. Researchers have long understood the value in measuring gene expression to investigate cellular physiology; however, it was not until high-throughput sequencing became readily available that transcriptomic studies became commonplace. Although genomic and transcriptomic studies of marine protists are generally lagging behind those of other microorganisms, recent studies have used transcriptomics to probe the ecology, physiology, and evolution of these organisms (Caron et al. 2016). For example, transcriptomic approaches have allowed researchers to explore the diatom response to a variety of environmental variables, including the availability of iron (Marchetti et al. 2012; Durkin et al. 2012), phosphorus (Dyhrman et al. 2012; Cruz de Carvalho et al. 2016), silicon (Mock et al. 2008; Smith et al. 2016b), and nitrogen (Yang et al. 2013; Levitan et al. 2015).

Studies have shown that *P. tricornutum* differentially expresses approximately half of its genes in response to nitrogen starvation (Yang et al. 2013; Levitan et al. 2015), suggesting that this organism utilizes widespread genomic regulation in response to nitrogen stress. Generally, genes involved in nitrogen assimilation were upregulated. The likely result is that *P. tricornutum* is able to respond quickly when nitrogen is reintroduced to the environment. One-third of genes involved in lipid biosynthesis were downregulated in response to nitrogen starvation, possibly because nitrogen starvation impedes growth and cells that are growing very slowly have a lower requirement for cellular and organellar membrane components. The most strongly downregulated genes were involved in photosynthesis, suggesting that nitrogen starvation negatively affects photosynthetic processes. Indeed, nitrogen deficiency is known to decrease light absorption by modulating concentrations of pigment proteins and decreasing synthesis of chloroplastic proteins (Berges et al. 1996). Additionally, photosynthetic efficiency in *T. pseudonana* has been shown to decline rapidly following nitrogen starvation (Jiang et al. 2012). Genes involved in glycolysis and gluconeogenesis were downregulated, while genes involved in the tricarboxylic acid (TCA) cycle were upregulated in response to nitrogen starvation. Upregulation of genes involved in the TCA cycle may increase flux through the cycle to compensate for lower rates of carbon fixation due to downregulation of photosynthesis. Genes involved in the glyoxylate cycle were significantly downregulated in response to nitrogen starvation (Yang et al. 2013).

Phosphorous starvation resulted in differential expression of over half of the genome in *P. tricornutum* (Cruz de Carvalho et al. 2016). Generally, genes encoding ribosomal proteins were downregulated in response to phosphorous starvation. RNA

synthesis represents a significant sink for phosphorous, so phosphorous starvation could lead to decreased rates of transcription. However, it is also possible that the downregulation of ribosomal proteins is the result of the severely reduced growth rate rather than the lack of phosphorous (Dyhrman et al. 2012). The most strongly regulated genes were alkaline phosphatases, heat shock proteins, heat shock factors, and several proteins of unknown function. Heat shock factors are important transcriptional regulators in diatoms (Rayko et al. 2010), so the strong regulation of these genes may suggest that diatoms have a strong adaptive response to phosphorous starvation. Less than 2% of the differentially expressed genes were shared between *T. pseudonana* and *P. tricornutum*, highlighting the species-specific nature of their genomes and their stress responses. Genes found exclusively in *P. tricornutum* accounted for close to 20% of differentially expressed genes, suggesting that the response of diatoms experiencing phosphorous stress involves genes from diverse origins. This could provide support to the idea that diatoms' unique mosaic genomes are partially responsible for their adaptability and success. Many genes involved in photosynthesis were significantly downregulated, indicating that diatoms respond to nutrient stress by downregulating photosynthesis.

Unlike nitrogen and phosphorous starvation, silicon starvation has little direct effect on most metabolic processes. However, most diatoms (with the notable exception of *P. tricornutum*) require silicon to grow, and transfer to silicon-free medium has been shown to immediately arrest cell division (Smith et al. 2016b). Thus, silicon starvation can be used to disentangle the effects of cell cycle arrest from adaptive responses to nutrient stress. Smith et al. (2016) found that in response to cell cycle arrest caused by silicon starvation, *T. pseudonana* differentially expressed approximately 66% of its

genes. Genes that were differentially expressed were involved in silicon transport, lipid metabolism, carbohydrate metabolism, carbon metabolism, and pigment synthesis. Mock et al. (2008) also found that silicon starvation increased expression of genes involved in silicon transport and precipitation; however, the majority of regulated genes encode proteins with no known function. This is in contrast with genes that are regulated under nitrogen stress, which typically encode known proteins. This finding underscores the differences in cellular responses to starvation of essential nutrients (*e.g.*, nitrogen and phosphorous) and growth rate limitation caused by lack of a necessary growth factor (*e.g.*, silicon, light, *etc.*).

Steady-state light limitation

Changes in transcript abundance can be attributed to an adaptive response to a specific environmental variable or to a coordinated regulatory program associated with growth (Smith et al. 2016). Generally, adaptive responses are associated with *starvation* conditions, which cause metabolic imbalances that require adaptation to resolve; in contrast, *limitation* conditions, which determine rates of growth, are more generally associated with coordinated regulatory programs (Vonshak and Torzillo 2004; MacIntyre and Cullen 2005). In studies involving culturing, it is particularly important to clearly define whether the treatment conditions are starvation or limitation conditions, because cells that are in the process of acclimating (*i.e.*, showing an adaptive response) are fundamentally different from cells that are fully acclimated and have achieved balanced growth (MacIntyre and Cullen 2005). Long exposure to highly stable culture conditions allows cells to reach physiological equilibrium; cells that have reached this equilibrium

are considered to be fully acclimated to their environment. In this study, *T. pseudonana* was acclimated to different growth irradiances over ten generations. Physiological measurements indicated that cells had achieved balanced growth (see *Methods* for more details); therefore, we concluded that any observed differences in transcript abundance between treatment groups could be attributed to a *limitation* response. Thus, we believed that any differentially expressed genes would be associated with a growth program rather than an adaptive response to a particular growth irradiance. A previous study focusing on steady-state growth in yeast found a clear transcriptional response to growth rate independent of nutrient stress, suggesting that eukaryotic cells do exercise growth-rate control at the transcriptional level and that this response can be distinguished from adaptive responses to starvation conditions (Castrillo et al. 2007).

We are interested in diatoms' responses to light-mediated growth rates because 1) light availability is one of the major factors controlling growth and productivity of photosynthetic organisms, 2) diatoms have been shown to be more tolerant of very low light levels than other types of algae (Quigg and Beardall 2003), raising questions about the metabolic and molecular mechanisms underlying this tolerance, and 3) information about the regulated growth programs that diatoms use to optimize growth under different light conditions can be used to refine global climate and biogeochemical models. Although the highly dynamic marine light environment contrasts starkly with the continuous light regime used in this study, we agree with MacIntyre and Cullen (2005), who wrote "...the fully acclimated state remains the expression of an idealized condition from whose contours we can discern the underlying elements of the cell's expression in response to a mutable world." We therefore believe that studying fully acclimated cells

growing in continuous light allows us to explore how cells grow in fluctuating light environments. Thus far, no contemporary studies have explored the effect of steady-state, light-limited diatom metabolism using transcriptomic methods. Older studies have explored diatoms' responses to light-limited growth rates (Admiraal and Peletier 1979; Harrison et al. 1990), and several recent studies have used transcriptomics to explore how growth and metabolism are affected by light fluctuation (Nymark et al. 2009; Domingues et al. 2012), but no transcriptomic studies have focused on steady-state growth.

Fisher and Halsey (2016) showed that diatoms are able to maintain maximal photosynthetic efficiency under steady-state, light-limited growth. Photosynthesis begins with the light-dependent generation of chemical energy in chloroplasts via linear electron flow (LEF). In LEF, electrons derived from hydrolysis in photosystem II (PS II) are passed through photosystem I (PS I) before ultimately reducing NADP^+ to NADPH. In the process, electrons are passed through the cytochrome b_6/f complex, which, together with the water-splitting complex associated with PSII, generates a proton gradient across the thylakoid membrane. This proton gradient drives the synthesis of ATP. The NADPH and ATP produced by the photosynthetic light reactions are used to supply energy for the Calvin-Benson cycle, in which inorganic carbon is fixed to produce organic compounds. The metabolic demand for ATP and NADPH varies under different physiological and environmental conditions. Therefore, a finely controlled balance of ATP and NADPH must always be maintained in order for cells to optimize their metabolic machinery. The water-water cycle and cyclic electron flow around PS I play important roles in maintaining an appropriate NADPH to ATP ratio (Asada 1999; Behrenfeld et al. 2004). Both processes are involved in balancing light absorption with carbon fixation. If the

generation of ATP and NADPH through photosynthesis do not satisfy the requirements for carbon fixation, additional ATP may be generated through alternate pathways including light-dependent respiration and cyclic electron transport around PS I. Despite the great amount of effort that has gone into understanding photosynthesis and primary productivity in the ocean, the regulatory networks and alternative pathways controlling the balance of ATP and NADPH are not yet well understood (Wilhelm et al. 2006).

Despite the current lack of understanding regarding molecular controls on photosynthesis, much is already known about photosynthetic organisms' general physiological responses to changes in growth irradiance. The growth rate of phytoplankton increases with increasing irradiance until a point at which further increases in photon flux no longer lead to an increase in growth rate (Figure 1) (Behrenfeld et al. 2008). After this light saturation point, the rate of the Calvin-Benson cycle limits the overall rate of photosynthesis (Formighieri 2015). When irradiance exceeds photosynthetic requirements, photoinhibition can occur. Photoinhibition is associated with damage to PS II reaction centers and the formation of damaging reactive oxygen species (Alderkamp et al. 2010). Phytoplankton utilize photoprotective mechanisms such as non-photochemical quenching (NPQ) to guard against oxidative damage in high light conditions. NPQ involves quenching singlet chlorophyll and dissipating excess energy as heat, thereby preventing the formation of triplet chlorophyll and damaging singlet oxygen. Under high light conditions, phytoplankton may also increase their concentration of secondary carotenoids, which are photoprotective (Hu 2004). When irradiance is low and growth rate is limited, phytoplankton increase the efficiency of light absorption by

upregulating the synthesis of chlorophyll *a* and other light-harvesting pigments (Hu 2004; Fisher and Halsey 2016).

At the molecular level, the proportion of photosynthetically fixed carbon allocated to three macromolecular pools—carbohydrate, protein, and lipid—also tends to differ based on growth irradiance. This is because the ultimate fate of glyceraldehyde-3-phosphate (GAP), which is the end product of carbon fixation via the Calvin-Benson cycle, is determined by cellular requirements which are responsive to environmental variables (Halsey and Jones 2015). Fisher and Halsey (2016) measured macromolecular composition of *T. pseudonana* growing at both high (200 $\mu\text{mol photons m}^{-2} \text{ s}^{-2}$) and low light (5 $\mu\text{mol photons m}^{-2} \text{ s}^{-2}$) using radiolabeled carbon. Measurements of the amount of carbon allocated to the carbohydrate, protein, and lipid pools were taken after 20 minutes (20-minute pulse) and 24 hours (24-hour biomass) (Figure 2). Under both low and high light, cells allocated most newly fixed carbon to carbohydrates and protein—less than 5% of newly fixed carbon was allocated to the lipid pool in both growth conditions. After 24 hours, cells grown in low light increased the percentage of total carbon allocated to protein while decreasing the percentage of carbon allocated to carbohydrates and lipids. The retention of carbon in the protein pool has previously been observed in nutrient-limited cultures (Morris 1980; Halsey et al. 2011). In contrast, cells growing in high light decreased the percentage of carbon allocated to protein and increased the percentage of carbon allocated to carbohydrates and lipids after 24 hours. The increase in the proportion of carbon allocated to lipids in cells growing under high light likely reflects the higher demand for membrane components in fast-growing cells.

The differential carbon allocation strategies observed by Fisher and Halsey (2016) are underpinned by unique pathway gating strategies that serve to optimize growth at different light-mediated growth rates. *T. pseudonana* is known to grow particularly efficiently at low light; Fisher and Halsey concluded that this efficiency is achieved via the redistribution of energy to more efficient ATP-generating pathways under low light conditions. However, there is still a need to understand the molecular mechanisms that control this growth optimization. Genomics approaches are valuable in the pursuit of this understanding because they allow for fine-scale insights into how these phytoplankton regulate pathway gating and allow us to compare specific molecular mechanisms across genomes to better understand how photosynthetic organisms as a whole adapt to different light conditions.

Recently, genomic techniques have been employed to study the molecular mechanisms underlying diatoms' responses to shifts in light availability (Bailleul et al. 2010). These studies involving fluctuating light are focused on the adaptive response of the organism of interest (*i.e.*, the immediate metabolic adjustments the cells make before they are acclimated to their environment). In this study, we sought to understand the gene expression patterns underlying physiological responses to light-limited growth in *T. pseudonana*. This study focuses on the effects of light *limitation*, in which the cells experience balanced, steady-state growth. Thus, we were interested in the strategies and regulatory programs that *T. pseudonana* employs to achieve the high growth efficiencies that have been observed in previous studies. Because central carbon metabolism (*i.e.*, glycolysis, the pentose phosphate pathway, and the TCA cycle) converts sugars into metabolic precursors that then go on to produce all cellular biomass, we hypothesized

that genes involved in these metabolic pathways would be differentially expressed between cells growing at different light-limited rates. We also hypothesized that a large number of genes would be differentially expressed. The growth irradiances used in this study led to a large range in growth rate—cells growing under low light had a specific growth rate of 0.2 d^{-1} , while cells growing under high light had a specific growth rate of 1.54 d^{-1} . We believed that this large difference in growth rate would lead to a corresponding large number of differentially expressed genes. The differential carbon allocation strategies observed by Fisher and Halsey (2016) suggest that cells growing at different light-mediated rates utilize different combinations of metabolic pathways, which lends support to our hypothesis that a large number of genes would be differentially expressed. Additionally, a previous study in *Escherichia coli* found that growth rate changes lead to genome-wide changes in expression (Barenholz et al. 2016); another found that *Saccharomyces cerevisiae* differentially expresses approximately half of its transcriptome in response to different steady-state growth rates (Regenberg et al. 2006). However, other studies in diatoms have shown that diatoms are capable of constitutively expressing certain photosystem genes in order to adapt quickly to changes in light availability (Broman et al. 2017). If diatoms constitutively express genes for the purpose of facilitating rapid adaptation to changing environmental conditions, we would expect that, in contrast to our original hypothesis, a relatively small number of genes would be differentially expressed between the treatment conditions.

Chapter 3: Methods

Culture conditions

The RNA used in this study was extracted from cells grown for a separate, culture-based study by Fisher and Halsey (2016). In that study, *Thalassiosira pseudonana* CCMP1335 was grown in a 300 ml continuous culturing system at 18°C in f/2 + Si medium supplemented with 0.17 μM Na_2SeO_3 . Cells were grown under three conditions: high light (200 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$), medium light (60 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$), and low light (5 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$). Cells were exposed to constant light rather than diel cycles to randomize their cell cycles. The growth rates and measurements of basic cell characteristics obtained under continuous light were similar to those obtained in *T. pseudonana* and *T. oceanica* under 12-hour light/dark cycles with the same average growth irradiances, suggesting that continuous light does not significantly affect the physiological response of these diatoms to low or high light.

The chosen light levels established growth rates of 1.54 d^{-1} , 0.85 d^{-1} , and 0.20 d^{-1} , respectively. Cultures were grown in triplicate under each growth irradiance. To maintain growth rates, peristaltic pumps were set to administer fresh media according to the equation $\mu = D/V$, where μ is the growth rate (d^{-1}), D is the flow rate (ml d^{-1}), and V is the culture volume (ml). Cultures were bubbled continuously to ensure that all cells were exposed to the same average light intensity and that CO_2 would not be limiting. All cultures were acclimated to their growth conditions for ten generations. Steady-state growth was verified by ensuring that cell concentration varied by less than 5% over three days before physiological measurements were taken (see Fisher and Halsey 2016 for more details about physiological measurements). Additionally, fluorescence

measurements (F_v/F_m) were taken to determine whether the cells were fully acclimated to their light conditions. F_v/F_m was consistent across all growth rates, indicating that photosynthetic efficiency was maximal under all light conditions and cells were fully acclimated. C:N ratios across growth rates were consistent, suggesting that nitrogen was not limiting and that all measurements were the result of light limitation rather than nitrogen limitation. Following physiological measurements, cells were filtered onto 0.2 μm filters. Filters were placed in 1.5 ml centrifuge tubes, flash-frozen in liquid nitrogen, and stored at -80°C until the time of RNA extraction.

RNA extraction and sequencing

RNA extractions were performed such that triplicates for each treatment were processed in tandem. Total RNA was extracted from filters using a Qiagen RNeasy Midi Kit following manufacturer's instructions, with a few additional steps. Silica beads (0.5 mm) were added to the lysis buffer. Filters, beads, and lysis buffer were vortexed together; supernatant was then filtered through Qias shredder columns to remove large particles. Eluted RNA was treated off-column with RNase-free DNase according to manufacturer's recommendations. mRNA isolation and library preparation were performed using the PrepX PolyA mRNA Isolation Kit and RNA-Seq for Illumina Library Kit from Takara Bio USA, Inc. These steps were carried out on an Apollo 324 robotic system by the Center for Genome Research and Computing at Oregon State University. RNA-derived reads were sequenced as a 150 bp single-end library via Illumina HiSeq 3000.

Quality trimming and sequence alignment

FastQC (Andrews 2010) was used to check GC content and per-base quality scores in each sample. Raw reads were quality trimmed and filtered using Sickle v1.33 (Joshi and Fass 2011) with the options “-q 33 -l 50”. The first option sets the quality score threshold for trimming to 33, while the second option discards all reads that are shorter than 50 base pairs after trimming. Sequencing adapters were not removed, as previous studies have shown that adapters do not significantly affect alignments to reference genomes and that aggressive trimming can lead to bias and poorer results (MacManes 2014; Williams et al. 2016). The *T. pseudonana* reference genome was obtained from NCBI in GFF format (accession number GCA_000149405.2) and converted to GTF using gffread. Reads were aligned to the reference genome using HISAT2 v2.1.0 (Kim et al. 2015) with the option “--dta”. HISAT2 is a splice-aware aligner that generates alignments more quickly and using less memory than older alignment software such as BowTie or BWA (Pertea et al. 2016). The “--dta” option tells the program to report alignments in a format that is tailored to transcriptome assemblers like StringTie. HISAT2 created sequence alignment/map (SAM) files for each sample. These files contain the alignments in the order that the sequences occurred in the input FASTQ files. SAMtools v1.3 (Li et al. 2009) was used to sort the alignments with respect to their genomic positions and then convert the files to binary alignment/map (BAM) files, which facilitate more rapid computation. Coverage depth of the genome was calculated using the “genomeCoverageBed” command with the “-ibam” option in BEDTools v2.25.0 (Quinlan and Hall 2010). StringTie v1.3.3 (Pertea et al. 2015) was used to assemble putative transcripts and generate counts tables for use in DESeq2. First, putative transcripts were assembled using the reference annotation file downloaded from

NCBI. Then, data from all samples was merged into a single file using the command “stringtie --merge.” Finally, “stringtie -eB” was used to calculate transcript abundances and generate count tables structured for Ballgown (DESeq2 also accepts these count tables as input). DESeq2 v1.20.0 (Love et al. 2014) was used to perform differential expression analysis and create exploratory visualizations.

Quality control and differential expression analysis

DESeq2 is an R package that is now commonly used for differential expression analysis. Prior to the development of DESeq and other similar programs, the Poisson distribution was commonly used to test for differential expression (Wang et al. 2010). However, the Poisson distribution’s single parameter requires that variance and mean are equal under this distribution. It has been shown that the Poisson distribution consistently predicts smaller variations than are observed in counts data. This overdispersion problem means that the Poisson distribution does not effectively control type I error. DESeq addresses the overdispersion problem by using the negative binomial (NB) distribution. Although other packages such as edgeR also use the NB distribution, DESeq is unique because it allows for data-driven relationships between mean and variance (Anders and Huber 2010).

Prior to differential expression analysis, we transformed the count data to make it homoscedastic. DESeq2 offers three transformations: the shifted logarithm transformation, the regularized-logarithm (rlog) transformation, and the variance-stabilizing transformation. Figure 2 plots the standard deviation of each sample against the mean using each of these transformations. The shifted logarithm transformation

showed higher standard deviations at the lower end of the range of counts, while the variance stabilizing transformation showed slightly elevated standard deviations at the higher end of the dynamic range. The rlog transformation showed the flattest curve of standard deviation relative to mean; therefore, we used rlog-transformed data for downstream exploratory analyses and visualizations.

DESeq2 tests for differential expression by 1) normalizing sequencing depth between samples using an estimate of “size factors,” 2) estimating dispersion across all samples, and 3) fitting a negative binomial generalized linear model. DESeq2 uses a Wald test to calculate a p-value for differential expression; these p-values are then adjusted for multiple testing using the Benjamini-Hochberg correction. DESeq2 also assigns \log_2 fold change values for each gene. Under the default parameters in DESeq2, the null hypothesis is that a gene is *not* differentially expressed between treatments. The alternative hypothesis is that a gene *is* differentially expressed between treatments. These default parameters were used to test for differential expression of genes between the high light and low light treatment groups. However, it cannot be assumed that genes with B-H adjusted p-values larger than our significance threshold are stably expressed. In order to test for stable expression between the two treatments, a different test must be performed. DESeq2 allows users to alter its parameters to test for stable expression. The parameters “altHypothesis = lessAbs” and “lfcThreshold = x ” were added to the results function to test for stable expression. The parameter x defines the threshold for \log_2 fold change. Unfortunately, setting a \log_2 fold change threshold of 0.5 (fold change = 1.4) did not yield any significant genes. Increasing the \log_2 fold change threshold to 1 (fold change = 2) yielded 4,270 significant genes; however, we determined that a \log_2 fold change

threshold of one was not sufficient to conclude that these genes were in fact stably expressed across treatments.

Functional annotation

The Bioconductor package KEGGREST (Tenenbaum 2018) was used to access gene annotations from the KEGG database; functional annotation was limited to KEGG pathways. Pathview Web (Luo et al. 2017) was used to visualize relative expression of genes within KEGG pathways. The protein structure and function prediction programs Phyre2 (Kelley et al. 2015) and I-TASSER (Zhang 2008) were used to try to identify the functions of unannotated genes with large fold changes.

Chapter 4: Results & Discussion

Quality control analyses

We received between 31 million and 43 million raw reads for each of nine samples (Table 1). The filtering step described in *Methods* removed approximately three million reads from each sample. Previous studies have shown that simple gene-level expression profiling requires about 5–50 million single-end reads per sample (SEQC/MAQC-III Consortium 2014); thus, our experiment produced a sufficient number of reads for differential expression analysis. The overall alignment rate (the proportion of reads that mapped to the reference genome) and the unique alignment rate (the proportion of reads that mapped to the reference genome *only once*) indicated that the medium light samples may have suffered from quality issues. Read alignment was inconsistent across replicates in this treatment; while the high and low light replicates all had unique alignment rates between 69% and 76%, the medium light replicates had unique alignment rates ranging from 47% to 75%. This variability in the unique alignment rate could have stemmed from inconsistent RNA degradation or from issues during RNA extraction, library preparation, or sequencing. Based on the observed inconsistencies, we chose to exclude the medium light replicates from downstream analyses. We concluded that a comparison of expression in the high and low light samples would still allow for an understanding of how *T. pseudonana* responds at the transcriptional level to changes in growth rate due to light limitation. The high and low light samples showed overall read alignment between 71–77%, with unique read alignment between 69–76%. Other RNA-Seq studies have observed overall alignment rates over 90% (Giannoukos et al. 2012;

Borozan et al. 2013), but unique alignment rates close to 70% are typical (Engström et al. 2013; Dong et al. 2014; Pertea et al. 2016).

Exploratory visualizations of rlog-transformed count data were used as quality control tools to check the degree of similarity between samples of the same treatment (Figures 4 & 5). The sample distance matrix (Figure 4) showed that high light triplicates were more similar to each other than to any of the low light triplicates (and vice versa). The PCA plot (Figure 5) showed that the high and low light samples clustered separately from each other on PC1, which explained 77% of the variation in the data. One high light sample, HLC, clustered separately from the other high light samples on PC2, which explained 9% of the variation in the data. We concluded that PC1 represents light level; however, we are unsure what factor is represented by PC2. Although the separation of one high light replicate from the others suggests that there may be some noise in the data, overall these exploratory visualizations indicated that the high and low light samples were of sufficient quality for downstream differential expression analysis.

Quality control steps also involved removing genes with little or no expression. Other studies have shown that for weakly expressed genes (genes with low read counts between all samples), it is very difficult to test for differential expression due to high Poisson noise that obscures biological effects. Essentially, Poisson noise dominates the variance at low count levels, introducing uncertainty into whether low counts accurately represent expression level (Busby et al. 2011; Todd et al. 2016). To investigate this, we created a histogram to show the relationship between expression levels and small p-values (Figure 6). The resulting figure shows that genes with very weak expression are never called significant, and genes with higher mean counts are called significant more

frequently. These data underscore that genes with low mean counts have little or no power and should be excluded from testing. We used DESeq2 to remove these low-count genes in a process known as independent filtering. These weakly expressed genes would not have been called significant; however, removal of low-count genes generally improves differential expression analyses by facilitating discovery of significant genes amongst the remaining gene set.

Overview

The *T. pseudonana* genome (assembly ASM14940v2) contains 11,673 protein-coding genes. 11,695 genes were represented in our initial data set, but 257 genes were removed for having either 0 or 1 total reads across all samples. We considered these genes to be not expressed. The remaining 11,438 genes were tested for differential expression. DESeq2 calculated \log_2 fold changes and Benjamini-Hochberg adjusted p-values for each gene. These values were then used to determine whether genes were differentially expressed between the high and low light samples.

Studies disagree about the best cutoffs to use—or whether or use cutoffs at all—when evaluating differential expression. Transcriptomic studies in phytoplankton often report genes with \log_2 fold changes greater than one or less than negative one (absolute fold change > 2) as significantly differentially expressed (Shin et al. 2016; Roth et al. 2017; Harke et al. 2017; Amato et al. 2017; Nan et al. 2018; Vorobev et al. 2018), but such low cutoffs could lead to the discovery of false positives caused by experimental or transcriptional noise. One study found that changing the arbitrary fold change and p-value cutoffs in a transcriptome profiling experiment significantly changed the biological interpretation of the data (Dalman et al. 2012). Although methods have been developed to

specifically test for biologically meaningful genes (McCarthy and Smyth 2009) and to minimize experimental noise (Peixoto et al. 2015), these methods do not appear to be widely utilized in gene expression studies. One study suggests that using fold-change rankings and less stringent p-value cutoffs may lead to more reproducible results. (MAQC Consortium 2006).

In this study, we chose to compare our results using fold change cutoffs of 2 and 4. In both cases, the adjusted p-value cutoff was 0.05. Table 2 shows the summarized differential expression analyses using both of these cutoffs. Under the more stringent cutoff, we found 598 genes—only 5% of genes in the genome—were called differentially expressed between the high light and low light cells. 343 (57%) of these genes are upregulated in low light relative to high light while 255 (43%) genes were downregulated in low light relative to high light. Out of these 598 genes, 75 genes had available KEGG annotations. 96% of genes that are upregulated in low light relative to high light are unannotated, while 77% of genes downregulated in low light are unannotated. Using the less stringent cutoff, we found 2,904 genes (25% of the genome) were called differentially expressed. 1,410 (49%) of these genes were upregulated in low light relative to high light and 1,494 (51%) were downregulated in low light relative to high light. 94% of the 1,410 upregulated genes were unannotated, while 76% of the 1,494 downregulated genes were unannotated.

One study focusing on growth-rate control in yeast found that genes that are downregulated at lower growth rates tend to be essential genes that encode proteins of known function, while genes with higher expression in slow-growing cells are more likely to have no known function (Castrillo et al. 2007). We observed a similar trend;

approximately 95% of genes upregulated in low light were unannotated in comparison to approximately three-quarters of genes downregulated in low light. Castrillo et al. (2007) also found that 15% of total protein-coding genes in the yeast genome were regulated by growth rate. This is in line with our own study, which found that between 5 and 25% of genes are differentially expressed at different growth rates (depending on which fold change cutoff is used). These findings contrast with studies in diatoms examining adaptive responses to nutrient stress, which found that at least half of all genes were differentially expressed in *T. pseudonana* and *P. tricornutum* growing under nitrogen starvation (Yang et al. 2013; Levitan et al. 2015). Although we hypothesized that the large range in growth rates between cells growing at low and high light would lead to a large number of genes being differentially expressed, the data suggests that cells' regulated growth programs actually involve fewer genes than their adaptive responses. This could be because cells that are undergoing an adaptive response to nutrient starvation are recalibrating their physiologies via numerous metabolic adjustments, while cells that are growing in steady state under limitation conditions have only to maintain their acclimated metabolisms. It is possible that the process of metabolic adjustment requires a higher degree of regulation than does steady-state growth; however, neither this study nor Castrillo et al. (2007) explores other regulatory mechanisms (*i.e.*, post-transcriptional, post-translational, or redox poise) that may be involved in eukaryotic growth rate control. Thus, it is difficult to know whether steady-state growth actually necessitates a lower degree of regulation or if cells growing in steady state simply rely more heavily on post-transcriptional controls to optimize growth under limitation conditions.

Patterns of differential gene expression

Using the conservative cutoff of $|FC| > 4$, a total of 256 genes—representing 2.2% of the full *T. pseudonana* genome—were downregulated in slow-growing cells relative to fast-growing cells. Only 60 downregulated genes were annotated in the KEGG database—77% of downregulated genes are unannotated. Figure 7A shows the distribution of annotations for the set of downregulated genes. Represented in the set of downregulated genes are genes encoding ribosomal proteins and genes involved in carbon metabolism, protein processing and export, RNA processing, oxidative phosphorylation, and biosynthesis of secondary metabolites. Downregulated genes related to carbon metabolism were involved in glycolysis, the TCA cycle, glycine, serine, and threonine metabolism, and cysteine and methionine metabolism. The downregulation of genes encoding ribosomal proteins is reflective of lower rates of protein synthesis in slower growing cells. Previous studies in yeast and bacteria have found the transcript levels of genes encoding ribosomal proteins and cellular concentration of ribosomes decreases with decreasing growth rate (Regenberg et al. 2006; Dressaire et al. 2008; Klumpp et al. 2009). In growth-limiting conditions, synthesis of other necessary proteins takes precedence over the synthesis of ribosomal proteins; under favorable conditions that permit fast growth, ribosomal proteins are synthesized at the expense of all other proteins.

342 genes—2.9% of the genome—were upregulated (fold change > 4) in slow-growing cells relative to fast-growing cells. 96% of upregulated genes are unannotated. Figure 7B shows the breakdown of annotations of upregulated genes. Genes with known functions were involved in fatty acid metabolism and valine, leucine, and isoleucine

degradation. Only two of these upregulated genes, *PDK1_2* and *PCK1*, were involved in carbon metabolism. *PDK1_2* encodes pyruvate-phosphate dikinase (PPDK), an enzyme that catalyzes the conversion of pyruvate to phosphoenolpyruvate (PEP). PPDK is canonically considered an essential enzyme of the C₄ pathway. The C₄ pathway is a photosynthetic process that overcomes the tendency of rubisco to catalyze a reaction between ribulose 1,5-bisphosphate and oxygen instead of the desired reaction with carbon dioxide. In diatoms, however, the role of the C₄ pathway is still being debated. Previous studies concluded that *Thalassiosira weissflogii* uses a C₃–C₄ intermediate pathway while *T. pseudonana* exclusively uses C₃ photosynthesis (Roberts et al. 2007; Kroth et al. 2008). However, a more recent study suggests that *T. pseudonana* does use the C₄ pathway (Kustka et al. 2014). We propose that upregulation of PPDK in slow-growing *T. pseudonana* is unrelated to the C₄ pathway, since C₄ photosynthesis is unlikely to be beneficial to cells growing under light-limited conditions. Instead, slow-growing *T. pseudonana* upregulate PPDK to increase availability of PEP, which is an important intermediate of glycolysis and gluconeogenesis. Increasing the availability of PEP will increase flux through gluconeogenesis, which may be important for maintaining carbon for growth in slow-growing cells. This idea is supported by upregulation of *PCK1*, which encodes phosphoenolpyruvate carboxykinase (PEPCK). PEPCK catalyzes the conversion of oxaloacetate (OAA) to PEP and CO₂. We propose that PPDK and PEPCK play a role carbon partitioning. These enzymes function at the intersection of glycolysis and gluconeogenesis with the TCA cycle, fatty acid biosynthesis, and amino acid metabolism (Smith et al. 2012). This intersection point is a strong candidate for regulation of carbon flux, and strong upregulation of *PDK1_2* (FC = 6.65, $p = 6.96 \times 10^{-20}$) and *PCK1* (FC =

7.26, $p = 4.28 \times 10^{-19}$) in *T. pseudonana* grown under low light suggests that these enzymes together play an important role in pathway gating that optimizes carbon use under light-limited conditions. The role of these two enzymes in light-limited growth rate control is discussed in further detail in a later section.

Using the less conservative cutoff ($|FC| > 2$), 1,494 genes—representing 13% of the full genome—were downregulated in low light relative to high light. 360 downregulated genes (24%) were annotated in the KEGG database. Figure 8A shows the distribution of KEGG annotations for downregulated genes based on the lower cutoff. Again, genes encoding ribosomal proteins made up a significant fraction of downregulated transcripts. However, the less conservative cutoff did lead to other changes in the overall distribution of KEGG annotations. We observed a greatly increased number of significantly downregulated genes with annotations relating to carbon metabolism. While the more conservative cutoff resulted in only 7 genes involved in carbon metabolism, the less conservative cutoff resulted in 40 downregulated genes involved in carbon metabolism; 22 of these genes were involved in glycolysis.

The less conservative cutoff also resulted in 1,410 genes—12% of the genome—that were upregulated in low light relative to high light. 86 upregulated genes (6.1%) had annotations in the KEGG database. Figure 8B shows the distribution of KEGG annotations for upregulated genes using the lower cutoff. The most highly represented categories were fatty acid metabolism, endocytosis, amino sugar and nucleotide sugar metabolism, and valine, leucine, and isoleucine degradation. Previous studies have suggested a link between metabolism of valine, leucine, and isoleucine and central carbon metabolism in diatoms (Smith et al. 2016a; Levering et al. 2017; Broddrick et al. 2019),

and the lower cutoff increased the number of significant genes involved in this process from three to seven. While the lower cutoff did not significantly increase the number of differentially expressed genes belonging to a given category, it did increase the number of categories represented in the pool of significantly differentially expressed genes.

As expected, lowering the fold-change threshold at which genes are called significant greatly increased the number of differentially expressed genes. However, many of the observed trends in gene expression stayed the same. Genes that were upregulated in slow-growing cells were more likely to have no known function, genes encoding ribosomal proteins were highly represented in the set of downregulated genes, and differentially expressed genes involved in carbon metabolism were more likely to be downregulated than upregulated. In their study investigating growth rate regulated genes in yeast, Regenberg et al. (2006) also found that slower growth rates resulted in downregulation of genes encoding ribosomal proteins and that genes that are upregulated at higher growth rates tend to have no known function. While using a lower fold-change threshold increases the number of genes of interest, it also increases the chance of false positives (genes called differentially expressed when they are not). Whenever possible, transcriptomic studies should provide supporting evidence of biological significance for genes called differentially expressed. Integration of transcriptomic and proteomic approaches would be helpful in addressing the problem of biological significance in transcriptomic studies—proteomics data can provide insight into post-transcriptional regulatory processes and help determine whether or not changes in transcript abundance affect cellular functioning.

Magnitude of differential expression

The average fold change of genes with $p < 0.05$ was 2.4. This is in line with other studies in diatoms that observed similar average fold changes (Shrestha et al. 2012; Dyhrman et al. 2012; Cheng et al. 2014; Bender et al. 2014), although the number of differentially expressed genes tends to differ based on study conditions. Based on the large difference in growth rate between slow-growing cells and fast-growing cells, we initially hypothesized that differentially expressed genes would have larger-than-average fold changes. However, most previous studies focusing on genes associated with growth rate have not reported fold changes, making it difficult to ascertain the general magnitude of differential expression in response to growth rate. The heatmap in Figure 9 shows a broad-scale view of the magnitude of differential expression observed in this study. The heatmap shows that there are differences in expression level between samples within a treatment—a phenomenon that is common in transcriptional studies and may be attributed to either biological or experimental noise—but that these differences are dwarfed by the differences in expression level between the two treatments. Additionally, although there are clear patterns of differential expression between treatments, the magnitude of the differential expression is relatively small—most rows are represented by paler colors, which correspond to less deviation from the mean.

Genes with the highest levels of differential expression were not annotated in the KEGG database. To investigate their function, we used Phyre 2 and I-TASSER to explore the predicted protein structure of these genes. These tools, however, were not able to predict the function of these highly differentially expressed proteins. Figure 10 shows the Phyre 2 prediction of the protein structure for *THAPSDRAFT_8336*. This gene

experienced a 4,090-fold increase in expression in slow-growing cells; however, Phyre 2 was only able to model 14 residues (3% of the total sequence) with 43.4% accuracy. Similarly, I-TASSER identified structural analogs with low identity (between 4 and 10%) to the protein sequence of *THAPSDRAFT_8336*, and all model predictions had low confidence scores (between -3.86 and -2.45). This observation was seen in many of the genes with higher than average fold change values. I-TASSER and Phyre both rely on homology to predict protein structure and function (Zhang 2008; Kelley et al. 2015); therefore, their inability to provide adequate predictions of protein structure suggest that these highly differentially expressed genes are diatom- or species-specific. Further investigation of these highly regulated genes could uncover previously unknown functions that play important roles in the coordinated growth programs of diatoms.

The TCA and glyoxylate cycles

Lack of annotations for a majority of differentially expressed genes presented a major obstacle to this study. To facilitate analysis, we focused on central metabolic pathways and looked for differentially expressed genes using the same B-H adjusted p-value cutoff of 0.05 but without using a fold change threshold to define differential expression. Using this approach, we observed that several genes involved the TCA and glyoxylate cycles were differentially expressed. Given the strong upregulation of phosphate-pyruvate dikinase, which works at the intersection of glycolysis and gluconeogenesis, the TCA cycle, and other metabolic pathways (Table 3 & Figure 13), and the role of the glyoxylate cycle in metabolism of photorespiratory glycolate (Figure 12B), we determined these differentially expressed genes could be involved the response

of *T. pseudonana* to light limitation even though in some cases the fold changes of the genes were lower than the previously discussed thresholds. The TCA cycle is used by all aerobic organisms to oxidize acetyl-CoA to produce ATP, carbon dioxide, and reducing agents that can be further converted into ATP via other metabolic pathways such as oxidative phosphorylation (Figure 12A). Acetyl-CoA entering the TCA cycle is derived from the oxidation of fatty acids, proteins, and carbohydrates. Thus, the TCA cycle is a central metabolic hub with important roles in both energy production and biosynthesis. The glyoxylate cycle is a variation of the TCA cycle that takes place in peroxisomes and does not produce CO₂ as a waste product; it achieves this via a two-step bypass (Figure 12A). In the first step, isocitrate lyase converts isocitrate into succinate and glyoxylate; in the second step, malate synthase condenses glyoxylate and acetyl-CoA to form malate. The resulting malate can replenish the glyoxylate cycle through the action of malate dehydrogenase, while the succinate can be used to replenish the TCA cycle or serve as precursors for carbohydrate or amino acid biosynthesis (Kunze et al. 2006). Thus, the glyoxylate cycle serves as a link between catabolism of fatty acids or amino acids for acetyl-CoA and biosynthesis of carbohydrates or amino acids. This link highlights the importance of the glyoxylate cycle in growth optimization in slow-growing cells. The pattern of differential expression we observed between the two treatment groups suggests that slow-growing cells preferentially shunt carbon through the glyoxylate cycle as a way to conserve carbon. Fold changes and p-values for differentially expressed genes are shown in Table 3.

Genes involved in branched-chain amino acid (BCAA) and fatty acid degradation were upregulated in slow-growing cells (Figure 11). Both of these processes produce

acetyl-CoA, which enters the glyoxylate and TCA cycles. Recent transcriptomic studies in *P. tricornutum* have shown that genes involved in BCAA metabolism and the TCA cycle are coregulated, suggesting a link between BCAA metabolism and central carbon metabolism in diatoms (Smith et al. 2016a; Levering et al. 2017; Broddrick et al. 2019). The three proteinogenic BCAAs are valine, leucine, and isoleucine. In our study, seven genes involved in degradation of these three amino acids were upregulated in slow-growing cells. *THAPSDRAFT_795*, *THAPSDRAFT_3692*, and *THAPSDRAFT_32067* putatively make up the E1 and E2 subunits of 2-oxoisovalerate dehydrogenase. *THAPSDRAFT_413* is a putative 3-hydroxyisobutyrate dehydrogenase that is involved specifically in the degradation of valine. This enzyme catalyzes the reaction between 3-hydroxyisobutyrate and NAD^+ to produce NADH and methylmalonate semialdehyde. *THAPSDRAFT_25495* is involved in the reversible conversion of propanoyl-CoA to methylmalonate semialdehyde. *HCD1* is specifically involved in the degradation of isoleucine, while *PCBI* is specifically involved in the degradation of valine. *ACD3*, a short/branched chain acyl-CoA dehydrogenase that is involved both in leucine, isoleucine, and valine degradation and in fatty acid degradation, is upregulated in slow-growing cells. Fatty acid degradation also produces acetyl-CoA that feeds in to the TCA and glyoxylate cycles. *FCLI*, a long-chain acyl-CoA synthetase involved in fatty acid degradation, is also upregulated in slow-growing cells.

Peroxisomes do not contain translation machinery; therefore, all peroxisomal enzymes must be imported from the cytosol (Kunze et al. 2006; Gonzalez et al. 2011). Gonzalez et al. (2011) found that *P. tricornutum* uses one of two known peroxisomal targeting signals (PTS) in combination with a series of peroxins (Pex) that recognize

these signals and transport cytosolically-synthesized proteins into the peroxisomal lumen. We found that three Pex genes (of the six known to exist in *T. pseudonana*) are upregulated in slow-growing cells relative to fast-growing cells: Pex5, Pex1, and Pex19 (Table 3 & Figure 12B). Pex5 is a cycling receptor that plays a critical role in protein transport; it is responsible for carrying proteins equipped with type 1 peroxisome targeting signal (PTS1) to the peroxisomal lumen (Williams and Stanley 2010). Pex1 is a AAA (ATPase associated with diverse cellular activities) protein involved in recycling PTS receptors (Williams and Stanley 2010; Cross et al. 2016). One study in *Arabidopsis thaliana* found that knockdown of Pex1 decreased import of proteins containing both known peroxisomal targeting signals (Nito et al. 2007). Pex19 binds and imports peroxisomal membrane proteins (Jones et al. 2004). Upregulation of these three genes suggests that slow-growing cells have higher peroxisomal activity; this supports our conclusion that slow-growing cells increase flux through the glyoxylate cycle.

Many genes directly involved in the TCA and glyoxylate cycles were also differentially expressed (Table 3 & Figure 12A). Genes involved in reactions bypassed by the glyoxylate cycle were all downregulated with the exception of isocitrate dehydrogenase, which catalyzes the reaction between isocitrate and α -ketoglutarate. The gene encoding isocitrate dehydrogenase was not significantly regulated in either direction. The gene encoding malate synthase, which condenses acetyl-CoA with glyoxylate to form malate in the second step of the glyoxylate cycle bypass, was upregulated in slow-growing cells. Malate dehydrogenase, which converts malate to oxaloacetate, was downregulated. Taken together, these data suggest that slow-growing cells increase flux through the glyoxylate cycle and decrease flux through the TCA cycle.

Davis et al. (2017) proposed that cells use the glyoxylate cycle to preserve carbon and limit excess energy production. We hypothesize that cells growing slowly under low growth irradiances increase flux through the glyoxylate cycle to conserve carbon. The rate of carbon fixation in slow-growing cells is limited by low light availability, and coupled with a higher demand for maintenance energy (Fisher and Halsey 2016), may cause the cells to impose additional regulatory controls on carbon flow. Diverting flux from the TCA cycle—which produces carbon dioxide as a byproduct—to the glyoxylate cycle may therefore help slow-growing cells overcome their carbon limitation.

Additionally, the glyoxylate cycle generates succinate that can be used for carbohydrate synthesis. Increasing synthesis of carbohydrates may be especially important for slow-growing cells, in which newly fixed carbon is more rapidly turned over compared to fast-growing cells (Halsey et al. 2013). The glyoxylate cycle allows slow-growing cells to conserve carbon and generate carbohydrates, thus maintaining the correct balance of macromolecules necessary for growth. This hypothesis is supported by flux balance models in *Arabidopsis*, which predicted that reactions between α -ketoglutarate and fumarate (steps bypassed by the glyoxylate cycle and downregulated in slow-growing cells) would experience no flux when the model is constrained only by the need to synthesize macromolecules in the correct proportions (Poolman et al. 2009). One challenge associated with using the glyoxylate cycle rather than the TCA cycle is that the glyoxylate cycle produces fewer NADH molecules than the full TCA cycle. However, a flux balance analysis in *P. tricornutum* found that the low levels of oxidative phosphorylation supported by the glyoxylate cycle and substrate-level phosphorylation in glycolysis are sufficient to meet the ATP demands of autotrophic cells (Kim et al. 2016).

The glyoxylate cycle also represents one of the two main fates of photorespiratory glycolate, which is produced by the oxygenation of ribulose 1,5-bisphosphate by the rubisco enzyme (Figure 12B). This glycolate can also be metabolized by the photorespiratory pathway to generate 3-phosphoglycerate (3-PGA), an intermediate of the Calvin-Benson cycle. However, one in four carbons entering the photorespiratory pathway is used to generate CO₂ rather than 3-PGA (Oliver 1998). Thus, the glyoxylate cycle and the photorespiratory pathway may play roles in regulating carbon loss (Schnitzler Parker et al. 2004). In marine algae, creation of photorespiratory glycolate has been shown to increase in high light conditions (Beardall 1989; Schnitzler Parker et al. 2004). Upregulation of the glyoxylate cycle in slow-growing cells may reflect the need for these cells to process photorespiratory glycolate created under fluctuating light conditions. Although cells were grown under continuous light for this study, the marine light environment is highly dynamic and even cells experiencing low average irradiances are likely subject to fluctuations in their light environment. Previous data showed that cells growing at low light essentially eliminate non-photochemical quenching (Bryce Penta, unpublished). We hypothesize that slow-growing diatoms, in the absence of NPQ, require a way to deal with excess light energy that leads to the creation of glycolate. Genes encoding two components of glycine decarboxylase and the gene encoding serine hydroxymethyltransferase were downregulated in slow-growing cells (Figure 12B). Both of these enzymes are involved in the conversion of glycine to serine in the photorespiratory pathway. Downregulation of genes involved in the photorespiratory pathway suggests that slow-growing cells preferentially shunt photorespiratory glycolate through the glyoxylate cycle. We propose that slow-growing cells may be

downregulating the photorespiratory pathway because they have a lower need for 3-PGA input to the Calvin-Benson cycle due to relatively slow rates of carbon fixation.

Downregulation of the photorespiratory pathway also conserves CO₂ in slow-growing cells, possibly to compensate for lower rates of carbon fixation.

Increasing flux through gluconeogenesis

Investigation of genes involved in the TCA and glyoxylate cycles led to the observation that several genes involved in processes that regulate flux through glycolysis, gluconeogenesis, and the TCA cycle were differentially regulated (Table 3 & Figure 13). As discussed above, genes encoding PPDK and PEPCK were upregulated in slow-growing cells. These genes are involved in processes that generate PEP, an intermediate of glycolysis and gluconeogenesis. These genes appear to be more strongly regulated than other differentially expressed genes, with fold changes of 6.65 (*PDK1_2*) and 7.26 (*PCK1*). This high degree of regulation suggests that this intersection of glycolysis, gluconeogenesis, and the TCA plays an important role in optimizing growth in response to light limitation in *T. pseudonana*. Further supporting this idea, we observed that *PYK1* and *PYK2* were also downregulated in slow-growing cells. These genes encode pyruvate kinase, which catalyzes the conversion of PEP to pyruvate (the reverse of the reaction catalyzed by PPDK). The coordinated downregulation of genes encoding pyruvate kinase and upregulation of the gene encoding PPDK suggests that slow-growing cells increase the availability of PEP in order to increase flux through gluconeogenesis. Additionally, the genes encoding PEP carboxylase (PEPC) and pyruvate carboxylase are downregulated. PEPC catalyzes the reaction between PEP and CO₂ that generates OAA,

while pyruvate carboxylase catalyzes also produces OAA by catalyzing a reaction between CO₂ and pyruvate.

These data suggest that slow-growing cells increase flux through gluconeogenesis. Cells may utilize this strategy as a way to conserve carbon, since slow-growing cells have slower rates of carbon fixation and experience high flux through glycolysis to supply ATP for maintenance demands. Fisher and Halsey (2016) showed that *T. pseudonana* grown in light-limited conditions allocated about 42% of *newly* fixed carbon (carbon fixed during a 20-minute incubation) to carbohydrates; however, the proportion of carbon allocated to this macromolecular pool decreased to about 19% of total fixed carbon after 24 hours. This decrease suggests that slow-growing cells experience high flux through glycolysis, likely in order to meet their high demand for ATP. However, slow-growing cells with high flux through glycolysis may have problems maintaining carbon for growth unless they have mechanisms to decrease catabolic flux and increase anabolic flux. Indeed, previous studies have shown that the lifetime of newly fixed carbon decreases with decreasing growth rate (Halsey et al. 2013). Our data suggest that slow-growing cells utilize the glyoxylate cycle and increase flux through gluconeogenesis in order to minimize carbon loss and maintain a balance of macromolecules that is conducive to growth.

Chapter 5: Conclusion

Conclusions

The goal of this study was to identify genes associated with physiological responses to steady-state, light-limited growth in *T. pseudonana*. We focused on light-limited growth because we were interested in genes associated with growth rate rather than with adaptive responses to particular environmental variables. We explored using different fold change thresholds to define differential expression; we found that changing the threshold changed the number of genes called differentially expressed but did not significantly alter the broad-scale trends in the distribution of gene functions. We found that *T. pseudonana*'s physiological responses to light-limited growth are controlled by relatively subtle changes in gene expression (between 5% and 25% of genes were up- or downregulated, depending on the cutoff used), in contrast with previous studies showing that diatoms differentially express around 50% of their genes in response to nutrient starvation (Yang et al. 2013; Levitan et al. 2015). The mean fold change was 2.4, which agreed with previous studies exploring the effect of nutrient starvation on diatoms. These data contrast with our original hypothesis that a large number of genes would experience large fold changes due to the wide range in growth rate (cells grown under high light grew almost eight times faster than cells grown under low light). Taken together, these data suggest that acclimation to changing environmental variables requires more widespread transcriptional control than optimizing and maintaining growth under resource-limited conditions. This is likely because cells in the process of acclimation are undergoing numerous metabolic adjustments, while cells growing in steady state are fully acclimated to their environment and are simply maintaining their physiological states.

We also found that light-limited cells differentially express a high proportion of genes involved in the glyoxylate cycle, TCA cycle, and photorespiratory pathway (Table 3). Genes involved in the glyoxylate cycle were upregulated in slow-growing cells, while genes involved in the TCA cycle and photorespiratory pathway were downregulated (Figure 12). These results, along with the observation that genes involved in catabolic processes producing acetyl-CoA were upregulated in slow-growing cells (Figure 11), show that light-limited cells use the glyoxylate cycle in order to conserve carbon and maintain an appropriate balance of macromolecules. Genes involved in regulating flux through gluconeogenesis, glycolysis, and the TCA cycle were also differentially expressed, suggesting that the intersection of these metabolic pathways is a central regulation point for cells growing in steady state. Genes involved in reactions that produce PEP were upregulated in slow-growing cells (Table 3 & Figure 13), indicating that these cells increase flux through gluconeogenesis. In addition to having high catabolic activity in order to meet their relatively high demand for maintenance energy, slow-growing cells also experience low rates of carbon fixation (Fisher and Halsey 2016). Increasing flux through gluconeogenesis may therefore be required to prevent carbon depletion.

While the differential expression of genes involved in these major metabolic pathways suggests that *T. pseudonana* does tightly regulate genes involved in carbon metabolism, the generally low number of differentially expressed genes might also indicate that *T. pseudonana* loosens transcriptional control over many other metabolic pathways in order to facilitate rapid acclimation to fluctuating conditions. Marchant et al. (2017) observed that denitrifying communities in coastal sediments did not exert

transcriptional control over denitrification genes even when O₂ concentrations fluctuated with tidal cycles. They proposed that the stable expression of genes involved in this key metabolic pathway allowed cells to rapidly exploit conditions favorable to denitrification. *T. pseudonana* may similarly relax transcriptional control over metabolic pathways that are not essential to regulating carbon flux and maintaining growth in order to facilitate rapid responses to fluctuating light. Previous work has shown that diatoms sometimes constitutively express genes involved in photosynthesis (Broman et al. 2017), supporting the idea that stable expression of some genes may be an adaptation to highly dynamic environments.

Impact

Elucidation of the molecular mechanisms that diatoms use to achieve high growth efficiencies under different light levels may contribute to the advancement of applied and ecosystems sciences. Diatoms are promising candidates for biofuel production and natural product discovery. They are highly productive, environmentally flexible, and can rapidly accumulate starch and lipids that can be used to produce bioethanol and biodiesel (Hildebrand et al. 2012; Yen et al. 2013). These factors make diatoms ideal candidates for large-scale production of algal biofuels; however, their complex evolutionary history has led to novel metabolic networks and growth strategies (Hildebrand et al. 2013) that have impeded efforts to optimize biofuel production. Algal productivity is ultimately dependent on the efficiency of carbon fixation and the rate of downstream metabolic processes, so an improved understanding of the genetic controls on diatom metabolism could significantly increase our ability to control algal productivity and make biofuel production economically viable.

Algae are also an important source of natural products (Cardozo et al. 2007). Recently, algal natural products have been recognized for their potential applications in food, cosmetics, and clinical drugs (Arad and Levy-Ontman 2010). Metabolites of interest are produced by a variety of metabolic pathways, and some are produced by modifications to pathways or by combinations of multiple pathways (Cardozo et al. 2007). Thus, understanding the controls on algal metabolism and the interactions between different metabolic pathways is integral to discovering new natural products and to determining whether known natural products can be efficiently exploited.

Additionally, diatom productivity is a major factor in many global models of marine systems (Aumont et al. 2003; Gregg et al. 2003; Moore et al. 2004). Elucidation of the molecular mechanisms controlling productivity in diatoms may provide finer resolution for these models, thereby improving our understanding of how carbon and other nutrients are cycled through marine environments. Growth rate is specifically included in these models (Moore et al. 2004), so a greater understanding of the mechanisms controlling growth rate in diatoms is also likely to be particularly beneficial.

Future directions

Today's technology has made sequencing diatom genomes faster and simpler than ever, but gene annotation continues to pose a problem to researchers. In our study, only 12.5% of differentially expressed genes and 13.6% of stably expressed genes had annotations in the KEGG database. Although comprehensive gene model annotations are crucial to accurate expression profiling (SEQC/MAQC-III Consortium 2014), the rate at which new diatom genomes are becoming available is exceeding the rate at which genes

are being annotated. The novelty of diatom genomes and metabolic networks makes annotation of diatom genomes relatively complicated, as many diatom-specific genes have no known orthologs. Future studies should focus on completing gene model annotations for the model diatoms *T. pseudonana* and *P. tricornutum*.

Future studies may also integrate transcriptomic and proteomic data to provide even deeper insights into cellular functioning. Investigating the connections between gene expression and protein abundance could provide systems-level information about transcriptional and regulatory networks (Fornie and Stitt 2012). In one study linking transcriptomics and proteomics in *T. pseudonana*, 60% of proteins with increased abundance had corresponding upregulated transcripts, while 30% of proteins with decreased abundance has corresponding downregulated transcripts (Dyhrman et al. 2012). This could be due to differences in the expression and turnover times for transcripts relative to proteins, or to variability in post-transcription and post-translational regulation. Further exploration of these differences may offer insight into diatoms' regulatory networks. The fine-scale data produced by this integrated approach may be very useful for refining both productivity models and quantitative models of cell growth.

Flux balance analysis (FBA) is another method that may provide important fine-scale physiological data. FBA utilizes metabolic models to study the biochemical networks of organisms. Network reconstructions contain all of the known metabolic reactions in each organism and can be used to calculate the flux of metabolites through each metabolic pathway (Orth et al. 2010). This information can then be used to predict growth rate as well as the production rate of any compound of interest. A metabolic model of *P. tricornutum* is available, and some studies have already utilized this model to

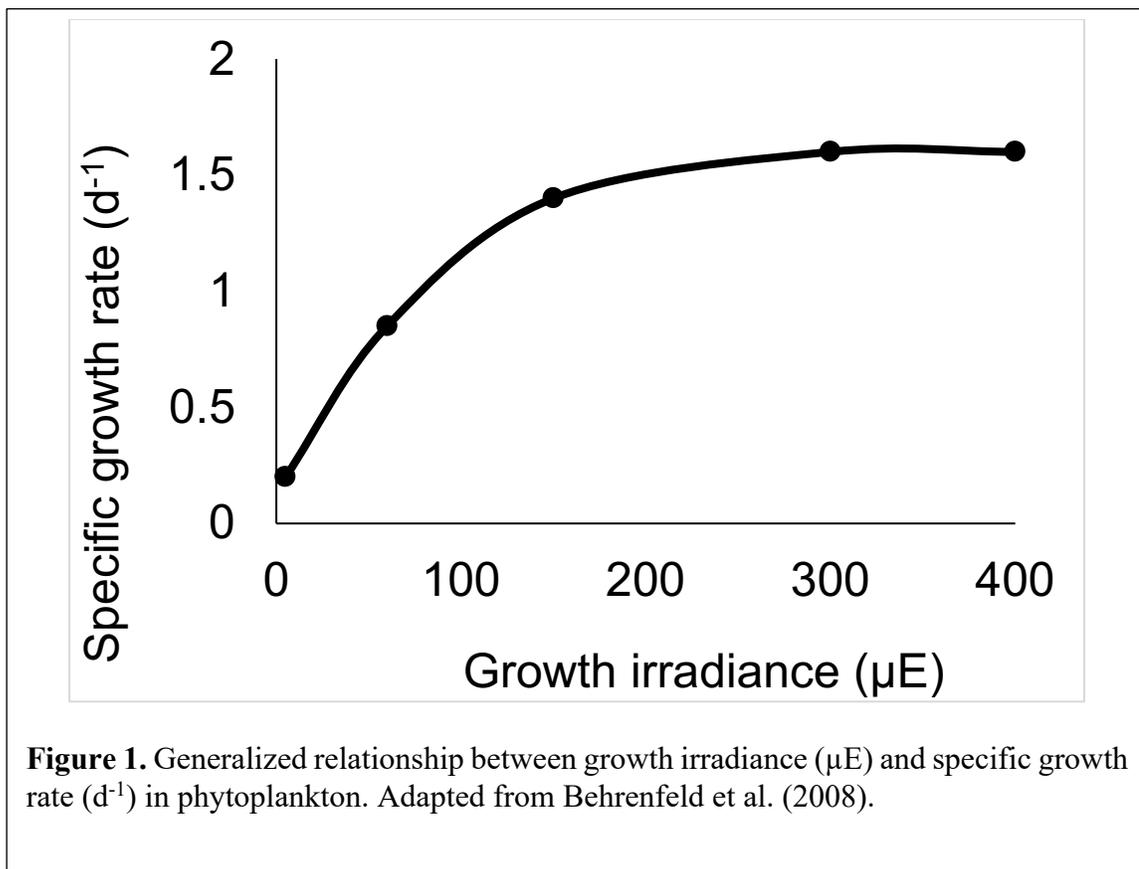
examine *P. tricornutum*'s metabolism and response to different environmental variables (Levitan et al. 2015; Kim et al. 2016; Levering et al. 2016). Currently, there is no complete metabolic model for *T. pseudonana*—a complete metabolic model would allow for a better understanding of the controls on growth rate and carbon allocation in *T. pseudonana*.

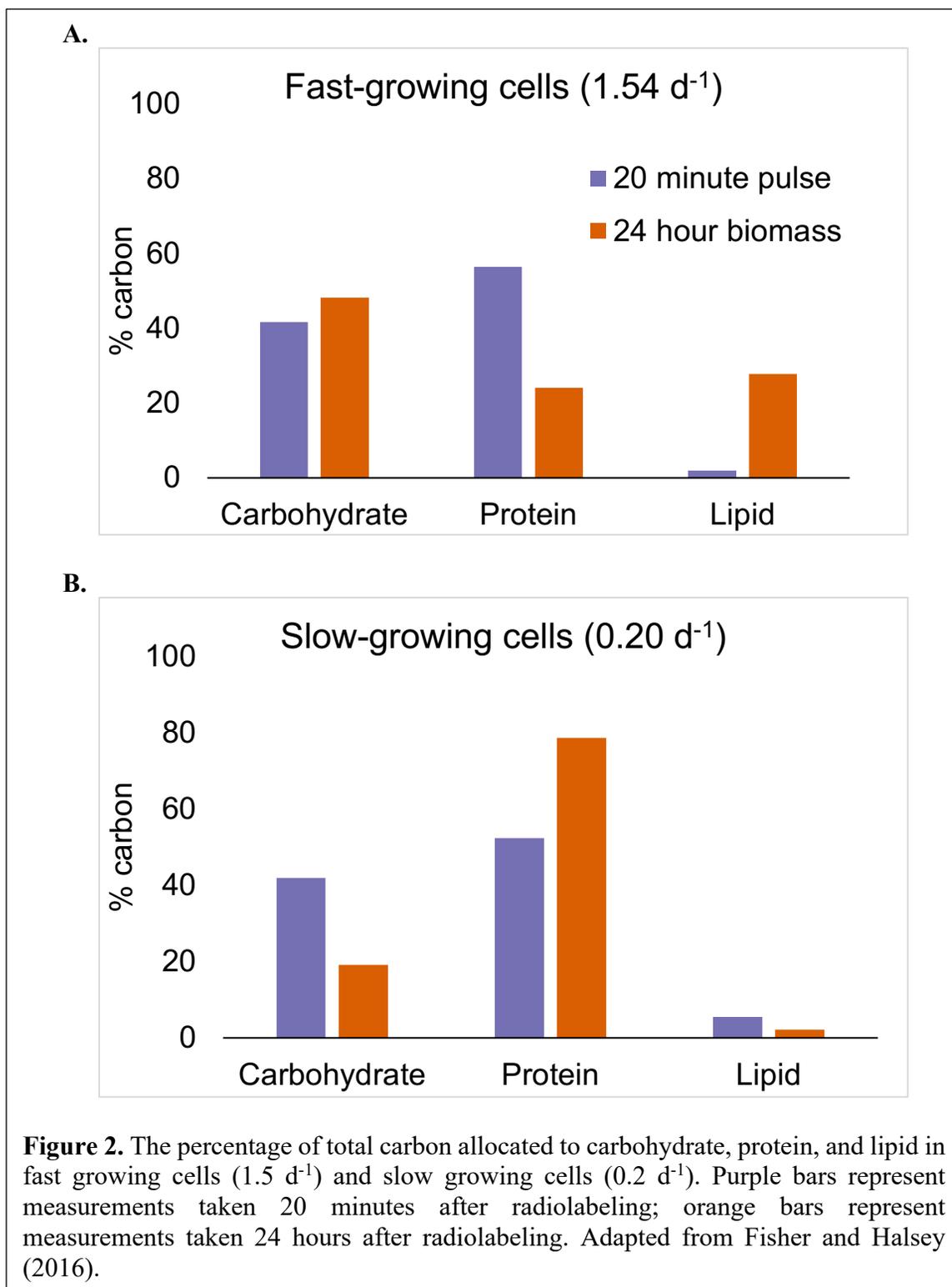
Future studies should also explore the effect of light quality on diatom growth. The factors that affect light availability (climate change, aerosols, cloud cover) are also likely to change light quality in the ocean. This study focuses solely on light intensity because previous studies showed that photosynthetic performance is more strongly affected by changes in light intensity than by variations in light quality (Morel et al. 1987; Falkowski and LaRoche 1991). However, more recent studies suggest that phytoplankton are able to perceive light quality through the use of photoreceptors, and that light quality may regulate important processes such as photoacclimation, pigment synthesis, and transcriptional regulation (Coesel et al. 2008, 2009; Schellenberger Costa et al. 2013).

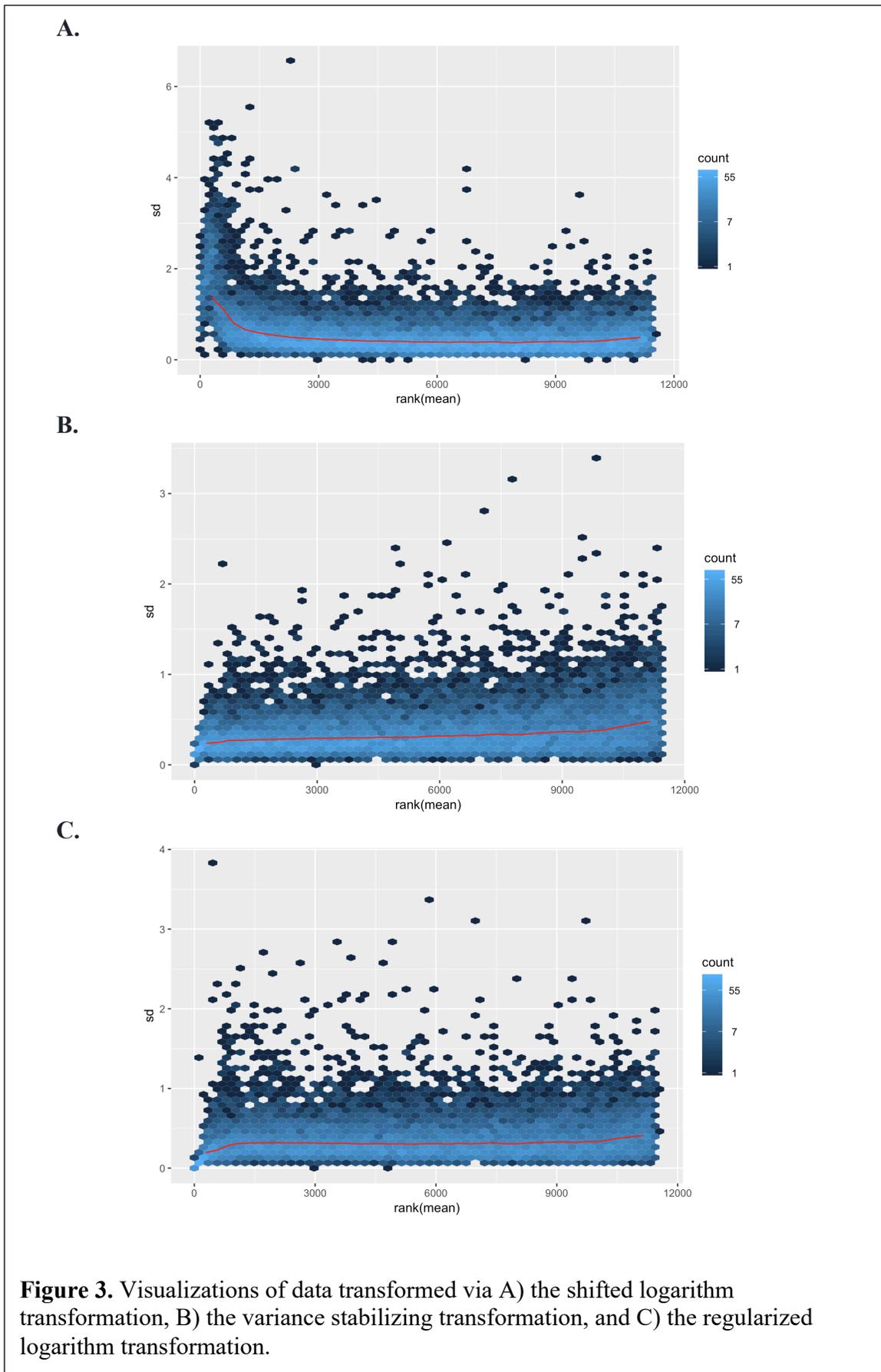
Finally, future studies should explore whether conclusions derived from lab-grown cultures of *T. pseudonana* are generalizable to natural populations. Although *T. pseudonana* was chosen as a model organism for sequencing and deeper study based on the assumption that this species is representative of the marine genus *Thalassiosira*, it has been suggested that *T. pseudonana* is actually descended from freshwater diatoms (Alverson et al. 2011). This ancestral history may have shaped several important physiological or life history traits, raising the possibility that *T. pseudonana* may not be as representative of the genus as once thought; further studies should explore this

possibility. Additionally, *T. pseudonana* has been in culture for over 50 years. It is quite possible that wild populations of *T. pseudonana* now differ in significant ways from the clone that is widely available to researchers. The genome, in particular, may be subject to change in response to different selective pressures experienced by cultivated and wild populations.

Chapter 6: Figures & Tables

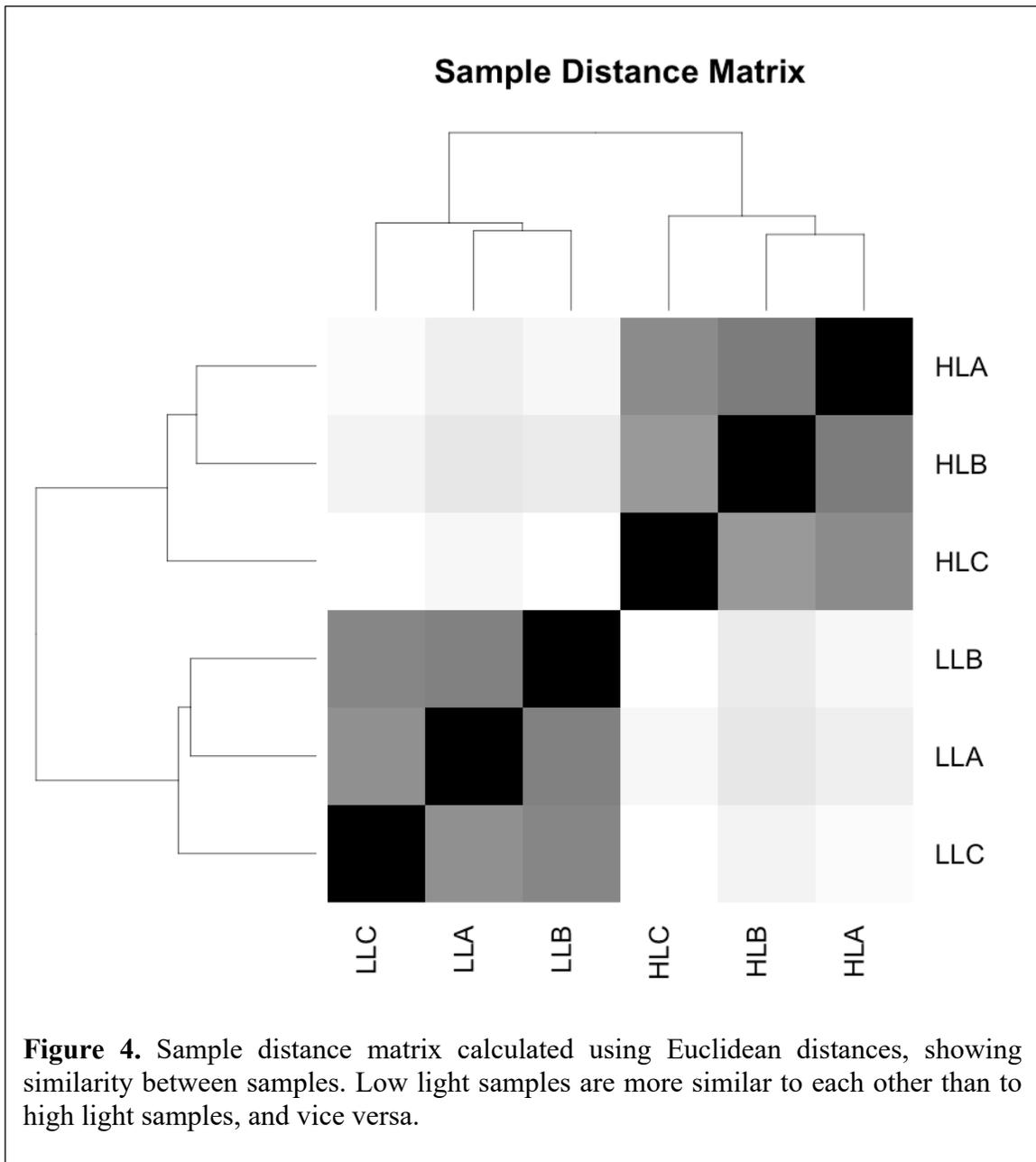


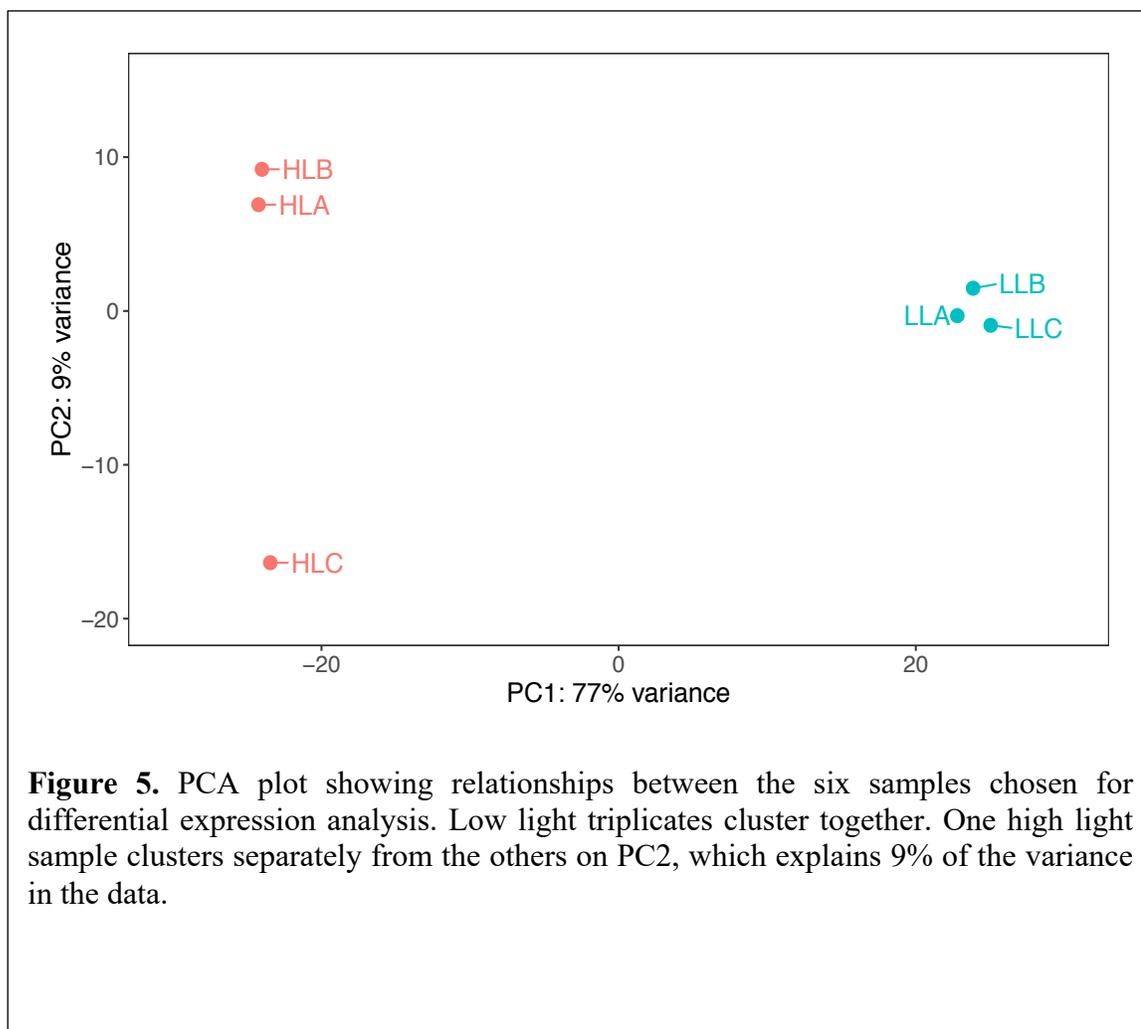


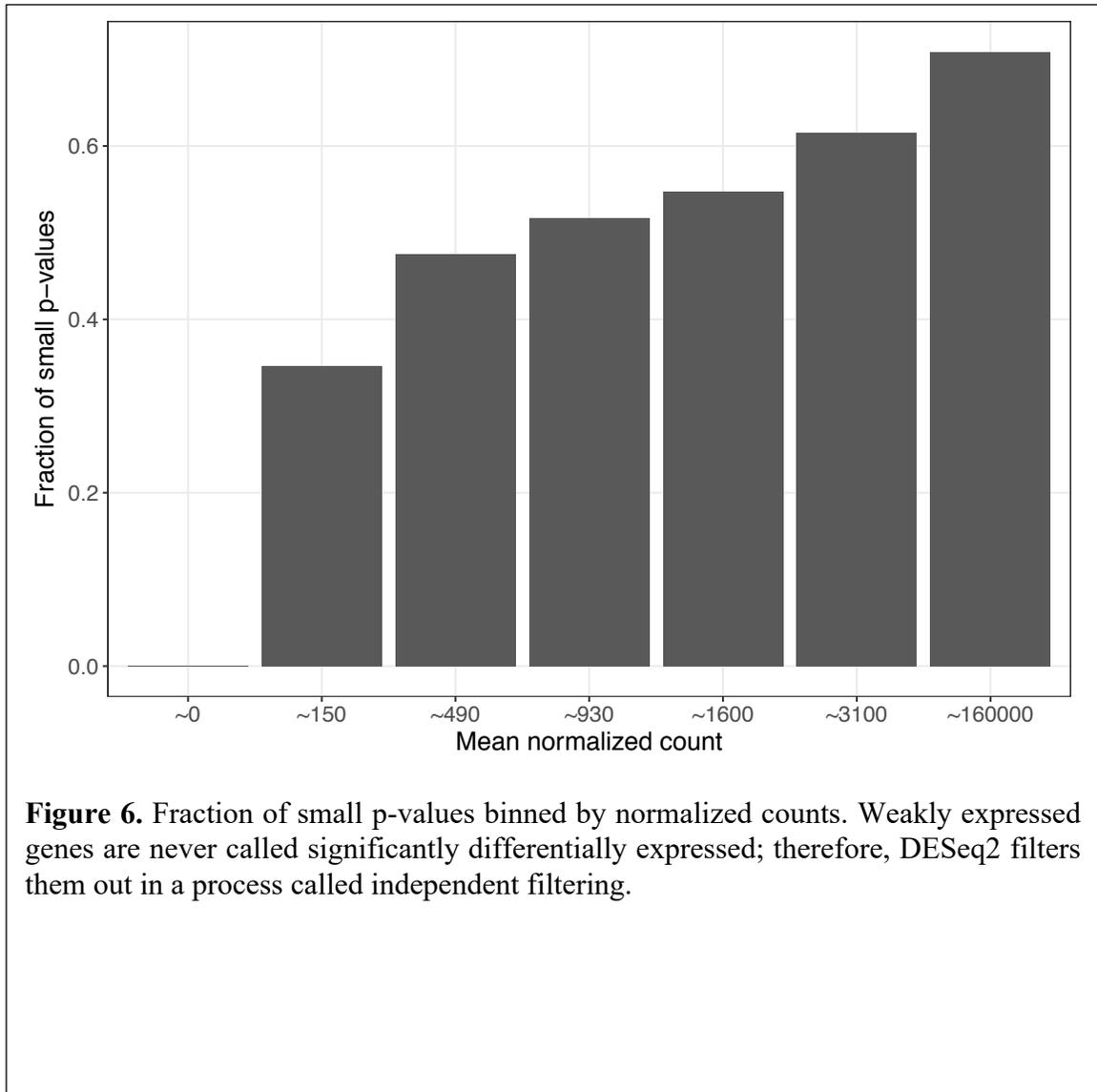


	# raw reads	# filtered reads	Difference	Overall alignment rate	Unique alignment rate
HLA	31,899,550	28,943,564	2,955,986	72.00%	70.24%
HLB	43,029,531	39,243,254	3,786,277	72.12%	70.45%
HLC	31,899,054	28,986,423	2,912,631	77.93%	76.32%
MLA	33,008,985	30,082,760	2,926,225	60.28%	58.74%
MLB	31,674,006	28,845,422	2,828,584	76.95%	75.32%
MLC	37,597,108	34,505,209	3,091,899	48.85%	47.51%
LLA	40,815,176	37,090,495	3,724,681	76.21%	74.29%
LLB	40,946,123	37,259,305	3,686,818	73.65%	70.15%
LLC	37,913,906	34,499,441	3,414,465	71.45%	69.79%

Table 1. Alignment statistics for nine original samples. High light replicates: HLA, HLB, HLC. Medium light replicates: MLA, MLB, MLC. Low light replicates: LLA, LLB, LLC. Alignment rates are inconsistent in medium light samples, suggesting quality issues. These samples were excluded from downstream analyses.



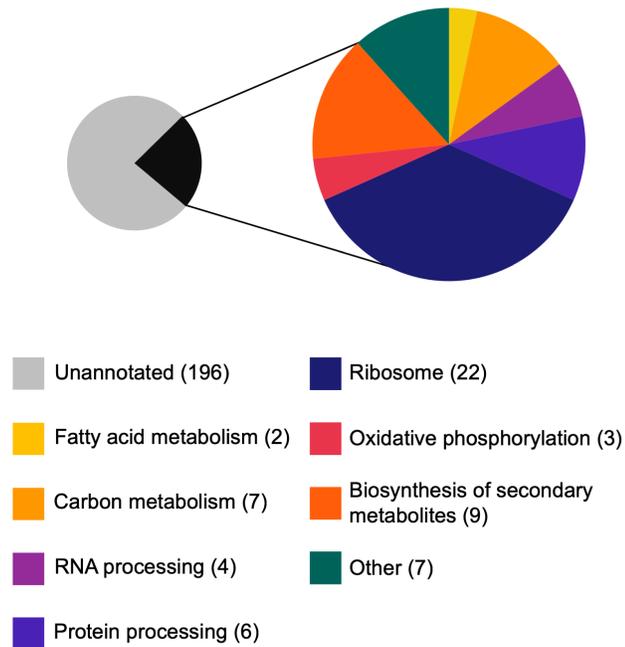




Fold change cutoff	Number of DE genes	% of genome DE	Number of upregulated genes	Number of downregulated genes
$ \text{FC} > 4$	598 (13)	5	343 (4)	255(23)
$ \text{FC} > 2$	2904 (15)	25	1410 (6)	1494 (24)

Table 2. Comparison of the number of genes that are called significantly differentially expressed using two fold-change cutoffs. Number in parentheses is the percentage of genes in each category that are annotated in the KEGG database. Upregulated genes are genes that have higher expression in slow-growing cells; downregulated genes have lower expression in slow-growing cells. Although lowering the fold-change cutoff increases the number of genes that are called significantly differentially expressed, the proportion of genes with annotations in the KEGG database stays relatively stable.

A.



B.

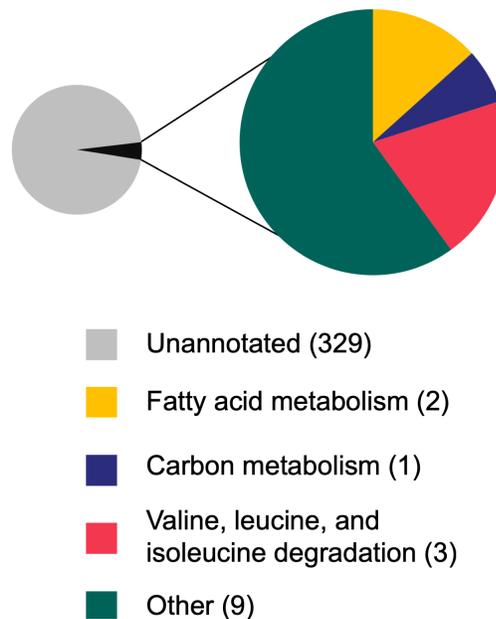
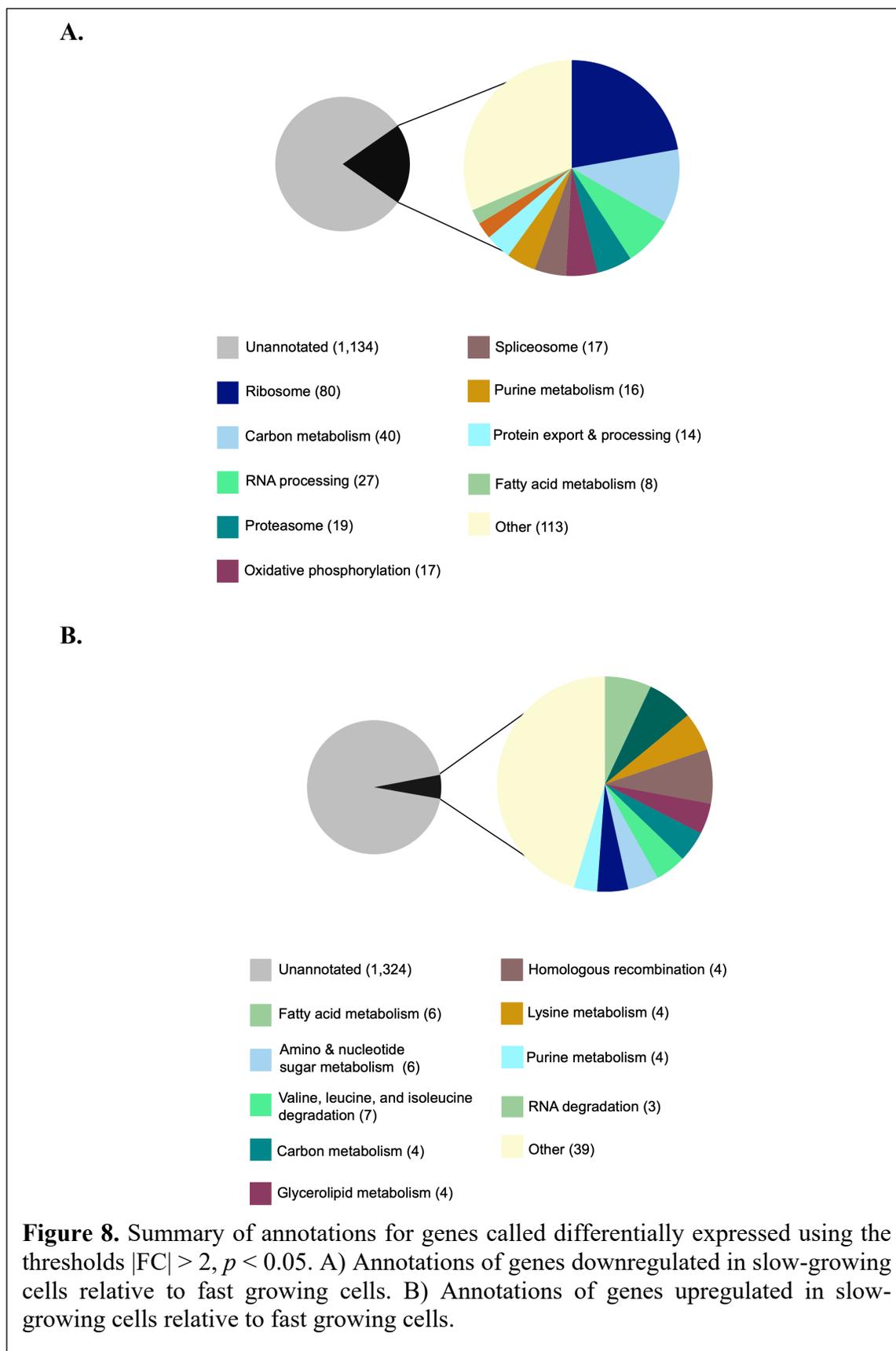
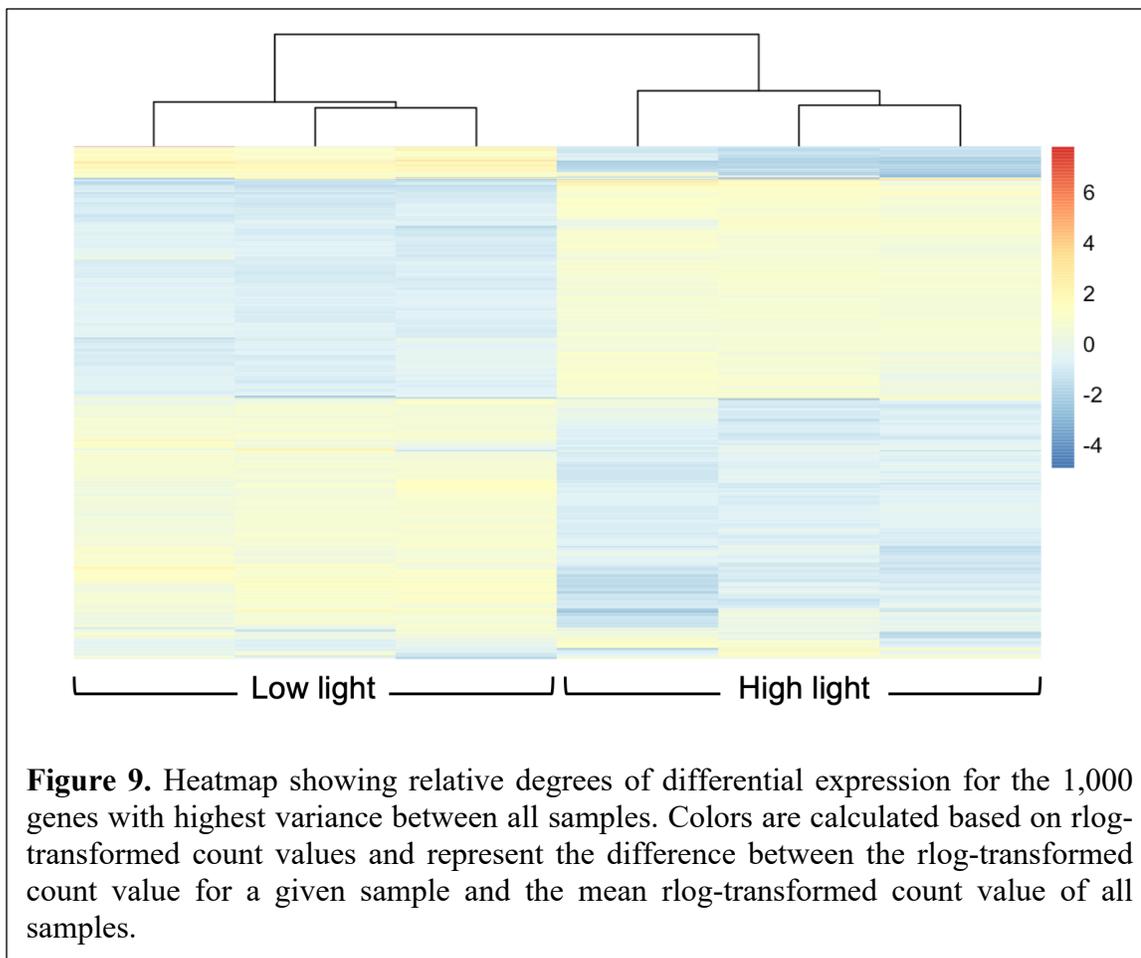
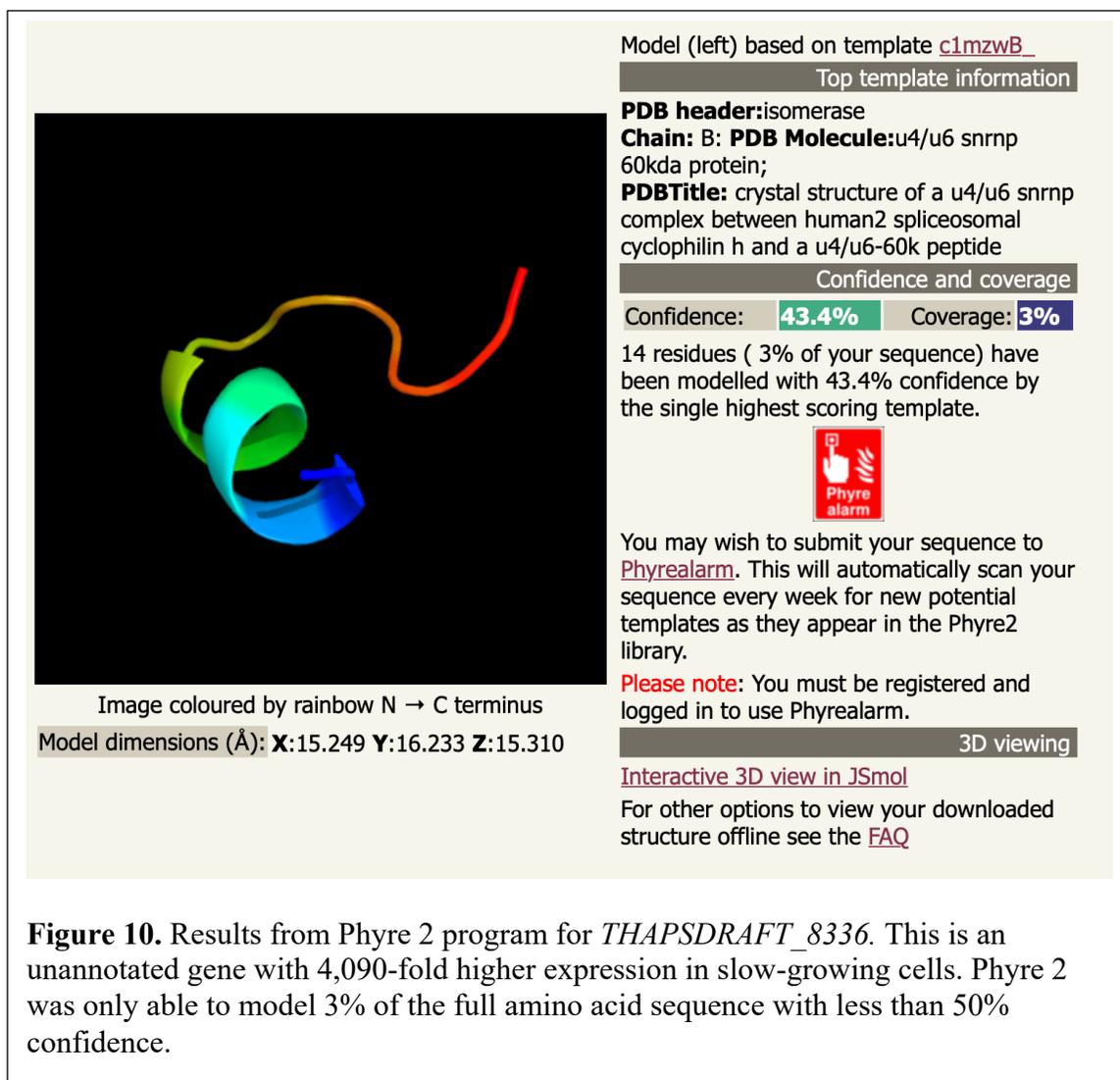


Figure 7. Summary of annotations for genes called differentially expressed using the thresholds $|FC| > 4$, $p < 0.05$. A) Annotations of genes downregulated in slow-growing cells relative to fast growing cells. B) Annotations of genes upregulated in slow-growing cells relative to fast growing cells.



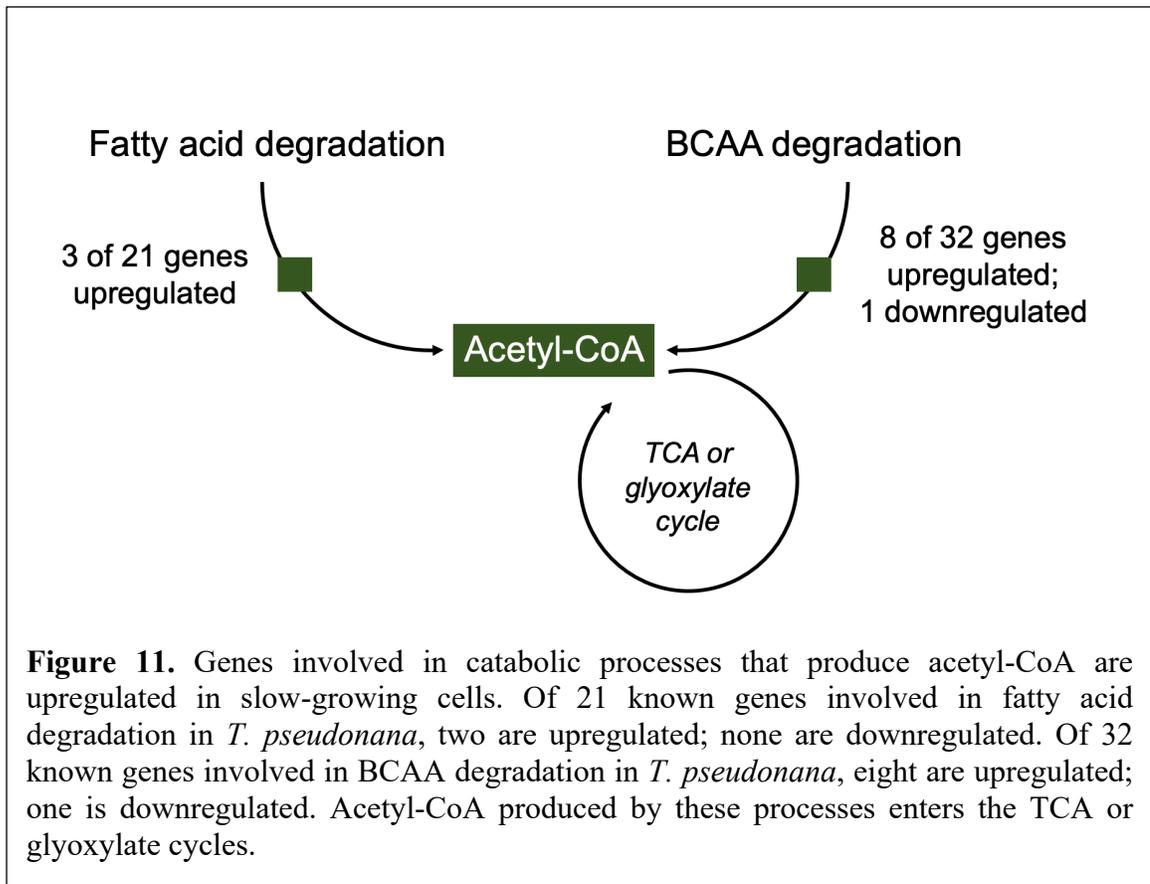


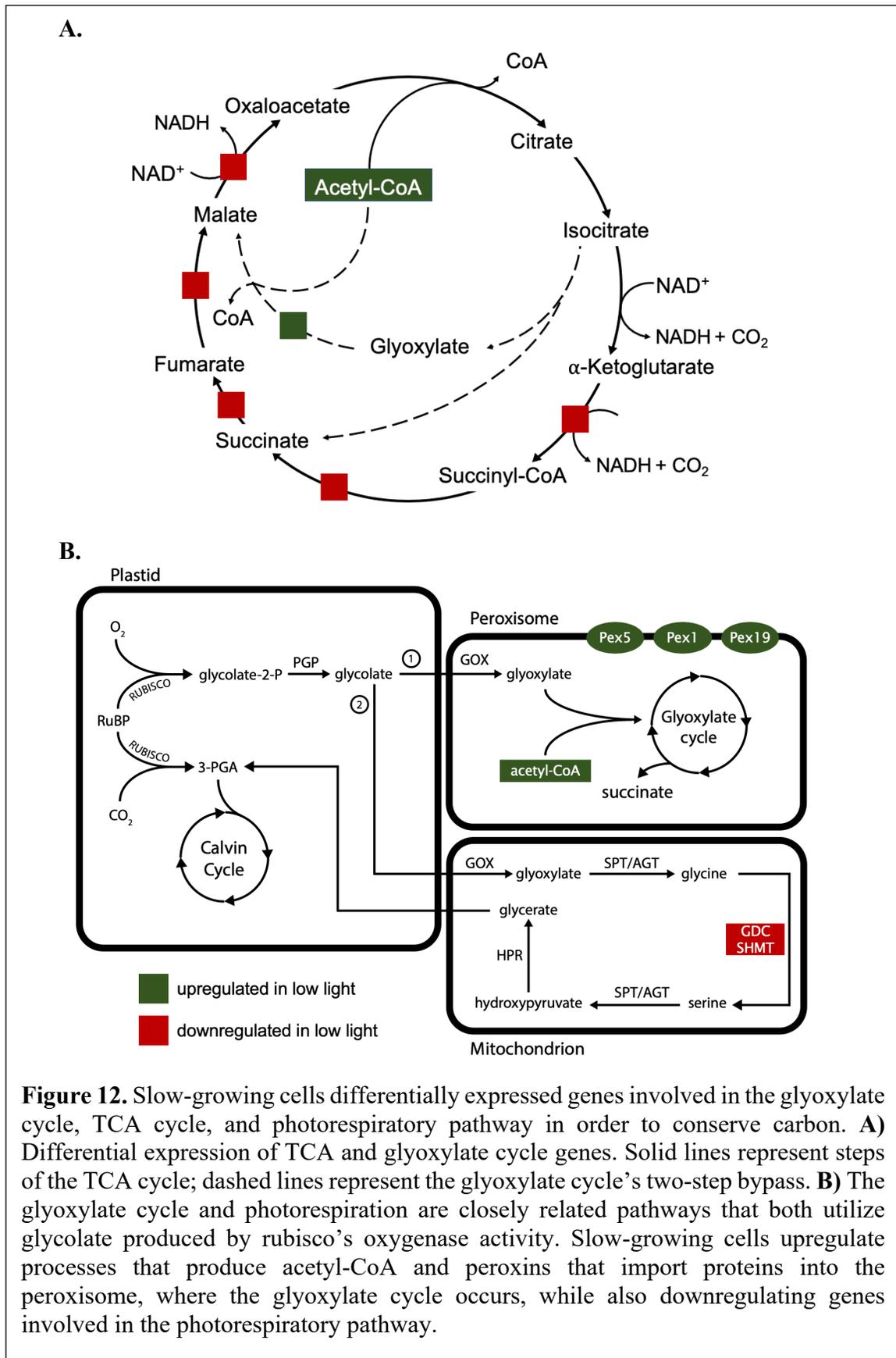


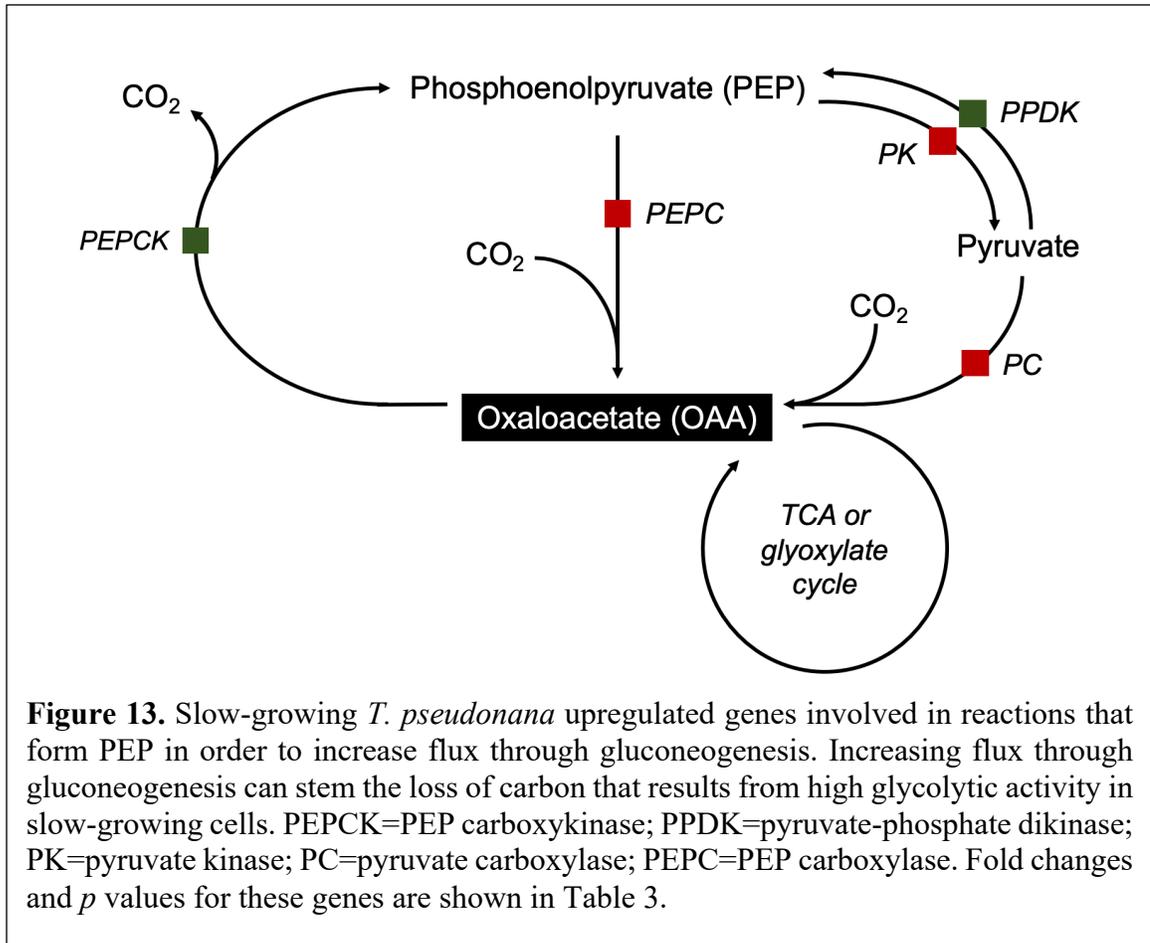
Gene name	Function	Pathway	Fold change	Adjusted p-value
THAPSDRAFT_795	2-oxoisovalerate E1 component alpha subunit	Valine, leucine, and isoleucine degradation	19.8	1.25E-29
THAPSDRAFT_36291	2-oxoisovalerate E2 component	Valine, leucine, and isoleucine degradation	10.3	4.88E-59
THAPSDRAFT_413	3-hydroxyisobutyrate dehydrogenase	Valine, leucine, and isoleucine degradation	6.42	8.52E-20
THAPSDRAFT_25495	Malonate-semialdehyde dehydrogenase (acetylating)	Valine, leucine, and isoleucine degradation	2.34	1.04E-04
THAPSDRAFT_32067	2-oxoisovalerate E1 component beta subunit	Valine, leucine, and isoleucine degradation	2.08	8.46E-04
HCD1	3-hydroxyacyl-CoA dehydrogenase	Valine, leucine, and isoleucine degradation	2.57	4.36E-06
PCB1	3-methylcrotonyl-CoA carboxylase beta subunit	Valine, leucine, and isoleucine degradation	2.53	9.73E-04
KCT2	Acetyl-CoA acyltransferase	Valine, leucine, and isoleucine degradation	-2.08	1.11E-05
ACD3	Acyl-CoA dehydrogenase	Fatty acid degradation; Valine, leucine, and isoleucine degradation	5.29	4.94E-15
GCD1	Glutaryl-CoA dehydrogenase	Fatty acid degradation	2.29	5.07E-05
FCL1	Long-chain fatty acid CoA-ligase	Fatty acid degradation	8.76	4.88E-20
THAPSDRAFT_262934	Probable malate synthase	Glyoxylate cycle	1.93	4.76E-02
MDH1	Malate dehydrogenase	TCA cycle	-3.36	3.01E-19
THAPSDRAFT_269718	Alpha-ketoglutarate dehydrogenase E1 component	TCA cycle	-1.46	1.73E-03
THAPSDRAFT_36971	Alpha-ketoglutarate dehydrogenase E2 component	TCA cycle	-1.78	9.27E-03
SDH1	Succinate dehydrogenase (ubiquinone) iron-sulfur subunit	TCA cycle	-1.29	2.40E-13
FUM1	Fumarate hydratase	TCA cycle	-2.45	1.80E-02

Gene name	Function	Pathway	Fold change	Adjusted p-value
SCS1	Succinyl-CoA synthetase beta subunit	TCA cycle	-3.26	9.36E-19
IDH1	Putative isocitrate dehydrogenase	TCA cycle	-2.43	1.14E-10
PPC1	Phosphoenolpyruvate carboxylase	TCA cycle	-1.45	9.21E-3
THAPSDRAFT_269908	Pyruvate carboxylase	TCA cycle	-1.86	2.05E-06
THAPSDRAFT_264438	Peroxin 5	Peroxisomal transport	1.79	2.04E-04
THAPSDRAFT_261078	Peroxin 1	Peroxisomal transport	1.87	1.35E-03
THAPSDRAFT_37854	Peroxin 19	Peroxisomal transport	1.60	2.19E-02
GDCT	Glycine decarboxylase t-protein (GDC)	Photorespiration	-1.49	1.06E-02
GDCP	Glycine decarboxylase p-protein (GDC)	Photorespiration	-2.18	3.09E-05
THAPSDRAFT_26031	Serine hydroxymethyltransferase (SHMT)	Photorespiration	-3.63	3.34E-18
PDK1_2	Pyruvate-phosphate dikinase (PPDK)	Gluconeogenesis	6.65	6.96E-20
PCK1	Phosphoenolpyruvate carboxykinase	Gluconeogenesis	7.26	4.82E-19
PYK1	Pyruvate kinase	Glycolysis	-6.47	6.63E-15
PYK2	Pyruvate kinase	Glycolysis	-2.27	8.83E-04

Table 3. Summary of differentially expressed genes involved in the glyoxylate cycle, TCA cycle, and related pathways.







Chapter 7: References

- Acuña J, López-Alvarez M, Nogueira E, González-Taboada F (2010) Diatom flotation at the onset of the spring phytoplankton bloom: an in situ experiment. *Marine Ecology Progress Series* 400:115–125. doi: 10.3354/meps08405
- Admiraal W, Peletier H (1979) Influence of organic compounds and light limitation on the growth rate of estuarine benthic diatoms. *British Phycological Journal* 14:197–206. doi: 10.1080/00071617900650211
- Alderkamp A-C, de Baar HJW, Visser RJW, Arrigo KR (2010) Can photoinhibition control phytoplankton abundance in deeply mixed water columns of the Southern Ocean? *Limnology and Oceanography* 55:1248–1264. doi: 10.4319/lo.2010.55.3.1248
- Alexander H, Jenkins BD, Rynearson TA, Dyhrman ST (2015) Metatranscriptome analyses indicate resource partitioning between diatoms in the field. *Proceedings of the National Academy of Sciences* 112:E2182–E2190. doi: 10.1073/pnas.1421993112
- Allen AE, Dupont CL, Oborník M, et al (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature* 473:203–207. doi: 10.1038/nature10074
- Alverson AJ, Beszteri B, Julius ML, Theriot EC (2011) The model marine diatom *Thalassiosira pseudonana* likely descended from a freshwater ancestor in the genus *Cyclotella*. *BMC Evolutionary Biology* 11:125. doi: 10.1186/1471-2148-11-125
- Amato A, Dell’Aquila G, Musacchia F, et al (2017) Marine diatoms change their gene expression profile when exposed to microscale turbulence under nutrient replete conditions. *Scientific Reports* 7:. doi: 10.1038/s41598-017-03741-6
- Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* 11:12
- Andrews S (2010) FastQC: A quality control tool for high throughput sequence data
- Arad S (Malis), Levy-Ontman O (2010) Red microalgal cell-wall polysaccharides: biotechnological aspects. *Current Opinion in Biotechnology* 21:358–364. doi: 10.1016/j.copbio.2010.02.008
- Armbrust EV (2009a) The life of diatoms in the world’s oceans. *Nature* 459:185–192. doi: 10.1038/nature08057
- Armbrust EV (2009b) The life of diatoms in the world’s oceans. *Nature* 459:185–192. doi: 10.1038/nature08057

- Armbrust EV, Berges JA, Bowler C, et al (2004) The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism. *Science* 306:79–86. doi: 10.1126/science.1091317
- Asada K (1999) The water-water cycle in chloroplasts: Scavenging of active oxygens and dissipation of excess photons. *Annual Review of Plant Physiology and Plant Molecular Biology* 50:601–639. doi: 10.1146/annurev.arplant.50.1.601
- Aumont O, Maier-Reimer E, Blain S, Monfray P (2003) An ecosystem model of the global ocean including Fe, Si, P colimitations. *Global Biogeochemical Cycles* 17:. doi: 10.1029/2001GB001745
- Bailleul B, Rogato A, de Martino A, et al (2010) An atypical member of the light-harvesting complex stress-related protein family modulates diatom responses to light. *Proceedings of the National Academy of Sciences* 107:18214–18219. doi: 10.1073/pnas.1007703107
- Barenholz U, Keren L, Segal E, Milo R (2016) A minimalistic resource allocation model to explain ubiquitous increase in protein expression with growth rate. *PLOS ONE* 11:e0153344. doi: 10.1371/journal.pone.0153344
- Basu S, Patil S, Mapleson D, et al (2017) Finding a partner in the ocean: molecular and evolutionary bases of the response to sexual cues in a planktonic diatom. *New Phytologist* 215:140–156. doi: 10.1111/nph.14557
- Beardall J (1989) Photosynthesis and photorespiration in marine phytoplankton. *Aquatic Botany* 34:105–130. doi: 10.1016/0304-3770(89)90052-1
- Behrenfeld MJ, Halsey KH, Milligan AJ (2008) Evolved physiological responses of phytoplankton to their integrated growth environment. *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2687–2703. doi: 10.1098/rstb.2008.0019
- Behrenfeld MJ, O'Malley RT, Siegel DA, et al (2006) Climate-driven trends in contemporary ocean productivity. *Nature* 444:752–755. doi: 10.1038/nature05317
- Behrenfeld MJ, Prasil O, Babin M, Bruyant F (2004) In search of a physiological basis for covariations in light-limited and light-saturated photosynthesis. *Journal of Phycology* 40:4–25. doi: 10.1046/j.1529-8817.2004.03083.x
- Bender SJ, Durkin CA, Berthiaume CT, et al (2014) Transcriptional responses of three model diatoms to nitrate limitation of growth. *Frontiers in Marine Science* 1:. doi: 10.3389/fmars.2014.00003
- Berges J, Charlebois D, Mauzerall D, Falkowski P (1996) Differential effects of nitrogen limitation on photosynthetic efficiency of photosystems I and II in microalgae. *Plant Physiology* 100:689–696

- Bhattacharya D, Archibald JM, Weber APM, Reyes-Prieto A (2017) How do endosymbionts become organelles? Understanding early events in plastid evolution. *BioEssays* 29:1239–1246. doi: 10.1002/bies.20671
- Borozan I, Watt SN, Ferretti V (2013) Evaluation of alignment algorithms for discovery and identification of pathogens using RNA-Seq. *PLoS ONE* 8:e76935. doi: 10.1371/journal.pone.0076935
- Bowler C, Allen AE, Badger JH, et al (2008) The *Phaeodactylum* genome reveals the evolutionary history of diatom genomes. *Nature* 456:239–244. doi: 10.1038/nature07410
- Broddrick JT, Du N, Smith SR, et al (2019) Cross-compartment metabolic coupling enables flexible photoprotective mechanisms in the diatom *Phaeodactylum tricornutum*. *New Phytologist* 222:1364–1379. doi: 10.1111/nph.15685
- Broman E, Sachpazidou V, Dopson M, Hylander S (2017) Diatoms dominate the eukaryotic metatranscriptome during spring in coastal ‘dead zone’ sediments. *Proceedings of the Royal Society B: Biological Sciences* 284:20171617. doi: 10.1098/rspb.2017.1617
- Busby MA, Gray JM, Costa AM, et al (2011) Expression divergence measured by transcriptome sequencing of four yeast species. *BMC Genomics* 12:. doi: 10.1186/1471-2164-12-635
- Cardozo KHM, Guaratini T, Barros MP, et al (2007) Metabolites from algae with economical impact. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology* 146:60–78. doi: 10.1016/j.cbpc.2006.05.007
- Caron DA, Alexander H, Allen AE, et al (2016) Probing the evolution, ecology and physiology of marine protists using transcriptomics. *Nature Reviews Microbiology* 15:6–20. doi: 10.1038/nrmicro.2016.160
- Carpenter EJ, Janson S (2000) Intracellular cyanobacterial symbionts in the marine diatom *Climacodium frauenfeldianum* (Bacillariophyceae). *Journal of Phycology* 36:540–544. doi: 10.1046/j.1529-8817.2000.99163.x
- Castrillo JI, Zeef LA, Hoyle DC, et al (2007) Growth control of the eukaryote cell: a systems biology study in yeast. *Journal of Biology* 6:4. doi: 10.1186/jbiol54
- Cheng R, Feng J, Zhang B-X, et al (2014) Transcriptome and gene expression analysis of an oleaginous diatom under different salinity conditions. *BioEnergy Research* 7:192–205. doi: 10.1007/s12155-013-9360-1
- Coesel S, Mangogna M, Ishikawa T, et al (2009) Diatom PtCPF1 is a new cryptochrome/photolyase family member with DNA repair and transcription regulation activity. *EMBO Reports* 10:655–661. doi: 10.1038/embor.2009.59

- Coesel S, Oborník M, Varela J, et al (2008) Evolutionary origins and functions of the carotenoid biosynthetic pathway in marine diatoms. *PLoS ONE*. doi: 10.1371/journal.pone.0002896
- Crombet Y, Leblanc K, Quéguiner B, et al (2011) Deep silicon maxima in the stratified oligotrophic Mediterranean Sea. *Biogeosciences* 8:459–475. doi: 10.5194/bg-8-459-2011
- Cross LL, Ebeed HT, Baker A (2016) Peroxisome biogenesis, protein targeting mechanisms and PEX gene functions in plants. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1863:850–862. doi: 10.1016/j.bbamcr.2015.09.027
- Cruz de Carvalho MH, Sun H-X, Bowler C, Chua N-H (2016) Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytologist* 210:497–510. doi: 10.1111/nph.13787
- Dalman MR, Deeter A, Nimishakavi G, Duan Z-H (2012) Fold change and p-value cutoffs significantly alter microarray interpretations. *BMC Bioinformatics* 13:. doi: 10.1186/1471-2105-13-S2-S11
- Davis A, Abbriano R, Smith SR, Hildebrand M (2017) Clarification of photorespiratory processes and the role of malic enzyme in diatoms. *Protist* 168:134–153. doi: 10.1016/j.protis.2016.10.005
- Depauw FA, Rogato A, Ribera d'Alcala M, Falciatore A (2012) Exploring the molecular basis of responses to light in marine diatoms. *Journal of Experimental Botany* 63:1575–1591. doi: 10.1093/jxb/ers005
- Diner RE, Schwenck SM, McCrow JP, et al (2016) Genetic manipulation of competition for nitrate between heterotrophic bacteria and diatoms. *Frontiers in Microbiology* 7:. doi: 10.3389/fmicb.2016.00880
- Domingues N, Matos AR, Marques da Silva J, Cartaxana P (2012) Response of the diatom *Phaeodactylum tricornutum* to photooxidative stress resulting from high light exposure. *PLoS ONE* 7:e38162. doi: 10.1371/journal.pone.0038162
- Dong H-P, Huang K-X, Wang H-L, et al (2014) Understanding strategy of nitrate and urea assimilation in a Chinese strain of *Aureococcus anophagefferens* through RNA-Seq analysis. *PLoS ONE* 9:e111069. doi: 10.1371/journal.pone.0111069
- Dressaire C, Redon E, Milhem H, et al (2008) Growth rate regulated genes and their wide involvement in the *Lactococcus lactis* stress responses. *BMC Genomics* 9:343. doi: 10.1186/1471-2164-9-343
- Durkin CA, Marchetti A, Bender SJ, et al (2012) Frustule-related gene transcription and the influence of diatom community composition on silica precipitation in an iron-

- limited environment. *Limnology and Oceanography* 57:1619–1633. doi: 10.4319/lo.2012.57.6.1619
- Dyhrman ST, Jenkins BD, Rynearson TA, et al (2012) The transcriptome and proteome of the diatom *Thalassiosira pseudonana* reveal a diverse phosphorus stress response. *PLoS ONE* 7:e33768. doi: 10.1371/journal.pone.0033768
- Engström PG, Steijger T, Sipos B, et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. *Nature Methods* 10:1185–1191. doi: 10.1038/nmeth.2722
- Falkowski PG, LaRoche J (1991) Acclimation to spectral irradiance in algae. *Journal of Phycology* 27:8–14. doi: 10.1111/j.0022-3646.1991.00008.x
- Falkowski PG, Wirick CD (1981) A simulation model of the effects of vertical mixing on primary productivity. *Marine Biology* 65:69–75
- Fernie AR, Stitt M (2012) On the discordance of metabolomics with proteomics and transcriptomics: coping with increasing complexity in logic, chemistry, and network interactions. *American Society of Plant Biologists* 159:1139–1145
- Fisher NL, Halsey KH (2016) Mechanisms that increase the growth efficiency of diatoms in low light. *Photosynthesis Research* 129:183–197. doi: 10.1007/s11120-016-0282-6
- Formighieri C (2015) Light Saturation of Photosynthesis. In: *Solar-to-fuel conversion in algae and cyanobacteria*. Springer International Publishing, Cham, pp 55–58
- Fortunato AE, Jaubert M, Enomoto G, et al (2016) Diatom phytochromes reveal the existence of far-red-light-based sensing in the ocean. *The Plant Cell* 28:616–628. doi: 10.1105/tpc.15.00928
- Geider RJ, Osbonie BA, Raven JA (1986) Growth, photosynthesis and maintenance metabolic cost in the diatom *Phaeodactylum tricorutum* at very low light levels. *Journal of Phycology* 22:39–48
- Giannoukos G, Ciulla DM, Huang K, et al (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology* 13:r23. doi: 10.1186/gb-2012-13-3-r23
- Gonzalez NH, Felsner G, Schramm FD, et al (2011) A single peroxisomal targeting signal mediates matrix protein import in diatoms. *PLoS ONE* 6:e25316. doi: 10.1371/journal.pone.0025316
- Gregg WW, Ginoux P, Schopf PS, Casey NW (2003) Phytoplankton and iron: validation of a global three-dimensional ocean biogeochemical model. *Deep Sea Research Part II: Topical Studies in Oceanography* 50:3143–3169. doi: 10.1016/j.dsr2.2003.07.013

- Guidi L, Chaffron S, Bittner L, et al (2016) Plankton networks driving carbon export in the oligotrophic ocean. *Nature* 532:465–470. doi: 10.1038/nature16942
- Halsey KH, Jones BM (2015) Phytoplankton strategies for photosynthetic energy allocation. *Annual Review of Marine Science* 7:265–297. doi: 10.1146/annurev-marine-010814-015813
- Halsey KH, Milligan AJ, Behrenfeld MJ (2011) Linking time-dependent carbon-fixation efficiencies in *Dunaliella tertiolecta* (Chlorophyceae) to underlying metabolic pathways. *Journal of Phycology* 47:66–76. doi: 10.1111/j.1529-8817.2010.00945.x
- Halsey KH, O'Malley RT, Graff JR, et al (2013) A common partitioning strategy for photosynthetic products in evolutionarily distinct phytoplankton species. *New Phytologist* 198:1030–1038. doi: 10.1111/nph.12209
- Harke MJ, Juhl AR, Haley ST, et al (2017) Conserved transcriptional responses to nutrient stress in bloom-forming algae. *Frontiers in Microbiology* 8:. doi: 10.3389/fmicb.2017.01279
- Harrison PJ, Thompson PA, Calderwood GS (1990) Effects of nutrient and light limitation on the biochemical composition of phytoplankton. *Journal of Applied Phycology* 2:45–56. doi: 10.1007/BF02179768
- Hildebrand M, Abbriano RM, Polle JE, et al (2013) Metabolic and cellular organization in evolutionarily diverse microalgae as related to biofuels production. *Current Opinion in Chemical Biology* 17:506–514. doi: 10.1016/j.cbpa.2013.02.027
- Hildebrand M, Dahlin K, Volcani BE (1998) Characterization of a silicon transporter gene family in *Cylindrotheca fusiformis* : sequences, expression analysis, and identification of homologs in other diatoms. *Molecular and General Genetics* MGG 260:480–486. doi: 10.1007/s004380050920
- Hildebrand M, Davis AK, Smith SR, et al (2012) The place of diatoms in the biofuels industry. *Biofuels* 3:221–240. doi: 10.4155/BFS.11.157
- Hoegh-Guldberg O, Bruno JF (2010) The impact of climate change on the world's marine ecosystems. *Science* 328:1523–1528. doi: 10.1126/science.1189930
- Hu Q (2004) Environmental Effects on Cell Composition. In: *Handbook of Microalgal Culture: Biotechnology and Applied Phycology*. pp 83–93
- Huisman J, Sharples J, Stroom JM, et al (2004) Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* 85:2960–2970
- Huysman MJ, Martens C, Vandepoele K, et al (2010) Genome-wide analysis of the diatom cell cycle unveils a novel type of cyclins involved in environmental signaling. *Genome Biology* 11:R17. doi: 10.1186/gb-2010-11-2-r17

- Jiang Y, Yoshida T, Quigg A (2012) Photosynthetic performance, lipid production and biomass composition in response to nitrogen limitation in marine microalgae. *Plant Physiology and Biochemistry* 54:70–77. doi: 10.1016/j.plaphy.2012.02.012
- Jones JM, Morrell JC, Gould SJ (2004) PEX19 is a predominantly cytosolic chaperone and import receptor for class 1 peroxisomal membrane proteins. *The Journal of Cell Biology* 164:57–67. doi: 10.1083/jcb.200304111
- Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle> 2011
- Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics* 9:605–618. doi: 10.1038/nrg2386
- Kelley LA, Mezulis S, Yates CM, et al (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* 10:845–858. doi: <https://doi.org/10.1038/nprot.2015.053>
- Kim D, Langmead B, Salzberg SL (2015) HISAT: A fast spliced aligner with low memory requirements. *Nature Methods* 12:357–360. doi: 10.1038/nmeth.3317
- Kim J, Fabris M, Baart G, et al (2016) Flux balance analysis of primary metabolism in the diatom *Phaeodactylum tricornutum*. *The Plant Journal* 85:161–176. doi: 10.1111/tpj.13081
- Klumpp S, Zhang Z, Hwa T (2009) Growth rate-dependent global effects on gene expression in bacteria. *Cell* 139:1366–1375. doi: 10.1016/j.cell.2009.12.001
- Knutson TR, McBride JL, Chan J, et al (2010) Tropical cyclones and climate change. *Nature Geoscience* 3:157–163. doi: 10.1038/ngeo779
- Kroth PG, Chiovitti A, Gruber A, et al (2008) A model for carbohydrate metabolism in the diatom *Phaeodactylum tricornutum* deduced from comparative whole genome analysis. *PLoS ONE* 3:e1426. doi: 10.1371/journal.pone.0001426
- Kunze M, Pracharoenwattana I, Smith SM, Hartig A (2006) A central role for the peroxisomal membrane in glyoxylate cycle function. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1763:1441–1452. doi: 10.1016/j.bbamcr.2006.09.009
- Kustka AB, Milligan AJ, Zheng H, et al (2014) Low CO₂ results in a rearrangement of carbon metabolism to support C₄ photosynthetic carbon assimilation in *Thalassiosira pseudonana*. *New Phytologist* 204:507–520. doi: 10.1111/nph.12926

- Levering J, Broddrick J, Dupont CL, et al (2016) Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLOS ONE* 11:e0155038. doi: 10.1371/journal.pone.0155038
- Levering J, Dupont CL, Allen AE, et al (2017) Integrated regulatory and metabolic networks of the marine diatom *Phaeodactylum tricornutum* predict the response to rising CO₂ levels. *mSystems* 2:. doi: 10.1128/mSystems.00142-16
- Levitan O, Dinamarca J, Zelzion E, et al (2015) Remodeling of intermediate metabolism in the diatom *Phaeodactylum tricornutum* under nitrogen stress. *Proceedings of the National Academy of Sciences* 112:412–417. doi: 10.1073/pnas.1419818112
- Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. doi: 10.1093/bioinformatics/btp352
- Litchman E (2000) Growth rates of phytoplankton under fluctuating light. *Freshwater Biology* 44:223–235. doi: 10.1046/j.1365-2427.2000.00559.x
- Lommer M, Specht M, Roy A-S, et al (2012) Genome and low-iron response of an oceanic diatom adapted to chronic iron limitation. *Genome Biology* 13:R66. doi: 10.1186/gb-2012-13-7-r66
- Long SP, Humphries S, Falkowski PG (1994) Photoinhibition of photosynthesis in nature. *Annual Review of Plant Physiology and Plant Molecular Biology* 45:633–662
- Love MI, Anders S, Huber W (2014) Differential analysis of count data - the DESeq2 package. *Genome Biology* 15:550. doi: 110.1186/s13059-014-0550-8
- Luo W, Pant G, Bhavnasi YK, et al (2017) Pathview Web: User friendly pathway visualization and data integration. *Nucleic Acids Research* 45:W501–W508. doi: <https://doi.org/10.1093/nar/gkx372>
- MacIntyre HL, Cullen JJ (2005) Using Cultures to Investigate the Physiological Ecology of Microalgae. In: *Algal Culturing Techniques*. Academic Press, pp 287–326
- MacManes MD (2014) On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics* 5:. doi: 10.3389/fgene.2014.00013
- Maheswari U, Jabbari K, Petit J-L, et al (2010) Digital expression profiling of novel diatom transcripts provides insight into their biological functions. *Genome Biology* 11:R85. doi: 10.1186/gb-2010-11-8-r85
- Malone TC, Falkowski PG, Hopkins TS, et al (1983) Mesoscale response of diatom populations to a wind event in the plume of the Hudson River. *Deep Sea Research Part A, Oceanographic Research Papers* 30:149–170. doi: 10.1016/0198-0149(83)90066-3

- MAQC Consortium (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology* 24:1151–1161. doi: 10.1038/nbt1239
- Marchant HK, Ahmerkamp S, Lavik G, et al (2017) Denitrifying community in coastal sediments performs aerobic and anaerobic respiration simultaneously. *The ISME Journal* 11:1799–1812. doi: 10.1038/ismej.2017.51
- Marchetti A, Schrueth DM, Durkin CA, et al (2012) Comparative metatranscriptomics identifies molecular bases for the physiological responses of phytoplankton to varying iron availability. *Proceedings of the National Academy of Sciences* 109:E317–E325. doi: 10.1073/pnas.1118408109
- Margalef R (1978) Phytoplankton communities in upwelling areas. *Oecologia Aquatica* 3:97–132
- Marinov I, Doney SC, Lima ID (2010) Response of ocean phytoplankton community structure to climate change over the 21st century: partitioning the effects of nutrients, temperature and light. *Biogeosciences* 7:3941–3959. doi: 10.5194/bg-7-3941-2010
- Marra J (1980) Time course of light intensity adaptation in a marine diatom. *Marine Biology Letters* 1:175–183
- Martens C, Vandepoele K, Van de Peer YV de P (2008) Whole-genome analysis reveals molecular innovations and evolutionary transitions in chromalveolate species. *Proceedings of the National Academy of Sciences* 105:3427–3432. doi: 10.1073/pnas.0712248105
- McCarthy DJ, Smyth GK (2009) Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics* 25:765–771. doi: 10.1093/bioinformatics/btp053
- Mitrovic SM, Howden CG, Bowling LC, Buckney RT (2003) Unusual allometry between in situ growth of freshwater phytoplankton under static and fluctuating light environments: possible implications for dominance. *Journal of Plankton Research* 25:517–526. doi: 10.1093/plankt/25.5.517
- Mock T, Otiillar RP, Strauss J, et al (2017) Evolutionary genomics of the cold-adapted diatom *Fragilariopsis cylindrus*. *Nature* 541:536–540. doi: 10.1038/nature20803
- Mock T, Samanta MP, Iverson V, et al (2008) Whole-genome expression profiling of the marine diatom *Thalassiosira pseudonana* identifies genes involved in silicon bioprocesses. *Proceedings of the National Academy of Sciences* 105:1579–1584. doi: 10.1073/pnas.0707946105
- Montsant A, Jabbari K, Maheswari U, Bowler C (2005) Comparative Genomics of the Pennate Diatom *Phaeodactylum tricornutum*. *Plant Physiology* 137:500–513. doi: 10.1104/pp.104.052829

- Moore JK, Doney SC, Lindsay K (2004) Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles* 18:. doi: 10.1029/2004GB002220
- Morel A, Lazzara L, Gostan J (1987) Growth rate and quantum yield time response for a diatom to changing irradiances (energy and color). *Limnology and Oceanography* 32:1066–1084. doi: 10.4319/lo.1987.32.5.1066
- Morris I (1980) Paths of Carbon Assimilation in Marine Phytoplankton. In: Falkowski PG (ed) *Primary Productivity in the Sea*. Springer US, Boston, MA, pp 139–159
- Nan F, Feng J, Lv J, et al (2018) Transcriptome analysis of the typical freshwater rhodophytes *Sheathia arcuata* grown under different light intensities. *PLOS ONE* 13:e0197729. doi: 10.1371/journal.pone.0197729
- Nearing M a, Pruski FF, O’Neal MR (2004) Expected climate change impacts on soil erosion rates: A review. *Journal of Soil and Water Conservation* 59:43–50
- Nicolaus M, Katlein C, Maslanik J, Hendricks S (2012) Changes in Arctic sea ice result in increasing light transmittance and absorption. *Geophysical Research Letters* 39:. doi: 10.1029/2012GL053738
- Nito K, Kamigaki A, Kondo M, et al (2007) Functional classification of Arabidopsis peroxisome biogenesis factors proposed from analyses of knockdown mutants. *Plant and Cell Physiology* 48:763–774. doi: 10.1093/pcp/pcm053
- Nunn BL, Faux JF, Hippmann AA, et al (2013) Diatom Proteomics Reveals Unique Acclimation Strategies to Mitigate Fe Limitation. *PLoS ONE* 8:e75653. doi: 10.1371/journal.pone.0075653
- Nymark M, Valle KC, Brembu T, et al (2009) An integrated analysis of molecular acclimation to high light in the marine diatom *Phaeodactylum tricornutum*. *PLoS ONE* 4:e7743. doi: 10.1371/journal.pone.0007743
- Oliver DJ (1998) Photorespiration and the C2 cycle. In: Raghavendra AS (ed) *Photosynthesis: A comprehensive treatise*. Cambridge University Press, Cambridge, pp 173–182
- Oliver MJ, Petrov D, Ackerly D, et al (2007) The mode and tempo of genome size evolution in eukaryotes. *Genome Research* 17:594–601. doi: 10.1101/gr.6096207
- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nature Biotechnology* 28:245–248. doi: 10.1038/nbt.1614
- Peixoto L, Risso D, Poplawski SG, et al (2015) How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets. *Nucleic Acids Research* 43:7664–7674. doi: 10.1093/nar/gkv736

- Perteau M, Kim D, Perteau GM, et al (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols* 11:1650–1667. doi: 10.1038/nprot.2016.095
- Perteau M, Perteau GM, Antonescu CM, et al (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology* 33:290–295. doi: 10.1038/nbt.3122
- Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiology* 151:1570–1581. doi: 10.1104/pp.109.141267
- Poulsen N, Sumper M, Kröger N (2003) Biosilica formation in diatoms: Characterization of native silaffin-2 and its role in silica morphogenesis. *Proceedings of the National Academy of Sciences* 100:12075–12080. doi: 10.1073/pnas.2035131100
- Quigg A, Beardall J (2003) Protein turnover in relation to maintenance metabolism at low photon flux in two marine microalgae. *Plant, Cell, and Environment* 26:693–703
- Quinlan AR, Hall IM (2010) The BEDTools manual. *Genome* 16:1–77. doi: 10.1093/bioinformatics/btq033
- Raven JA, Waite AM (2004) The evolution of silicification in diatoms: Inescapable sinking and sinking as escape? *New Phytologist* 162:45–61. doi: 10.1111/j.1469-8137.2004.01022.x
- Rayko E, Maumus F, Maheswari U, et al (2010) Transcription factor families inferred from genome sequences of photosynthetic stramenopiles. *New Phytologist* 188:52–66. doi: 10.1111/j.1469-8137.2010.03371.x
- Regenberg B, Grotkj T, Åkesson M, et al (2006) Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in *Saccharomyces cerevisiae*. *Genome Biology* 13
- Roberts K, Granum E, Leegood RC, Raven JA (2007) C3 and C4 Pathways of photosynthetic carbon assimilation in marine diatoms Are under genetic, not environmental, control. *Plant Physiology* 145:230–235. doi: 10.1104/pp.107.102616
- Roth MS, Cokus SJ, Gallaher SD, et al (2017) Chromosome-level genome assembly and transcriptome of the green alga *Chromochloris zofingiensis* illuminates astaxanthin production. *Proceedings of the National Academy of Sciences* 114:E4296–E4305. doi: 10.1073/pnas.1619928114
- Sarmiento JL, Slater R, Barber R, et al (2004) Response of ocean ecosystems to climate warming. *Global Biogeochemical Cycles* 18:. doi: 10.1029/2003GB002134

- Sarthou G, Timmermans KR, Blain S, Tréguer P (2005) Growth physiology and fate of diatoms in the ocean: A review. *Journal of Sea Research* 53:25–42. doi: 10.1016/j.seares.2004.01.007
- Scharek R, Tupas LM, Karl DM (1999) Diatom fluxes to the deep sea in the oligotrophic North Pacific gyre at Station ALOHA. *Inter-Research* 182:55–67
- Schellenberger Costa B, Jungandreas A, Jakob T, et al (2013) Blue light is essential for high light acclimation and photoprotection in the diatom *Phaeodactylum tricornutum*. *Journal of Experimental Botany* 6:483–493. doi: 10.1093/jxb/ers340
- Schnitzler Parker M, Armbrust EV, Piovia-Scott J, Keil RG (2004) Induction of photorespiration by light in the centric diatom *Thalassiosira weissflogii* (Bacillariophyceae): Molecule characterization and physiological consequences. *Journal of Phycology* 40:557–567. doi: 10.1111/j.1529-8817.2004.03184.x
- SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nature Biotechnology* 32:903–914. doi: 10.1038/nbt.2957
- Shin H, Hong S-J, Yoo C, et al (2016) Genome-wide transcriptome analysis revealed organelle specific responses to temperature variations in algae. *Scientific Reports* 6:. doi: 10.1038/srep37770
- Shrestha R, Tesson B, Norden-Krichmar T, et al (2012) Whole transcriptome analysis of the silicon response of the diatom *Thalassiosira pseudonana*. *BMC Genomics* 13:499. doi: 10.1186/1471-2164-13-499
- Smith SR, Abbriano RM, Hildebrand M (2012) Comparative analysis of diatom genomes reveals substantial differences in the organization of carbon partitioning pathways. *Algal Research* 1:2–16. doi: 10.1016/j.algal.2012.04.003
- Smith SR, Gillard JTF, Kustka AB, et al (2016a) Transcriptional orchestration of the global cellular response of a model pennate diatom to diel light cycling under iron limitation. *PLOS Genetics* 12:e1006490. doi: 10.1371/journal.pgen.1006490
- Smith SR, Glé C, Abbriano RM, et al (2016b) Transcript level coordination of carbon pathways during silicon starvation-induced lipid accumulation in the diatom *Thalassiosira pseudonana*. *New Phytologist* 210:890–904. doi: 10.1111/nph.13843
- Tanaka T, Maeda Y, Veluchamy A, et al (2015) Oil accumulation by the oleaginous diatom *Fistulifera solaris* as revealed by the genome and transcriptome. *The Plant Cell Online* 27:162–176. doi: 10.1105/tpc.114.135194
- Tenenbaum D (2018) KEGGREST: Client-side REST access to KEGG

- Thamatrakoln K, Korenovska O, Niheu AK, Bidle KD (2012) Whole-genome expression analysis reveals a role for death-related genes in stress acclimation of the diatom *Thalassiosira pseudonana*: Death genes and stress acclimation in a marine diatom. *Environmental Microbiology* 14:67–81. doi: 10.1111/j.1462-2920.2011.02468.x
- Thomas DN, Dieckmann GS (2002) Antarctic Sea Ice--a Habitat for Extremophiles. *Science* 295:641–644. doi: 10.1126/science.1063391
- Tirichine L, Rastogi A, Bowler C (2017) Recent progress in diatom genomics and epigenomics. *Current Opinion in Plant Biology* 36:46–55. doi: 10.1016/j.pbi.2017.02.001
- Todd EV, Black MA, Gemmell NJ (2016) The power and promise of RNA-seq in ecology and evolution. *Molecular Ecology* 25:1224–1241. doi: 10.1111/mec.13526
- Traller JC, Cokus SJ, Lopez DA, et al (2016) Genome and methylome of the oleaginous diatom *Cyclotella cryptica* reveal genetic flexibility toward a high lipid phenotype. *Biotechnology for Biofuels* 9:. doi: 10.1186/s13068-016-0670-3
- Tréguer PJ, De La Rocha CL (2013) The world ocean silica cycle. *Annual Review of Marine Science* 5:477–501. doi: 10.1146/annurev-marine-121211-172346
- Uitz J, Claustre H, Gentili B, Stramski D (2010) Phytoplankton class-specific primary production in the world's oceans: Seasonal and interannual variability from satellite observations. *Global Biogeochemical Cycles* 24:. doi: 10.1029/2009GB003680
- Vardi A, Thamatrakoln K, Bidle KD, Falkowski PG (2009) Diatom genomes come of age. *Genome Biology* 9:245. doi: 10.1186/gb-2008-9-12-245
- Vonshak A, Torzillo G (2004) Environmental Stress Physiology. In: *Handbook of Microalgal Culture: Biotechnology and Applied Phycology*. Blackwell Science, pp 57–83
- Vorobev A, Sharma S, Yu M, et al (2018) Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals: Labile DOM in a coastal ocean. *Environmental Microbiology* 20:3012–3030. doi: 10.1111/1462-2920.14344
- Wagner H, Jakob T, Wilhelm C (2005) Balancing the energy flow from captured light to biomass under fluctuating light conditions. *New Phytologist* 169:95–108. doi: 10.1111/j.1469-8137.2005.01550.x
- Wang L, Feng Z, Wang X, et al (2010) DESeq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26:136–138. doi: 10.1093/bioinformatics/btp612

- Wilhelm C, Büchel C, Fisahn J, et al (2006) The regulation of carbon and nutrient assimilation in diatoms is significantly different from green algae. *Protist* 157:91–124. doi: 10.1016/j.protis.2006.02.003
- Williams CP, Stanley WA (2010) Peroxin 5: A cycling receptor for protein translocation into peroxisomes. *The International Journal of Biochemistry & Cell Biology* 42:1771–1774. doi: 10.1016/j.biocel.2010.07.004
- Williams CR, Baccarella A, Parrish JZ, Kim CC (2016) Trimming of sequence reads alters RNA-Seq gene expression estimates. *BMC Bioinformatics* 17:. doi: 10.1186/s12859-016-0956-2
- Yang Z-K, Niu Y-F, Ma Y-H, et al (2013) Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnology for Biofuels* 6:67. doi: 10.1186/1754-6834-6-67
- Yen H-W, Hu I-C, Chen C-Y, et al (2013) Microalgae-based biorefinery – From biofuels to natural products. *Bioresource Technology* 135:166–174. doi: 10.1016/j.biortech.2012.10.099
- Zehr JP, Carpenter EJ, Villareal TA (2000) New perspectives on nitrogen-fixing microorganisms in tropical and subtropical oceans. *Trends in Microbiology* 8:68–73. doi: 10.1016/S0966-842X(99)01670-4
- Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:. doi: 10.1186/1471-2105-9-40