TRIACYLGLYCEROL BIOSYNTHESIS IN *CHLAMYDOMONAS REINHARDTII*

By

Rachel Elizabeth Miller

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Cell and Molecular Biology – Doctor of Philosophy

2015

ABSTRACT

TRIACYLGLYCEROL BIOSYNTHESIS IN *CHLAMYDOMONAS REINHARDTII*

By

Rachel Elizabeth Miller

One of the main problems facing America today is the reliance on fossil fuels for power and transportation. The combination of CO_2 emission and declining petroleum reserves means that an alternative needs to be found. Biofuels are one of the main alternatives being considered, as they can act as a carbon sink during plant growth, minimizing the net CO_2 released when consumed, and they can be used for transportation fuel. Biodiesel has many good fuel properties, such as high energy density and ability to be used with existing technologies, but obtaining a high enough yield to meet current transportation needs is still an issue. One approach to address this is to use algae for biodiesel production, as they are capable of yielding much higher oil (in the form of triacylglycerol, or TAG) per land used, compared to traditional land-based oil crops. However, research is still needed to identify effective production strains, and modify them to improve overall yield.

Chlamydomonas reinhardtii was used as a model microalga to study TAG synthesis, as it is already well studied and has various molecular tools available. Additionally, TAG synthesis can be induced by nitrogen deprivation, making it easy to study in laboratory conditions. Two approaches were used to study TAG synthesis, the first focusing on the enzymes directly responsible for synthesis and the second looking at the larger metabolic changes that contribute to TAG synthesis, along with potential regulators of the process.

The first project discovered five putative type 2 diacylglycerol acyltransferases (DGATs). Testing of these DGATs, named *DGTT1-5*, in yeast confirmed *DGTT2*, *DGTT3* and *DGTT4*

have DGAT activity, and *DGTT5* is likely a pseudogene. Testing in yeast also revealed a potential difference in fatty acyl preference between DGTT2, DGTT3 and DGTT4. Expression of *DGTT2* in Arabidopsis further confirmed its DGAT activity, demonstrated its ability to synthesize TAG in vegetative tissues, and also showed its ability to incorporate long-chain fatty acyls into TAG. Testing of *DGTT2* and *DGTT3* in *Chlamydomonas* by over-expression and amiRNA was attempted, but repeatable changes in TAG phenotype were not detected.

In the second project, a transcript profiling experiment was performed using high-throughput sequencing, comparing populations grown under nitrogen replete and nitrogen deprived conditions. Many metabolic pathways showed consistent changes in transcript level, indicative of the changes occurring within the cell during nitrogen deprivation, which may contribute to TAG accumulation. In addition, various genes of interest for further study were discovered, including several candidate transcription factors. These candidate genes were cloned and tested via over-expression in *Chlamydomonas*. However, after testing, these candidates were found not to have a detectable role in TAG synthesis. The transcript profile dataset continues to be a valuable resource for future research in *Chlamydomonas*.

This dissertation is dedicated to my parents, Martyn and Linda Miller, for their support

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

DAG:	Diacylglycerol
DGAT:	Diacylglycerol acyltransferase
DGTA:	DAG:DAG transacylase
DW:	Dry weight
ER:	Endoplasmic reticulum
ESI-MS:	Electrospray ionization mass spectrometry
FAME:	Fatty acid methyl ester
GC-FID:	Gas chromatography-flame ionization detector
GPAT:	Glycerol phosphate acyltransferase
LPAAT:	Lysophosphatidic acid acyltransferase
PA:	Phosphatidic acid
PAP:	Phosphatidic acid phosphatase
PDAT:	Phospholipid:diacylglycerol acyltransferase
TAG:	Triacylglycerol
TEM:	Transmission electron microscopy
TF:	Transcription factor
TLC:	Thin layer chromatography
VLCFA:	Very long chain fatty acids

CHAPTER 1: INTRODUCTION

America runs on fuel. Personal transportation, movement of material goods, and electricity generation all depend largely on the availability of fuel to power engines and power stations. For over a century, the fuel of choice has been either coal or oil, hydrocarbon-based fossil fuels deposited underground up to 650 million years ago (Mann, P. et al. 2003). Fossil fuels are an effective fuel source, being energy-dense and portable. In the United States, almost a third of the total energy consumed is used for transportation, with the vast majority coming from fossil fuel sources (Figure 1-1).

In recent decades, concerns have been raised over the long-term use of fossil fuels. While many have feared petroleum shortages in the near future, due to the non-renewable nature of fossil fuels (Murray, J. and King, D. 2012), recent discoveries of oil deposits and development of non-traditional petroleum sources indicates that global oil demands will be met for at least the next 25 years (U.S.Energy Information Administration 2011). Another, more recent concern is the effect of the carbon released by burning fossil fuels, in the form of CO_2, and the effect this might have on global climate. With CO_2 concentrations reaching record levels, governments are beginning to take drastic steps to reduce greenhouse gas emissions. Low-carbon fuel standards, such as the Global Warming Solutions Act of 2006 enacted by California, are designed to reduce the amount of CO_2 emitted by encouraging the use of alternative fuels, including natural gas and biofuels, as well as new technologies, such as electric and hybrid vehicles (California Air Resources Board 2013). Both the United Kingdom and the European Union have enacted similar standards. On the national scale, the US has enacted regulation, such as the Energy Independence and Security Act of 2007, to reduce gasoline usage, decrease greenhouse gas emissions, and encourage the development of alternative fuels. Despite these concerns and regulations, demand

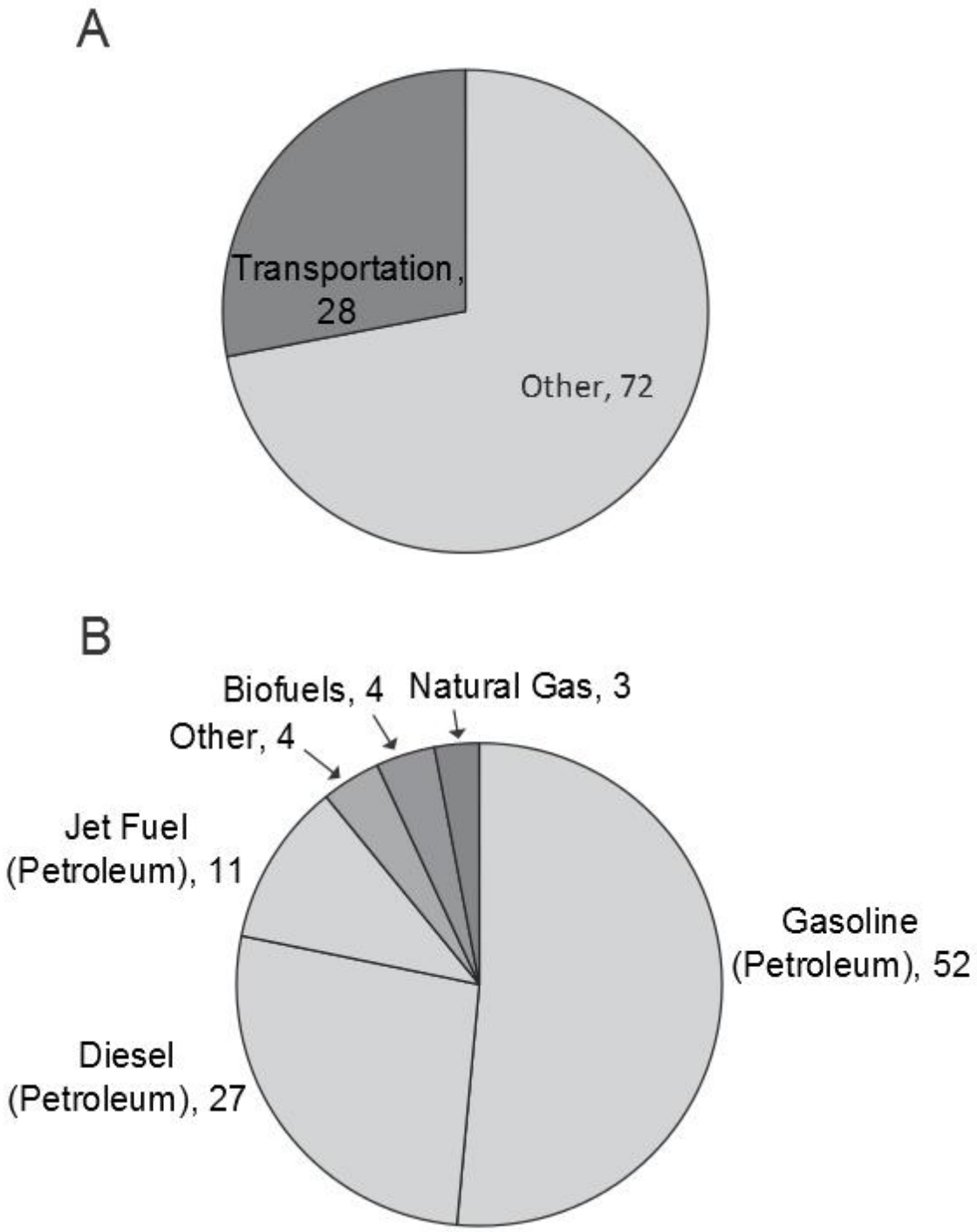


Figure 1-1. Transportation fuel usage in the United States. (A) The percentage of total energy consumption used for transportation purposes in the US in 2011. **(B)** The types of fuel used in the US in 2011, and their sources. "Other" includes motor gasoline and aviation gas used in light planes, excludes ethanol. "Biofuels" does not include biodiesel. Data from U.S.Energy Information Administration (2012b).

for oil in the US is likely to keep increasing as the population grows, since our transportation infrastructure is based on internal combustion vehicles. While alternatives to reduce fuel usage, including electric cars, increased mass transit, and cycling, are being encouraged, it will take time and pressure to overcome current cultural preferences.

For these reasons, the use of renewable energy sources for both transportation and power generation is steadily increasing. Various forms of power generation are currently being tried out, ranging from solar to wind and wave energy. However, for transportation purposes, fuel needs to be portable, and solar and wind are therefore inadequate for these purposes. Electric cars are being produced, but, due to current battery technology, have several disadvantages, including lengthy recharges and shorter travel time between recharges, which make them inferior to conventionally powered cars or hybrid cars for most situations.

Biofuels as an alternative source of energy

One major area of interest is the use of biological material for fuel, either burned directly as biomass or refined into compounds such as ethanol or biodiesel (Antoni, D. et al. 2007, Durrett, T. P. et al. 2008, Ohlrogge, J. et al. 2009). Biofuels depend primarily on photosynthetic organisms, plants and algae, as feedstock. Plants are able to act as living solar collectors, using energy from the sun to fix carbon into useful compounds. This has the double advantage of reducing the energy input required for producing the fuel, and reducing the amount of CO_2 in the atmosphere. Biomass has been used for fuel in the past, as wood or charcoal, and many early automobiles were designed to run on alcohol or vegetable oil, before gasoline or diesel became standard. In recent decades, various governments, including the US and EU, have enacted programs and offered funding to promote the scientific study of biofuels, and encourage their use

commercially. Many academic labs, as well as various companies, are currently engaged in the process of making biofuels an effective replacement for fossil fuels.

Currently, the vast majority of transportation fuel in the US is derived from fossil oil (U.S.Energy Information Administration 2012b). The amount of biofuels being used has been gradually increasing, however. The two most commonly discussed forms of biofuel are ethanol and biodiesel, with other potential fuels including biohydrogen and biogas. Today, the US is the largest producer of biofuels, primarily in the form of ethanol, ahead of Brazil (Renewable Fuels Association 2012). The EPA has set standards for the use of renewable fuel in automobiles, as a way to encourage the growth of biofuels production in the US. Other countries have similar mandates. The EPA's Renewable Fuel Standard, first enacted in 2005 and updated in 2007, give a target amount of biofuels to be added to the gasoline pool, with individual targets for specific biofuels (corn-based ethanol, cellulosic ethanol, biodiesel, and other advanced biofuels) (U.S.Environmental Protection Agency 2010). The amount is set to increase gradually, going from 4 billion gallons per year in 2006 to 36 billion gallons by 2022.

There are several challenges facing the use of biofuels. The first is yield. To have a significant effect on the amount of petroleum fuel used and CO_2 released, growers need to produce drastically more biofuels than are currently being produced each year. Various factors affect the yield, including biological limitations of the various crops used, various limiting resources like water and arable land, and the need to compete with other crops for said resources (Boddiger, D. 2007, Stein, K. 2007). Improvements via species and strain selection, combined with various engineering improvements, can increase the total yield. Another factor is cost (Chisti, Y. 2007, Pimentel, D. 2003). In order to be a viable alternative and speed acceptance, biofuels must be on par with fossil fuels cost-wise. Recent increases in the cost of gasoline have

made this goal easier, but there is still a ways to go. Increasing the yield will help, as will improving the efficiency of biofuels processing. A third limitation is the need to produce consistent, defined chemical compounds, in order to produce fuel with ideal properties (Canakci, M. and Sanli, H. 2008, Knothe, G. 2006). Fuel properties can be improved by metabolic engineering, taking advantage of existing plant enzymes to modify the end products.

Types of biofuels

Currently, ethanol is the most abundant biofuel available. Ethanol is produced through fermentation by various microbes, which are fed with sugars produced from plants or similar. Most biofuels currently being produced in the US are ethanol, and specifically corn ethanol, which constitutes the majority of biofuels being produced domestically (12 billion gallons, vs ~1 billion gallons of biodiesel and less of cellulosic ethanol) (U.S.Energy Information Administration 2012a). Corn ethanol is derived from the starch in corn grains, which is broken down into the sugars needed for fermentation (Bothast, R. J. and Schlicher, M. A. 2005, Hill, J. et al. 2006). Corn ethanol is a simple way to produce biofuels, but it has several flaws, most notably the fact that it competes directly with the use of corn as a food crop. Additionally, the overall yield of corn ethanol is too low, requiring too much arable land and other resources to be an effective long-term fuel source (Hill, J. et al. 2006). Sugarcane is also being used for ethanol production, primarily in Brazil, and has greater yield than corn, but can only be grown in limited regions.

Due to the limitations of starch- and sugar-based ethanol production, most research is focused on lignocellulosic ethanol, where the sugars are derived from the various polysaccharides found in plant cell walls (Jordan, D. B. et al. 2012, Pauly, M. and Keegstra, K. 2010, Saha, B. and Cotta, M. A. 2012). The total amount of usable biomass per land area is

significantly increased, as most of the plant can be harvested and used, rather than just the grains, and it reduces the impact on food supply. Various sources of biomass are being considered, including leftover corn stalks, fast-growing grass species like switchgrass and miscanthus, and trees like poplar and willow. Different species could be used in different regions of the country, to maximize the overall yield and efficiency of ethanol. The downside of lignocellulosics is that the process of extracting the sugars is much more involved, requiring greater inputs of energy and increased use of various enzymes. One major limitation is the presence of lignins, which are phenolic compounds that help strengthen cell walls. Lignins interfere with the sugar extraction method, and are costly to remove. Genetic engineering methods have been used to reduce the amount of lignin produced by the plant, but plants require a certain minimum amount of lignin to maintain their structure. In addition, facilities for the processing of cellulosic ethanol have been slow to start up. As a result, the EPA-mandated requirements for cellulosic ethanol have been reduced the past several years (U.S.Energy Information Administration 2012a). Research is being done to both improve the plant feedstock and increase the efficiency of the processing methods.

In 2012, ethanol fuel consumption reached ~12 billion gallons, almost 10% of the ~133 billion gallons of gasoline consumed. Most of the ethanol was used as a blend of 10% ethanol/90% gasoline, known as E10 (U.S.Energy Information Administration 2012a). E10 is compatible with gasoline-fueled vehicles, without modification. A slightly higher blend, E15, is currently being introduced for use in newer cars, but increased amounts, such as E85, require specially-modified vehicles, known as “flex-fuel” vehicles, that can run on either gasoline or ethanol.

Despite ethanol's popularity in the US, it has several flaws that may limit its large-scale use (Hill, J. et al. 2006, Pimentel, D. 2003). First, ethanol has about two-thirds the energy density of gasoline (21.2 MJ/L vs. 34.8 MJ/L), requiring a greater volume to be produced and consumed for the same amount of travel. This means more ethanol needs to be transported from the manufacturing plants to the distribution stations, decreasing the energy ratio. Second, ethanol is not compatible with current fuel systems, and requires either blending with gasoline or modified vehicles, as discussed earlier. Fuel storage and transportation would also have to be modified to handle ethanol in higher concentrations. Third, even with improved yields from lignocellulosics, large amounts of land would be required for growing the feedstock crops, affecting the growth of current food and fiber crops and using various natural resources such as fresh water, having a potentially negative impact on the environment.

Another alternative is hydrogen. Hydrogen is a clean burning fuel source that has long been considered as a replacement for gasoline, and can be used in internal combustion engines (Hirose, K. 2010). Hydrogen can also be used to power fuel cells and act as a source of power for electric cars. However, the use of hydrogen as a fuel is currently limited by existing production methods. The most common production method is steam reforming, where hydrocarbons such as methane are heated to high temperatures to produce H_2 gas (Holladay, J. D. et al. 2009). However, this process has several downsides: it requires a hydrocarbon source, usually natural gas; it releases CO_2 as a byproduct; and it requires significant amounts of energy to heat the reaction. An alternative is electrolytic breakdown of water, which releases H_2 and O_2. This technique requires more energy than is available in the resulting hydrogen, giving it a net energy loss, but it has potential when coupled with renewable energy sources like solar and wind power (Walter, M. G. et al. 2010).

Recently, more attention has been focused on hydrogen produced by living organisms. Hydrogen has been shown to be produced naturally by various photosynthetic microbes, including rhodobacter, cyanobacteria and green algae (Kruse, O. et al. 2005, McKinlay, J. B. and Harwood, C. S. 2010, Srirangan, K. et al. 2011). Hydrogen is normally produced in the absence of oxygen, as a byproduct of photosynthesis. The hydrogen itself is produced by various hydrogenase enzymes, which are only active in the absence of oxygen. Generating biohydrogen requires the use of enclosed photobioreactors, to control the environment and limit oxygen contamination. This necessarily increases the cost of production, but allows the bioreactors to be sited on land unsuitable for growing normal crops. Although this is a promising area of research, much work still needs to be done to overcome existing biological and engineering limitations and increase hydrogen yield, to make this approach economically viable.

Several other forms of biofuel are based primarily on the reaction of total biomass, including biogas production and straight combustion (Borjesson, P. and Mattiasson, B. 2008, Rittmann, B. E. 2008, Rottig, A. et al. 2010). These types of fuels are dependent more on total carbon assimilated into plant matter, rather than the specific compounds formed or enzymatic reactions occurring. As a result, most plant waste and by products can be used for fuel, along with fast growing energy crops and even manure. However, these forms of fuel are generally not suitable for transportation purposes, and are better suited for fueling power plants. Thus, the conversion or combustion of leftover biomass is an important side product stream of the other biofuels discussed, and can help reduce the costs and increase the net energy gain from these other methods (Jones, C. S. and Mayfield, S. P. 2012).

The third major form of biofuels being considered is biodiesel, derived from vegetable oils. Vegetable oil, or triacylglycerol (TAG), consists of three fatty acids esterified to a glycerol

backbone (Durrett, T. P. et al. 2008, Hu, Q. et al. 2008). The fatty acids can be transesterified in methanol, forming long hydrocarbon chain-methyl esters that have similar properties to diesel components. Used cooking oil has been successfully recycled into functional biodiesel, but providing enough fuel for the entire country will require dedicated biodiesel crops.

One of the advantages of biodiesel is that unlike ethanol or hydrogen, it can be used directly in existing diesel engines, and is compatible with current storage and transportation technology. In addition, the energy density of biodiesel is substantially higher than ethanol (35.7 MJ/L vs. 21.8 MJ/L), more like that of conventional diesel (40.3 MJ/L), making biodiesel combustion, transportation and storage more efficient. The energy return after processing is also significantly higher than ethanol, with biodiesel giving a calculated 93% return over the energy needed for production, while ethanol gives only 25% more energy (Hill, J. et al. 2006). Thus, biodiesel production is more efficient than ethanol, based on the inputs required. For these reasons, biodiesel is a promising replacement for fossil-fuel based transportation. However, the current yield of vegetable oil from oil crops is nowhere near enough to meet transportation demands (Chisti, Y. 2008). Given the yield of current oil seed crops, such as canola, soybean, or palm oil, there is not enough arable land in the US to meet demand (Table 1-1). Also, like with ethanol, oil crops provide an important source of food, which would be significantly impacted by their use as fuel sources. Clearly, another solution is needed.

Algae as a biodiesel source

An alternative to land-based oil crops is microalgae, single-celled photosynthetic algae capable of producing TAG. The advantages of microalgae as biodiesel feedstock are many: higher yield of oil per acre per year, due to their fast growing time and increased yield per dry weight (Table 1-1); lack of competition with food crops; and can be grown on marginal land,

Table 1-1. Comparison of land use required for oil crops to meet current fuel usage

Name	Percent land use
Continental US	100%
Soybean	330%
Jatropha	75%
Oil palm	23%
Algae	<4%

Data taken from Georgianna, D. R. and Mayfield, S. P. (2012)

with much less fresh water than conventional crops (Hu, Q. et al. 2008, Jones, C. S. and Mayfield, S. P. 2012, Williams, P. J. L. and Laurens, L. M. L. 2010). Much previous research has gone into studying the feasibility of algae as fuel sources, most notably the National Renewable Energy Lab's Aquatic Species Program, which ran from the 70's to the 90's and looked at both various algae strains, as well as various growth and harvesting methods to maximize yields (Sheehan, J. et al. 1998).

Although there are many similarities between algae and land plants, there are some differences as well. One major difference is how and why they synthesize TAG. In land plants, TAG is produced primarily in developing seeds, and acts as a carbon and energy source for developing seedlings. There are some exceptions, such as nut sedge, firewood or fruits like avocado, olives and oil palm, which produce oil in vegetative tissues such as tubers, stems and mesocarps, but most TAG production is linked to a specific developmental stage (Giannoulia, K. et al. 2000, Manaf, A. M. and Harwood, J. L. 2000, Stoller, E. W. and Weber, E. J. 1975, Takenaga, F. et al. 2008, Wang, G. et al. 2007). Recently, it has been shown that under certain stress conditions, such as freezing stress, Arabidopsis produces TAG as part of a membrane remodeling process (Moellering, E. R. et al. 2010). In algae, conversely, TAG is produced primarily in response to environmental conditions, usually ones that lead to significant reductions in growth rate, such as nutrient deprivation or high light stress (Roessler, P. G. 1988, Suen, Y. et al. 1987, Tornabene, T. G. et al. 1983, Zhekisheva, M. et al. 2002). While the exact purpose of this accumulation of TAG is unknown, several uses have been proposed, including material for later growth or as a highly-reduced compound to protect against oxidative stress produced by excess light absorption. Another major difference between land plants and microalgae is that land plants are multicellular, while microalgae are single-celled. Thus, while

the mechanisms behind TAG synthesis are largely conserved between the two groups, the regulation of TAG synthesis is likely to be significantly different.

While much progress has been made in understanding TAG synthesis in algae, much more work is needed to make microalgal biofuels a viable replacement for fossil fuels (Georgianna, D. R. and Mayfield, S. P. 2012, Williams, P. J. L. and Laurens, L. M. L. 2010). Numerous algal species have been identified that accumulate large amounts of TAG, but most species have a trade-off between growth and TAG synthesis that limits overall yield. Controlling TAG synthesis, along with the engineering problems of growth and harvesting, are still major obstacles (Halim, R. et al. 2012).

***Chlamydomonas reinhardtii* as a model for oil synthesis**

For most algae research, *Chlamydomonas reinhardtii* is the model species used. *Chlamydomonas* is a fresh-water green alga first isolated in the 1940's (Harris, E. H. 2001). *Chlamydomonas* are single-celled algae, roughly 10 mm in diameter, with two flagella. It has a single large chloroplast, containing a pyrenoid for carbon fixation and starch synthesis, but is capable of growing heterotrophically as well, using acetate in the medium. Like other microalgae, *Chlamydomonas* is capable of accumulating significant amounts of TAG in response to environmental stress, including nitrogen, phosphorus and sulfur deprivation, with TAG amounts reaching 30% of the total lipid in the case of nitrogen deprivation (Matthew, T. et al. 2009, Weers, P. M. M. and Gulati, R. D. 1997). As a green alga, *Chlamydomonas* is related to other studied microalgae, such as *Chlorella*, *Dunaliella*, *Botryococcus*, *Ostreococcus* and *Volvox* (Sheehan, J. et al. 1998, Wagner, M. et al. 2010). It is from a different lineage than other commonly-studied algae, such as diatoms or *Nannochloropsis*, which belong to the heterokont lineage (van den Hoek, C. et al. 1995).

As a model organism, *Chlamydomonas* has been used for multiple areas of research, including flagella development and function, photosynthesis, and gene silencing. A closely related species, *Volvox carteri*, has been used in conjunction with *Chlamydomonas* to study the development of multicellularity. Chlamydomonas has a large molecular toolkit, including a sequenced genome, reliable transformation methods for all three genomes, and the ability to perform sexual crosses (Fuhrmann, M. 2002, Lefebvre, P. A. and Silflow, C. D. 1999, Merchant, S. S. et al. 2007). Although its resources are fewer than for land plants, *Chlamydomonas* is the most studied microalgae, and thus the best existing model for researching TAG synthesis in microalgae.

Prior to the start of my project, relatively little research had been done on lipid synthesis in *Chlamydomonas*. The main enzymes involved in membrane lipid synthesis were deduced based on gene annotation, but relatively few enzymes have been confirmed biochemically (Riekhof, W. R. et al. 2005b). Based on gene annotation, *Chlamydomonas* lipid synthesis is the same as that in higher plants, with a few exceptions. One major difference is the absence of phosphatidylcholine in *Chlamydomonas*, which plays a major role in lipid synthesis and remodeling in Arabidopsis. The *Chlamydomonas*-specific betaine lipid diacylglycerol-*N,N,N*-trimethylhomoserine has a similar structure, and has been hypothesized to replace phosphatidylcholine functionally, but experiments to date have not been conclusive (Riekhof, W. R. et al. 2005a). Another major difference between land plants and *Chlamydomonas* is the absence of an endoplasmic reticulum (ER) to chloroplast lipid transport mechanism in *Chlamydomonas*. In land plants, fatty acids synthesized in the chloroplast are exported to the ER where they are modified and then reimported for incorporation into chloroplast membrane lipids (Benning, C. 2009, Xu, C. et al. 2003). The absence of this pathway indicates that fatty acid

modification and the fatty acid species available in *Chlamydomonas* differ from those in land plants. A third difference between *Chlamydomonas* and most land plants is the presence of more highly unsaturated fatty acids in *Chlamydomonas*, such as 16:4 and 18:4 (Tatsuzawa, H. et al. 1996, Weers, P. M. M. and Gulati, R. D. 1997). Other algae species, such as *Nannochloropsis*, have other unusual fatty acids, including 20:5, which may affect the properties of biodiesel produced from microalgae.

As mentioned, only a few of the enzymes involved in lipid synthesis have been tested and confirmed biochemically. In addition, very little is known about how lipid synthesis is regulated and controlled, how nutrient deprivation triggers TAG synthesis, how changes in other pathways in the cell are involved, etc. Thus, there is room for much more research to be done on TAG synthesis in *Chlamydomonas*.

Overview of my thesis project

My thesis project has focused on two main areas of TAG synthesis. The first area for study is the identification and characterization of the structural enzymes directly involved in TAG synthesis. These proteins affect not only the overall yield of TAG, but also its fatty acid composition, and thus the properties of the resulting biodiesel. As mentioned, most of the respective genes have been identified based on sequence similarity to known lipid genes, but only a few have confirmed activity. Although all the enzymes involved in lipid synthesis potentially affect TAG, the main enzymes responsible are the various acyltransferases involved in the Kennedy pathway: glycerol phosphate acyltransferases, lysophosphatidyl acyltransferases, and diacylglycerol acyltransferases (DGATs). Of these three, DGATs are the only TAG-specific enzymes, and catalyze the final step in TAG synthesis. Thus, these enzymes are important targets for future engineering attempts to increase TAG yields. In addition, in many land plant species,

DGATs are important for determining the fatty acid composition of TAG, which will be important for optimized biodiesel production. For these reasons, identification and characterization of the DGATs in *Chlamydomonas* is an important step in understanding TAG synthesis.

Multiple DGATs from various organisms, both plant and animal, have been identified (Liu, Q. et al. 2012). These genes are largely conserved between species, and thus it is likely that the same holds true for microalgae. Identification of potential DGAT candidates should be possible using BLAST, given the complete genome available. Once identified, the candidate genes can be tested by heterologous expression in other systems, as done in the past, to confirm their DGAT activity. The role of these genes in *Chlamydomonas* can be tested by ectopic expression or amiRNA-based knockdowns. My identification and analysis of the *Chlamydomonas* DGATs is described in Chapter 2, which also includes the work done in Arabidopsis, in collaboration with Dr. Sanjaya of the Benning lab.

The second area of study for my project is the regulation of TAG synthesis, and how it is affected by the overall cell metabolism. This is important for improving and controlling TAG synthesis and accumulation under biodiesel production systems. In *Chlamydomonas*, TAG synthesis is triggered by nutrient deprivation, which causes several major changes in metabolism. The cells slow down or stop dividing, and activate scavenging mechanisms to increase the amount of nutrient uptake, or adjust to growth on different sources of a particular nutrient (Grossman, A. 2000). Build-up of storage compounds, such as starch and TAG, also occur, indicating further changes in metabolism (Msanne, J. et al. 2012).

All of these changes are linked to changes in regulation, at various levels – transcriptional, post-transcriptional, and post-translational. These regulatory changes enact the

large- and small-scale metabolic changes that lead to TAG synthesis, among other changes. Determining how these various regulatory methods work together is key to understanding TAG synthesis, and determining how to engineer higher-yielding strains. However, the actual determination is hard, especially given the lack of information available for *Chlamydomonas*. The easiest way to start is to look at transcriptional regulation. This can be measured via transcript profile screens, using either microarray technology or more recent high-throughput sequencing methods. This approach will identify at least some of the key genes involved, which will allow for a better determination of what overall changes in metabolism are occurring. This global approach, versus a more targeted approach, will also allow for the detection of changes in unexpected genes and pathways. This approach won't identify any gene products modified post-transcriptionally, but will provide a foundation for further research. My experiment, described in Chapter 3, started by determining the best conditions for comparison via high throughput sequencing. I then discussed the analysis of the data, performed in collaboration with others both in the Benning lab and in other labs on campus, in particular with Guangxi Wu of the Shiu lab.

A major aspect of transcriptional regulation is the activity of transcription factors (TFs). These proteins are particularly important for understanding and manipulating the regulation of TAG, as they activate or repress multiple genes in a pathway. It is likely that one or more transcription factors are involved in the activation of TAG synthesis. By isolating and expressing these TFs, TAG synthesis can potentially be induced without the other changes in metabolism, allowing TAG accumulation without affecting the growth rate of the cells. Alternatively, by selectively activating or deactivating certain pathways, more carbon can be shunted into TAG versus other compounds (such as starch).

One possible result of the transcript profiling screen, in addition to the overall view of cellular metabolism and identification of potential structural genes of importance, is the identification of some of the TFs involved, either directly or indirectly, in TAG synthesis. Given that TAG synthesis is induced by nutrient deprivation, it is reasonable to assume that transcription factors that control TAG synthesis are also induced. Once potential candidates are identified in the transcription profile, they can be tested biochemically to confirm their role in TAG synthesis. Once identified, lipid-related TFs will be powerful tools for further identification of the regulatory networks involved in TAG synthesis, as well as useful for future engineering of microalgae for increased TAG yield. The analysis of some of the TFs identified in the transcript profiling screen is described in Chapter 4.

BIBLIOGRAPHY

BIBLIOGRAPHY

Antoni,D., Zverlov,V.V., and Schwarz,W.H. (2007) Biofuels from microbes. Appl.Microbiol.Biotechnol. **77**:23-35.

Benning,C. (2009) Mechanisms of lipid transport involved in organelle biogenesis in plant cells. Annu.Rev.Cell Dev.Biol. **25**:71-91.

Boddiger,D. (2007) Boosting biofuel crops could threaten food security. Lancet **370**:923-924.

Borjesson,P. and Mattiasson,B. (2008) Biogas as a resource-efficient vehicle fuel. Trends.Biotechnol **26**:7-13.

Bothast,R.J. and Schlicher,M.A. (2005) Biotechnological processes for conversion of corn into ethanol. Appl.Microbiol.Biotechnol. **67**:19-25.

California Air Resources Board. Low Carbon Fuel Standard Program. 5-23-2013. 5-30-2013. Ref Type: Online Source

Canakci,M. and Sanli,H. (2008) Biodiesel production from various feedstocks and their effects on the fuel properties. J.Ind.Microbiol.Biotechnol. **35**:431-441.

Chisti,Y. (2007) Biodiesel from microalgae. Biotechnol.Adv. **25**:294-306.

Chisti,Y. (2008) Biodiesel from microalgae beats bioethanol. Trends.Biotechnol **26**:126-131.

Durrett,T.P., Benning,C., and Ohlrogge,J. (2008) Plant triacylglycerols as feedstocks for the production of biofuels. Plant.J **54**:593-607.

Fuhrmann,M. (2002) Expanding the molecular toolkit for *Chlamydomonas reinhardtii*--from history to new frontiers. Protist. **153**:357-364.

Georgianna,D.R. and Mayfield,S.P. (2012) Exploiting diversity and synthetic biology for the production of algal biofuels. Nature **488**:329-335.

Giannoulia,K., Haralampidis,K., Poghosyan,Z., Murphy,D.J., and Hatzopoulos,P. (2000) Differential expression of diacylglycerol acyltransferase (DGAT) genes in olive tissues. Biochem.Soc.Trans. **28**:695-697.

Grossman,A. (2000) Acclimation of *Chlamydomonas reinhardtii* to its nutrient environment. Protist **151**:201-224.

Halim,R., Danquah,M.K., and Webley,P.A. (2012) Extraction of oil from microalgae for biodiesel production: A review. Biotechnol.Adv. **30**:709-732.

Harris,E.H. (2001) *Chlamydomonas* as a model organism. Annu.Rev.Plant.Phys **52**:363-406.

Hill,J., Nelson,E., Tilman,D., Polasky,S., and Tiffany,D. (2006) Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc.Natl.Acad.Sci.U.S.A **103**:11206-11210.

Hirose,K. (2010) Materials towards carbon-free, emission-free and oil-free mobility: hydrogen fuel-cell vehicles--now and in the future. Philos.Trans.A Math.Phys.Eng Sci. **368**:3365-3377.

Holladay,J.D., Hu,J., King,D.L., and Wang,Y. (2009) An overview of hydrogen production technologies. Catal.Today **139**:244-260.

Hu,Q., Sommerfeld,M., Jarvis,E., Ghirardi,M., Posewitz,M., Seibert,M., and Darzins,A. (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant.J **54**:621-639.

Jones,C.S. and Mayfield,S.P. (2012) Algae biofuels: versatility for the future of bioenergy. Curr.Opin.Biotechnol. **23**:346-351.

Jordan,D.B., Bowman,M.J., Braker,J.D., Dien,B.S., Hector,R.E., Lee,C.C., Mertens,J.A., and Wagschal,K. (2012) Plant cell walls to ethanol. Biochem.J. **442**:241-252.

Knothe,G. (2006) Analyzing biodiesel: Standards and other methods. J.Am.Oil.Chem.Soc **83**:823-833.

Kruse,O., Rupprecht,J., Mussgnug,J.H., Dismukes,G.C., and Hankamer,B. (2005) Photosynthesis: a blueprint for solar energy capture and biohydrogen production technologies. Photochem.Photobiol.Sci. **4**:957-970.

Lefebvre,P.A. and Silflow,C.D. (1999) *Chlamydomonas*: The cell and its genomes. Genetics **151**:9-14.

Liu,Q., Siloto,R.M., Lehner,R., Stone,S.J., and Weselake,R.J. (2012) Acyl-CoA:diacylglycerol acyltransferase: molecular biology, biochemistry and biotechnology. Prog.Lipid Res. **51**:350-377.

Manaf,A.M. and Harwood,J.L. (2000) Purification and characterisation of acyl-CoA: glycerol 3-phosphate acyltransferase from oil palm (*Elaeis guineensis*) tissues. Planta **210**:318-328.

Mann,P., Gahagan,L., and Gordon,M.B. (2003) Tectonic Setting of the World's Giant Oil and Gas Fields. In Giant Oil and Gas Fields of the Decade 1990-1999, J.C.Lorenz, ed (Kansas City, MO: The American Association of Petroleum Geologists), pp. 15-106.

Matthew,T., Zhou,W., Rupprecht,J., Lim,L., Thomas-Hall,S.R., Doebbe,A., Kruse,O., Hankamer,B., Marx,U.C., Smith,S.M., and Schenk,P.M. (2009) The metabolome of *Chlamydomonas reinhardtii* following induction of anaerobic H2 production by sulfur depletion. J.Biol.Chem. **284**:23415-23425.

McKinlay,J.B. and Harwood,C.S. (2010) Photobiological production of hydrogen gas as a biofuel. Curr.Opin.Biotechnol. **21**:244-251.

Merchant,S.S., Prochnik,S.E., Vallon,O., Harris,E.H., Karpowicz,S.J., Witman,G.B., Terry,A., Salamov,A., Fritz-Laylin,L.K., Marechal-Drouard,L., Marshall,W.F., Qu,L.H., Nelson,D.R., Sanderfoot,A.A., Spalding,M.H., Kapitonov,V.V., Ren,Q.H., Ferris,P., Lindquist,E., Shapiro,H., Lucas,S.M., Grimwood,J., Schmutz,J., Grigoriev,I.V., Rokhsar,D.S., and Grossman,A.R. (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science **318**:245-251.

Moellering,E.R., Muthan,B., and Benning,C. (2010) Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. Science **330**:226-228.

Msanne,J., Xu,D., Konda,A.R., Casas-Mollano,J.A., Awada,T., Cahoon,E.B., and Cerutti,H. (2012) Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae *Chlamydomonas reinhardtii* and *Coccomyxa* sp. C-169. Phytochemistry **75**:50-59.

Murray,J. and King,D. (2012) Climate policy: Oil's tipping point has passed. Nature **481**:433-435.

Ohlrogge,J., Allen,D., Berguson,B., Dellapenna,D., Shachar-Hill,Y., and Stymne,S. (2009) Energy. Driving on biomass. Science **324**:1019-1020.

Pauly,M. and Keegstra,K. (2010) Plant cell wall polymers as precursors for biofuels. Curr.Opin.Plant Biol. **13**:305-312.

Pimentel,D. (2003) Ethanol Fuels: Energy Balance, Economics, and Environmental Impacts Are Negative. Nat.Resour.Res **12**:127-134.

Renewable Fuels Association. World Fuel Ethanol Production. 2012. 5-10-2013.
Ref Type: Online Source

Riekhof,W.R., Andre,C., and Benning,C. (2005a) Two enzymes, BtaA and BtaB, are sufficient for betaine lipid biosynthesis in bacteria. Arch.Biochem.Biophys **441**:96-105.

Riekhof,W.R., Sears,B.B., and Benning,C. (2005b) Annotation of genes involved in glycerolipid biosynthesis in Chlamydomonas reinhardtii: discovery of the betaine lipid synthase BTA1Cr. Eukaryot.Cell **4**:242-252.

Rittmann,B.E. (2008) Opportunities for renewable bioenergy using microorganisms. Biotechnol.Bioeng. **100**:203-212.

Roessler,P.G. (1988) Effects of Silicon Deficiency on Lipid-Composition and Metabolism in the Diatom *Cyclotella cryptica*. J.Phycol **24**:394-400.

Rottig,A., Wenning,L., Broker,D., and Steinbuchel,A. (2010) Fatty acid alkyl esters: perspectives for production of alternative biofuels. Appl.Microbiol.Biot **85**:1713-1733.

Saha,B. and Cotta,M.A. (2012) Ethanol production from lignocellulosic biomass by recombinant Escherichia coli strain FBR5. Bioengineered. **3**:197-202.

Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P. A Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae. National Renewable Energy Laboratory . 1998.
Ref Type: Electronic Citation

Srirangan,K., Pyne,M.E., and Perry,C.C. (2011) Biochemical and genetic engineering strategies to enhance hydrogen production in photosynthetic algae and cyanobacteria. Bioresour.Technol. **102**:8589-8604.

Stein,K. (2007) Food vs biofuel. J.Am.Diet.Assoc **107**:1870-1876.

Stoller,E.W. and Weber,E.J. (1975) Differential cold tolerance, starch, sugar, protein, and lipid of yellow and purple nutsedge tubers. Plant.Physiol **55**:859-863.

Suen,Y., Hubbard,J.S., Holzer,G., and Tornabene,T.G. (1987) Total Lipid Production of the Green Alga *Nannochloropsis* Sp Qii Under Different Nitrogen Regimes. J.Phycol **23**:289-296.

Takenaga,F., Matsuyama,K., Abe,S., Torii,Y., and Itoh,S. (2008) Lipid and fatty acid composition of mesocarp and seed of avocado fruits harvested at northern range in Japan. J.Oleo.Sci. **57**:591-597.

Tatsuzawa,H., Takizawa,E., Wada,M., and Yamamoto,Y. (1996) Fatty acid and lipid composition of the acidophilic green alga *Chlamydomonas* sp. J.Phycol **32**:598-601.

Tornabene,T.G., Holzer,G., Lien,S., and Burris,N. (1983) Lipid Composition of the Nitrogen Starved Green Alga *Neochloris oleoabundans*. Enzyme.Microb.Tech **5**:435-440.

U.S.Energy Information Administration. International Energy Outlook 2011. 9-19-2011. 5-30-2013.
Ref Type: Online Source

U.S.Energy Information Administration. Biofuels Issues and Trends. www.eia.gov . 2012a. 5-8-2013a.
Ref Type: Electronic Citation

U.S.Energy Information Administration. Energy Use For Transportation. www.eia.gov . 6-14-2012b. 5-8-2013b.
Ref Type: Online Source

U.S.Environmental Protection Agency. EPA Finalizes Regulations for the National Renewable Fuel Standard Program for 2010 and Beyond. www.epa.gov . 3-26-2010. 5-8-2013.
Ref Type: Electronic Citation

van den Hoek,C., Mann,D.G., and Jahns,H.M. (1995) Heterkontophyta: Class Eustignatophyceae. In Algae: An introduction to phycology, (Cambridge, UK: University Press), pp. 130-132.

Wagner,M., Hoppe,K., Czabany,T., Heilmann,M., Daum,G., Feussner,I., and Fulda,M. (2010) Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (DGAT2) gene from the microalga *O. tauri*. Plant.Physiol.Bioch **48**:407-416.

Walter,M.G., Warren,E.L., Mckone,J.R., Boettcher,S.W., Mi,Q.X., Santori,E.A., and Lewis,N.S. (2010) Solar Water Splitting Cells. Chem.Rev **110**:6446-6473.

Wang,G., Lin,Q., and Xu,Y. (2007) *Tetraena mongolica* Maxim can accumulate large amounts of triacylglycerol in phloem cells and xylem parenchyma of stems. Phytochemistry **68**:2112-2117.

Weers,P.M.M. and Gulati,R.D. (1997) Growth and reproduction of *Daphnia galeata* in response to changes in fatty acids, phosphorus, and nitrogen in *Chlamydomonas reinhardtii*. Limnol.Oceanogr **42**:1584-1589.

Williams,P.J.L. and Laurens,L.M.L. (2010) Microalgae as biodiesel & biomass feedstocks: Review & analysis of the biochemistry, energetics & economics. Energ.Environ.Sci **3**:554-590.

Xu,C., Fan,J., Riekhof,W., Froehlich,J.E., and Benning,C. (2003) A permease-like protein involved in ER to thylakoid lipid transfer in Arabidopsis. EMBO.J **22**:2370-2379.

Zhekisheva,M., Boussiba,S., Khozin-Goldberg,I., Zarka,A., and Cohen,Z. (2002) Accumulation of oleic acid in *Haematococcus pluvialis* (Chlorophyceae) under nitrogen starvation or high light is correlated with that of astaxanthin esters. J.Phycol **38**:325-331.

CHAPTER 2: IDENTIFICATION AND ANALYSIS OF *CHLAMYDOMONAS REINHARDTII* DIACYLGLYCEROL ACYLTRANSFERASE GENES

This chapter contains descriptions of experiments published in (Sanjaya et al. 2013). Parts of the methods and results have been reproduced here. This chapter also described experiments not included in the paper.

Introduction

Lipid synthesis in plants is a complicated process, involving multiple interwoven pathways in two different compartments in the cell, the plastid and the endoplasmic reticulum (Chapman, K. D. and Ohlrogge, J. B. 2012). The flux of components through the pathways and between the two compartments is an important factor in regulating the amount and composition of lipids created, including TAG. TAG synthesis starts with fatty acid synthesis within the chloroplast. Fatty acids are created by the fatty acid synthase complex, where they are extended to 16 or 18 carbons, and then exported to the ER to become part of the acyl-CoA pool used for further membrane lipid synthesis. The main pathway for TAG synthesis in the ER is via the Kennedy pathway for glycerolipid synthesis. In this pathway, acyl-CoAs are added step-wise to glycerol-3-phosphate to create first lyso-phosphatidic acid, then phosphatidic acid (PA), catalyzed by GPAT and LPAAT acyltransferases, respectively. PA is then dephosphorylated by PAP to create DAG, which acts as a building block for various glycerolipids, including TAG.

The main enzyme responsible for TAG synthesis in eukaryotes is diacylglycerol acyltransferase, or DGAT (Athenstaedt, K. and Daum, G. 2006, Chapman, K. D. and Ohlrogge, J. B. 2012). This enzyme adds a third acyl-CoA to DAG, creating TAG. This is the final step in the Kennedy pathway, and is specific to TAG synthesis. As a result, DGATs have been widely studied as important targets for potential bioengineering schemes to increase TAG yields in

microalgae or land plants. One major use of increased TAG is for biofuels production, where DGATs can be used to create TAG throughout the plant, rather than simply in the seed (Andrianov, V. et al. 2010). In addition to simply increasing the overall amount of oil, much research has been done on creating oil incorporating specific types of fatty acids, such as hydroxy fatty acids, for use in industrial processes (Burgal, J. et al. 2008, Li, R. et al. 2010, Lu, C. et al. 2011). DGATs are also important for nutrition and health in humans and animals in diverse ways. Vegetable oils are a major source of calories throughout the world, and ensuring people get the proper amount and types of fatty acids is a major goal. In addition, DGAT activity has been linked to numerous health issues, including obesity and liver disease (Wang, Z. G. et al. 2010, Yen, C. L. E. et al. 2008). Thus, DGATs have been the most-studied TAG synthesis enzymes in recent years.

Research has revealed that TAG synthesis is not as straight forward as the Kennedy pathway shows. Other types of TAG synthesis enzymes exist, that aren't dependent on acyl-CoAs as a substrate. Phospholipid:diacylglycerol acyltransferases (PDATs) also synthesize TAG, transferring a fatty acid from a phospholipid donor to DAG to create TAG (Dahlqvist, A. et al. 2000). PDATs are found in many organisms, including yeast, land plants and algae. DAG:DAG transacylases (DGTA) are a third form of TAG synthesis enzymes, that act by transferring a fatty acid from one DAG molecule to another (Stobart, K. et al. 1997). These enzymes provide alternative ways for fatty acids to be added to TAG, making the determination of the total amount and make-up of TAG more complicated. Although examples of these enzymes have been found in several organisms, how they interact with DGATs and what role they play in TAG synthesis is still unclear, and likely varies depending on species and conditions.

DGATs can be divided into several families, based on their structure. The two main families are type 1 and type 2 DGATs. These two families have the same activity, but differ in structure, with different enzymatic domains and number of transmembrane domains, and appear to be the result of convergent evolution (Turchetto-Zolet, A. C. et al. 2011). Both families can be found throughout eukaryotic species, and virtually all organisms have at least one copy of each within their genome. A third family of DGAT is the type 3 DGAT. Unlike the first two families, the type 3 DGAT is soluble and localized to the cytosol. The first example of this type of DGAT was isolated from peanut (*Arachis hypogaea*), and recently a putative homolog has been identified in Arabidopsis (Peng, F. Y. and Weselake, R. J. 2011, Saha, S. et al. 2006). A fourth type of DGAT is a bifunctional wax ester synthase/*DGAT* gene, which can create both TAG and waxes. Examples of this gene have been found in fungi and Arabidopsis (Kalscheuer, R. and Steinbuchel, A. 2003, Li, F. et al. 2008).

DGAT activity was detected in extracts from various species for many years before the genes responsible were isolated (Stymne, S. and Stobart, A. K. 1984). The first DGATs were identified in mice and humans in 1998 (Cases, S. et al. 1998). The first plant DGAT was identified in Arabidopsis a year later (Hobbs, D. H. et al. 1999, Routaboul, J. M. et al. 1999, Zou, J. T. et al. 1999). Both of these were type 1 DGATs. Type 2 DGATs were first identified in a fungus, *Mortierella ramanniana*, and later found in mammals (Cases, S. et al. 2001, Lardizabal, K. D. et al. 2001). Since then, both type 1 and type 2 DGATs have been identified in several other plant species, such as castor bean (*Ricinus communis*), tung tree (*Vernicia fordii*), olives (*Olea europaea*), and maize (*Zea mays*) (Banilas, G. et al. 2011, He, X. H. et al. 2004b, Kroon, J. T. M. et al. 2006, Shockey, J. M. et al. 2006, Zheng, P. et al. 2008). More recently, DGATs have

been identified and characterized in the green alga *Ostreococcus tauri*, and in the diatom *Phaeodactylum tricornutum* (Guiheneuf, F. et al. 2011, Wagner, M. et al. 2010).

The localization of DGATs is still somewhat unknown. TAG accumulates in lipid bodies, which are hypothesized to develop from the ER membrane. Most DGATs are assumed to be localized to the ER, and have been shown to contain ER-targeting motifs (Shockey, J. M. et al. 2006). However, there is evidence of DGAT activity in chloroplast extracts, and TAG-containing plastoglobules have been found in chloroplasts (Kessler, F. and Vidi, P. A. 2007, Lung, S. C. and Weselake, R. J. 2006). This suggests that DGATs may be dual-targeted to both compartments, or that different DGAT proteins in each organism may go to different organelles.

As mentioned before, most organisms seem to have at least one copy each of *DGAT1* and *DGAT2*. One of the few exceptions is yeast, which only has a *DGAT2* (Sorger, D. and Daum, G. 2002, Turchetto-Zolet, A. C. et al. 2011). The role of the two types of DGATs seems to vary depending on the species. In Arabidopsis, mutation in the *DGAT1* locus was shown to significantly reduce the amount of seed oil produced, with expression of the *DGAT1* cDNA restoring normal seed oil levels (Jako, C. et al. 2001, Routaboul, J. M. et al. 1999, Zou, J. T. et al. 1999). However, mutations in the *DGAT2* locus have no effect on the amount of seed oil that accumulates, and double mutants of *DGAT1* and *DGAT2* show no additional effect, suggesting that *DGAT2*, despite being expressed in the seed, has a limited role in Arabidopsis (Chapman, K. D. and Ohlrogge, J. B. 2012). *PDAT1*, on the other hand, does play an important role in TAG synthesis in Arabidopsis, with double mutants of *DGAT1* and *PDAT1* showing severe phenotypes in pollen and embryo formation (Zhang, M. et al. 2009). In olives, transcript analysis of *DGAT1* and *DGAT2* show different, though overlapping, expression patterns in flowers and developing seeds, with *DGAT1* being expressed predominantly during the periods of maximum

TAG synthesis (Banilas, G. et al. 2011). This indicates that, like in Arabidopsis, DGAT1 is the enzyme primarily responsible for TAG synthesis in the seed.

In other plants, DGAT2 plays a larger role, and in some plants is apparently involved in the synthesis of TAG with unusual fatty acids. In castor beans, both DGAT1 and DGAT2 were shown to have DGAT activity in the seed (He, X. H. et al. 2004a, Kroon, J. T. M. et al. 2006). However, comparison of the expression levels of the two genes show that *DGAT1* is expressed roughly equally in leaves and seed, and peaks early during seed development, while *DGAT2* is expressed highly in the seed, and later in seed development, correlating with the main period of TAG synthesis. Kroon et al also showed that DGAT2 was capable of using the hydroxy fatty acid ricinoleic acid to create triricinolein, the most abundant TAG species in castor beans (~90%). This evidence indicates that DGAT2 is the enzyme primarily responsible for the synthesis of TAG in the seed. The substrate specificity of DGAT2 versus DGAT1 was also demonstrated in Arabidopsis. When transformed with a fatty acid hydroxylase (FAH12), Arabidopsis accumulated ~17% hydroxyl fatty acids in its TAG. Addition of the castor bean DGAT2 increased the incorporation of hydroxyl fatty acids to ~30%, while castor bean DGAT1 had little effect (Burgal, J. et al. 2008, Snyder, C. L. et al. 2009). This demonstrates that multiple *DGAT* genes can play different roles, and also is an indication of the importance of DGATs when engineering oil production in plants (Cahoon, E. B. et al. 2007, Napier, J. A. 2007).

In tung tree, *DGAT1* and *DGAT2* follow a similar pattern to that of castor bean. *DGAT1* is expressed throughout the plants, while *DGAT2* is expressed primarily in the seed, and shows a preference for the unusual fatty acid eleostearic acid, the major component of tung tree TAG (Shockey, J. M. et al. 2006). Additionally, it was shown that DGAT1 and DGAT2 are localized to different subdomains of the ER membrane, further suggesting that they play separate roles.

These observations indicate that DGAT1 and DGAT2 in plants are regulated differentially and fill different roles, in order to synthesize the proper TAG species in the proper amounts.

In green algae, the situation is somewhat different. Most green algae that have been sequenced contain multiple *DGAT2*s, and one or no *DGAT1*. For example, *Ostreococcus tauri* contains three *DGAT2* homologs and no *DGAT1* (Wagner, M. et al. 2010). *Micromonas* and *Chlorella* show similar patterns, based on the results from searching with BLAST, with 3-5 *DGAT2*s and 0-1 *DGAT1*. One potential explanation for this difference is the difference in TAG synthesis between land plants and algae. In land plants, TAG is produced primarily in developing seeds, where it acts as an energy source during germination and seedling growth. Conversely, in algae, TAG is synthesized primarily in response to environmental stresses, such as nutrient deprivation or oxidative stress. The exact function of algal TAG is unknown, but it may act as protection against oxidative species, or as a source of fatty acids for rapid membrane synthesis after nutritional conditions allowing for growth are restored.

Because of DGATs' function as the final step in TAG synthesis, along with their role in determining the composition of TAG, understanding how they work is important for any attempt to engineer better oil-producing crops, whether for biofuel feedstock, industrial compounds, or nutraceuticals. This is especially important for microalgae, as their lipid metabolism pathways are much less studied than in land plants. As *Chlamydomonas reinhardtii* is the model organism for microalgae, identifying and characterizing their DGATs, showing how they work together to create TAG and possibly addressing the question of why microalgae have multiple copies of type 2 DGATs, is an important step in understanding their metabolism. That is the goal of this project. I started by identifying potential DGATs in the *Chlamydomonas* genome, and then confirmed their activity via expression in various organisms. Parts of the experiments I performed have

been published, along with work performed by a postdoctoral researcher in the lab, Dr. Sanjaya, who performed the expression of DGTT2 in Arabidopsis. I also collaborated with Dr. Tim Durrett of the Ohlrogge lab to quantify the amount of TAG produced by exogenous expression of the DGTTs.

Methods

Phylogenetic analysis

The protein sequences for the *Chlamydomonas DGTT* genes (DGTT1: XP_001702848.1; DGTT2: XP_001694904.1; DGTT3: XP_001691447.1; DGTT4: XP_001693189.1; DGTT5: XP_001701667.1) were retrieved by sequence comparison with Arabidopsis DGAT2 (NP_566952.1) using BLAST (Altschul, S. F. et al. 1997). Additional protein sequences included in the analysis were VfDGAT2 (DQ356682.1), RcDGAT2 (DQ923084.1), ScDGA1 (NP_014888.1), and OtDGAT2B (XP_003083539.1). The sequences were aligned using the ClustalW software in MEGA5 (Tamura, K. et al. 2011). The alignment was then used to construct a phylogenetic tree using the Neighbor-Joining method in MEGA5, with the tree being tested by bootstrapping with 1000 replicates.

Plasmid construction

Chlamydomonas strain dw15.1 (cw15, nit1, mt+), provided by Arthur Grossman, was grown under continuous light (~80 μm/m2/s) and at 22ºC in liquid TAP media (Harris, E. H. 1989) until mid-log phase, and then pelleted. Total RNA was extracted from the cells using Qiagen (http://www.qiagen.com) RNeasy Plant Mini kit. cDNA was synthesized with Invitrogen (http://www.invitrogen.com) SuperScriptIII and oligo-dT primer, and used as a template for PCR. The primers used are listed in Table 2-1. The amplified regions were digested with *Hin*dIII and *Sph*I, for *DGTT2* and *DGTT4*, and *Hin*dIII and *Xho*I, for *DGTT3* and *DGTT5*. The gene

Table 2-1. Primers used in this experiment

Name		5’ Sequence 3’
DGTT2-his	F	GCGCAAGCTTAGCATGGGTCATCATCACCATCACCATGCGATTGATAAAGCA
	R	GCGCGCGCATGCTCAGCTGATGACCAGCGG
DGTT3-his	F	GCGCGCAAGCTTAGCATGGGTCATCATCACCATCACCATGCAGGTGGAAAGTCA
	R	GCGCGCCTCGAGCTACTCGATGGACAGCGG
DGTT4-his	F	GCGCGCAAGCTTAGCATGGGTCATCATCACCATCACCATCCGCTCGCAAAGCTG
	R	GCGCGCGCATGCCTACATTATGACCAGCTC
DGTT5-his	F	GCGCGCAAGCTTAGCATGGGTCATCATCACCATCACCATCCGCGGGATCCGCCGG
	R	GCGCGCCTCGAGTCAGCACACCTCCAGCGG
p9-35S-DGTT2	F	CCTAggatccATGGCGATTGATAAAGC
	R	CCCGgaattctAGCTGATGACCAG
DGTT2-HA	F	GCGCGCACTAGTATGGCGATTGATAAAGCA
	R	GCGCGCGAATTCTCAGTAGTCGGGCACGTCGTAGGGGTAGCTGATGACCAGCGG
DGTT3-HA	F	GCGCGCACTAGTATGGCAGGTGGAAAGTCA
	R	GCGCGCGAATTCTCAGTAGTCGGGCACGTCGTAGGGGTACTCGATGGACAGCGG
PsaD terminator		CCCGTATCAATCAGCGAAAT
DGTT2-amiRNA	F	ctagtGTGGAAGGTTATGGCCACTTAtctcgctgatcggcac catgggggtggtggtgatcagcgctaTAAGAGGCCATAACCTTCCACg
	R	ctagcGTGGAAGGTTATGGCCTCTTAtagcgctgatcaccaccacc cccatggtgccgatcagcgagaTAAGTGGCCATAACCTTCCACa
DGTT3-amiRNA	F	ctagtCTCCGTTTGATTGGGTAGATAtctcgctgatcggcac catgggggtggtggtgatcagcgctaTATCAACCCAATCAAACGGAGg
	R	ctagcCTCCGTTTGATTGGGTTGATAtagcgctgatcaccaccacc cccatggtgccgatcagcgagaTATCTACCCAATCAAACGGAGa

sequences were then sub-cloned into the Invitrogen pYES2 vector to form pYES2-DGTT2-5 for expression in yeast.

DGTT2 cDNA was amplified from pYES2-DGTT2 by PCR using gene specific primers (Table 2-1). A fragment of 975 bp containing the complete open reading frame was digested with *Bam*HI and *Eco*RI. This fragment was then placed between the CaMV 35S promoter and OCS terminator of vector p9-35S-OCS (DNA Cloning Service, Hamburg, Germany) to form 35S:DGTT2.

Yeast expression

Yeast strain H1266 (*are2*Δ *lro1*Δ *dga1*Δ) (Sandager, L. et al. 2002) was grown to mid-log phase in YPD media (Sherman, F. 2002) and transformed with the DGTT constructs, along with an empty pYES2 vector as a negative control, according to Gietz, R. D. et al. (1995). The transformants were selected on SC media (Sherman, F. 2002) with 2% glucose and the uracil omitted (SC-U). Colonies were picked and grown overnight in SC-U + 2% glucose, before being transferred to SC-U + 2% galactose, 1% raffinose. After 48 hours, 30 mL of the cultures were collected and pelleted by centrifugation for 5 min at 3000 x g. Lipids were extracted and analyzed as described below.

Microsome DGTT2 assays

Transformed yeast colonies were picked and grown in SC-U + 2% glucose overnight, and then transferred to SC-U + 2% galactose, 1% raffinose. The cells were harvested after 12 hours, and the microsomes were prepared as described in Milcamps, A. et al. (2005). The total protein concentration was measured using Bio-Rad Protein Assay Dye Reagent (http://www.bio-rad.com/), by adding 900 µL of the reagent to 20 µL of BSA standards or microsome samples. 50 ng of the microsome fraction were added to a mix containing 100 mM Tris, 8 mM $MgCl_2$, 1

mg/mL BSA, 20% glycerol, 0.25 mg/mL DAG, and 1.725 nmol [1-^{14}C]-16:0-acyl-CoA or [1-^{14}C]-18:1-acyl-CoA (Moravek Biochemicals, http://www.moravek.com/). The reaction was incubated at room temperature for one hour. The lipids were extracted with chloroform:methanol (1:1 v/v) and phase separated with 0.2 M H_3PO_4 and 1 M KCl. The organic layer was extracted and separated on a silica TLC plate using 80:20:1 (v/v/v) petroleum ether:ether:acetic acid as the solvent. The TLC plate was exposed to film for 72 hours to visualize the radiolabeled lipids.

For the competition assay, 1.725 nmol unlabeled 16:0-acyl-CoA, 18:1-acyl-CoA, or 22:1-acyl-CoA were added. The separated TAG bands were scraped from the plate, and counted in a scintillation counter to quantify the amount of radioactivity incorporated into the lipid.

Plant material and generation of Arabidopsis transgenic plants

Arabidopsis wild-type (Col2) seeds or transgenic seeds were surface sterilized and grown on ½ Murashige and Skoog (MS) (Murashige, T. and Skoog, F. 1962) agar plates containing 1% sucrose in a growth chamber adjusted to 16 h light/8 h dark (100 µm/m2/s) at 22ºC after 3 days of stratification at 4ºC. Fifteen-day-old wild type and transgenic plants were transferred onto soil and grown in a growth chamber at 16 h light/8 h dark (100 µm/m2/s) and 22ºC. These plants were harvested for lipid and metabolic assays (6-week-old). Starch analysis was performed using a Megazyme kit according to the manufacturer's instructions (Megazyme, Wicklow, Ireland). For insect assays, transgenic and wild type plants were grown in a short-day growth chamber at 12 h light/12 h dark (100 µm/m2/s) at 22ºC. The binary vector p35S-DGTT2-nos was introduced into Agrobacterium strain GV3101 by electroporation. Arabidopsis was transformed using the flower-dip method (Clough, S. J. and Bent, A. F. 1998). Transgenic plants (T1) were selected on ½ MS agar plates containing 1% sucrose and 100 mg/L kanamycin. Wild type and homozygous

transgenic seedlings were grown on the same shelf of the growth chamber when used for lipid, metabolite, and insect assays.

Lipid isolation and quantification

Soil grown 6-week-old and 15-day-old (grown on ½ MS agar plates containing 1% sucrose) wild-type and transgenic Arabidopsis plants were freeze-dried. Neutral lipids were extracted from dried samples using chloroform:methanol (1:1, v/v) with 100 μM internal standard tri15:0 TAG and separated on silica plate using a mixture of solvents consisting of petroleum ether:ethyl ether:acetic acid (80:20:1, by volume). TAG bands were isolated from the TLC plate after thin layer chromatography (TLC), dissolved in toluene with 10 μM tri13:0 TAG internal standard and assayed using ESI-MS as previously described (Durrett, T. P. et al. 2010). Distinct lipid and TAG bands were scraped from the TLC plates and used to prepare fatty acid methyl esters (FAMEs) by acid-catalyzed transmethylation. Identification and quantification of FAMEs were performed as previously described (Xu, C. et al. 2003). The amounts of lipids were calculated based on the content of fatty acids derived from GC using C15:0 as an internal standard.

To quantify the amount of TAG accumulating in yeast expressing the DGTT constructs, neutral lipids were extracted from the yeast pellets and submitted for ESI-MS, following the method described by Durrett, T. P. et al. (2010).

Microscopy

For oil droplet visualization and TEM, the leaf samples from 6-week-old soil grown transgenic and wild-type plants were used. Whole leaf samples for TEM were fixed in a mixture of 2.5% Glutaraldehyde and 2.5% Paraformaldehyde in 0.1 M cacodylate buffer at 40C for 24 hours, postfixed in 1% osmium tetroxide and dehydrated in a graded acetone series. Samples

were infiltrated and embedded in Spurr resin (Polysciences, http://www.polysciences.com). Thin sections (70 nm thickness) were obtained with a PTXL ultramicrotome (RMC, Boeckeler Instruments, Tucson, AZ) on 200 mesh copper grids stained with uranyl acetate and lead citrate. Sections were imaged using a JEOL 100CX Transmission Electron Microscope (Japan) at a 100kV accelerating voltage. Freshly harvested leaf samples were used for oil droplet visualization by confocal microscopy as previously described (Sanjaya et al. 2011).

Overexpression of DGTTs in Chlamydomonas

For over-expression, *DGTT2* and *DGTT3* were cloned with HA tags (Table 2-1), digested with restriction enzymes *Eco*RI and *Spe*I, and inserted into modified pJR38 vector (Neupert, J. et al. 2009), which uses the *PSAD* promoter and terminator (Fischer, N. and Rochaix, J. D. 2001). The constructs were transformed into *Chlamydomonas* using the glass bead method (Harris, E. H. 1989) and selected on TAP media containing 10 μg/ml paromomycin. Transformants were picked and grown in 10 ml liquid culture. The cells were pelleted for future experiments.

To test whether the transgenes were being expressed, total RNA was extracted using TRIzol reagent, following the method recommended by the manufacturers. cDNA was synthesized using Invitrogen (http://www.invitrogen.com) SuperScript III and oligo-dT primer. The transgenes were amplified using the forward cloning primers, which hybridized to the *DGTT* coding sequence, and the *PSAD* terminator sequence, to distinguish between the native version of the DGTTs and the overexpressed version (Table 2-1). For Western blotting, the *Chlamydomonas* pellets were resuspended in SDS buffer and boiled for 10 minutes. The samples were loaded on a 12% SDS-PAGE gel and separated. The proteins were transferred to membrane, and then blotted with anti-HA antibody to detect the transgene.

For lipid analysis, the *Chlamydomonas* pellets were extracted with 1:1 chloroform:methanol. The lipid samples were separated on a TLC plate using 80:20:1 petroleum ether:ether:acetic acid, and the TAG bands were scraped off and quantified via FAME and GC-FID, along with an equal amount of total fatty acids. The ratio of TAG-derived fatty acids to total fatty acids was used to quantify the amount of TAG produced.

amiRNA knockdowns

amiRNA knockdown was performed using vectors created by Molnar, A. et al. (2009). Constructs targeting *DGTT2* and *DGTT3* separately were designed using WMD 2 (http://wmd2.weigelworld.org/cgi-bin/mirnatools.pl) (Ossowski, S. et al. 2008). Oligonucleotides were synthesized by MSU's Research Technology Support Facility and cloned into pChlamiRNA3int, an amiRNA expression vector (Table 2-1). The constructs were transformed into *Chlamydomonas*, selecting with paromomycin. The transformants were grown in TAP-N, to induce TAG synthesis. Total RNA was extracted from the transformants, and used as a template for QRT-PCR to quantify the level of the *DGTT* mRNAs. *RACK1* was used to normalize RNA levels.

Results

Multiple putative Chlamydomonas DGAT2 isoforms

Sequence similarity to Arabidopsis DGAT2 was used to identify potential DGAT orthologs, using BLAST against Version 3 of the *Chlamydomonas* genome. The BLAST search results returned five gene models with high sequence similarity to the Arabidopsis type-2 DGAT. These five candidates were named Diacylglycerol Acyltransferase Type Two (DGTT) 1-5 and compared to type-2 DGATs from *Arabidopsis thaliana*, *Ostreococcus tauri*, yeast (*Saccharomyces cerevisiae)*, castor bean (*Ricinus communis)*, and tung tree (*Vernicia fordii)*.

When analyzed with MEGA5 (Tamura, K. et al. 2011), DGTT1 was the most closely related to the ScDGA1. *DGTT2*, *DGTT3* and *DGTT5* formed their own clade, as did the three land plant DGATs. *OtDGAT2B* and *DGTT4* were both more similar to the land plant DGATs than to the other *Chlamydomonas* DGTTs (Figure 2-1A).

Structural analysis of the gene model-translated protein sequences was performed in silico using the TMHMM Server 2.0 (http://www.cbs.dtu.dk/services/TMHMM/) to predict transmembrane sequences. All five candidates had 1-3 transmembrane domains in the N-terminal half of the protein, consistent with other type-2 DGATs. SignalP V3.0 (http://www.cbs.dtu.dk/services/SignalP/) and TargetP V1.1 (http://www.cbs.dtu.dk/services/TargetP/) failed to identify potential cellular localization signals. Searching against the NCBI Conserved Domain Database revealed a DAGAT enzymatic domain in the C-terminal half of all five candidates, although *DGTT5* has an apparent insertion in its domain (Figure 2-1B). When the data from my transcript profiling experiment was analyzed, all five genes were detected, with expression patterns as shown in Figure 2-2.

Expression of DGTT constructs in yeast

To test the function of the predicted DGTT genes, the coding sequences of *DGTT2-5* were isolated and expressed in yeast strain H1266, a triple knockout mutant for DGA1, LRO1, and ARE2 (Sandager, L. et al. 2002). This strain has very little native DGAT activity, providing a suitable background for testing the *Chlamydomonas* DGAT2 candidates. I was unable to isolate a *DGTT1* cDNA based on the available gene model. I confirmed the presence of the DGTT2-5 proteins in yeast by immunoblotting (see Figure 2-3). To quantify the amount of TAG produced, electrospray ionization mass spectrometry (ESI-MS) was performed on lipid extracts from the transgenic yeast (Figure 2-1C). Both DGTT2- and DGTT3-producing yeast accumulated

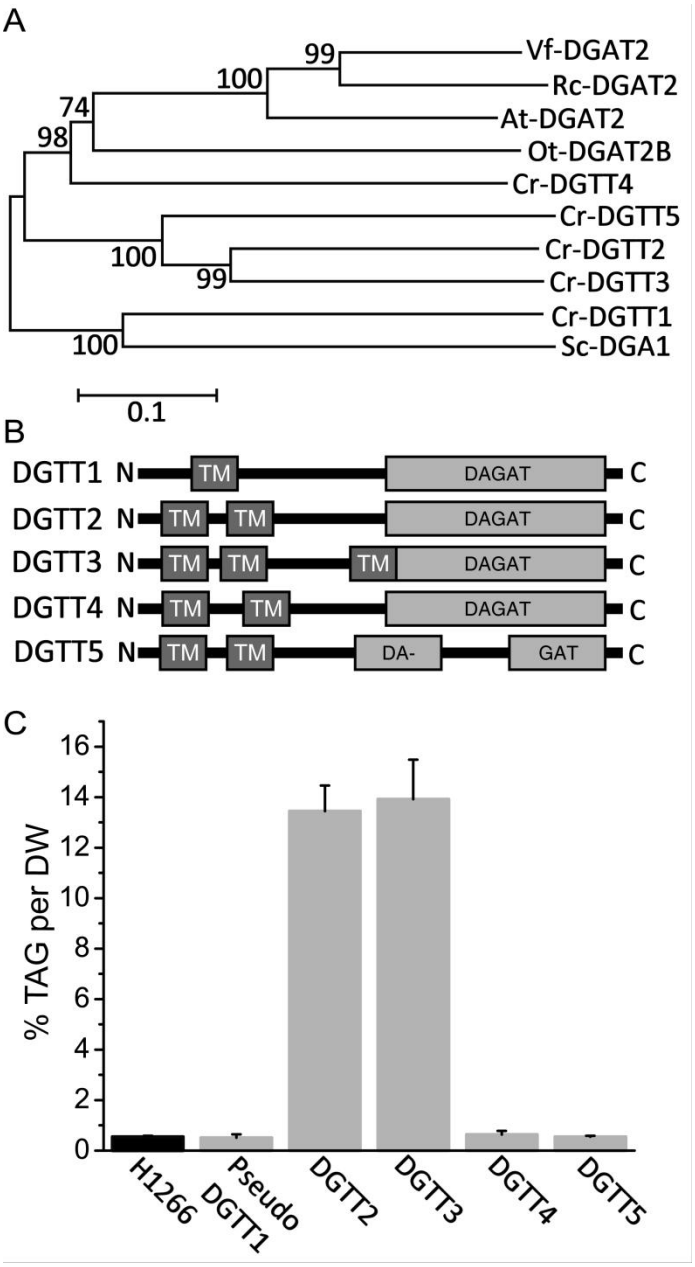


Figure 2-1. Identification of *C. reinhardtii* Type 2 DGATs and Expression of DGTT Constructs in Yeast. (A) Phylogenetic tree of *C. reinhardtii DGTT1-5* compared with DGAT2s from other species: *Arabidopsis* (NP_566952.1), *O. tauri* (XP_003083539.1), *S. cerevisiae* (NP_014888.1), castor bean (DQ923084.1), and tung tree (DQ356682.1). **(B)** Predicted structure of DGTT1-5. Dark-gray boxes indicate predicted transmembrane domains (TM), and light-gray boxes indicate the conserved enzymatic domain (DAGAT). **(C)** ESI-MS quantification of TAG levels extracted from transformed yeast. The total amount of TAG was normalized based on the dry weight of the yeast. Tritridecanoin (tri13:0) and tripentadecanoin (tri15:0) TAGs were added as internal standards (n = 4; average ± sd). H1266 indicates the empty vector control. Pseudo DGTT1 refers to an inactive clone of DGTT1 based on a faulty gene model.

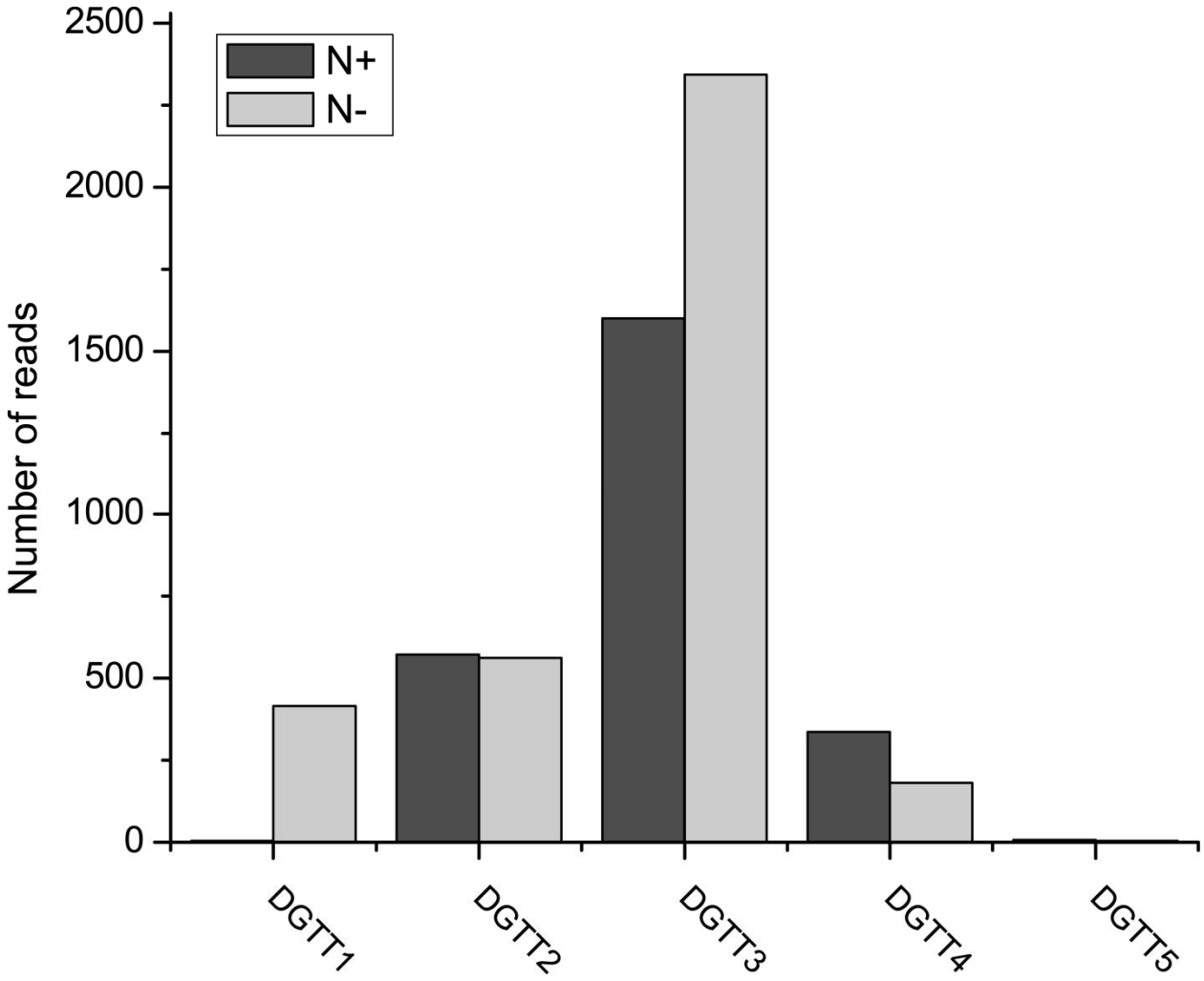


Figure 2-2. Expression levels of *DGTT1-5*. The transcript levels of *Chlamydomonas* cultures grown under N-replete and N-deprived conditions were compared using high-throughput sequencing. The total number of reads assigned to each DGTT gene are graphed. Data taken from the Illumina sequencing experiment described in (Miller, R. et al. 2010).

Figure 2-3. Confirmation of DGTT2-5 protein expression in yeast by immunoblotting analysis. The DGTT proteins run level with the 26 kDA marker, as expected. The * indicates a non-specific yeast protein found in all samples.

considerable levels of TAG (13.4 and 13.9% per dry weight (DW), respectively) compared to the empty vector control (0.5% per DW). Both DGTT4- and DGTT5-producing strains showed TAG levels equivalent to the empty vector (0.6 and 0.5% per DW, respectively), indicating that these may not have DGAT activity, or lack appropriate co-factors or substrates in yeast necessary for proper DGAT activity.

In vitro DGAT activity of recombinant proteins

The substrate preference of the DGTT proteins was estimated in an in vitro reaction with isolated microsomes. DGTT-containing microsomes were incubated with radiolabeled palmitoyl-CoA (16:0) and oleoyl-CoA (18:1), along with dioleoyl DAG. The level of DGAT activity was estimated from their incorporation of radiolabel into isolated TAGs. Comparison of the results using the two different substrate acyl-CoAs revealed an apparent difference in substrate specificity, with DGTT2 and DGTT3 preferring 16:0-CoA and DGTT4 favoring 18:1-CoA (Figure 2-4A and 2-4B). These assays also confirmed that DGTT5 has essentially no activity under these assay conditions. This result, in conjunction with the apparent lack of expression of its gene in *Chlamydomonas* and disruption in the DAGAT domain, suggests that *DGTT5* is a pseudogene.

Although *DGTT2* and *DGTT3* belong to the same clade, following N-deprivation the level of *DGTT2* transcripts remained unchanged, whereas *DGTT3* transcript levels increased (Boyle, N. R. et al. 2012, Miller, R. et al. 2010). When expressed in yeast, DGTT2 activity in vitro was higher than that of the other isoforms. Based on the observed activity levels, we focused on the characterization of this enzyme as a potential tool for the manipulation of cytosolic acyl-CoAs in vegetative tissues of plants, the primary goal of this study. A microsome-based competition assay was used to further probe the substrate specificity of DGTT2. When

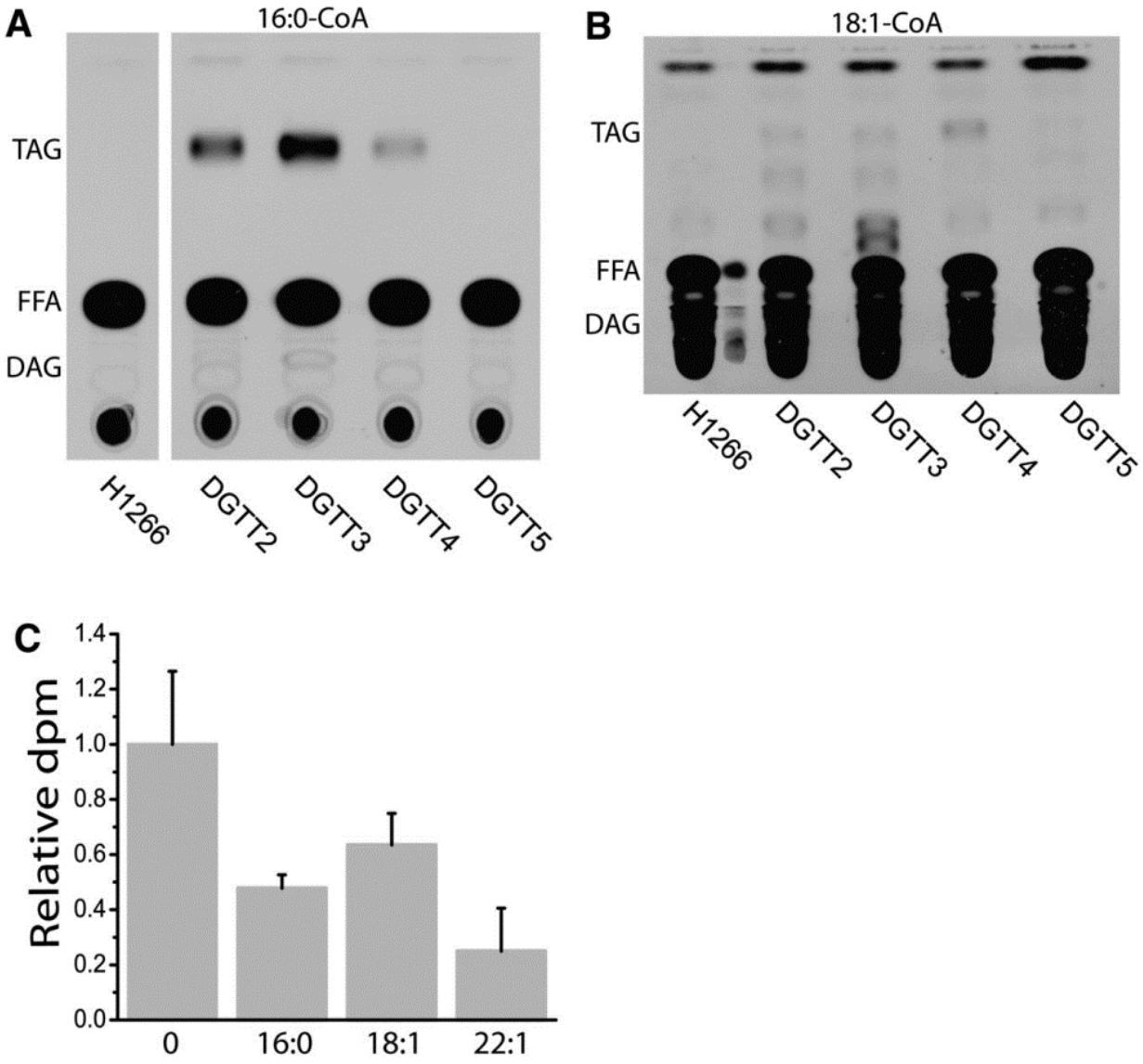


Figure 2-4. DGTT Activity in Transgenic Yeast Microsomes. (A) and **(B)** Autoradiograph of [1-^{14}C]-16:0-acyl-CoA and [1-^{14}C]-18:0-acyl-CoA radiolabeled lipids separated on a TLC plate. Signals representing TAG, free fatty acids (FFA), and diacylglycerol (DAG) are indicated. H1266 indicates the empty vector control. **(C)** Competition assay. For each reaction, 1.725 nmol of [1-^{14}C]-16:0-acyl-CoA was added. Competitors consisting of 1.725 nmol of unlabeled 16:0-acyl-CoA, 18:1-acyl-CoA, or 22:1-acyl-CoA was added to the respective reactions and compared with a no-competitor control (0). The newly synthesized ^{14}C-labeled TAGs were separated and quantified by scintillation counting. The results were normalized by subtracting the background (in dpm) and then dividing all the results by the no-competitor control (n = 3; average ± sd).

incubated with equimolar amounts of ^{14}C-labeled and unlabeled 16:0-CoA, the radioactivity in the TAG band was reduced to ~50% compared to the reactions with only ^{14}C-labeled 16:0-CoA (Figure 2-4C). When incubated with equal moles of unlabeled 18:1-CoA, the decrease was only ~40%. However, when incubated with equal moles of unlabeled 22:1-CoA, the decrease was ~75%. Together, these data suggested that DGTT2 is able to incorporate varying acyl-CoA species into TAG, with an apparent preference for very long chain acyl groups, which will become relevant for the interpretation of lipid data for the transgenic plants described below.

Production of DGTT2 in Arabidopsis affects seedling growth

To explore DGTT2 as a tool to manipulate acyl-CoA pools in plants and to engineer TAGs in vegetative tissues, we expressed the full length *Chlamydomonas DGTT2* coding sequence in Arabidopsis under the control of the constitutive 35S cauliflower mosaic virus (CaMV) promoter, 35S:DGTT2. Five homozygous lines, 14, 22, 46, 52 and 57, were carried forward to the T4 generation and used for detailed analysis.

The hypocotyls of transgenic 10-to-12-day-old seedlings grown on MS agar plates supplemented with 1% sucrose were more elongated than those of control plants, and slightly pale in color (Figure 2-5A, B). Consistent with the pale green phenotype, the total chlorophyll content in the transgenic lines was about 30% lower (Figure 2-5C). On soil, all transgenic lines followed wild-type growth and development patterns. To determine the abundance of *DGTT2* mRNA in the seedlings of overexpressors (15-day-old), we used q-RT-PCR. The expression of *DGTT2* relative to *ACTIN*2 ranged from 2-to-7 (ratio of *DGTT2/ACTIN2*) in the independent transgenic lines tested (Figure 2-5D), while no transcripts were detected in the wild type.

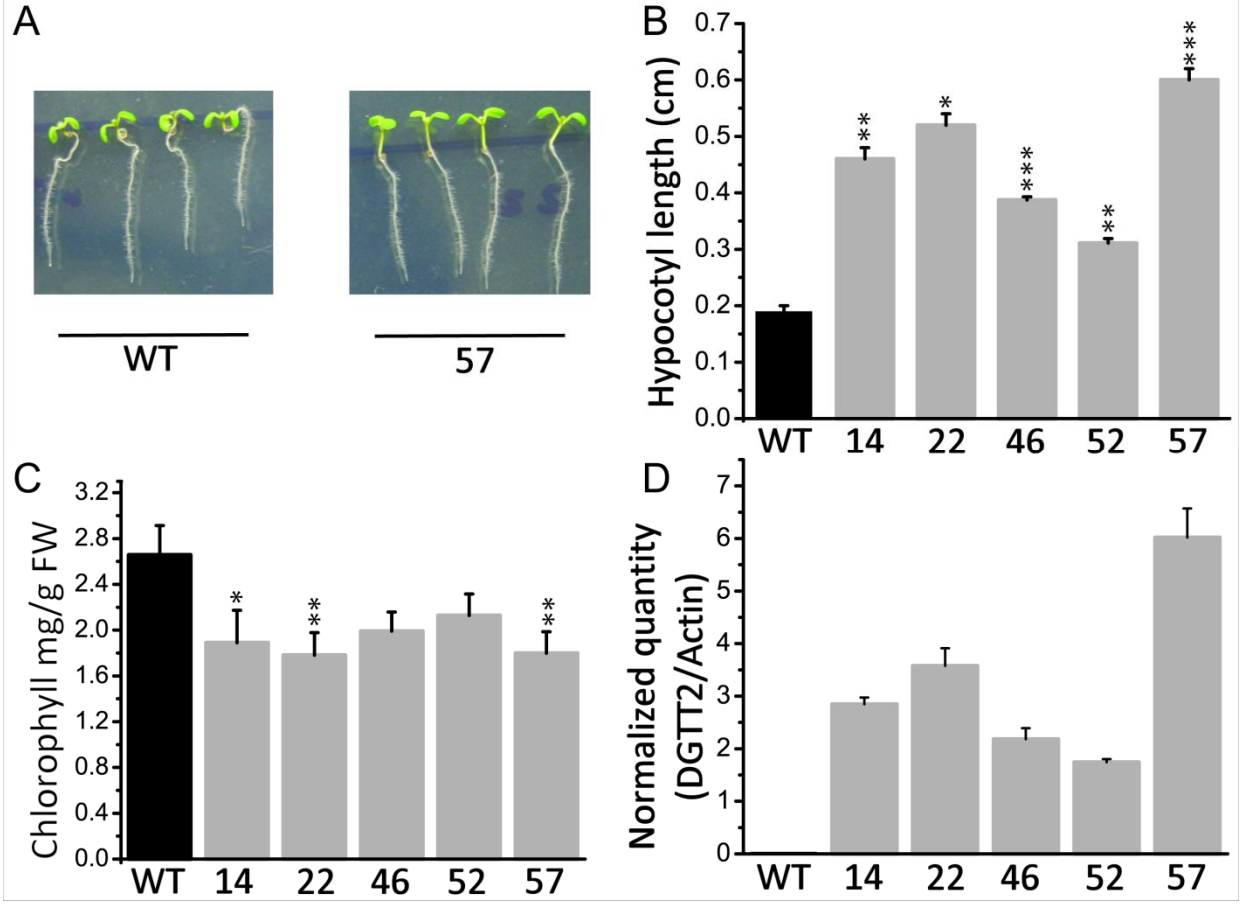


Figure 2-5. Overexpression of *DGTT2* affects the seedling phenotype. (A and B) Growth habit and hypocotyl length of 10-12-day-old seedlings of wild type (WT) and *DGTT2* transgenic lines (14, 22, 46, 52, 57) grown on ½ MS medium supplemented with 1% sucrose. **(C)** Chlorophyll content of WT and *DGTT2* transgenic lines (14, 22, 46, 52, 57) grown on ½ MS medium supplemented with 1% sucrose. Chlorophyll was extracted from 15-day-old seedlings with 80% acetone, and its concentration determined spectrophotometrically, **t*-test significant at $p < 0.05$ or ** $p < 0.01$ versus wild type (n=3; average ± SD). **(D)** Quantification of *DGTT2* mRNA by qRT-PCR in 15-day-old seedlings of WT and *DGTT2* transgenic lines (14, 22, 46, 52, 57) grown on ½ MS medium supplemented with 1% sucrose (n=3; average ± SD). For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

DGTT2 production causes accumulation of TAGs with VLCFAs

To determine the effect of DGTT2 production on the accumulation of TAG levels, we analyzed 15-day-old whole seedlings by ESI-MS. The TAG levels in 15-day seedlings of transgenic lines were increased up to 20-fold (Figure 2-6A). Lines 22 and 57 contained 6.3 and 7.6% TAG per dry weight (DW), respectively, compared to 0.32% TAG per DW in the wild type. The increase in TAG levels correlated well with the relative abundance of *DGTT2* transcript in each line (Figure 2-5D). Comparison of ESI-MS spectra of neutral lipid extracts from the wild type and line 57 indicated an abundance of TAG molecular species containing very long chain fatty acids (VLCFAs) in the latter (Figure 2-6B, C). The presence of long and very long chain fatty acid-containing TAG molecular species in the seedling extracts of line 57 was further confirmed by ESI-MS/MS (Figure 2-6D), which produced product ions with masses consistent with the loss of VLCFAs. These spectra were consistent with each TAG molecule containing only one VLCFA.

We also used ESI-MS to analyze the TAG content in the leaves of 6-week-old soil-grown plants before bolting. As shown in Figure 2-6A, TAG levels in transgenic lines 22 and 57 were 1.2 and 2.1% of DW, a 6- and 10-fold increase relative to wild-type plants (0.2% TAG per DW), respectively. These increases were considerably lower compared to those observed with young seedlings and might reflect turnover of TAG in more mature leaves. The ESI-MS spectrum of neutral lipid extracts from 6-week-old leaves of line 57 confirmed the presence of TAG molecular species containing VLCFAs, which were not detected in the wild type. The presence of long and VLCFAs in neutral lipid extracts of soil-grown line 57 was also confirmed by ESI-MS/MS. Interestingly, the level of TAGs with VLCFAs (e.g., C24:0) in leaves of soil-grown plants of line 57 was considerably higher (~8-10 mol%) than wild type (~2 mol%) (Figure 2-7).

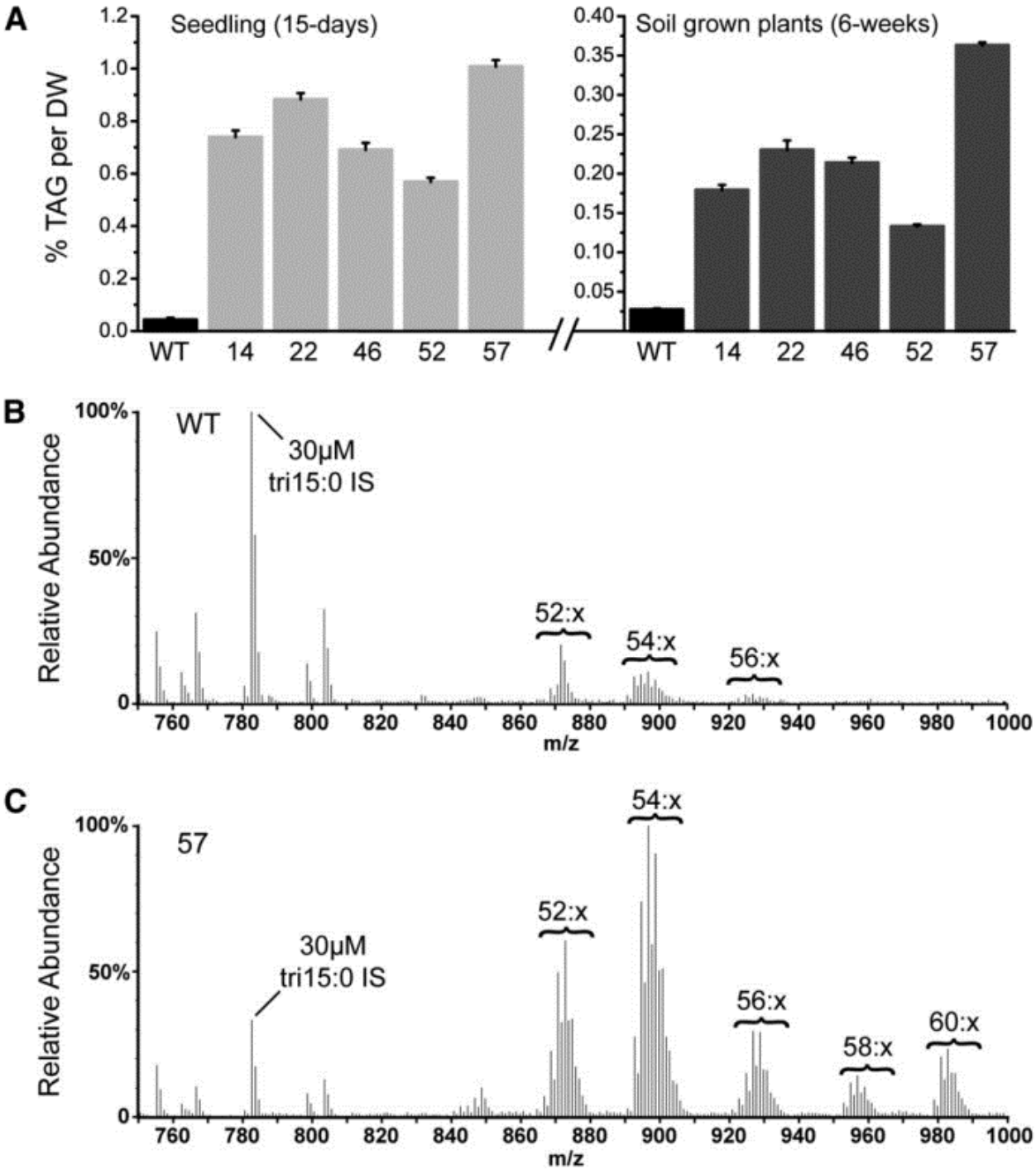
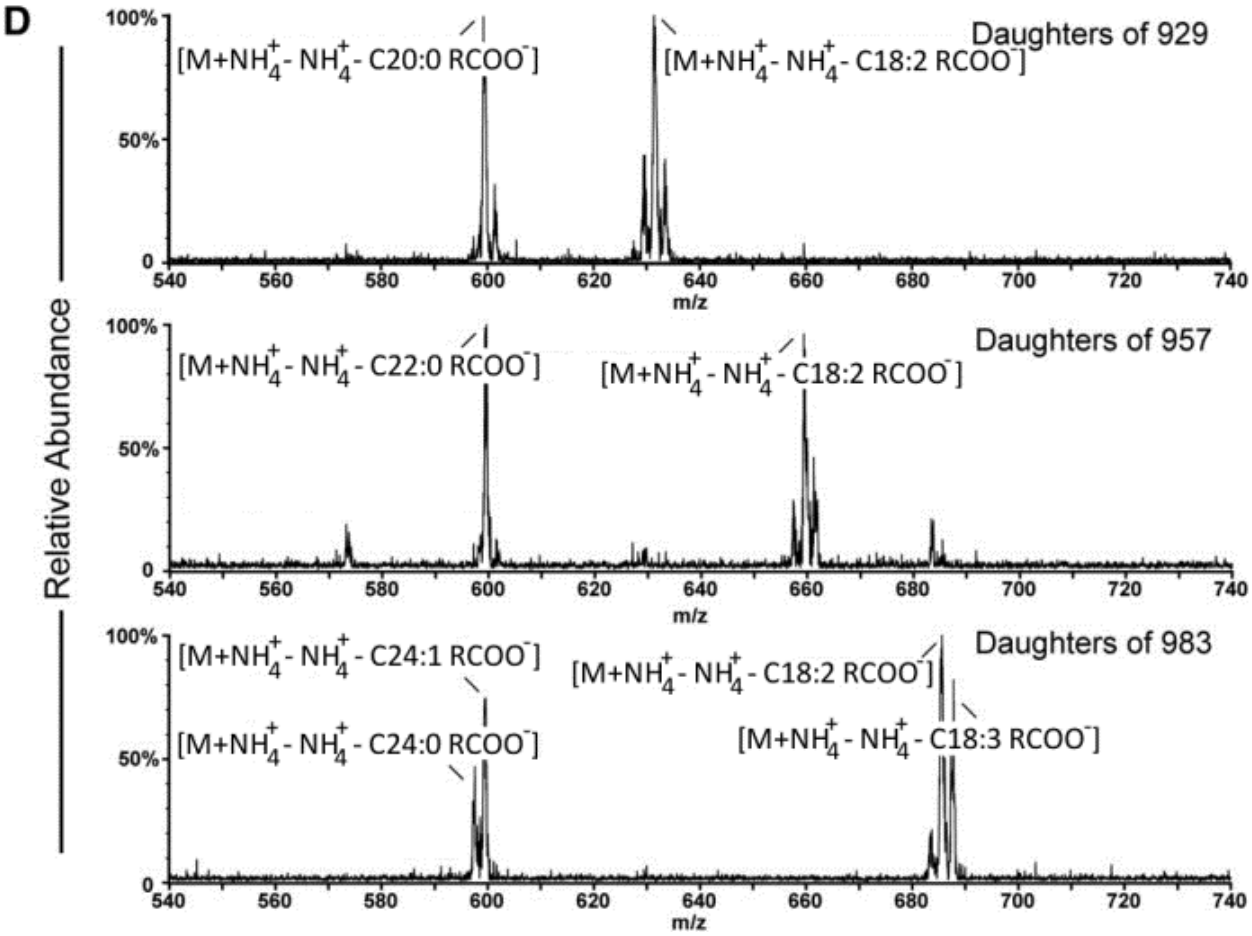


Figure 2-6. DGTT2 Leads to the Accumulation of TAG with VLCFAs in *Arabidopsis* Seedlings (15 d Old) and Soil-Grown Plants (6 Weeks Old).

Figure 2-6 (cont'd)



(A) ESI-MS quantification of TAGs in neutral lipid extracts of seedlings and soil-grown plants of wild-type (WT) and homozygous transgenic plants producing DGTT2 (n = 4; average ± sd). **(B)** and **(C)** Positive-ion electrospray ionization mass spectra of neutral lipid extracts from seedlings of the wild type and transgenic line 57. Tritridecanoin (tri13:0) and tripentadecanoin (tri15:0) TAGs were added as internal standards. **(D)** ESI-MS/MS analysis of neutral lipid extracts of line 57. Shown are the daughter fragment ions from TAGs with $[M + NH_4]^+$ adducts with mass-to-charge ratio values of 929, 957, and 983. The mass-to-charge values were rounded up to the nearest nominal mass.

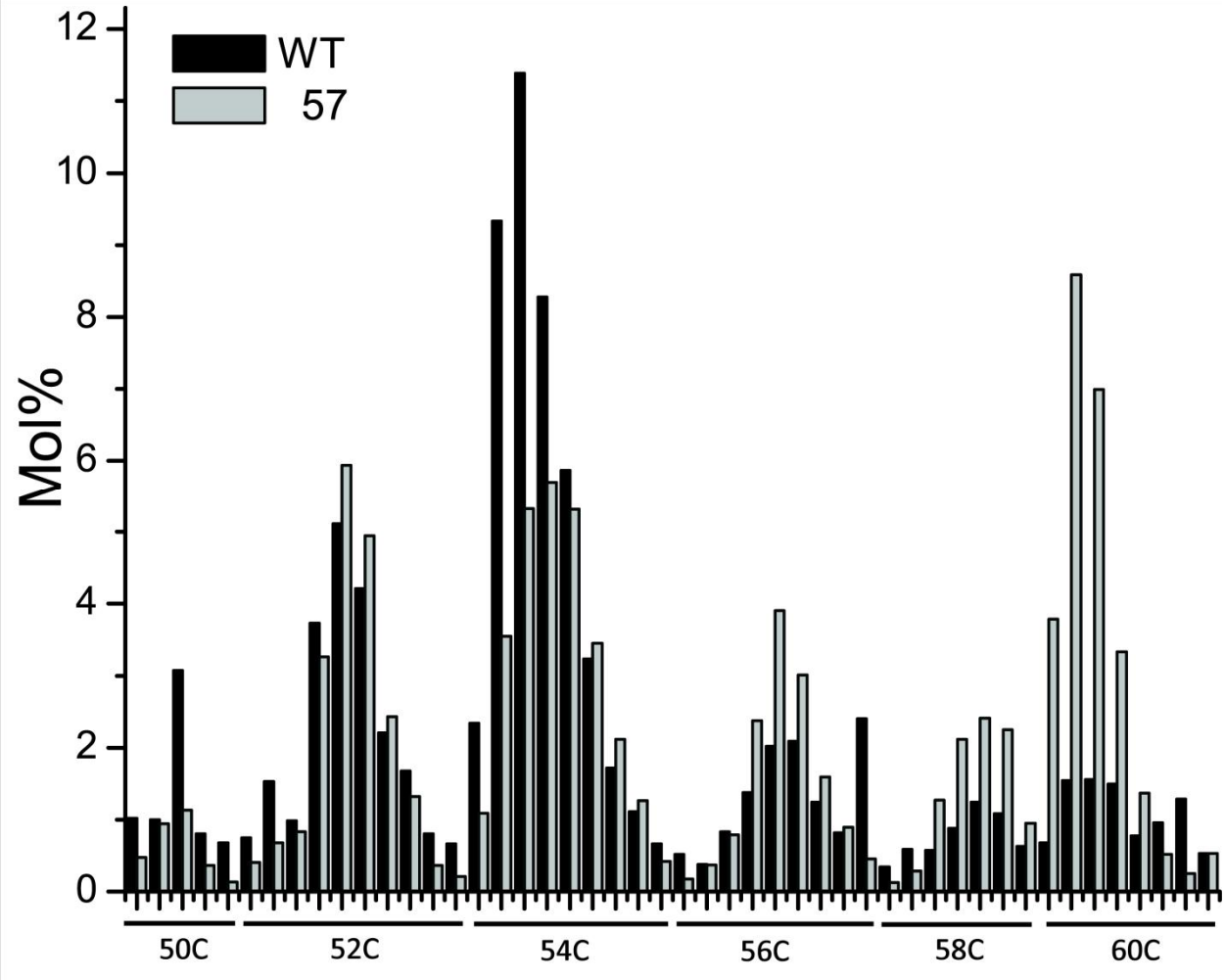


Figure 2-7. Confirmation of long and very long chain fatty acids in 6-week-old soil grown *DGTT2* line 57 and wild-type plants by ESI-MS/MS. Legends at the bottom indicate the total number of carbons in the TAG species.

Using an independent method, we also confirmed the levels of TAG in transgenic soil-grown plants by transmethylation of TLC-separated TAG and GC-FID analysis of the resulting fatty acid methyl esters (Figure 2-8).

An abundance of oil droplets in the leaves of transgenic lines

Leaves from 6-week-old TAG-accumulating line 57 and of wild-type were compared by confocal microscopy following Nile Red staining and by transmission electron microscopy (TEM). Oil droplets were abundant in line 57 and were distributed in the proximity of the chloroplasts (Figure 2-9A, B, C). In contrast, few or no oil droplets were observed in the wild-type leaf sample (Figure 2-9D, E, F). We used TEM to analyze the location of oil droplets in leaf sections. In line 57, large and distinct electron dense oil droplets were observed outside the chloroplast in mesophyll cells, most likely associated with the ER (Figure 2-9H). No oil droplets were observed in cells of wild-type sections (Figure 2-9G). The shape of chloroplasts and distribution of starch granules were similar in both wild-type and transgenic lines.

Overexpression of DGTTs in Chlamydomonas

50 lines of each construct were tested, along with several empty vector controls. When screened by RT-PCR, ~50% of the selected transformants showed bands corresponding to the transgene. Western blot analysis of the positive transformants gave negative results, however. Lipid analysis of the positive transformants showed no significant changes in either TAG or total fatty acid levels, compared to wild type *Chlamydomonas*.

amiRNA knockdowns in Chlamydomonas

50 *DGTT2* and 50 *DGTT3* transformants were screened by Q-RT-PCR. Some of the transformants showed differences in mRNA level compared to wild type, but these differences

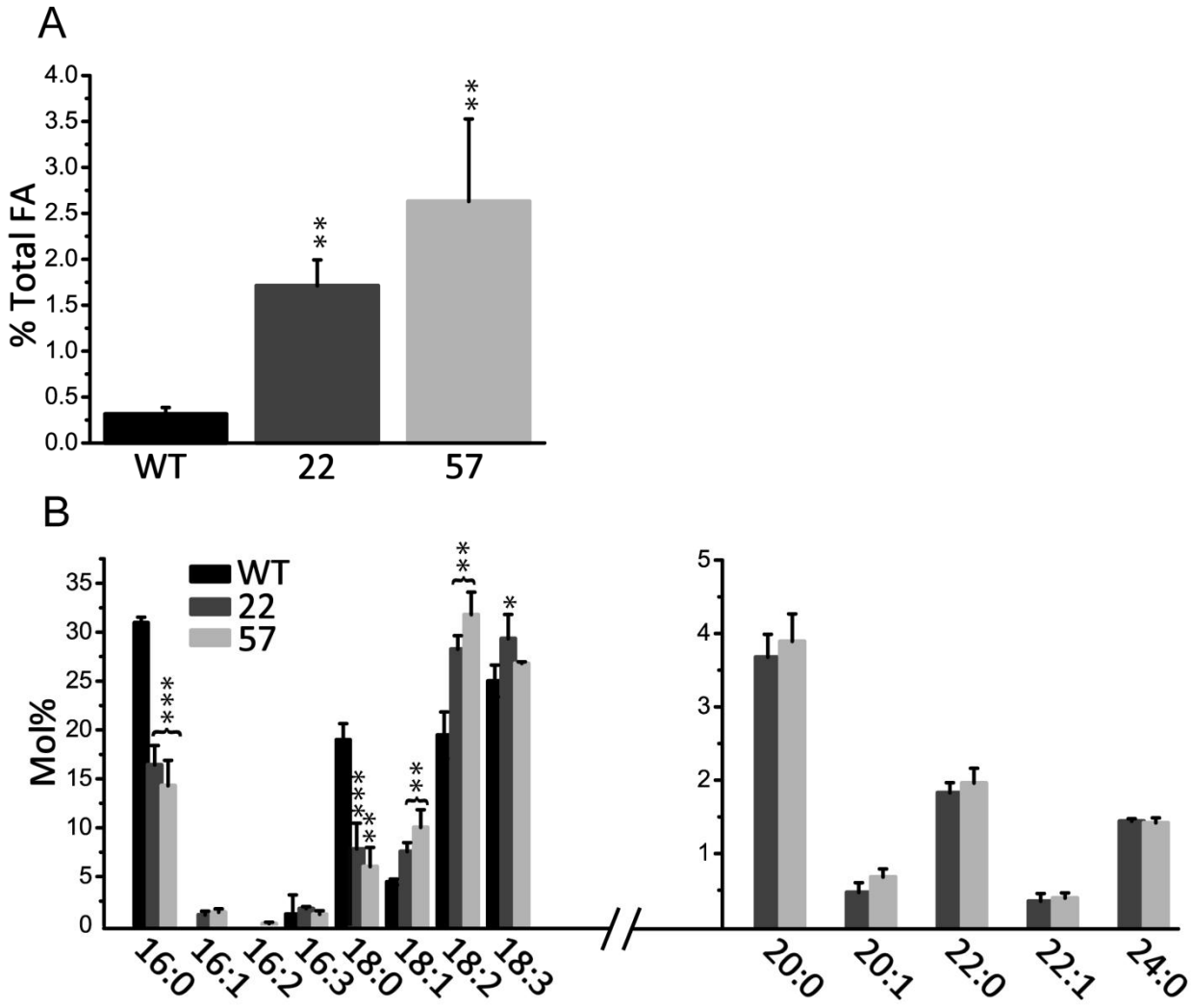


Figure 2-8. Analysis of TAG in soil grown 6-week-old leaves of wild type and homozygous transgenic plants expressing DGTT2 by GC-FID. (A) TAG content in leaves relative to total % of fatty acids in wild type and *DGTT2* transgenic lines 22 and 57. Shading of the bars as defined in **(B). (B)** Fatty acid composition of TAG in the wild type and the *DGTT2* transgenic lines 22 and 57. **t*-test significant at $p < 0.05$ or ** $p < 0.01$ or *** $p < 0.001$ versus wild type (n=3; average ± SD).

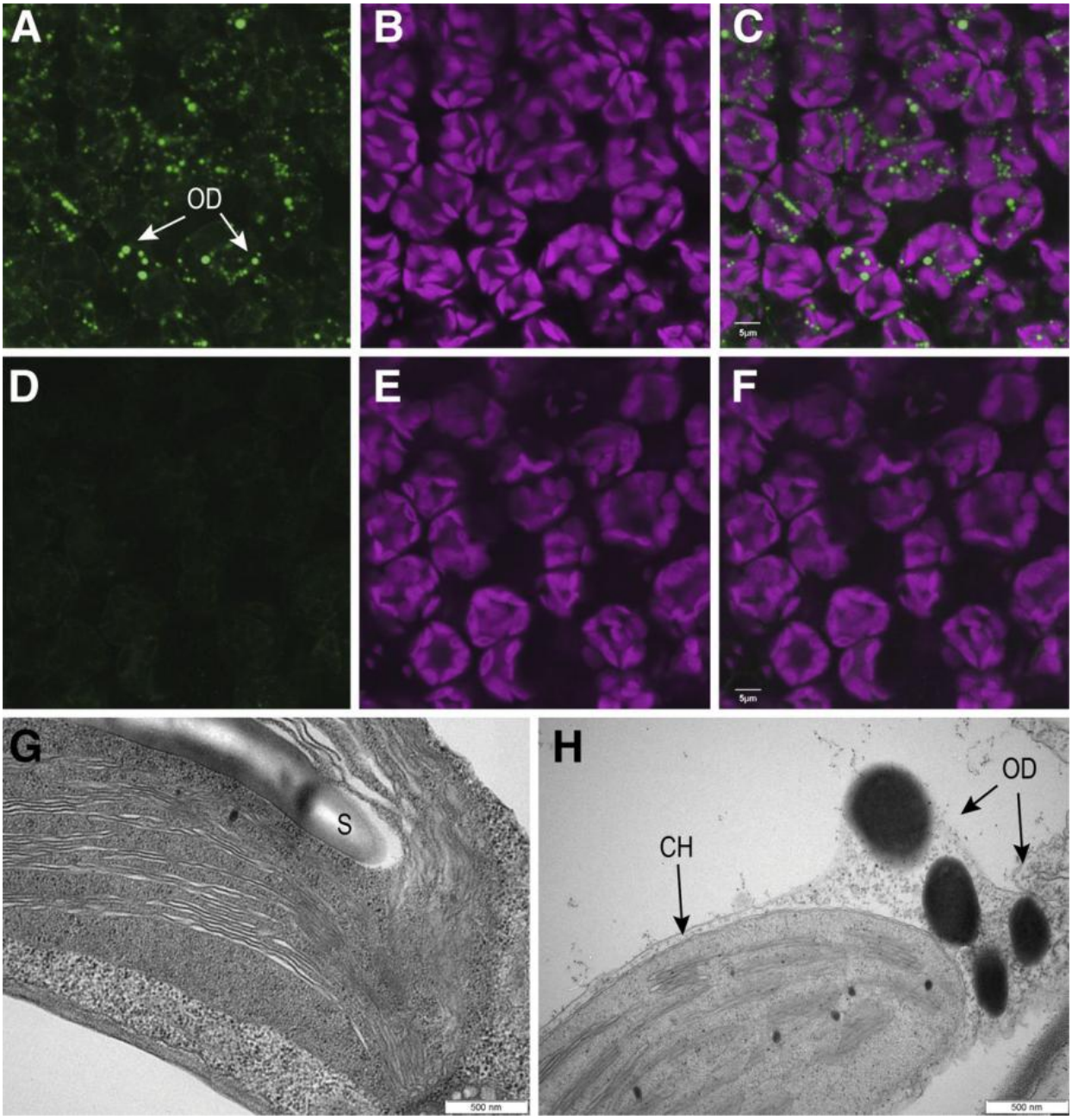


Figure 2-9. Oil Droplets Are Abundant in Leaves of *DGTT2* Transgenic Line 57. (A) to **(C)** Confocal fluorescence image of leaf mesophyll cells of line 57, showing chloroplasts (red) and oil droplets (OD; arrows, green) stained with Nile red. Bar = 5 µm. **(D)** to **(F)** Confocal fluorescence image of leaf samples from the wild type of the same age as line 57 (6 weeks old, soil-grown). Bar = 5 µm. **(G)** and **(H)** TEM analysis of the first leaf pair from 6-week-old wild type **(G)** and transgenic line 57 **(H)**. Oil droplets (OD), chloroplasts (CH), and starch granules (S) are indicated. Bars = 500 µm.

were not consistent when retested. Lipid analysis showed no significant differences in TAG or total fatty acid levels between the transformants and wild types.

Discussion

DGATs are important for TAG synthesis. The goal of this project was to identify DGATs in the model alga *Chlamydomonas*, and characterize their activity, in order to advance our understanding of TAG synthesis in microalgae, and illuminate how it differs from higher plants. Two main types of DGATs have been characterized in other species. When the Arabidopsis DGAT protein sequences were used to search with BLAST, multiple type 2 DGATs (*DGTT1-5*) were identified in *Chlamydomonas*, whereas no type 1 DGAT was identified at the time. When examined structurally, the DGTT genes all showed conserved type 2 DGAT features, including 1-3 predicted transmembrane domains and the DAGAT enzymatic domain. Although no targeting or signaling sequences were predicted, this may be due to the fact that the prediction programs were trained on land plants, and may not be able to detect all *Chlamydomonas* signals. The most likely location for the DGTTs is the ER, which is the main site of TAG synthesis in organisms, but it is possible that some are targeted to the chloroplast in addition to or instead of the ER. There have been reports of DGAT activity in the chloroplast from land plant species, and recent reports indicate that *Chlamydomonas* may have a distinct chloroplast-based TAG synthesis pathway, in addition to the ER-based pathway (Fan, J. et al. 2011, Goodson, C. et al. 2011).

The finding of multiple type 2 DGATs matches the results from other green algae species, and contrasts with most higher organisms. Although a few species of plants, such as soybean, have multiple type 2 DGATs, most plants have one copy of each of the two types, as do mice and humans. On the other hand, yeast have only a single type 2 DGAT. Structurally, the

two types of DGATs differ significantly, but they appear to function similarly. How multiple DGATs interact to synthesize TAG is still unknown. It is possible that the different enzymes have different substrate specificities, and thus help regulate the TAG species created. There is evidence from castor beans that its type 2 DGAT is specifically responsible for the incorporation of the unusual fatty acid ricinoleic acid into TAG, while the type 1 DGAT generates primarily standard TAG species. The *Chlamydomonas* DGTTs may also discriminate among fatty acids, although to date only DGTT2 has been tested for specificity.

My transcript profiling experiment identified all five DGTT genes in the Illumina dataset. *DGTT1* showed the greatest amount of regulation, increasing ~100-fold under nitrogen deprivation. *DGTT2* and *DGTT3* were the most highly expressed overall, with *DGTT3* showing a less than 2-fold increase under N-deprivation. *DGTT4* showed a less than 2-fold decrease, and *DGTT5* showed very little expression under either condition. The different regulation patterns suggest that the multiple genes play different roles in TAG synthesis, and may be active under different conditions.

To confirm that the identified genes were in fact active DGATs, their activity needed to be tested. There were two ways to test this: heterologous expression in other organisms, and ectopic expression or knock-down of expression in *Chlamydomonas*. The first method was done using yeast to test all five candidates. Yeast was chosen over *E. coli* for expression because all five DGTTs had predicted transmembrane domains, and the internal membranes in yeast would better mimic those found in *Chlamydomonas*. In addition, the availability of knock-out lines with essentially no background TAG activity simplified the analysis of the *Chlamydomonas* DGTTs.

In my initial efforts, I was unable to successfully clone the *DGTT1* cDNA. This is due at least partially to a poor gene model that was available in the Version 3 genome, which was used

for primer design. Later PCR attempts by others based on an updated gene model were successful, and subsequent cloning showed that *DGTT1* had the ability to complement a DGAT-deficient yeast strain (Boyle, N. R. et al. 2012). The other four DGTTs were readily cloned, with the full-length cDNA being inserted into a yeast vector. Expression in yeast showed clear differences between the genes, with DGTT2- and DGTT3-expressing yeast able to accumulate significant amounts of TAG, while DGTT4- and DGTT5-expressing yeast were not. The reason for the differences may be due to differences in overall activity of the genes, or it could be due to the differences between yeast and *Chlamydomonas*. For example, yeast contains only saturated and mono-unsaturated fatty acids, while *Chlamydomonas* contains poly-unsaturated FAs as well. Alternatively, post-transcriptional or post-translational changes may be needed that aren't available in yeast, although this is less likely, given that DGTT4 shows activity in the *in vitro* assays. Another potential issue with this assay is that it measures TAG accumulation, rather than TAG synthesis. It is possible that the TAG generated by the expression of DGTT4 and DGTT5 is being degraded before it can be detected, due to the FA make-up of the synthesized TAG or to regulatory differences not seen in DGTT2- and DGTT3-expressing yeast.

To test the actual synthesis of TAG by yeast expressing DGTTs, I conducted microsome-based assays. Isolating the DGTT-containing microsomes and using radiolabeled fatty acyl-CoAs allows for a more precise measurement of DGAT activity in the transformed yeast. Because the protein isolation was done crudely, there are still many proteins left in addition to the DGTTs, as well as various lipids, but comparison to the empty vector control should show the effect of the DGTT protein.

The results from the microsome assay confirmed the initial results for DGTT2 and DGTT3. Yeast expressing both proteins were able to incorporate both C16:0 and C18:1 fatty

acyl-CoAs into TAG, further demonstrating their DGAT activity. In contrast to the earlier results, yeast expressing DGTT4 were also able to incorporate the two fatty acyl-CoAs, indicating that it, too, has DGAT activity. DGTT5, on the other hand, did not incorporate the two fatty acyl-CoAs, suggesting that it lacks DGAT activity. It is possible that it requires different substrates, but the lack of detected activity, combined with the lack of gene expression under either nitrogen-replete or nitrogen-deprived conditions (Miller, R. et al. 2010) and the apparent insertion in the conserved enzymatic domain suggest that *DGTT5* is a pseudogene. The results from the assay also show a potential difference in substrate specificity between the three active DGTTs. Both DGTT2 and DGTT3 produced strong radiolabeled TAG bands when fed with 16:0-acyl-CoA, and weak bands when fed with 18:1-acyl-CoA. DGTT4 showed the opposite pattern, with a stronger 18:1-acyl-CoA band and a weaker 16:0-acyl-CoA band. More precise experiments would be needed to confirm this apparent difference in substrate specificity, however.

To further test the activity of the DGTTs, and to determine the potential utility of these genes in future engineering schemes, one of the genes was chosen for transformation into Arabidopsis. Because DGTT2 had the highest activity in yeast, it was chosen and placed under the control of the 35S promoter. This non-specific promoter was chosen to stimulate TAG synthesis in vegetative, instead of primarily in the seed (as is the case with Arabidopsis). My hypothesis was that ectopic expression of a DGAT in the leaves could lead to the accumulation of significant amount of TAG.

Several T4 transgenic lines were examined at both the seedling stage and at 6 weeks on soil, and in both conditions showed significant increases in TAG compared to the wild type. The microscopy confirmed this, with clear lipid droplets visible in the leaves of transgenic plants

under both confocal and electron microscopy. Quantification by mass spectrophotometry further confirmed this, and also revealed the accumulation of TAG species containing very long chain fatty acids, C20 to C24, in the leaves. This was unexpected, since *Chlamydomonas* TAG contains C16 and C18 fatty acids almost exclusively. This finding could indicate that DGTT2 has a preference for these acyl-CoAs, or that DGTT2 has no particular preference and is incorporating all available acyl-CoAs, including those normally used for synthesis of other lipid compounds.

To test these hypotheses, the DGTT2-containing yeast microsomes were used for a competition assay, to try to measure the substrate specificity of the enzyme. The affinity of DGTT2 for 16:0-acyl-CoA was compared to that for 18:1-acyl-CoA and 22:1-acyl-CoA. 18:1 is one of the most abundant fatty acids in *Chlamydomonas* TAG, while 22:1 is one of the VLCFA seen in TAG in the transgenic Arabidopsis lines. Additionally, both 16:0 and 18:1 have been shown to be incorporated by DGTT2-expressing microsomes, but with apparent varying degrees of efficiency, as shown in Figures 2-4A and 2-4B.

Using 16:0 as a baseline, 18:1 was less able to compete, resulting in less displacement of the radiolabeled 16:0-acyl-CoA. This matches the results seen earlier, and strongly suggests a difference in preference between the two fatty acyl-CoAs for DGTT2. In contrast, 22:1 shows a much greater ability to inhibit incorporation of the radiolabeled 16:0, suggesting a preference of DGTT2 for the VLC fatty acyl-CoA. The enzyme preference can explain the results seen in Arabidopsis. However, it is possible that the 22:1-acyl-CoA is acting as an inhibitor of DGAT activity without actually being incorporated, as acyl-CoAs have been shown to do in the past.

The data from both yeast and Arabidopsis is enough to conclude that at least three of the five putative type 2 DGATs (DGTT2, DGTT3 and DGTT4) are indeed DGAT enzymes, while

work done by others has shown similar results for DGTT1. However, the role of these genes in *Chlamydomonas* is still unknown. As mentioned earlier, the multiple DGATs may help the cell regulate TAG species by incorporating different substrates, they may synthesize TAG in different compartments, or some combination. To attempt to address these questions, I tried to assay the function of the DGTT genes within *Chlamydomonas*. This part of the experiment was focused on *DGTT2* and *DGTT3*, as they were the most active in yeast, and the most highly expressed. Two approaches were tried: overexpression and amiRNA knockdown.

About 50% of the *DGTT2* and *DGTT3* overexpression lines gave positive results for the mRNA test. This indicated that the transgenes were being successfully transcribed. However, the Western blot screens were all negative, indicating the absence of the transgene proteins. This suggests that expression of the transgenes is being disrupted post-transcriptionally. Silencing of transgenes is a known problem when attempting expression experiments in *Chlamydomonas*, and may be responsible. Recently, a German group (La Russa, M. et al. 2012) performed similar experiments, attempting to over-express *DGTT1*, *DGTT2* and *DGTT3*. They also saw expression at the mRNA level, but found no lipid phenotype. It is possible that regulatory factors prevent the DGTTs from functioning properly under N-replete conditions, or other limiting steps may exist before the DGAT step, which prevent the DGTTs from showing an increase in TAG under N-replete conditions.

When the amiRNA lines were tested, no repeatable decrease in mRNA was detected. The TAG levels were tested as before and also showed no change. Recently, Deng, X. D. et al. (2012) performed knockdowns on the *DGTT* genes, and found that knocking down *DGTT1* and *DGTT3* led to a reduction in total TAG accumulated under N-deprived conditions. Knocking down *DGTT4*, on the other hand, actually increased the amount of TAG that accumulated. The

knockdown of *DGTT2* showed no phenotype, although whether this is due to compensation from the other *DGTT*s or lack of function within *Chlamydomonas* during N-deprivation is unclear.

Although my initial screen for DGATs did not discover a type 1 DGAT in the *Chlamydomonas* genome, one was discovered after further gene annotation (Boyle, N. R. et al. 2012). A PDAT has also been found, raising the total number of TAG synthesis genes to six. Determining which of these genes are the major contributors to TAG synthesis, and how they interact and are regulated, remains a goal for future research.

BIBLIOGRAPHY

BIBLIOGRAPHY

Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J.H., Zhang,Z., Miller,W., and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic.Acids.Res **25**:3389-3402.

Andrianov,V., Borisjuk,N., Pogrebnyak,N., Brinker,A., Dixon,J., Spitsin,S., Flynn,J., Matyszczuk,P., Andryszak,K., Laurelli,M., Golovkin,M., and Koprowski,H. (2010) Tobacco as a production platform for biofuel: overexpression of Arabidopsis *DGAT* and *LEC2* genes increases accumulation and shifts the composition of lipids in green biomass. Plant.Biotechnol.J **8**:277-287.

Athenstaedt,K. and Daum,G. (2006) The life cycle of neutral lipids: synthesis, storage and degradation. Cell.Mol.Life.Sci **63**:1355-1369.

Banilas,G., Karampelias,M., Makariti,I., Kourti,A., and Hatzopoulos,P. (2011) The olive *DGAT2* gene is developmentally regulated and shares overlapping but distinct expression patterns with *DGAT1*. J.Exp.Bot **62**:521-532.

Boyle,N.R., Page,M.D., Liu,B., Blaby,I.K., Casero,D., Kropat,J., Cokus,S., Hong-Hermesdorf,A., Shaw,J., Karpowicz,S.J., Gallaher,S., Johnson,S., Benning,C., Pellegrini,M., Grossman,A., and Merchant,S.S. (2012) Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J.Biol.Chem. **289**:15811-15825.

Burgal,J., Shockey,J., Lu,C., Dyer,J., Larson,T., Graham,I., and Browse,J. (2008) Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. Plant.Biotechnol.J. **6**:819-831.

Cahoon,E.B., Shockey,J.M., Dietrich,C.R., Gidda,S.K., Mullen,R.T., and Dyer,J.M. (2007) Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux. Curr.Opin.Plant Biol. **10**:236-244.

Cases,S., Smith,S.J., Zheng,Y.W., Myers,H.M., Lear,S.R., Sande,E., Novak,S., Collins,C., Welch,C.B., Lusis,A.J., Erickson,S.K., and Farese,R.V. (1998) Identification of a gene encoding an acyl CoA : diacylglycerol acyltransferase, a key enzyme in triacylglycerol synthesis. Proc.Natl.Acad.Sci.U.S.A **95**:13018-13023.

Cases,S., Stone,S.J., Zhou,P., Yen,E., Tow,B., Lardizabal,K.D., Voelker,T., and Farese,R.V. (2001) Cloning of DGAT2, a second mammalian diacylglycerol acyltransferase, and related family members. J.Biol.Chem **276**:38870-38876.

Chapman,K.D. and Ohlrogge,J.B. (2012) Compartmentation of triacylglycerol accumulation in plants. J.Biol.Chem. **287**:2288-2294.

Clough,S.J. and Bent,A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant.J **16**:735-743.

Dahlqvist,A., Stahl,U., Lenman,M., Banas,A., Lee,M., Sandager,L., Ronne,H., and Stymne,H. (2000) Phospholipid : diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. Proc.Natl.Acad.Sci.U.S.A **97**:6487-6492.

Deng,X.D., Gu,B., Li,Y.J., Hu,X.W., Guo,J.C., and Fei,X.W. (2012) The Roles of acyl-CoA: Diacylglycerol Acyltransferase 2 Genes in the Biosynthesis of Triacylglycerols by the Green Algae *Chlamydomonas reinhardtii*. Mol.Plant **5**:945-947.

Durrett,T.P., McClosky,D.D., Tumaney,A.W., Elzinga,D.A., Ohlrogge,J., and Pollard,M. (2010) A distinct DGAT with *sn*-3 acetyltransferase activity that synthesizes unusual, reduced-viscosity oils in *Euonymus* and transgenic seeds. Proc.Natl.Acad.Sci.U.S.A **107**:9464-9469.

Fan,J., Andre,C., and Xu,C. (2011) A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. FEBS.Lett **585**:1985-1991.

Fischer,N. and Rochaix,J.D. (2001) The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. Mol.Genet.Genomics **265**:888-894.

Gietz,R.D., Schiestl,R.H., Willems,A.R., and Woods,R.A. (1995) Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. Yeast **11**:355-360.

Goodson,C., Roth,R., Wang,Z.T., and Goodenough,U. (2011) Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost. Eukaryot.Cell **10**:1592-1606.

Guiheneuf,F., Leu,S., Zarka,A., Khozin-Goldberg,I., Khalilov,I., and Boussiba,S. (2011) Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom *Phaeodactylum tricornutum*. FEBS.J **278**:3651-3666.

Harris,E.H. (1989) *The Chlamydomonas Source Book*. San Diego: Academic.

He,X.H., Chen,G.Q., Lin,J.T., and Mckeon,T.A. (2004a) Regulation of diacylglycerol acyltransferase in developing seeds of castor. Lipids **39**:865-871.

He,X.H., Turner,C., Chen,G.Q., Lin,J.T., and Mckeon,T.A. (2004b) Cloning and characterization of a cDNA encoding diacylglycerol acyltransferase from castor bean. Lipids **39**:311-318.

Hobbs,D.H., Lu,C.F., and Hills,M.J. (1999) Cloning of a cDNA encoding diacylglycerol acyltransferase from *Arabidopsis thaliana* and its functional expression. FEBS.Lett **452**:145-149.

Jako,C., Kumar,A., Wei,Y.D., Zou,J.T., Barton,D.L., Giblin,E.M., Covello,P.S., and Taylor,D.C. (2001) Seed-specific over-expression of an Arabidopsis cDNA encoding a

diacylglycerol acyltransferase enhances seed oil content and seed weight. Plant.Physiol **126**:861-874.

Kalscheuer,R. and Steinbuchel,A. (2003) A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. J.Biol.Chem. **278**:8075-8082.

Kessler,F. and Vidi,P.A. (2007) Plastoglobule lipid bodies: their functions in chloroplasts and their potential for applications. Adv.Biochem.Eng Biotechnol. **107**:153-172.

Kroon,J.T.M., Wei,W.X., Simon,W.J., and Slabas,A.R. (2006) Identification and functional expression of a type 2 acyl-CoA : diacylglycerol acyltransferase (DGAT2) in developing castor bean seeds which has high homology to the major triglyceride biosynthetic enzyme of fungi and animals. Phytochemistry **67**:2541-2549.

La Russa,M., Bogen,C., Uhmeyer,A., Doebbe,A., Filippone,E., Kruse,O., and Mussgnug,J.H. (2012) Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. J.Biotechnol **162**:13-20.

Lardizabal,K.D., Mai,J.T., Wagner,N.W., Wyrick,A., Voelker,T., and Hawkins,D.J. (2001) DGAT2 is a new diacylglycerol acyltransferase gene family - Purification, cloning, and expression in insect cells of two polypeptides from *Mortierella ramanniana* with diacylglycerol acyltransferase activity. J.Biol.Chem **276**:38862-38869.

Li,F., Wu,X., Lam,P., Bird,D., Zheng,H., Samuels,L., Jetter,R., and Kunst,L. (2008) Identification of the wax ester synthase/acyl-coenzyme A: diacylglycerol acyltransferase WSD1 required for stem wax ester biosynthesis in Arabidopsis. Plant.Physiol **148**:97-107.

Li,R., Yu,K., Hatanaka,T., and Hildebrand,D.F. (2010) *Vernonia* DGATs increase accumulation of epoxy fatty acids in oil. Plant.Biotechnol.J **8**:184-195.

Lu,C., Napier,J.A., Clemente,T.E., and Cahoon,E.B. (2011) New frontiers in oilseed biotechnology: meeting the global demand for vegetable oils for food, feed, biofuel, and industrial applications. Curr.Opin.Biotechnol. **22**:252-259.

Lung,S.C. and Weselake,R.J. (2006) Diacylglycerol acyltransferase: a key mediator of plant triacylglycerol synthesis. Lipids **41**:1073-1088.

Milcamps,A., Tumaney,A.W., Paddock,T., Pan,D.A., Ohlrogge,J., and Pollard,M. (2005) Isolation of a gene encoding a 1,2-diacylglycerol-sn-acetyl-CoA acetyltransferase from developing seeds of *Euonymus alatus*. J.Biol.Chem **280**:5370-5377.

Miller,R., Wu,G., Deshpande,R.R., Vieler,A., Gartner,K., Li,X., Moellering,E.R., Zauner,S., Cornish,A.J., Liu,B., Bullard,B., Sears,B.B., Kuo,M.H., Hegg,E.L., Shachar-Hill,Y., Shiu,S.H., and Benning,C. (2010) Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. Plant.Physiol **154**:1737-1752.

Molnar,A., Bassett,A., Thuenemann,E., Schwach,F., Karkare,S., Ossowski,S., Weigel,D., and Baulcombe,D. (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. Plant.J **58**:165-174.

Murashige,T. and Skoog,F. (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol.Plantarum **15**:473-497.

Napier,J.A. (2007) The production of unusual fatty acids in transgenic plants. Annu.Rev.Plant.Biol **58**:295-319.

Neupert,J., Karcher,D., and Bock,R. (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant.J **57**:1140-1150.

Ossowski,S., Schwab,R., and Weigel,D. (2008) Gene silencing in plants using artificial microRNAs and other small RNAs. Plant.J **53**:674-690.

Peng,F.Y. and Weselake,R.J. (2011) Gene coexpression clusters and putative regulatory elements underlying seed storage reserve accumulation in Arabidopsis. BMC.Genomics **12**:286.

Routaboul,J.M., Benning,C., Bechtold,N., Caboche,M., and Lepiniec,L. (1999) The TAG1 locus of Arabidopsis encodes for a diacylglycerol acyltransferase. Plant.Physiol.Biochem **37**:831-840.

Saha,S., Enugutti,B., Rajakumari,S., and Rajasekharan,R. (2006) Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. Plant.Physiol **141**:1533-1543.

Sandager,L., Gustavsson,M.H., Stahl,U., Dahlqvist,A., Wiberg,E., Banas,A., Lenman,M., Ronne,H., and Stymne,S. (2002) Storage lipid synthesis is non-essential in yeast. J.Biol.Chem. **277**:6478-6482.

Sanjaya, Durrett,T.P., Weise,S.E., and Benning,C. (2011) Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. Plant.Biotechnol.J **9**:874-883.

Sanjaya, Miller,R., Durrett,T.P., Kosma,D.K., Lydic,T.A., Muthan,B., Koo,A.J., Bukhman,Y.V., Reid,G.E., Howe,G.A., Ohlrogge,J., and Benning,C. (2013) Altered lipid composition and enhanced nutritional value of Arabidopsis leaves following introduction of an algal diacylglycerol acyltransferase 2. Plant Cell **25**:677-693.

Sherman,F. (2002) Getting started with yeast. Methods.Enzymol **350**:3-41.

Shockey,J.M., Gidda,S.K., Chapital,D.C., Kuan,J.C., Dhanoa,P.K., Bland,J.M., Rothstein,S.J., Mullen,R.T., and Dyer,J.M. (2006) Tung tree DGAT1 and DGAT2 have nonredundant functions in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. Plant Cell **18**:2294-2313.

Snyder,C.L., Yurchenko,O.P., Siloto,R.M.P., Chen,X., Liu,Q., Mietkiewska,E., and Weselake,R.J. (2009) Acyltransferase action in the modification of seed oil biosynthesis. New.Biotechnol **26**:11-16.

Sorger,D. and Daum,G. (2002) Synthesis of triacylglycerols by the acyl-coenzyme A:diacyl-glycerol acyltransferase Dga1p in lipid particles of the yeast *Saccharomyces cerevisiae*. J.Bacteriol **184**:519-524.

Stobart,K., Mancha,M., Lenman,M., Dahlqvist,A., and Stymne,S. (1997) Triacylglycerols are synthesised and utilized by transacylation reactions in microsomal preparations of developing safflower (*Carthamus tinctorius* L) seeds. Planta **203**:58-66.

Stymne,S. and Stobart,A.K. (1984) The biosynthesis of triacylglycerols in microsomal preparations of developing cotyledons of sunflower (*Helianthus annuus* L.). Biochem.J **220**:481-488.

Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M., and Kumar,S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol.Biol.Evol **28**:2731-2739.

Turchetto-Zolet,A.C., Maraschin,F.S., de Morais,G.L., Cagliari,A., Andrade,C.M., Margis-Pinheiro,M., and Margis,R. (2011) Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. BMC.Evol.Biol **11**:263.

Wagner,M., Hoppe,K., Czabany,T., Heilmann,M., Daum,G., Feussner,I., and Fulda,M. (2010) Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (DGAT2) gene from the microalga *O. tauri*. Plant.Physiol.Bioch **48**:407-416.

Wang,Z.G., Yao,T., and Song,Z.Y. (2010) Involvement and mechanism of DGAT2 upregulation in the pathogenesis of alcoholic fatty liver disease. J.Lipid.Res **51**:3158-3165.

Xu,C., Fan,J., Riekhof,W., Froehlich,J.E., and Benning,C. (2003) A permease-like protein involved in ER to thylakoid lipid transfer in Arabidopsis. EMBO.J **22**:2370-2379.

Yen,C.L.E., Stone,S.J., Koliwad,S., Harris,C., and Farese,R.V. (2008) DGAT enzymes and triacylglycerol biosynthesis. J.Lipid.Res **49**:2283-2301.

Zhang,M., Fan,J., Taylor,D.C., and Ohlrogge,J.B. (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in Arabidopsis triacylglycerol biosynthesis and are essential for normal pollen and seed development. Plant Cell **21**:3885-3901.

Zheng,P., Allen,W.B., Roesler,K., Williams,M.E., Zhang,S., Li,J., Glassman,K., Ranch,J., Nubel,D., Solawetz,W., Bhattramakki,D., Llaca,V., Deschamps,S., Zhong,G.Y., Tarczynski,M.C., and Shen,B. (2008) A phenylalanine in DGAT is a key determinant of oil content and composition in maize. Nat.Genet **40**:367-372.

Zou,J.T., Wei,Y.D., Jako,C., Kumar,A., Selvaraj,G., and Taylor,D.C. (1999) The *Arabidopsis thaliana* TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. Plant.J **19**:645-653.

CHAPTER 3: TRANSCRIPT PROFILING OF *CHLAMYDOMONAS REINHARDTII* UNDER NITROGEN DEPRIVATION

This chapter focuses on the experiments published in (Miller, R. et al. 2010). The methods and results are partially reproduced.

Introduction

As established previously, *Chlamydomonas reinhardtii*, like other microalgae, synthesizes TAG in response to nutrient stress. This accumulation of TAG is part of a larger response, involving changes in the level of many metabolites, and alterations of regulatory networks within the algae (Bolling, C. and Fiehn, O. 2005, Grossman, A. 2000). Due to their environment, algae are regularly short on key nutrients, and need ways to adapt and preserve themselves until conditions change. Most algal blooms are related to an increase in nutrient content, due to weather conditions, and are indicative of the rapid adaptation to changes in environment that algae are capable of. The response varies depending on the nutrient being deprived, and includes certain common elements, as well as divergent elements. Both nitrogen, phosphorus and sulfur deprivation all lead to the accumulation of TAG, among other changes (Matthew, T. et al. 2009, Weers, P. M. M. and Gulati, R. D. 1997). N, P and S deprivation also lead to the accumulation of significant amounts of starch in the cell, an up to 18-fold increase in the case of nitrogen deprivation (Ball, S. G. et al. 1990, Msanne, J. et al. 2012). Nitrogen deprivation is the strongest, fastest and easiest of the three to use in the laboratory, leading to significant amounts of TAG after 24 hours, and thus was the method used.

Nitrogen deprivation leads to the activation of several other pathways, as well as TAG synthesis. The immediate response is the activation of various nitrogen-scavenging mechanisms, allowing the cell to look for alternative sources of nitrogen (Grossman, A. 2000). Other changes

include the breakdown of the photosynthetic machinery, including chlorophyll and the thylakoid membranes (Moellering, E. R. and Benning, C. 2010, Msanne, J. et al. 2012). Starch is accumulated, in addition to TAG (Fan, J. et al. 2012). Finally, nitrogen deprivation triggers gametogenesis, preparing the cells to mate (Goodenough, U. et al. 2007).

Given the large changes in metabolism occurring, it is reasonable to conclude that significant changes in transcription levels for many genes are occurring. These changes are partially responsible for regulating the changes in overall metabolism, and thus are important for understanding how TAG synthesis is controlled. The best way to measure this is by a global transcript profiling experiment, which will give a general overview of all the changes in transcripts in response to nitrogen deprivation. The advantage to this is that it will show changes in all genes, not just the obvious lipid-related targets. Regulators, such as transcription factors or kinases, aren't easy to associate with specific activities based on data mining, but may be found through a transcript profiling experiment. This experiment will therefore give an overall picture of changes in the cell, in addition to just the obvious lipid synthesis pathways. Changes in other pathways will alter how carbon is channeled, which may be a significant factor in TAG accumulation. One downside of this approach is that it will miss any post-transcriptional changes, thus giving an incomplete picture of the regulatory changes that are occurring. Additionally, the large number of processes activated by nitrogen deprivation will make it harder to sort out the lipid-related changes from those related to other processes. However, the transcriptional changes will give a good place to start.

In the past, most global transcription experiments used microarrays to measure transcript levels (Girke, T. et al. 2000). This has been done with *Chlamydomonas* as well, but microarray technology is limited for *Chlamydomonas*. Only part of the genome is represented by most

microarray chips (~10,000 of the estimated 15,000), and generating chips is difficult and time-consuming (Erickson, B. et al. 2005, Kucho, K. et al. 2005, Miura, K. et al. 2004, Zhang, Z. D. et al. 2004). A recent alternative is using high-throughput sequencing to sequence the mRNA and quantify the level of expression of all genes.

High-throughput sequencing has several advantages over microarray chips. A single run can give a lot of data, more than with microarray, and requires no prior knowledge of gene models. This is useful for *Chlamydomonas*, as many of the gene models were still poorly annotated at the time of this experiment. The availability of a genome sequence allows the match-up of the reads created during the sequencing to annotated or unknown genes, making the analysis easier, allowing the quantification of transcript levels and helping to annotate gene models.

High-throughput sequencing is a recent technique, invented in the last decade. Initially, sequencing was done via two methods, the chemical-based Maxam-Gilbert and chain-termination-based Sanger sequencing (Maxam, A. M. and Gilbert, W. 1977, Sanger, F. et al. 1977). In the Maxam-Gilbert method, DNA is modified at specific bases and cleaved. While in the Sanger method, the sequencing proceeds from a primer and is terminated randomly using modified bases (dideoxynucleotides). In both cases, the sequence is read by separating the DNA fragments on a gel and reading the bands. Later, fluorescent bases allowed reading the Sanger method to be automated, speeding up the process, but sequencing was still slow, as only one strand could be sequenced at a time.

High-throughput sequencing uses new technology to monitor sequencing reactions as they occur, allowing detection and determination of the sequence during synthesis, rather than after. DNA bases are added one at a time, and incorporation into the growing chain measured.

Unincorporated bases are removed, and then replaced with the next base. This process can be done in parallel, using multi-well plates or flow cells designed to hold multiple samples. This allows for many pieces of DNA to be read at once, drastically decreasing the time needed for full coverage of a genome or transcriptome. In addition, specific primers are no longer necessary, thus allowing the use of this technique with unknown sequences or new species. Several different methods of high-throughput sequencing have been developed, all following the same general principles but with different details. The two methods I used are 454 technology and Illumina technology (formerly known as Solexa). 454 sequencing relies on the release of pyrophosphate after DNA base incorporation to give a signal. The intensity of the signal indicates if there are multiple bases in a row. Illumina, on the other hand, uses fluorescent bases, and removes the fluorescent marker before addition of the next base. Both technologies are useful for many different purposes, including transcript profiling.

Methods

Growth conditions

The *Chlamydomonas* strain used was dw15.1 (cw15, nit2, mt+), kindly provided by Arthur Grossman. The cells were grown in liquid cultures under continuous light (~80 μmole photons m-2 s-1). For N-replete growth, TAP medium (Harris, E. H. 1989) with 10 mM NH_4^+ (TAP+N) was used. For preliminary experiments, N-deprivation was applied by two methods: continuous growth in TAP with 0.5 mM NH_4^+; or growth in TAP+N to $5x10^6$ cells/mL, followed by transfer to TAP with no NH_4^+ (TAP-N) for an additional 24 or 48 hours. For further experiments, N-deprivation was defined as growth in TAP+N to $5x10^6$ cells/mL, followed by transfer to TAP-N for 48 hours.

To generate material for high-throughput sequencing, cells were grown in 100 mL TAP+N to $5x10^6$ cells/mL. The cultures were split in half and cells were collected by centrifugation, with one pellet being resuspended in 50 mL TAP+N, and the other in 50 mL TAP-N. After 48 hours, the total RNA was harvested using a QIAGEN RNeasy Plant Mini kit (QIAGEN, Valencia, CA, USA). The RNA samples were treated with QIAGEN RNase-free DNase I during extraction.

Preparation of samples

For 454 sequencing, full-length cDNA pools were generated with Clontech SMART cDNA library construction kit (Clontech, Mountain View, CA, USA). cDNA was synthesized using a modified cDNA synthesis primer (5'TAGAGACCGAGGCGGCCGACATGTTTTGTTTTTTTTTCTTTTTTTTTTTVN3'). Full-length cDNAs were amplified by PCR and pooled to increase their concentration. An *Sfi*I digest was performed, followed by size fractionation. Fractions with the highest intensity and size distribution were pooled and purified. The resulting cDNA pools were then submitted to MSU-Research Technologies Service Facility (RTSF) for sequencing on a 454 GSFLX Titanium Sequencer (454 Life Sciences, Branford, CT, USA). For Illumina sequencing, total RNA was submitted directly to the MSU-RTSF for sequencing on an Illumina Genome Analyzer II (Illumina, Inc., San Diego, CA, USA).

Assembly of sequencing results

Default parameters were used to pass reads using 454 and Illumina quality control tools. The filtered sequence data were deposited in the NCBI Short Read Archive with the reference series number GSE24367 and subseries numbers GSE24365 and GSE24366 for the Illumina and the 454 datasets, respectively. The filtered 454 sequencing reads were mapped to

Chlamydomonas v4.0 assembly from the Joint Genome Institute with GMAP (Wu, T. D. and Watanabe, C. K. 2005). In GMAP, the maximum intron length was set at 980bp which is at the 95 percentile of annotated *Chlamydomonas* intron lengths. The Illumina reads were mapped with Bowtie (Langmead, B. et al. 2009) using parameters as follows: ≤ 2 mismatches, sum of Phred quality values at all mismatched positions ≤ 70, and excluding reads mapped to > 1 locations. Because the sequence qualities of Illumina reads degrade quickly toward the 3'-end, an alternative mapping data was generated with reads trimmed from the 3'-end (until the 3'-end most position with Phred-equivalent score was ≥ 20). Trimmed reads < 30bp were excluded from further analysis. In addition to the sequence quality issue, some reads may span two exons and would not be mapped by Bowtie correctly. Tophat (Trapnell, C. et al. 2009) was used to identify these exon spanning reads to generate another set of read mapping. The information from Tophat was used for assembling mapped reads into transcribed fragments (transfrags) with Cufflinks (Trapnell, C. et al. 2010). In Cufflinks, the maximum intron length was set at 1855bp (99 percentile of all the intron lengths), 5% minimum isoform fraction, and 5% pre-mRNA fraction. Transfrags within 1855bp of an existing *Chlamydomonas* v4.0 gene model were regarded as potential missing exons of annotated genes. The rest were regarded as intergenic exons and adjacent transfrags < 1855bp apart were joined into "transcriptional units".

Northern blot

Total RNA was harvested from N-replete or N-deprived cells as described above, and 4 μg of each total RNA was separated on a 1% formaldehyde gel and transferred to a Hybond-N+ nylon membrane (GE Healthcare, Piscataway, NJ, USA). Probes were synthesized from cDNA and labeled with [^{32}P] using Amersham Megaprime labeling kit (GE Healthcare). The blots were hybridized with the labeled probes in Ambion ULTRAhybe (Applied Biosystems/Ambion,

Austin, TX, USA) at 42°C overnight. The blots were washed twice for 5 min with low-stringency buffer (1X SSC, 0.1% SDS) at 60°C, and then twice for 5 min with high-stringency buffer (0.1X SSC, 0.1% SDS) at 60°C. The blots were exposed to a Molecular Dynamics phosphor screen (GE Healthcare) overnight, and visualized with a Storm 820 phosphoimager (GE Healthcare). Probes were synthesized from cDNA for *AMT4* (5'GTATTGCGTCCGATCTGC3', 5'CGTGGAAATGCTGTAGGG3'), *DGTT2* (5'TAAAGCACCGACAAATGTGC3', 5'CATGATCTGGCATTCTGTGG3'), and *DGTT3* (5'GGTGGTGCTCTCCTACTGGA3', 5'CCATGTACATCTCGGCAATG3').

Analysis of expression data

Differential signals between *Chlamydomonas* cultures grown in N-replete and N-depleted medium were determined using the numbers of mapped reads overlapped with annotated *Chlamydomonas* genes as inputs to EdgeR (Robinson, M. D. et al. 2010). In the Joint Genome Institute database, multiple sets of *Chlamydomonas* v.4 gene models are available. The "filtered" gene models, which contain the best gene model for each locus, were used. Genes were regarded as differentially expressed if they have ≥ 2-fold change between N-replete and N-deprived samples and ≤ 5% False Discovery Rate (FDR). Differentially expressed genes were regarded as up-regulated if their expression levels in N-deprived samples were significantly higher than those in N-replete samples. Conversely, down-regulated genes were those with significantly lower levels of expression following N-deprivation.

In addition to EdgeR, three other methods to evaluate differential expression were used, including Fisher's exact test (Bloom, J. S. et al. 2009), likelihood ratio test (Marioni, J. C. et al. 2008), and a method based on intensity ratio and average intensity (MARS, (Wang, L. et al. 2010)). All three methods were implemented in the DEGexp package (Wang, L. et al. 2010). We

found that among 4004 differentially expressed genes called by EdgeR, 99.7% to 100% were regarded as differentially expressed by the other three methods. On the other hand, EdgeR calls overlap with 96.6%, 94.7%, and 95.8% of calls by Fisher's exact test, likelihood ratio test, and MARS, respectively. These findings indicate that the EdgeR is more conservative than the other methods but the overall differential expression calls are highly similar among methods. It should be noted that methods other than EdgeR did not explicitly consider variance between replicates and, as a result, will likely have a higher false positive differential expression call rate than that of EdgeR. Therefore, in all subsequent analysis, only EdgeR-based differential expression calls were used.

Gene ontology (GO) annotation for *Chlamydomonas* v4.0 genome was acquired from Joint Genome Institute. Enrichment of differentially regulated genes in each GO category was determined using Fisher's exact test. To account for multiple testing, the p values from Fisher's exact tests were adjusted (Storey, J. D. 2003) and an FDR value of 5% was used as the threshold for enriched GO terms.

Results

Determination of conditions

Ideally, one would like to use finely spaced time course experiments to distinguish rapid versus long-term changes in gene expression following N-deprivation. However, because resources were limited, the experiment focused on two conditions, N-replete and N-deprived. Independent biological replicates allowed for statistically sound interpretations of the data. To determine the time point for N-deprivation that was most likely to provide an accurate snapshot of readjustment of transcript steady-state levels following N-deprivation, Northern blot hybridization was used to compare expression of genes known or expected to be regulated in

Chlamydomonas following N-deprivation. An ammonium transporter, *AMT4*, which has been previously shown to be activated by nitrogen deprivation (Mamedov, T. G. et al. 2005), and two putative diacylglycerol acyltransferases, *DGTT2* and *DGTT3*, were monitored to test the various conditions. RNA was isolated from cells grown in standard (10 mM NH_4^+) and low nitrogen (0.5 mM NH_4^+) TAP medium to impose N-limitation, as well as cells grown to mid-log phase in standard TAP medium, then transferred to either standard or no nitrogen (0 mM NH_4^+) TAP, with samples taken at 24 and 48 hours, to accomplish more drastic N-deprivation (Figure 3-1). For standardization, equal amounts of RNA were loaded and the 18S rRNA abundance was examined. Although *Chlamydomonas* ribosomes turn over following N-deprivation (Martin, N. C. et al. 1976, Siersma, P. W. and Chiang, K. S. 1971), their abundance drops no lower than 50%. *AMT4* mRNA was absent from the uninduced cells, and present at a high level in N-limited or -deprived cells under the conditions tested. *DGTT2* mRNA was present at low levels in all conditions. *DGTT3* mRNA was present at low levels and increased slightly following N-deprivation. N-deprivation for 48 hours showed the greatest difference in mRNA levels compared to the N-replete cultures. Based on this basic analysis and a previous time course study of lipid droplet formation and TAG accumulation (Moellering, E. R. and Benning, C. 2010), a 48 hour period of N-deprivation was chosen to compare global transcript levels in N-replete and N-deprived cells.

Sequencing results

To determine the differential expression of genes in *Chlamydomonas* under N-replete and N- deprived conditions, two sequencing approaches, 454 and Illumina were applied. Read length is longer for 454 but the number of reads per run is lower. As shown in Table 3-1, 60-85 fold more sequence tags were generated with Illumina than with 454 sequencing. Among the 454

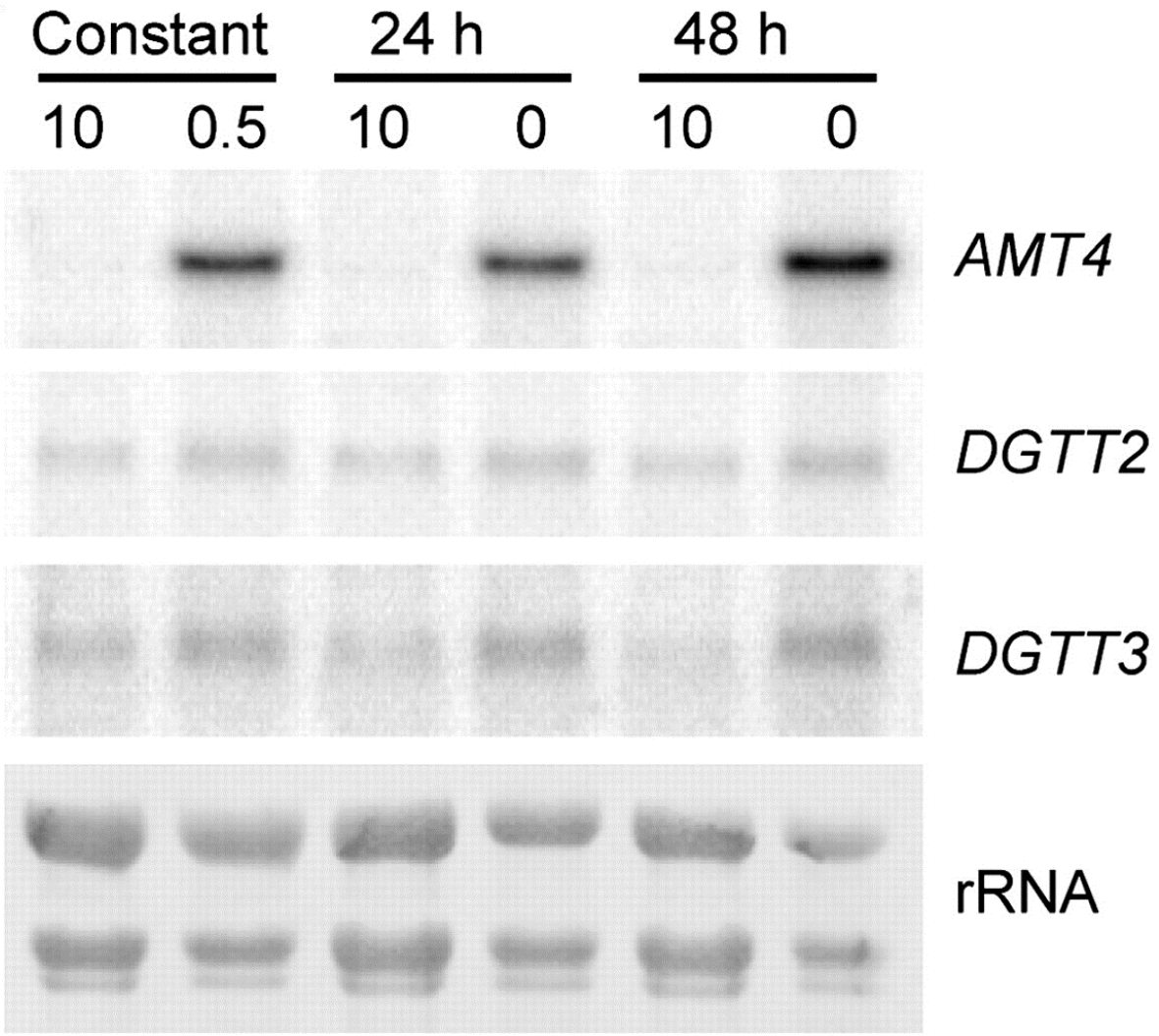


Figure 3-1. Transcript levels of specific genes. Cultures were grown in TAP medium that was N replete (10 mm NH_4^+; 10), continual N limited (0.5 mm NH_4^+; 0.5), or N deprived (0 mm NH_4^+; 0) for 24 or 48 h. The expression levels of *AMT4*, *DGTT2*, and *DGTT3* were measured by RNA-DNA hybridization, and rRNA was visualized as a loading control.

Table 3-1. Summary of expression tags generated using two different sequencing methods

Sequencing methods	**454**		**Illumina**					
Treatment[a]	N-R	N-D	T1	T2	T3	N1	N2	N3
Total[b]	$2.51\times10^{+05}$	$2.15\times10^{+05}$	$1.69\times10^{+07}$	$1.83\times10^{+07}$	$1.78\times10^{+07}$	$1.77\times10^{+07}$	$1.79\times10^{+07}$	$1.52\times10^{+07}$
Mapped[c]	$2.01\times10^{+05}$	$1.68\times10^{+05}$	$1.09\times10^{+07}$	$1.24\times10^{+07}$	$1.15\times10^{+07}$	$1.07\times10^{+07}$	$1.13\times10^{+07}$	$9.86\times10^{+06}$
Genic[d]	$1.08\times10^{+05}$	$1.32\times10^{+05}$	$8.96\times10^{+06}$	$9.43\times10^{+06}$	$1.02\times10^{+07}$	$8.33\times10^{+06}$	$8.82\times10^{+06}$	$7.81\times10^{+06}$
Intergenic[e]	$9.30\times10^{+04}$	$3.61\times10^{+04}$	$1.93\times10^{+06}$	$2.97\times10^{+06}$	$1.31\times10^{+06}$	$2.38\times10^{+06}$	$2.48\times10^{+06}$	$2.05\times10^{+06}$

[a] Treatment types: (N-R) or T1, T2 and T3 represent N-replete and (N-D) or N1,N2, N3 represent N-deprived growth. For Illumina, there are three biologically independent replicates for each treatment (T1-3 and N1-3).

[b] Number of sequencing reads after filtering out low quality reads based on 454 and Illumina base-calling methods.

[c] Number of sequencing reads after mapping to *Chlamydomonas* v4.0 genome

[d] Number of mapped sequencing reads overlapping with *Chlamydomonas* v4.0 filtered gene models

[e] Number of mapped sequencing reads not overlapping with any *Chlamydomonas* v4.0 filtered gene models

reads, 78-80% mapped to the *Chlamydomonas* genome. For the Illumina data, we mapped in three different ways, with varying stringency depending on whether 3'-end read quality and exon-spanning reads were considered. Without filtering reads, a substantially smaller proportion of Illumina reads (63-68%) were mapped compared to 454 reads. Trimming low quality 3' regions of reads resulted in a further 2.7% decrease in the number of mapped reads. Despite the large number of unmapped Illumina reads, out of 16,710 *Chlamydomonas* gene models, 15,505 (92%) had ≥1 reads. In contrast, only 6372 gene models (38.1%) were supported by the 454 transcriptome data set. In addition, nearly all genes covered by 454 were also covered by Illumina. Therefore, the sequencing data cover most annotated genes enabling the interrogation of differential expression under normal conditions and following N-deprivation. In addition, as expected, Illumina data provided a better coverage of the gene space than 454 sequences.

To determine differential gene expression following N-deprivation, count data were modeled with a moderated negative binomial distribution (see Materials and Methods). Using thresholds of ≤ 5% False Discovery Rates (FDR) and ≥ 2-fold change for the Illumina dataset, 2,128 and 1,875 genes were categorized as up- and down-regulated, respectively, following N-deprivation. To see if fold changes inferred based on 454 and Illumina datasets were consistent, the statistical correlation in fold change between these two datasets (Figure 3-2) was determined, and found to be moderately strong (Pearson's correlation coefficient, $r^2 = 0.57$, $p < 2.2 \times 10^{-16}$; Figure 3-2, blue data points) when comparing only the genes that had a significant number of hits in the 454 data set (2,056 genes with ≥10 reads combined and ≥1 read in both conditions, out of 6369 genes with ≥ 1 reads from both datasets). The initial comparison of all genes found in both datasets was rather weak ($r^2 = 0.10$, $p < 2.2 \times 10^{-16}$), due to a large number of genes, 4313 out of 6369, that had a high degree of up- and down-regulation that was only observed with the 454

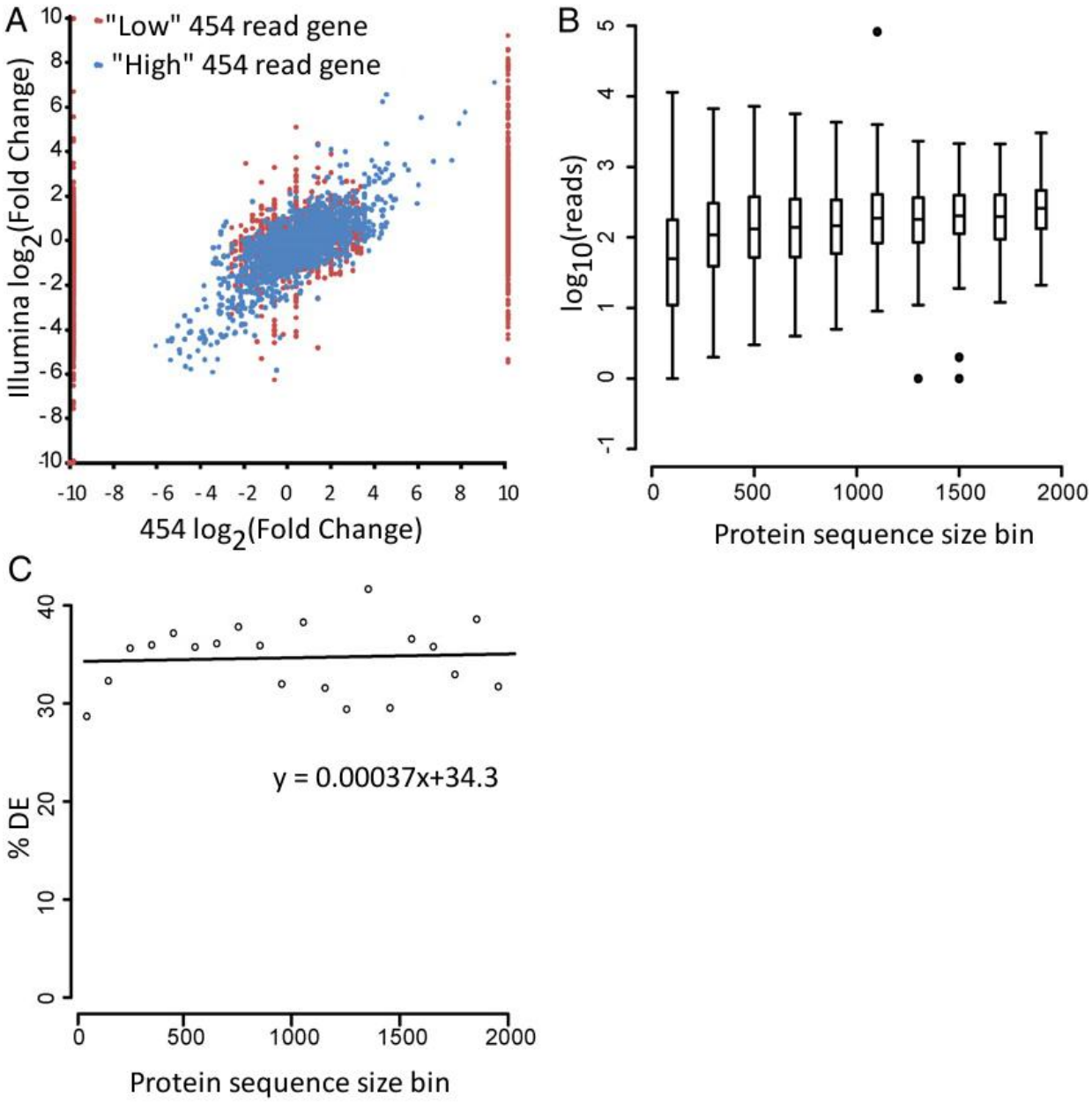


Figure 3-2. Fold change correlation between Illumina and 454 data sets, and impacts of Illumina length bias on differential expression call. (A) Only genes with one or more 454 and Illumina reads under either N-replete (+N) or N-deprived (−N) conditions were plotted. Fold change is determined by the number of reads following N deprivation divided by the number of reads under N-replete conditions for each gene. For genes with 2^{10}-fold or greater or 2^{-10}-fold or lesser changes, the fold change values were set to 10. Blue circles ("high" 454 read genes) indicate genes with 10 or greater 454 reads (+N and –N combined) and one or more 454 reads in both +N and –N. Red circles ("low" 454 read genes) indicate genes that did not satisfy one or both of the above criteria. **(B)** Each box plot depicts the numbers of reads for protein-coding genes (log base 10) in a protein sequence size bin (0–2,000 amino acids, bin size of 100 amino acids). All proteins of 2,000 or more amino acids are classified as 2,000 amino acids. Outliers are

Figure 3-2 (cont'd)
shown in black circles. **(C)** Percentage of genes that are regarded as differentially expressed (DE) in each protein sequence size bin. The line indicates the linear fit, and the equation for the line is shown as well.

but not the Illumina dataset (Figure 3-2A). Most of these genes with extreme responses based on 454 had very low counts (<10 reads combined in both conditions or 0 read in one of the conditions; Figure 3-2A, red data points). As a result, high and likely inaccurate fold change values were assigned to those genes.

One important consideration in identifying differentially regulated genes is that there is a considerable transcript length bias in Illumina data. A longer transcript tends to have more reads than a shorter transcript expressed at the same level (Bullard, J. H. et al. 2010, Oshlack, A. and Wakefield, M. J. 2009). Consistent with earlier findings, a significant correlation between the number of reads assigned to a protein sequence and its length (Spearman's rank ρ=0.33, $p<2.2e-16$, Figure 3-2B) was observed. Because of this length bias, longer transcripts may have more significant differences in differential gene expression studies and in some cases will lead to false positive differential expression calls (Bullard, J. H. et al. 2010, Oshlack, A. and Wakefield, M. J. 2009). However, unlike previously published studies, a significant correlation between percent genes differentially expressed and sequence length ($\rho = 0.09$, $p = 0.7$, Figure 3-2C) was not found. This finding indicates that, although length bias remains an issue, the differential expression call in this case may not be as significantly affected as previous reported.

Approximately 7-14% of the Illumina reads mapped to the “intergenic regions” (Table 3-1). Illumina reads were assembled into 42,574 transcribed fragments (transfrags). Among them, 17,095 transfrags did not map with, or within the vicinity (1855 bases, 99 percentile intron length) of current gene models. With the same conservative criterion, transfrags were joined into 1828 “intergenic transcriptional units”. Most importantly, 287 of these intergenic transcriptional units were up-regulated and 176 were down-regulated following N-deprivation. These transfrags are unannotated genes that require further analysis to establish their authenticity.

Gene ontology (GO) annotation was used to coarsely identify major categories of genes involved in particular biological processes to assess trends in their transcriptional regulation following N-deprivation. Multiple GO categories were found with significant enrichment in their numbers of differentially regulated genes (Table 3-2). Particularly, genes associated with lipid metabolism tend to be up-regulated, while those involved in photosynthesis and DNA replication initiation tend to be down-regulated.

Changes in primary metabolism

Changes in transcript abundance of genes encoding enzymes of primary metabolism, excluding genes involved in fatty acid metabolism described elsewhere, are depicted in Figure 3-3 and summarized in Table 3-3. Transcripts encoding key enzymes of the glyoxylate cycle, gluconeogenesis and the photosynthetic carbon fixation cycle markedly decrease following N-deprivation. Transcript abundance for glyoxylate cycle enzymes isocitrate lyase and malate synthase decreased more than 16-fold. In addition, mRNA abundance of the cytosolic (predicted) phosphoenolpyruvate carboxykinase (PCK), which catalyzes the committed reaction of gluconeogenesis, dropped to 25% of the levels in N-replete cells, as did transcripts encoding enzymes involved in carbon fixation and reduction, ribulose-bisphosphate carboxylase, sedoheptulose 1,7 bisphosphate aldolase and sedoheptulose-bisphosphatase. In contrast, there is a considerable increase in the transcript abundance of the cytosolic enzyme pyruvate phosphate dikinase. This is a key enzyme in the C4 photosynthetic pathway and is regulated by light. It has also been associated with suppressed PCK activity (Osteras, M. et al. 1997) and salt stress (Fisslthaler, B. et al. 1995). Recently this enzyme has been shown to play an important role in N-remobilization (Taylor, L. et al. 2010). Likewise, the transcript abundances for enzymes of the pentose phosphate cycle predicted to be localized in the cytosol, glucose-6-phosphate

Table 3-2. Gene Ontology categories significantly enriched in differentially regulated *Chlamydomonas* genes

GO	Annotation	GO R[a]	No GO R[b]	GO U[c]	No GO U[d]	Reg[e]	p[f]	q[g]
GO:0006270 bp[h]	DNA replication initiation	9	926	0	6633	down	6.48×10^{-09}	4.51×10^{-06}
GO:0015995 bp	chlorophyll biosynthetic process	8	927	0	6633	down	5.29×10^{-08}	2.76×10^{-05}
GO:0033014 bp	tetrapyrrole biosynthetic process	6	929	0	6633	down	3.51×10^{-06}	7.32×10^{-04}
GO:0009765 bp	photosynthesis, light harvesting	22	913	14	6619	down	5.72×10^{-12}	5.98×10^{-09}
GO:0015979 bp	photosynthesis	23	912	0	6633	down	1.02×10^{-21}	2.13×10^{-18}
GO:0005576 cc[i]	extracellular region	11	924	14	6619	down	8.28×10^{-05}	1.44×10^{-02}
GO:0009522 cc	photosystem I	4	931	0	6633	down	2.32×10^{-04}	3.23×10^{-02}
GO:0009538 cc	photosystem I reaction center	5	930	0	6633	down	2.85×10^{-05}	5.42×10^{-03}
GO:0009654 cc	oxygen evolving complex	6	929	0	6633	down	3.51×10^{-06}	7.32×10^{-04}
GO:0019898 cc	extrinsic to membrane	4	931	0	6633	down	2.32×10^{-04}	3.23×10^{-02}
GO:0003755 mf[j]	peptidyl-prolyl cis-trans isomerase activity	21	914	32	6601	down	4.39×10^{-07}	1.83×10^{-04}
GO:0004600 mf	cyclophilin	19	916	30	6603	down	2.29×10^{-06}	5.98×10^{-04}
GO:0016851 mf	magnesium chelatase activity	4	931	0	6633	down	2.32×10^{-04}	3.23×10^{-02}
GO:0030051 mf	FK506-sensitive peptidyl-prolyl cis-trans isomerase	19	916	30	6603	down	2.29×10^{-06}	5.98×10^{-04}
GO:0042027 mf	cyclophilin-type peptidyl-prolyl cis-trans isomerase activity	19	916	30	6603	down	2.29×10^{-06}	5.98×10^{-04}
GO:0006006 bp	glucose metabolic process	5	944	1	6618	up	1.65×10^{-04}	3.83×10^{-02}
GO:0006468 bp	protein amino acid phosphorylation	117	832	501	6118	up	2.37×10^{-06}	2.47×10^{-03}
GO:0006629 bp	lipid metabolic process	18	931	38	6581	up	9.92×10^{-05}	2.59×10^{-02}

Table 3-2 (cont'd)

GO:0004672 mf	protein kinase activity	113	836	479	6140	up	2.08×10^{-06}	2.47×10^{-03}
GO:0004713 mf	protein-tyrosine kinase activity	85	864	342	6277	up	8.04×10^{-06}	4.20×10^{-03}
GO:0004674 mf	protein serine/threonine kinase activity	92	857	391	6228	up	2.50×10^{-05}	7.46×10^{-03}

[a]GO R indicates number of significantly up or down (R) regulated genes with the GO annotation in question. [b]No GO R, number of significantly up or down regulated genes without the GO annotation. [c]GO U, number of genes without significant expression change with the GO annotation . [d]No GO U, number of genes with no significant expression change that do not have the GO annotation. [e]Reg, direction of regulation (nitrogen deprived compared to nitrogen replete). [f]Fisher's exact test *p* value. [g]*q* value is calculated using R package qvalue. [h]bp, biological process. [i]cc, cellular component. [j]mf, molecular function.

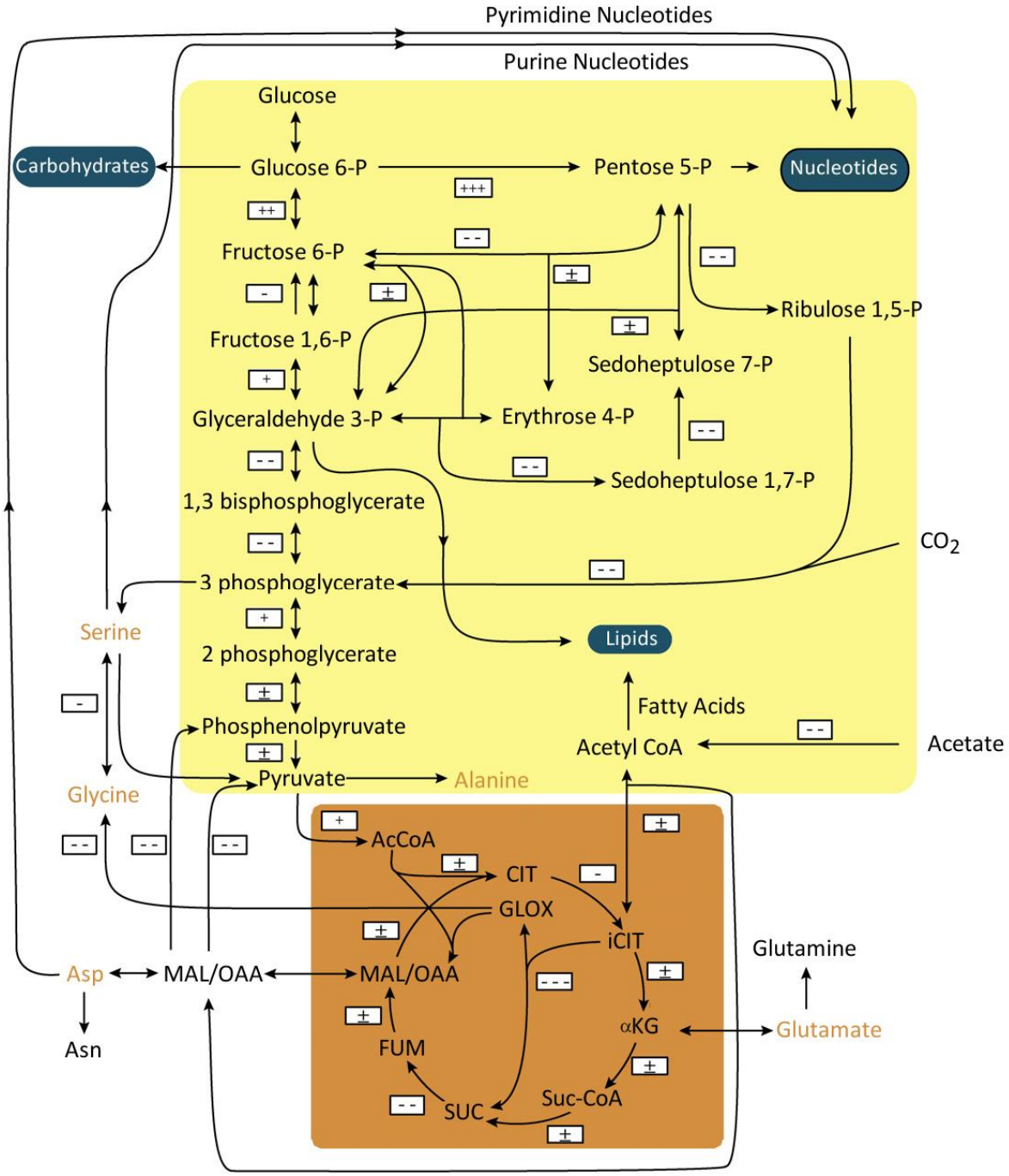


Figure 3-3. Regulation of genes involved in primary metabolism. The figure indicates the central metabolic pathways of *Chlamydomonas* and gives the differential regulation of gene expression following N deprivation. Symbols represent $\log_2$ fold change as follows: +++, greater than 5; ++, greater than 2 and less than 5; +, greater than 1; ±, less than 1 and greater than −1; −, less than −1; −−, less than −2 and greater than −5; −−−, less than −5.

Table 3-3. Illumina analysis of transcripts encoding proteins of central metabolism, excluding fatty acid metabolism

PID[a]	Annotation[b]	TID[a]	T +N[c]	T -N[c]	Log_2 FC[d]	FDR[e]
416863	Carbonic Anhydrase	410580	8	0	-28.0	3.2×10^{-02}
390723	Carbonic Anhydrase	422364	3	0	-26.6	4.2×10^{-01}
191668	Isocitrate Lyase, ICL1, Isocitrate lyase	191668	96807	1559	-5.8	3.5×10^{-49}
128726	Carbonic Anhydrase	128726	43	1	-5.3	3.2×10^{-08}
129019	Glyceraldehyde 3 Phsophate Dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase?	129019	70635	2046	-5.0	1.4×10^{-50}
196328	Malate Synthase, Malate synthase	196328	59457	2735	-4.3	5.4×10^{-36}
132210	Phosphoglycerate kinase	132210	28200	1401	-4.2	2.9×10^{-37}
194541	Alanine-glyoxylate aminotransferase, Alanine-glyoxylate transaminase 1 (AGT1)	194541	14763	791	-4.1	1.8×10^{-25}
415395	Carbonic Anhydrase	411637	3574	223	-3.9	1.1×10^{-23}
196351	Malic Enzyme	196351	1076	71	-3.8	1.3×10^{-26}
82986	Ribulose-bisphosphate Carboxylase	82986	7545	632	-3.4	1.1×10^{-13}
206640	Ribulose-bisphosphate Carboxylase	206647	62688	5722	-3.3	9.3×10^{-25}
152892	Fructose Bisphosphate Aldolase, Fructose 1,6 Bisphosphate Aldolase/Sedoheptulose 1,7 Bisphosphate Aldolase	152892	7226	682	-3.3	1.1×10^{-20}
142120	Carbonic Anhydrase	142120	28	3	-3.1	5.0×10^{-04}
24459	Fructose Bisphosphate Aldolase, Fructose 1,6 Bisphosphate Aldolase/Sedoheptulose 1,7 Bisphosphate Aldolase	24459	84500	9385	-3.0	4.8×10^{-23}
319904	Carbonic Anhydrase	319911	1235	143	-3.0	6.2×10^{-17}
330937	Carbonic Anhydrase	330944	112	14	-2.9	7.1×10^{-08}
391320	Carbonic Anhydrase	422369	974	131	-2.8	1.3×10

						-13

Table 3-3 (cont'd)

196612	PEP Carboxykinase	196617	69162	10305	-2.6	3.5×10 -15
206465	Carbonic Anhydrase	206470	823	128	-2.6	3.7×10 -10
195910	Phosphoribulokinase	195910	20645	3285	-2.5	2.0×10 -14
24120	Carbonic Anhydrase	24120	70159	11772	-2.5	8.1×10 -09
191010	Carbonic Anhydrase	191010	782	138	-2.4	2.3×10 -09
189186	Sedoheptulose-bisphosphatase	189186	32075	5798	-2.3	4.2×10 -14
34358	Carbonic Anhydrase	34358	18254	4870	-1.8	3.8×10 -07
206184	Alanine aminotransferase	206189	14228	3825	-1.8	4.1×10 -07
146322	Phosphoglycerate dehydrogenase	146322	65	18	-1.7	1.6×10 -03
286944	Carbonic Anhydrase	286951	907	250	-1.7	3.2×10 -07
406746	Aldehyde Dehydrogenase	391748	7	2	-1.7	3.5×10 -01
141319	Transketolase	141319	97282	28516	-1.6	6.1×10 -08
325517	Phosphoglycerate mutase	325524	13	4	-1.6	1.5×10 -01
29185	Fructose Bisphosphate Aldolase, Fructose 1,6 Bisphosphate Aldolase/Sedoheptulose 1,7 Bisphosphate Aldolase	29185	1495	474	-1.5	1.1×10 -06
129025	Aconitate Hydratase, Aconitate hydratase	129025	70342	23313	-1.4	3.7×10 -05
196354	Glycine/serine hydroxymethyltransferase, serine hydroxymethyltransferase 2	196354	12105	4042	-1.4	1.6×10 -03
287436	Transaldolase	287443	1807	613	-1.4	1.5×10 -05
55838	Ribose 5 Phosphate isomerase	55838	8796	3045	-1.4	1.5×10 -05
137163	Malate Dehydrogenase, MDH1, Malate dehydrogenase	137163	25524	8848	-1.4	6.1×10 -05
26265	Triosephosphate Isomerase	26265	10127	3600	-1.4	2.3×10 -05

Table 3-3 (cont'd)

128310	Glyoxylate/hydroxypyruvate reductase	128310	6199	2418	-1.2	6.8×10^{-04}
394775	Succinate Dehydrogenase, SDH1	402518	5993	2329	-1.2	1.6×10^{-04}
194461	Glycine/serine hydroxymethyltransferase, serine hydroxymethyltransferase	194461	15168	6210	-1.2	6.9×10^{-04}
194609	Serine-glyoxylate aminotransferase, Serine glyoxylate aminotransferase	194609	15914	6627	-1.1	7.4×10^{-03}
196298	Phosphoglycolate Phosphatase	196298	4399	1863	-1.1	3.6×10^{-03}
18029	Phosphoglucomutase	18029	8013	3463	-1.1	2.3×10^{-03}
24084	Fructose-1,6-Phosphatase	24084	16293	7134	-1.1	6.0×10^{-04}
184506	Glycolate dehydrogenase	184506	18563	8339	-1.0	1.1×10^{-02}
146346	Succinate Dehydrogenase	146346	707	321	-1.0	2.7×10^{-03}
136854	Pyruvate Kinase	136854	8430	14901	1.0	2.4×10^{-03}
205967	Alanine-glyoxylate aminotransferase, Alanine-glyoxylate transaminase 3 (AGT3)	205972	61	114	1.0	2.8×10^{-02}
196400	Glycine/serine hydroxymethyltransferase, serine hydroxymethyltransferase 3	196400	5551	10521	1.1	7.9×10^{-04}
119861	Glucose-6-phosphate 1-Dehydrogenase	119861	430	836	1.1	1.2×10^{-03}
127786	Pyruvate Dehydrogenase	127786	5380	10670	1.1	5.1×10^{-03}
146801	Pyruvate Formate-Lyase	146801	40671	86859	1.2	8.4×10^{-05}
176076	Fructose 6 Phosphate Aldolase	176076	405	952	1.4	2.4×10^{-05}
96789	Glycerol 3-Phosphate Dehydrogenase	96789	692	1735	1.4	1.6×10^{-05}
196832	Malic Enzyme	196837	713	1765	1.4	9.2×10^{-06}
8761	Phosphoglycerate mutase	8761	54	139	1.5	2.1×10^{-04}
192686	Phosphoglycerate mutase, phosphoribosylanthranilate isomerase (ASB2)	192686	611	1863	1.7	2.7×10^{-07}
394095	Carbonic Anhydrase	396265	5	28	2.6	1.6×10

						-03

Table 3-3 (cont'd)

146574	Transaldolase	146574	40761	226224	2.6	1.8×10^{-15}
405437	Phosphoglucomutase	393139	1316	9094	2.9	8.0×10^{-19}
135220	Glucose 6 Phosphate Isomerase	135220	18154	203434	3.6	1.4×10^{-29}
158911	Phosphogluconate dehydrogenase (decarboxylating).	158911	7909	277704	5.3	8.6×10^{-51}
206694	Pyruvate Phosphate dikinase	206701	20	1363	6.2	1.6×10^{-46}
141972	Carbonic Anhydrase	141972	0	1	25.1	$1.0\times10^{+00}$
185846	Carbonic Anhydrase	185846	0	2	26.1	$1.0\times10^{+00}$

[a]PID and TID are the protein and transcript ID for each gene model, for *Chlamydomonas* genome v4.0. [b]Annotation indicates gene name or predicted function. [c]T +N and T –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [d]Log_2FC is the log to base 2 of the difference in hits between the two conditions, for each gene model. [e]FDR refers to the false discovery rate for each gene model.

1-dehydrogenase and phosphogluconate dehydrogenase (decarboxylating), were increased under those conditions. The mRNA encoding one of the pyruvate dehydrogenase subunits represented in the data set was also increased in abundance following N-deprivation. The pyruvate dehydrogenase complex converts pyruvate to acetyl-CoA, which is a precursor of fatty acid biosynthesis. Genes for other enzymes of the glycolytic pathway such as pyruvate kinase did not show very drastic changes in response to the N-deprivation.

Changes in photosynthesis

In *Chlamydomonas*, photosynthetic efficiency decreases following N-deprivation, at least partially due to a reduction in abundance of light harvesting complexes (Peltier, G. and Schmidt, G. W. 1991, Plumley, F. G. and Schmidt, G. W. 1989) and selective degradation of the cytochrome b6f complex (Bulte, L. and Wollman, F. A. 1992, Majeran, W. et al. 2000). Likewise, the abundance of transcripts encoding photosynthesis-related proteins was substantially reduced following N-deprivation. This regulation was not restricted to light harvesting complexes and cytochromes, but extended to the two photosystems as well (Table 3-4). Following N-deprivation, the steady-state level of all nuclear-encoded PSI genes decreased by at least 6-fold, while abundance of transcripts from genes encoding the corresponding light harvesting proteins was decreased even further, resulting in a 19-to-43-fold decrease relative to N-replete conditions. Only four of the cytochrome subunits are encoded by the nuclear genome, and three of them showed a considerable down-regulation (6-fold) following N-deprivation. In contrast, the transcript levels of *PetO* were weakly increased (2-fold). This observation supports the hypothesis that this protein may have a regulatory role as opposed to being a functional cytochrome b6f subunit (Hamel, P. et al. 2000), because the PetO protein is only loosely bound to the complex, and its function is not required for the oxidoreductase activity. Expression of all

Table 3-4. Illumina analysis of transcripts related to photosynthesis

PID[a]	Annotation[b]	TID[a]	T +N[c]	T -N[c]	Log_2 FC[d]	FDR[e]
Cytochrome b6f						
177155	Subunit of the chloroplast cytochrome b6f complex	177155	21892	3053	-2.7	1.2×10^{-19}
151833	Rieske iron-sulfur protein, cytochrome b6f complex iron-sulfur subunit (PetC)	151833	38633	5779	-2.6	4.3×10^{-17}
147968	Cytochrome b6f complex PetM subunit	147968	15961	2591	-2.5	4.1×10^{-15}
185971	Cytochrome b6f complex subunit V (PetO)	185971	10538	16289	0.8	2.1×10^{-02}
Photosystem I						
186531	Photosystem I reaction center subunit N	186531	26768	1045	-4.6	3.4×10^{-42}
165416	Photosystem I reaction center subunit V (PSI-G) (P35 protein)	165416	41221	1668	-4.5	1.5×10^{-38}
193847	Similar to 10-kDa subunit of eukaryotic photosystem I	193847	8807	448	-4.2	1.2×10^{-33}
182959	Formerly called protein P28. See PMID: 2693938.	182959	37431	2409	-3.8	2.9×10^{-32}
76146	Photosystem I reaction center subunit IV, chloroplast precursor (PSI-E) (Photosystem I 8.1 kDa protein)	76146	31352	2041	-3.8	9.0×10^{-30}
130914	Plastocyanin binding subunit, photosystem I reaction center subunit III	130914	102986	8456	-3.5	6.8×10^{-29}
192478	8.4 kD subunit of photosystem I (polypeptide 37) photosystem I reaction center subunit	192478	39226	3316	-3.4	3.6×10^{-28}
205935	Photosystem I reaction center subunit XI	205940	54979	4737	-3.4	8.4×10^{-26}
144056	Chloroplast-targeted; usually encoded in chloroplast genome	144056	36231	4058	-3.0	3.3×10^{-21}
184971	Photosystem I reaction center subunit II, chloroplast precursor (Photosystem I 20 kDa subunit)	184971	94708	13228	-2.7	4.7×10^{-19}
Light harvesting complex of PSI						
186299	Light-harvesting protein of photosystem I	186299	61414	1327	-5.4	6.0×10^{-46}
168073	Photosynthesis light harvesting	168073	53920	1206	-5.4	2.2×10^{-42}
184471	Light-harvesting protein of photosystem I	184471	74134	1742	-5.3	2.6×10^{-49}

Table 3-4 (cont'd)

192961	Light-harvesting protein of photosystem I	192961	48908	1401	-5.0	1.4×10^{-39}
183363	Light-harvesting protein of photosystem I	183363	72458	2704	-4.6	1.9×10^{-43}
187025	Light-harvesting protein of photosystem I	187025	48227	1882	-4.6	7.4×10^{-41}
153678	Chlorophyll a-b binding protein, chloroplast precursor	153678	45360	1892	-4.5	1.5×10^{-39}
144609	Light-harvesting protein of photosystem I	144609	80227	3378	-4.5	2.0×10^{-39}
136294	Encodes a less abundant light-harvesting antenna protein for PSI	136294	78899	3758	-4.3	1.7×10^{-38}
Photosystem II						
33411	Oxygen-evolving enhancer protein 2 of photosystem II	33411	125716	6362	-4.2	6.2×10^{-36}
182015	Encoding a 4.1 kDa subunit for PSII with a single TMH	182015	78088	4044	-4.2	5.8×10^{-33}
130316	Oxygen-evolving enhancer protein 1 33 kDa subunit of oxygen evolving complex of photosystem II	130316	133435	12995	-3.2	1.2×10^{-24}
182560	Ycf32-related subunit of photosystem II	182560	27700	2795	-3.2	4.8×10^{-23}
127879	Conserved expressed lumen targeted protein	127879	2726	298	-3.1	5.2×10^{-19}
153656	Oxygen evolving enhancer protein 3 (PsbQ)	153656	94481	10514	-3.0	4.2×10^{-23}
155150	Photosystem II reaction center W protein	155150	49301	8923	-2.3	8.3×10^{-15}
182896	Photosystem II subunit 28	182896	4736	1028	-2.1	4.0×10^{-11}
205916	Lumenal PsbP-like protein	205921	1079	339	-1.5	1.7×10^{-05}
205877	Shows similarity to PSBP OEE2, oxygen-evolving enhancer protein 2 of photosystem II, OEC23	205882	912	340	-1.3	8.7×10^{-05}
98021	Similar to At1g79040, chloroplast-targeted	98021	34499	13590	-1.2	6.8×10^{-05}
191340	Thylakoid-anchored PsbP-like protein	191340	188	98	-0.8	5.4×10^{-02}
184451	Related to the 23 kDa PsbP subunit of the oxygen evolving system of photosystem II	184451	112	65	-0.7	1.3×10^{-01}
Light harvesting complex of PSII						
184071	Chlorophyll a-b binding protein of LHCII	184071	122368	2165	-5.7	1.5×10^{-42}

Table 3-4 (cont'd)

184067	Encoding a light-harvesting antenna protein for PSII	184067	68968	2358	-4.8	1.5×10^{-40}
184397	Minor chlorophyll a-b binding protein of photosystem II	184397	128319	9439	-3.7	9.2×10^{-28}
185533	Chlorophyll a-b binding protein of LHCII	185533	272555	31261	-3.0	1.5×10^{-19}
195162	Light-harvesting complex II chlorophyll a-b binding protein M3	195162	87662	11039	-2.9	1.9×10^{-15}
187295	Chloroplast protein required for the integration of the light-harvesting complex	187295	3299	636	-2.3	1.5×10^{-10}
184810	Chlorophyll a-b binding protein of photosystem II	184810	149315	29101	-2.2	1.0×10^{-12}
205903	Putative light-harvesting protein; (LHCII type I CAB) (LHCP)	205908	644	644	0.1	8.3×10^{-01}
298826	Photosynthesis light harvesting (PSBS2)	298833	8	383	5.7	1.7×10^{-29}
196341	Photosynthesis light harvesting (PSBS1)	196341	6	648	6.9	4.4×10^{-42}

[a]PID and TID are the protein and transcript ID for each gene model, for *Chlamydomonas* genome v4.0. [b]Annotation indicates gene name or predicted function. [c]T +N and T –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [d]Log_2FC is the log to base 2 of the difference in hits between the two conditions, for each gene model. [e]FDR refers to the false discovery rate for each gene model.

nuclear genes encoding PSII components also decreased following N-deprivation (Table 3-4), although the two least abundant transcripts decreased only slightly. The PSII light harvesting complex transcripts showed a comparable change in abundance. Most of the transcripts levels were reduced, while the weakly expressed *LHCB7* gene showed no alteration in transcript levels.

Changes in lipid metabolism

N-deprivation has been demonstrated to lead to the accumulation of TAG in specialized organelles as well as to structural changes and breakdown of the intracellular membrane systems such as the thylakoids and the ER (Martin, N. C. et al. 1976, Moellering, E. R. and Benning, C. 2010). Therefore, we expected this to be reflected in the expression of genes encoding enzymes of lipid metabolic pathways. However, changes in transcript levels of genes encoding fatty acid metabolism (Table 3-5) were modest (Figure 3-4). A 2-fold increase in transcript levels for ketoacyl-ACP synthetase was observed. This enzyme is part of the fatty acid synthase pathway that catalyses the acyl-acyl carrier protein (acyl-ACP) dependent elongation steps from C4 to C14 in higher plants. The gene for acyl-ACP thioesterase (*FAT1*) also showed elevated transcript levels following N-deprivation (about 4-fold). Its reaction terminates fatty acid synthesis by cleaving the acyl chain from ACP. This reaction competes with the direct transacylation of ACP by glycerol-3-phosphate acyltransferases for the formation of phosphatidate. An increase in FAT1 activity could therefore be indicative of increased fatty acid export from the chloroplast to the ER, where TAG assembly occurs, as acyl-ACPs have to be hydrolyzed prior to export (Pollard, M. and Ohlrogge, J. 1999).

A strong increase in transcript levels was observed for the gene encoding the committing step of TAG synthesis. Out of the five putative diacylglycerol acyltransferases genes identified in the version 4.0 genome dataset, only four were expressed under either or both growth conditions

Table 3-5. Illumina analysis of transcripts related to lipid metabolism

PID[a]	Annotation[b]	TID[a]	T +N[c]	T -N[c]	Log_2FC[d]	FDR[e]
Fatty Acid Synthesis						
205887	KAS1 - 3-Ketoacyl-ACP synthase	205892	10692	20433	1.1	4.8×10^{-04}
196283	FAT1 - Acyl-ACP thiolase	196283	1881	6771	2.0	6.9×10^{-11}
Glycerolipid Metabolism						
196311	Acyl-CoA synthetase	196311	77173	8404	-3.1	9.4×10^{-21}
182408	SAS1 - AdoMet synthetase	182408	75344	9686	-2.8	3.2×10^{-13}
425901	Long-chain acyl-CoA synthetase (LCS)	397197	5998	1851	-1.5	1.3×10^{-06}
391501	Acyl-CoA synthetase	425400	5183	1661	-1.5	5.1×10^{-06}
96283	Phosphatidate phosphatase (PAP)	96283	86	165	1.1	9.2×10^{-03}
411256	Phosphatidate phosphatase (PAP)	415646	101	209	1.2	3.5×10^{-03}
380622	Long-chain acyl-CoA synthetase (LCS)	380629	3927	8198	1.2	1.3×10^{-04}
293423	Phosphatidylinositol synthase (INO)	293430	62	184	1.7	1.2×10^{-05}
285889	acylCoA DAG acyltransferase (DGAT)	285896	3	403	7.2	1.5×10^{-37}
Desaturase						
59933	C5-sterol-desaturase	59933	3469	1516	-1.1	1.3×10^{-03}
205753	FAB2 - Acyl-ACP d9 desaturase	205758	8156	19128	1.4	8.2×10^{-06}
179158	FAD5b - MGDG palmitate-delta7-desaturase	179158	47	131	1.6	1.4×10^{-04}
135825	delta12-desaturase (ER-origin)	135825	967	3170	1.8	9.9×10^{-05}
Lipase						
141065	Esterase/lipase/thioesterase	141065	359	25	-3.7	2.7×10^{-16}
319684	Patatin (AtSDP1 homolog)	319691	1237	357	-1.7	1.3×10^{-06}
416990	Esterase/lipase/thioesterase	409254	945	314	-1.5	4.2×10^{-05}
187121	Lipase, active site (Peptidase S41)	187121	490	170	-1.4	5.4×10^{-05}
205939	Esterase/lipase/thioesterase	205944	552	205	-1.3	8.0×10^{-05}

Table 3-5 (cont'd)

131867	Esterase/lipase/thioesterase	131867	375	151	-1.2	1.9×10^{-03}
189662	Esterase/lipase/thioesterase (Peptidase S10, serine carboxypeptidase)	189662	3465	1432	-1.2	2.1×10^{-04}
286103	Esterase/lipase/thioesterase (Prolylcarboxypeptidase (angiotensinase C))	286110	615	256	-1.2	4.8×10^{-04}
126671	Esterase/lipase/thioesterase (Peptidase S10, serine carboxypeptidase)	126671	2475	1052	-1.1	3.3×10^{-04}
378177	Lipase, class 3	378184	6012	2646	-1.1	7.2×10^{-04}
206048	Phospholipase A2	206053	49295	22454	-1.0	2.8×10^{-03}
294914	Patatin	294921	1216	2287	1.0	1.8×10^{-03}
191980	Esterase/lipase/thioesterase	191980	118	232	1.0	7.0×10^{-03}
186222	Patatin	186222	2678	5216	1.1	2.3×10^{-03}
378954	Lipase, class 3	378961	309	608	1.1	1.9×10^{-03}
174129	Patatin	174129	258	525	1.1	3.9×10^{-03}
71905	Esterase/lipase/thioesterase	71905	159	322	1.1	3.5×10^{-03}
281487	Esterase/lipase/thioesterase (Prolylcarboxypeptidase)	281494	837	1689	1.1	8.4×10^{-04}
425829	Esterase/lipase/thioesterase	399296	180	366	1.1	2.6×10^{-03}
294269	Esterase/lipase/thioesterase	294276	49	101	1.1	1.4×10^{-02}
191796	Lipase, class 3 (FAP199)	191796	5276	11048	1.2	3.7×10^{-04}
406022	Lipase, class 3	391940	149	335	1.3	9.0×10^{-04}
154269	Esterase/lipase/thioesterase	154269	241	541	1.3	2.8×10^{-04}
118056	Esterase/lipase/thioesterase	118056	70	161	1.3	1.3×10^{-03}
166201	Phospholipase A2	166201	33	76	1.3	8.0×10^{-03}
422685	Esterase/lipase/thioesterase	415395	259	592	1.3	1.9×10^{-04}

Table 3-5 (cont'd)

423389	Esterase/lipase/thioesterase (Peptidase S10, serine carboxypeptidase)	426864	426	998	1.3	3.3×10^{-05}
122836	Esterase/lipase/thioesterase	122836	233	543	1.3	9.7×10^{-04}
311312	Patatin	311319	1274	3060	1.4	2.1×10^{-05}
147983	Lipase, active site (Peroxisomal membrane protein MPV17 and related proteins)	147983	18	43	1.4	2.0×10^{-02}
147390	Patatin	147390	671	1633	1.4	6.9×10^{-05}
148415	Esterase/lipase/thioesterase	148415	791	1981	1.4	5.9×10^{-06}
390475	Phospholipase A2	422596	140	379	1.5	2.6×10^{-05}
284764	Esterase/lipase/thioesterase	284771	201	549	1.6	3.8×10^{-06}
191373	Lipase, class 3 (CrLip2)	191373	570	1560	1.6	1.5×10^{-05}
399316	Esterase/lipase/thioesterase	424001	198	545	1.6	3.5×10^{-06}
396930	Phospholipase A2	404320	163	463	1.6	1.2×10^{-05}
345094	Phospholipase A2	345101	15	45	1.7	2.7×10^{-03}
205917	Lipase, active site (Choline transporter-like protein)	205922	1809	5528	1.7	1.6×10^{-07}
287111	Lipase, class 3	287118	96	296	1.7	1.3×10^{-06}
282435	Esterase/lipase/thioesterase	282442	791	2908	2.0	7.3×10^{-10}
424904	Esterase/lipase/thioesterase	397159	328	1246	2.0	1.4×10^{-10}
205924	Predicted lipase/calmodulin-binding heat-shock protein	205929	3235	18756	2.7	2.3×10^{-13}
344422	Esterase/lipase/thioesterase	344429	366	2347	2.8	9.9×10^{-19}
157360	Esterase/lipase/thioesterase	157360	130	1271	3.4	1.2×10^{-22}
148864	Lipase, class 3	148864	16	744	5.6	4.1×10^{-33}
Lipase/Acyltransferase						
381604	Phospholipid/glycerol acyltransferase	381611	435	175	-1.2	9.1×10^{-04}

Table 3-5 (cont'd)

423568	Lecithin:cholesterol acyltransferase (LCAT)/Acyl-ceramide synthase	398425	267	721	1.6	1.5×10^{-06}
421263	Predicted phosphate acyltransferase, contains PlsC domain	394515	116	348	1.7	2.0×10^{-06}
Beta Oxidation						
394387	acyl-CoA oxidase	396229	2606	622	-1.9	9.4×10^{-10}
138637	ATO1 - 3-oxoacyl CoA thiolase	138637	9156	2299	-1.9	4.6×10^{-09}
395830	acyl-CoA oxidase	406257	898	378	-1.1	1.2×10^{-03}
183312	ECH1 - enoylCoA hydratase/isomerase	183312	3060	6117	1.1	4.9×10^{-04}

[a]PID and TID are the protein and transcript ID for each gene model, for *Chlamydomonas* genome v4.0. [b]Annotation indicates gene name or predicted function. [c]T +N and T –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [d]Log_2FC is the log to base 2 of the difference in hits between the two conditions, for each gene model. [e]FDR refers to the false discovery rate for each gene model.

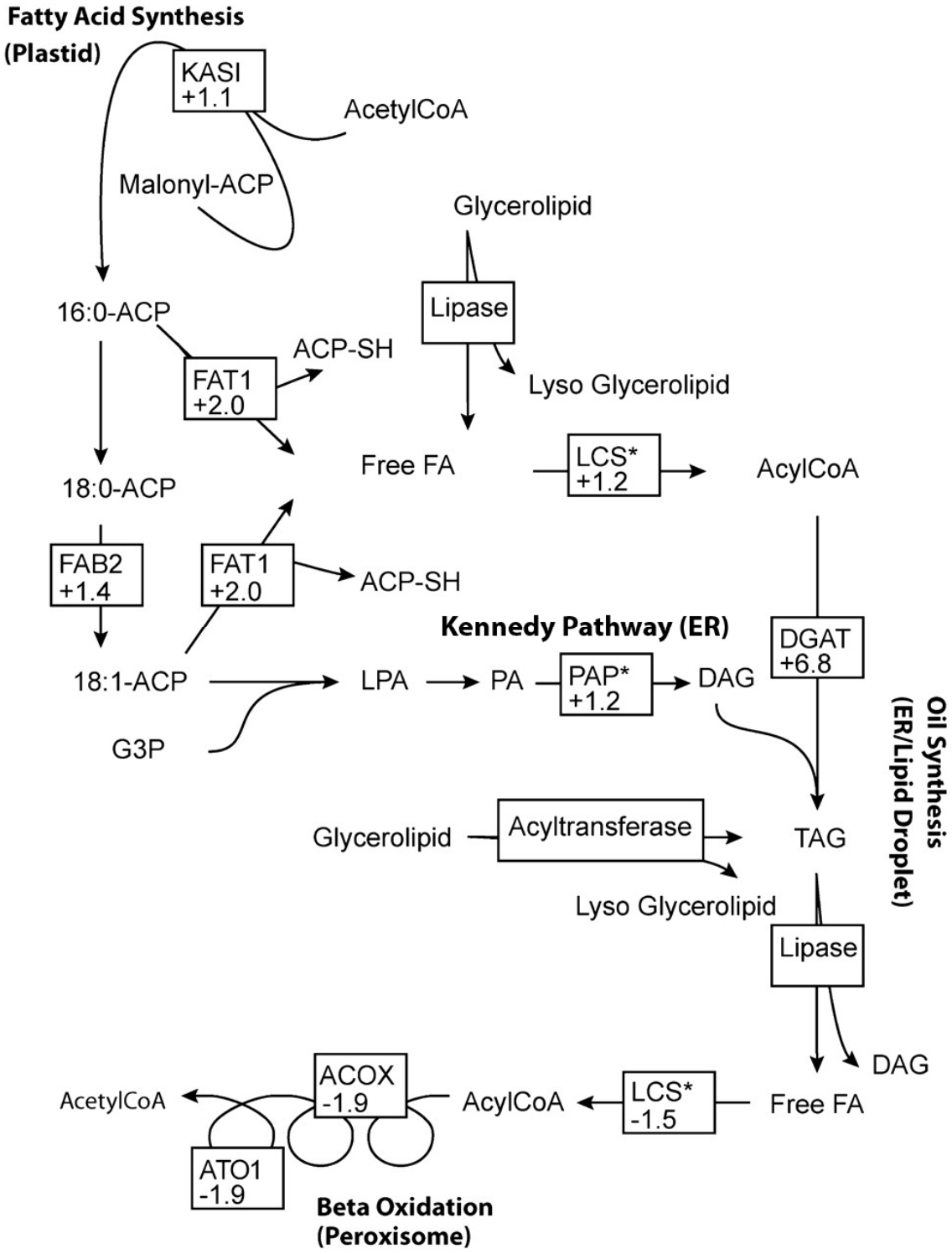


Figure 3-4. Selected changes in glycerolipid metabolism transcript abundance. Numbers indicate $\log_2$ fold change of transcript abundance following N deprivation. Enzymes labeled with an asterisk cannot be unequivocally assigned to a specific step in the metabolic pathway and are hypothetical.

(Table 3-5). One of these genes, *DGTT1*, was almost completely suppressed under N-replete conditions, but showed a large increase in transcript abundance following N-deprivation. However, its overall transcript abundance was too low to be detected by Northern blot compared to other genes encoding putative diacylglycerol acyltransferases, which were much less differentially expressed consistent with the initial RNA-DNA hybridization analysis (Figure 3-1).

Phosphatidic acid phosphatase takes part in the Kennedy pathway of glycerolipid and TAG synthesis (Figure 3-4). Both of the two candidate genes for phosphatidic acid phosphatase in *Chlamydomonas* annotated in the version 4.0 dataset showed increased transcript levels following N-deprivation. Both are part of the PAP2 family, which is thought to have broad substrate specificity (Carman, G. M. and Han, G. S. 2006). The increase in the expression of the presumed phosphatidic acid phosphatase genes is consistent with the notion that DAG is generated from phosphatidic acid for further TAG biosynthesis following N-deprivation.

Out of a total of 16 putative membrane-bound desaturase and hydroxylase encoding genes found in *Chlamydomonas*, only three showed a change in transcript abundance that is greater than 2-fold (Table 3-5). Transcript abundance for microsomal Δ12-desaturase is more than 3-fold higher following N-deprivation as well as that for the plastidic acyl-ACP-Δ9-desaturase, which introduces the first double bond in an acyl chain. Other microsomal desaturase encoding transcripts such as that encoding *FAD13* - an ω13/Δ5-desaturase - are also slightly increased in abundance whereas the plastid desaturase encoding genes are not affected.

Of all lipid-related genes those encoding putative lipases showed the strongest differences in transcript abundance between the two conditions tested. By searching for "lipase", "phospholipase", or "patatin" through the version 4.0 genome sequence data, 130 proteins

containing the GXSXG lipase motif common to hydrolases were identified. Among the respective genes, 35 (27%) showed increased and 11 (8.5%) decreased transcript levels by 2-fold or more following N-deprivation. Table 3-5 lists these 46 most strongly differentially regulated lipase candidates. In addition, many potential lipases may be considered constitutively expressed. 74 out of 130 (57%) lipase candidates were expressed at slightly higher levels following N-deprivation. Some of these genes may encode lipases that are important for the turnover or degradation of membranes during cell growth or gamete fusion.

Changes in transcription factor expression

The Plant Transcription Factor Database (Perez-Rodriguez, P. et al. 2010) was used to identify 386 genes encoding putative transcription factors and transcriptional regulators in the *Chlamydomonas* transcript data set, which could be sorted into 53 families (Table 3-6). Of the 386 genes, 83 showed a 2-fold or greater change in transcript abundance following N-deprivation, with 46 being up-regulated, and 37 being down-regulated.

To date, only a few of the putative transcription factors identified in the *Chlamydomonas* genome have a known function. Transcript abundance for the gene encoding *NIT2*, a transcription factor regulating nitrate metabolism, was increased 6-fold, while that for *NAB1*, a transcription factor regulating light harvesting proteins, was decreased 16-fold, consistent with previously described physiological changes in response to N-deprivation (Camargo, A. et al. 2007, Mussgnug, J. H. et al. 2005). The transcript level for *GSP1* mt+ gamete-specific transcription factor was decreased 3-fold at the 48 hour sample point. When looking at the changes in RNA abundance of putative transcription factor genes, no obvious trends emerged. However, transcripts falling into the AP2-EREBP and bHLH families were generally more abundant following N-deprivation, while those of the FHA family were generally decreased.

Table 3-6. Illumina analysis of transcripts encoding transcription factors and regulators following N-deprivation

PID[a]	Annotation[b]	TID[a]	T +N[c]	T -N[c]	Log_2FC[d]	FDR[e]
Transcription Factors						
115124	SBP	115124	3	0	-26.5	4.2×10^{-01}
425069	CPP	399885	332	5	-5.9	5.1×10^{-23}
126810	NAB1, involved in the light-regulated differential expression of the light-harvesting antenna, CSD	126810	23357	1253	-4.1	3.5×10^{-36}
119948	MYB-related	119948	50	3	-4.0	1.5×10^{-07}
191829	FHA	191829	221	15	-3.8	1.2×10^{-14}
147364	CCAAT	147364	119	9	-3.6	4.7×10^{-10}
290169	bZIP	290176	10444	895	-3.4	6.9×10^{-28}
154254	FHA	154254	302	26	-3.4	9.0×10^{-12}
119194	MYB-related	119194	11	1	-3.4	3.1×10^{-02}
285394	FHA	285401	2150	207	-3.3	4.3×10^{-23}
426624	bZIP	397080	10907	1271	-3.0	2.1×10^{-21}
119283	MYB-related	119283	96	12	-2.9	1.9×10^{-08}
194555	C3H	194555	1764	336	-2.3	2.9×10^{-09}
415440	VARL	411514	843	162	-2.3	6.2×10^{-11}
417527	C2C2-GATA	413714	103	23	-2.1	1.1×10^{-05}
96716	SBP	96716	12	3	-1.9	8.4×10^{-02}
426728	MYB-related	399133	22	6	-1.8	3.3×10^{-02}
186803	GSP1; Gamete-specific transcription factor, HB	186803	980	289	-1.7	1.3×10^{-04}
423729	FHA	397389	391	114	-1.7	1.4×10^{-05}
112237	MYB	112237	502	155	-1.6	3.5×10^{-06}
407701	S1Fa-like	425995	363	120	-1.5	2.0×10^{-04}
104804	CPP	104804	29	10	-1.4	3.8×10^{-02}
285882	bZIP	285889	1004	367	-1.3	2.0×10^{-04}
149734	C2H2	149734	6593	2748	-1.2	1.2×10^{-03}
171165	TIG	171165	16705	7723	-1.0	1.8×10^{-03}
415443	bHLH	409330	181	332	1.0	7.9×10^{-03}
393055	bZIP	421105	21	39	1.0	1.0×10^{-01}
393622	bHLH	420168	317	593	1.0	6.7×10^{-03}
399062	HSF	422722	1776	3323	1.0	1.8×10^{-03}
186995	RWP-RK	186995	3085	5907	1.0	7.9×10^{-04}
416794	AP2-EREBP	410803	313	612	1.1	2.4×10^{-03}
426843	PLATZ	397100	80	158	1.1	6.8×10^{-03}
298349	MYB-related	298356	451	912	1.1	2.8×10^{-03}
205561	MYB-related	205566	5271	11132	1.1	1.4×10^{-02}
147286	bZIP	147286	906	1889	1.2	2.8×10^{-04}
419111	C3H	412152	329	704	1.2	3.5×10^{-04}
342954	MYB-related	342961	757	1636	1.2	2.2×10^{-04}
130971	G2-like	130971	300	680	1.3	3.5×10^{-04}

Table 3-6 (cont'd)

417388	MYB-related	413124	1291	2917	1.3	1.5×10^{-04}
187531	bZIP	187531	803	1897	1.3	2.2×10^{-04}
189471	CCAAT	189471	79	189	1.4	3.8×10^{-04}
196473	G2-like	196473	551	1349	1.4	1.1×10^{-05}
415966	MYB-related	409909	4	10	1.4	2.7×10^{-01}
321493	TAZ	321500	9879	25105	1.4	5.9×10^{-04}
205919	AP2-EREBP	205924	23	61	1.5	4.1×10^{-03}
347478	none	347485	931	2593	1.6	1.8×10^{-05}
205871	C2C2-GATA	205876	1867	5390	1.6	3.7×10^{-07}
19751	AP2-EREBP	19751	297	925	1.7	1.1×10^{-06}
183777	bHLH	183777	6895	21726	1.8	1.3×10^{-08}
290479	SBP	290486	189	616	1.8	1.8×10^{-07}
195916	RWP-RK	195916	11	36	1.8	6.8×10^{-03}
284832	MYB-related	284839	142	532	2.0	6.3×10^{-09}
148404	AP2-EREBP	148404	1324	5634	2.2	1.4×10^{-10}
414856	SBP	411130	1520	6689	2.2	2.8×10^{-12}
146990	VARL	146990	14	61	2.2	3.8×10^{-05}
205642	Nit2, NIT2 NIT2, NIT2,Nitrate regulated transcription factor	205647	1366	7924	2.6	7.5×10^{-18}
117291	MYB	117291	117	740	2.8	1.1×10^{-13}
195838	RWP-RK	195838	108	718	2.8	3.4×10^{-16}
116658	MYB-related	116658	56	452	3.1	2.9×10^{-16}
153934	bHLH	153934	82	705	3.2	2.5×10^{-19}
195891	RWP-RK	195891	404	3825	3.3	8.9×10^{-25}
424240	C3H	397539	19	273	3.9	8.7×10^{-19}
118761	SBP	118761	1	42	5.5	2.4×10^{-08}
405949	bHLH	390447	10	1821	7.6	4.3×10^{-62}
177225	bHLH	177225	7	1666	8.0	1.2×10^{-62}
Transcriptional Regulators						
206670	SNF2	206677	106	7	-3.8	3.3×10^{-09}
143060	PHD	143060	332	24	-3.7	2.1×10^{-15}
401818	PHD	417802	7	2	-1.7	3.5×10^{-01}
325701	TRAF	325708	3740	1118	-1.6	3.2×10^{-07}
146398	TRAF	146398	372	113	-1.6	5.6×10^{-06}
169174	SNF2	169174	209	67	-1.5	8.1×10^{-05}
287740	PHD	287747	3	1	-1.5	7.9×10^{-01}
377090	GNAT	377097	821	290	-1.4	3.9×10^{-04}
151030	TRAF	151030	406	156	-1.3	4.1×10^{-04}
142283	HMG	142283	11647	4587	-1.2	1.7×10^{-03}
142152	GNAT	142152	189	80	-1.1	4.7×10^{-03}
172711	SET	172711	549	233	-1.1	3.5×10^{-03}
115484	SET	115484	39	71	1.0	5.8×10^{-02}
193146	TRAF	193146	1032	1899	1.0	2.5×10^{-03}
145759	TRAF	145759	503	1004	1.1	2.2×10^{-03}

Table 3-6 (cont'd)

154505	SET	154505	1304	2760	1.2	4.5×10^{-04}
308637	PHD	308644	20	49	1.4	1.3×10^{-02}
188181	TRAF	188181	265	723	1.5	3.5×10^{-06}
282628	GNAT	282635	584	1684	1.6	2.1×10^{-06}
192899	HMG	192899	1261	4097	1.8	3.7×10^{-08}
205788	GNAT	205793	215	782	2.0	1.4×10^{-09}
423513	HMG	398414	149	1178	3.1	2.5×10^{-19}
321619	PHD	321626	0	2	26.0	$1.0\times10^{+00}$
143723	GNAT	143723	0	3	26.6	2.5×10^{-01}

[a]PID and TID are the protein and transcript ID for each gene model, for *Chlamydomonas* genome v4.0. [b]Annotation indicates gene name or predicted function. [c]T +N and T –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [d]Log_2FC is the log to base 2 of the difference in hits between the two conditions, for each gene model. [e]FDR refers to the false discovery rate for each gene model.

Discussion

This experiment used transcript profiling to create a general overview of changes in cell metabolism during nitrogen deprivation, as well as to find genes of interest for further analysis. The first step was to determine the proper growth conditions for RNA library generation. Due to cost constraints, we were only able to sequence RNA from two conditions at a single time point, comparing N-replete and N-deprived cells. I used Northern blotting to compare the expression of three different genes under different conditions: 0.5M N continuously, switch to 0M N for 24 hours, or switch to 0M N for 48 hours. Of the genes used for testing, *AMT4* was chosen as a known regulated gene, while *DGTT2* and *DGTT3* were presumed to be regulated by N-deprivation, being potential TAG synthesis genes. In all three conditions, *AMT4* showed strong differences in expression level, while *DGTT2* and *DGTT3* showed little or no difference. Nitrogen deprivation for 48 hours was finally chosen, as it showed the greatest contrast in *DGTT3* expression levels. Although TAG synthesis and accumulation begin earlier, the 48 hour time point gives an overview of the steady-state condition of the cell after it has adjusted to the N-deprived condition, indicating the major metabolic changes that occur and contribute to TAG synthesis, among other physiological changes. Although most regulators that control responses to N-deprivation would likely change expression earlier, the change would likely last through the 48 hour time point, particularly in the case of secondary responses, such as TAG accumulation. Indeed, *NIT2*, the transcription factor that regulates nitrate uptake, shows a 6-fold increase after 48 hours of N-deprivation.

To sequence the transcriptomes from N-replete or N-deprived *Chlamydomonas* cultures, we used two different methods. The first method used 454 sequencing technology. This procedure generated considerable amounts of data, but had several limitations. First, despite the

large number of reads, only ~40% of the total estimated genes in *Chlamydomonas* had one or more hits. This means a large number of genes had no information about their expression level, and suggests that the transcripts sequenced were not evenly distributed, thus suggesting that the data may not be reliable. Second, and related, only one biological replicate for each condition was sequenced, limiting the statistical significance of the results. Despite this, certain trends were observed. Genes related to nitrogen metabolism were up regulated, while genes related to photosynthesis were down regulated. Both of these trends match previously reported physiological changes resulting from N-deprivation, and were expected. An unexpected result was that genes involved in several primary carbon metabolic pathways also showed significant changes in expression. Most notably, genes encoding malate synthase and isocitrate lyase, key enzymes in the glyoxylate cycle, were strongly down regulated during N-deprivation. These results suggest how carbon metabolism is altered by N-deprivation, to change the flow of carbon into starch and TAG.

Due to the statistical limitations of the 454 data set, the transcriptome sequencing was repeated using Illumina technology, with *Chlamydomonas* grown under the same conditions. This time, three biological replicates per condition were sequenced. Although the individual reads were shorter than with 454, the total number of reads was much greater. When matched to the genome, ~95% of predicted genes had at least one hit. In addition, the total number of hits was higher, and the statistics were more reliable, due to the replicates. Therefore, the Illumina data set was used for further analysis. Like the 454 data set, the Illumina data set showed certain expected trends, such as genes for N-metabolism being up regulated, while genes for photosynthesis were down regulated.

As with the 454 data set, the overall carbon metabolism pathways were examined for changes in gene expression. As before, isocitrate lyase and malate synthase were significantly down regulated. PCK, the committed step in gluconeogenesis, and several genes involved in carbon fixation were also down regulated. In contrast, a pyruvate dehydrogenase, which converts pyruvate to acetyl-CoA, was up regulated, along with several genes in the pentose phosphate pathway.

These results indicate a significant change in the way carbon is metabolized in the N-deprived cells. Carbon flux appears to be bypassing normal carbon fixation (due to a decrease in photosynthesis) and sugar generation pathways, in favor of generation of pyruvate and acetyl-CoA, precursors for fatty acid synthesis. A flux analysis experiment was performed by a collaborator, Dr. Rahul Deshpande from the Shachar-Hill lab, to confirm the results from the transcript profiling. The changes in the level of key metabolites matched the results expected from the transcript changes, showing that the changes in metabolism were real (Miller, R. et al. 2010).

The lipid pathways were also examined for changes in gene expression levels. Given the increase seen in TAG and the breakdown of thylakoid membranes that occurs during N-deprivation, it is logical to assume that major changes in the lipid synthesis and degradation pathways occur. However, only a few genes involved in lipid synthesis showed significant changes in expression. Those that do change include two genes involved in fatty acid synthesis, and one of the DGAT genes, indicating some increase in lipid synthesis, but it seems that most of the increase in TAG is due to changes in other pathways altering the carbon flux into lipid pathways.

The class of lipid gene that did show large changes in expression is the lipase genes. Many genes from this family were either up- or down-regulated under N-deprivation. Given the role of lipases in both membrane lipid and TAG breakdown, this expression pattern can be interpreted to account for the different roles of lipases. It is likely that the lipases that are up-regulated are specific for membrane lipids, and serve to break down the photosynthetic membranes and release fatty acids for incorporation into TAG. The down-regulated lipases, on the other hand, are likely TAG lipases, that ordinarily break down TAG, but which are deactivated to allow TAG to accumulate. Due to the lipase mechanism, it is difficult to determine the target of a lipase from its sequence, and thus biochemical analysis is required to confirm the role of the affected lipases. This work is being done by others in the Benning lab, and so far two lipases have been characterized. One lipase, up-regulated by N-deprivation, has been shown to target the glycolipid monogalactosyldiacylglycerol, and apparently acts to shuttle newly-synthesized 18:1 fatty acid to the ER for incorporation into TAG and other extra-plastidic lipids (Li, X. et al. 2012b). Another, down-regulated under N-deprivation, has been shown to complement a *tgl3Δtgl4Δ* yeast strain, indicating its function as a TAG lipase (Li, X. et al. 2012a).

Other metabolic pathways were examined in detail, and the results matched expectations fairly well. Photosynthetic genes were predominantly down-regulated, matching the decrease in photosynthetic activity, and breakdown of thylakoids, independently observed. The affected genes included chlorophyll synthesis, light harvesting complexes, and photosystem I and II, indicating a broad down-regulation. One exception to this is the cytochrome *PetO*, which was weakly up-regulated. This suggests a different function for this gene, possibly regulatory rather

than enzymatic, and serves as an example of the unexpected results that can be gleaned from this kind of global analysis.

When putative transcription factors were analyzed, many were found to be up- or down-regulated. Much like the lipases, the exact function of a transcription factor is impossible to determine solely from the protein sequence, and thus further biochemical analysis is needed to determine their role in N-deprivation. My attempts to characterize several of the regulated transcription factors are detailed in the next chapter.

BIBLIOGRAPHY

BIBLIOGRAPHY

Ball,S.G., Dirick,L., Decq,A., Martiat,J.C., and Matagne,R.F. (1990) Physiology of starch storage in the monocellular alga *Chlamydomonas reinhardtii*. Plant.Sci **66**:1-9.

Bloom,J.S., Khan,Z., Kruglyak,L., Singh,M., and Caudy,A.A. (2009) Measuring differential gene expression by short read sequencing: quantitative comparison to 2-channel gene expression microarrays. BMC.Genomics **10**:221.

Bolling,C. and Fiehn,O. (2005) Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. Plant.Physiol **139**:1995-2005.

Bullard,J.H., Purdom,E., Hansen,K.D., and Dudoit,S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC.Bioinformatics. **11**:94.

Bulte,L. and Wollman,F.A. (1992) Evidence for a selective destabilization of an integral membrane protein, the cytochrome b6/f complex, during gametogenesis in *Chlamydomonas reinhardtii*. Eur.J.Biochem. **204**:327-336.

Camargo,A., Llamas,A., Schnell,R.A., Higuera,J.J., Gonzalez-Ballester,D., Lefebvre,P.A., Fernandez,E., and Galvan,A. (2007) Nitrate signaling by the regulatory gene *NIT2* in *Chlamydomonas*. Plant Cell **19**:3491-3503.

Carman,G.M. and Han,G.S. (2006) Roles of phosphatidate phosphatase enzymes in lipid metabolism. Trends.Biochem.Sci. **31**:694-699.

Erickson,B., Stern,D.B., and Higgs,D.C. (2005) Microarray analysis confirms the specificity of a *Chlamydomonas reinhardtii* chloroplast RNA stability mutant. Plant.Physiol **137**:534-544.

Fan,J., Yan,C., Andre,C., Shanklin,J., Schwender,J., and Xu,C. (2012) Oil accumulation is controlled by carbon precursor supply for fatty acid synthesis in *Chlamydomonas reinhardtii*. Plant.Cell.Physiol **53**:1380-1390.

Fisslthaler,B., Meyer,G., Bohnert,H.J., and Schmitt,J.M. (1995) Age-dependent induction of pyruvate, orthophosphate dikinase in *Mesembryanthemum crystallinum* L. Planta **196**:492-500.

Girke,T., Todd,J., Ruuska,S., White,J., Benning,C., and Ohlrogge,J. (2000) Microarray analysis of developing Arabidopsis seeds. Plant.Physiol **124**:1570-1581.

Goodenough,U., Lin,H., and Lee,J.H. (2007) Sex determination in *Chlamydomonas*. Semin.Cell.Dev.Biol **18**:350-361.

Grossman,A. (2000) Acclimation of *Chlamydomonas reinhardtii* to its nutrient environment. Protist **151**:201-224.

Hamel,P., Olive,J., Pierre,Y., Wollman,F.A., and de,V.C. (2000) A new subunit of cytochrome b6f complex undergoes reversible phosphorylation upon state transition. J.Biol.Chem. **275**:17072-17079.

Harris,E.H. (1989) *The Chlamydomonas Source Book*. San Diego: Academic.

Kucho,K., Okamoto,K., Tabata,S., Fukuzawa,H., and Ishiura,M. (2005) Identification of novel clock-controlled genes by cDNA macroarray analysis in *Chlamydomonas reinhardtii*. Plant.Mol.Biol **57**:889-906.

Langmead,B., Trapnell,C., Pop,M., and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome.Biol **10**:R25.

Li,X., Benning,C., and Kuo,M.H. (2012a) Rapid Triacylglycerol Turnover in *Chlamydomonas reinhardtii* Requires a Lipase with Broad Substrate Specificity. Eukaryot.Cell **11**:1451-1462.

Li,X., Moellering,E.R., Liu,B., Johnny,C., Fedewa,M., Sears,B.B., Kuo,M.H., and Benning,C. (2012b) A Galactoglycerolipid Lipase Is Required for Triacylglycerol Accumulation and Survival Following Nitrogen Deprivation in *Chlamydomonas reinhardtii*. Plant Cell **24**:4670-4686.

Majeran,W., Wollman,F.A., and Vallon,O. (2000) Evidence for a role of ClpP in the degradation of the chloroplast cytochrome b(6)f complex. Plant Cell **12**:137-150.

Mamedov,T.G., Moellering,E.R., and Chollet,R. (2005) Identification and expression analysis of two inorganic C- and N-responsive genes encoding novel and distinct molecular forms of eukaryotic phosphoenolpyruvate carboxylase in the green microalga *Chlamydomonas reinhardtii*. Plant.J **42**:832-843.

Marioni,J.C., Mason,C.E., Mane,S.M., Stephens,M., and Gilad,Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome.Res **18**:1509-1517.

Martin,N.C., Chiang,K.S., and Goodenough,U.W. (1976) Turnover of Chloroplast and Cytoplasmic Ribosomes During Gametogenesis in *Chlamydomonas reinhardi*. Dev.Biol **51**:190-201.

Matthew,T., Zhou,W., Rupprecht,J., Lim,L., Thomas-Hall,S.R., Doebbe,A., Kruse,O., Hankamer,B., Marx,U.C., Smith,S.M., and Schenk,P.M. (2009) The metabolome of *Chlamydomonas reinhardtii* following induction of anaerobic H2 production by sulfur depletion. J.Biol.Chem. **284**:23415-23425.

Maxam,A.M. and Gilbert,W. (1977) A new method for sequencing DNA. Proc.Natl.Acad.Sci.U.S.A **74**:560-564.

Miller,R., Wu,G., Deshpande,R.R., Vieler,A., Gartner,K., Li,X., Moellering,E.R., Zauner,S., Cornish,A.J., Liu,B., Bullard,B., Sears,B.B., Kuo,M.H., Hegg,E.L., Shachar-Hill,Y., Shiu,S.H., and Benning,C. (2010) Changes in transcript abundance in *Chlamydomonas*

reinhardtii following nitrogen deprivation predict diversion of metabolism. Plant.Physiol **154**:1737-1752.

Miura,K., Yamano,T., Yoshioka,S., Kohinata,T., Inoue,Y., Taniguchi,F., Asamizu,E., Nakamura,Y., Tabata,S., Yamato,K.T., Ohyama,K., and Fukuzawa,H. (2004) Expression profiling-based identification of CO2-responsive genes regulated by CCM1 controlling a carbon-concentrating mechanism in *Chlamydomonas reinhardtii*. Plant.Physiol **135**:1595-1607.

Moellering,E.R. and Benning,C. (2010) RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. Eukaryot.Cell **9**:97-106.

Msanne,J., Xu,D., Konda,A.R., Casas-Mollano,J.A., Awada,T., Cahoon,E.B., and Cerutti,H. (2012) Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae *Chlamydomonas reinhardtii* and *Coccomyxa* sp. C-169. Phytochemistry **75**:50-59.

Mussgnug,J.H., Wobbe,L., Elles,I., Claus,C., Hamilton,M., Fink,A., Kahmann,U., Kapazoglou,A., Mullineaux,C.W., Hippler,M., Nickelsen,J., Nixon,P.J., and Kruse,O. (2005) NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. Plant Cell **17**:3409-3421.

Oshlack,A. and Wakefield,M.J. (2009) Transcript length bias in RNA-seq data confounds systems biology. Biol.Direct **4**:14.

Osteras,M., Driscoll,B.T., and Finan,T.M. (1997) Increased pyruvate orthophosphate dikinase activity results in an alternative gluconeogenic pathway in *Rhizobium* (*Sinorhizobium*) *meliloti*. Microbiology **143**:1639-1648.

Peltier,G. and Schmidt,G.W. (1991) Chlororespiration: an adaptation to nitrogen deficiency in *Chlamydomonas reinhardtii*. Proc.Natl.Acad.Sci.U.S.A **88**:4791-4795.

Perez-Rodriguez,P., Riano-Pachon,D.M., Correa,L.G., Rensing,S.A., Kersten,B., and Mueller-Roeber,B. (2010) PlnTFDB: updated content and new features of the plant transcription factor database. Nucleic.Acids.Res **38**:D822-D827.

Plumley,F.G. and Schmidt,G.W. (1989) Nitrogen-dependent regulation of photosynthetic gene expression. Proc.Natl.Acad.Sci.U.S.A **86**:2678-2682.

Pollard,M. and Ohlrogge,J. (1999) Testing models of fatty acid transfer and lipid synthesis in spinach leaf using in vivo oxygen-18 labeling. Plant.Physiol **121**:1217-1226.

Robinson,M.D., McCarthy,D.J., and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics **26**:139-140.

Sanger,F., Nicklen,S., and Coulson,A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc.Natl.Acad.Sci.U.S.A **74**:5463-5467.

Siersma,P.W. and Chiang,K.S. (1971) Conservation and Degradation of Cytoplasmic and Chloroplast Ribosomes in *Chlamydomonas reinhardtii*. J.Mol.Biol **58**:167-185.

Storey,J.D. (2003) The Positive False Discovery Rate: A Bayesian Interpretation and the *q*-Value. Ann.Stat. **31**:2013-2035.

Taylor,L., Nunes-Nesi,A., Parsley,K., Leiss,A., Leach,G., Coates,S., Wingler,A., Fernie,A.R., and Hibberd,J.M. (2010) Cytosolic pyruvate,orthophosphate dikinase functions in nitrogen remobilization during leaf senescence and limits individual seed growth and nitrogen content. Plant.J **62**:641-652.

Trapnell,C., Pachter,L., and Salzberg,S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. Bioinformatics **25**:1105-1111.

Trapnell,C., Williams,B.A., Pertea,G., Mortazavi,A., Kwan,G., van Baren,M.J., Salzberg,S.L., Wold,B.J., and Pachter,L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat.Biotechnol **28**:511-515.

Wang,L., Feng,Z., Wang,X., Wang,X., and Zhang,X. (2010) DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics. **26**:136-138.

Weers,P.M.M. and Gulati,R.D. (1997) Growth and reproduction of *Daphnia galeata* in response to changes in fatty acids, phosphorus, and nitrogen in *Chlamydomonas reinhardtii*. Limnol.Oceanogr **42**:1584-1589.

Wu,T.D. and Watanabe,C.K. (2005) GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics **21**:1859-1875.

Zhang,Z.D., Shrager,J., Jain,M., Chang,C.W., Vallon,O., and Grossman,A.R. (2004) Insights into the survival of *Chlamydomonas reinhardtii* during sulfur starvation based on microarray analysis of gene expression. Eukaryot.Cell **3**:1331-1348.

CHAPTER 4: TESTING OF PUTATIVE TRANSCRIPTION FACTORS

Introduction

Transcription factors are an important part of the regulation of metabolic processes. These proteins act to control the expression of multiple genes, usually as part of a cascade response to specific physiological signals. Transcription factors are important targets for engineering purposes, as controlling their activity allows control of the key genes in a given metabolic pathway. However, it is hard to find the specific targets of a given transcription factor, or conversely the transcription factor controlling a given gene, using bioinformatics only. Generally, biochemical or genetic experiments have to be performed to confirm the role of a given transcription factor.

Transcription factors are divided into many different families, based on their structure. All transcription factors contain DNA-binding domains, which target specific DNA sequences, and activation domains, which interact with activator or repressor proteins. Transcription factors themselves are regulated, and are activated or inactivated in response to physiological signals. The regulation may be through changes in transcription level, or changes in the location or function of the protein. One common regulation strategy is to modify the existing transcription factor protein, resulting in it being transported from the cytosol to the nucleus, where it can function.

In plants, triacylglycerol (TAG) synthesis occurs primarily in the seed, and thus the regulation of TAG synthesis is tied to seed development and embryogenesis. Work in Arabidopsis has uncovered many aspects of the regulatory network involved in the control of TAG synthesis, and how it is tied into the overall seed development regulatory network. The key transcription factor is Wrinkled1 (WRI1), which directly controls TAG synthesis in seeds. This

transcription factor was discovered over a decade ago (Focks, N. and Benning, C. 1998) in a mutant screen, and the mutant seeds have significantly less TAG compared to wild type, leading to smaller seeds with shrunken, wrinkled seed coats (hence the name). *WRI1* has been isolated and found to be part of the AP2/EREBP family, which contains many transcription factors involved in development (Cernac, A. and Benning, C. 2004). Overexpression of *WRI1* leads to an increase in seed TAG, confirming its role. Further research has shown that *WRI1* is correlated with the expression of several genes involved in carbon metabolism, specifically the later steps of glycolysis and fatty acid synthesis (Baud, S. et al. 2007b).

Several other transcription factors have been identified that are involved in seed development (Lotan, T. et al. 1998, Yang, Y. et al. 2011). One of them, Leafy Cotyledon 2 (LEC2), has been shown to lead to accumulation of TAG in vegetative tissues when expressed ectopically (Santos, Mendoza M. et al. 2005). Co-expression with *DGAT1* has also been shown to dramatically increase the amount of TAG induced. Comparison of the expression pattern of *WRI1* and *LEC2* showed that both were expressed at the same time. The relationship between the two was tested by expressing a construct containing the *WRI1* promoter attached to a reporter gene in a *LEC2* mutant (Baud, S. et al. 2007a). Analysis of the transformants shows that the reporter gene's expression is low in the mutant, indicating that *LEC2* is responsible for regulating *WRI1* expression, and thus indirectly controlling TAG synthesis. This result suggests that the TAG accumulation seen in *LEC2* overexpressors is due primarily to the resulting expression of *WRI1*. *LEC2* is itself controlled by the expression of *LEC1*, another transcription factor.

The regulatory network involved in TAG synthesis of Arabidopsis serves as a model for that in *Chlamydomonas reinhardtii*. The exact shape of the network, as well as the triggering factors and signals, will obviously be different. However, the multiple layers of control, with

transcription factors controlling specific pathways, like WRI1, being themselves regulated by more general factors, such as LEC2, are likely to be repeated in *Chlamydomonas*, with general transcription factors responding to environmental change, such as nitrogen deprivation, and then regulating downstream processes, including TAG synthesis, by activating transcription factors specific to those processes. In addition, the examples from Arabidopsis of overexpression of *WRI1* and *LEC2* show how powerful transcription factors can be in engineering biofuels crops. Ectopic expression of a single gene, under an inducible promoter, can activate the structural genes involved in TAG synthesis and lead to significant amounts of TAG accumulating, ideally shortly before harvesting. Thus, the identification of the transcription factors controlling TAG synthesis in *Chlamydomonas* could be an important step in increasing overall TAG yield.

To date, relatively few *Chlamydomonas* transcription factors have been identified and studied in detail. However, an overall survey of putative transcription factor candidates in *Chlamydomonas* identified 234 putative transcription factors and transcriptional regulators, from 39 gene families (Riano-Pachon, D. M. et al. 2008). Analysis of these genes reveals certain overall changes relating to the evolution of *Chlamydomonas* and other organisms. For example, the transcription factor families of C2H2 and HB (in animals) and MADS-box (in plants) are involved in the regulation of development, and contain many copies (~80-100) in these organisms. In *Chlamydomonas* and other unicellular algae, in contrast, these gene families contain only a few members (<10). Comparison of transcription factor families in *Chlamydomonas* and other species reveals several families involved in the various stages of acquisition of photosynthesis, including some families shared between red and green algae, the most ancient. Some other transcription factor families are found only in land plants, which evolved more recently. These land-plant-specific families are likely involved in multicellularity

or cellular differentiation, or adaptation to growth on land, and thus are not necessary in *Chlamydomonas*.

Although relatively few *Chlamydomonas* transcription factors are known, a few have been identified. One example is NIT2, a transcription factor that regulates nitrate signaling (Camargo, A. et al. 2007). This gene is from the RWP-RK family, and responds to the absence of nitrate by activating genes involved in the assimilation of nitrate. The presence of ammonia represses the expression of *NIT2*, allowing the *Chlamydomonas* cell to adjust its physiology to differing nitrogen sources in the environment. Another important transcription factor known in *Chlamydomonas* is MID, another RWP-RK transcription factor that is involved in mating (Goodenough, U. et al. 2007). The two mating types in *Chlamydomonas*, *plus* and *minus*, are controlled by the type-specific loci MT+ and MT-. MID is specific to the MT- locus, and is required for proper expression of *minus*-specific genes and repression of *plus*-specific genes, and subsequent successful mating of the gametes. A different form of transcriptional regulator is represented by NAB1, which regulates the light harvesting complex proteins post-transcriptionally, by binding to the mRNA and blocking translation (Mussgnug, J. H. et al. 2005). These three *Chlamydomonas* transcriptional regulators are examples of how various processes are regulated in *Chlamydomonas*. It is likely that TAG synthesis has similar regulatory elements, and their identification will be key to understanding the overall process.

In the previous chapter, I described a transcript profiling experiment performed on *Chlamydomonas*, comparing nitrogen replete and nitrogen deprived cells. One of the goals of this experiment was to identify possible transcriptional regulators of TAG synthesis. The rationale is based on the induction of TAG synthesis by nitrogen deprivation, a process that is likely to be regulated through transcriptional controls. Thus, transcription factors that show

altered transcription levels under our conditions would be targets for further analysis. Several candidate transcription factors were selected to be tested by expression in *Chlamydomonas*, to see if they have a measurable effect. Since the role of the selected transcription factors is unknown, several phenotypes were examined, not just TAG accumulation.

Methods

Selection of candidates

Two sets of expression data were generated, one using 454 technology and the other using Illumina technology, as previously described. Putative transcription factors were identified using the database compiled in (Riano-Pachon, D. M. et al. 2007). The 454 dataset, having fewer reads and fewer gene hits, gave five transcription factor (TF) candidates with significant increases in transcript level, and none with a decrease in transcript level (Table 4-1). These transcription factors were named TF1-5. In the Illumina dataset, many more transcription factor candidates were identified: ~83 candidates had significant differences between the nitrogen replete and nitrogen deprived conditions. The five transcription factors identified in the 454 dataset did not show significant differences in the Illumina dataset, likely due to the difference in total number of hits between the two sets.

Of the transcription factor genes identified in the Illumina dataset, multiple genes were either up- or down-regulated. Several of these genes were selected for further testing, based on their change in expression, their total amount of expression, and the length of their predicted gene model. Three were successfully cloned and fully analyzed, and named TFU1, TFD2 and TFD3 (Table 4-2). *TFU1* was up-regulated under nitrogen deprivation, and *TFD2* and *TFD3* were both down-regulated.

Table 4-1. Transcription factors identified in the 454 dataset and tested

name	PID[a]	TID[a]	+N[b]	-N[b]	Z score[c]
TF1	159133	159133	8	27	-3.7
TF2	185094	185094	12	29	-3.2
TF3	205561	205566	0	8	-3.1
TF4	114109	114109	0	7	-2.9
TF5	157388	157388	2	10	-2.6

[a]PID and TID are the protein and transcript ID for each gene model, for C. reinhardtii genome v4.0. [b]+N and –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [c]Z score is a statistical analysis of the fold change, and significance of the change, with ±2.5 being significant

Table 4-2. Transcription factors identified in the Illumina dataset and tested

name	PID[a]	TID[a]	T +N[b]	T -N[b]	Log_2FC[c]	FDR[d]
TFU1	177225	177225	6	1609	8.19	$4.72X10^{-63}$
TFD2	191829	191829	220	14	-3.83	$7.57X10^{-15}$
TFD3	147364	147364	119	9	-3.57	$1.80X10^{-9}$

[a]PID and TID are the protein and transcript ID for each gene model, for C. reinhardtii genome v4.0. [b]T +N and T –N are the total number of hits for each gene model, from either N-replete or N-deprived conditions. [c]Log_2FC is the log to base 2 of the difference in hits between the two conditions, for each gene model. [d]FDR refers to the false discovery rate for each gene model.

Cloning and expression of genes

TF1-5 were cloned from cDNA and ligated into vector pJR38 (Neupert, J. et al. 2009), which uses the *PSAD* promoter and terminator for transgene expression (Fischer, N. and Rochaix, J. D. 2001) (Figure 4-1). This vector was modified to contain a small multiple cloning site for easier cloning. *TFU1*, *TFD2* and *TFD3* were cloned and assembled half from cDNA and half from genomic DNA, with the resulting constructs containing at least one native intron, to improve expression rate (Lumbreras, V. et al. 1998). The transcription factor candidates were amplified using the primers listed in Table 4-3, from *Chlamydomonas* cDNA and genomic DNA. The gene fragments were digested with the appropriate restriction enzymes and cloned into the modified pJR38 vector. The vectors were transformed into *Chlamydomonas* strain UVM4 (Neupert, J. et al. 2009), using the glass beads method (Harris, E. H. 1989). The transformants were selected and maintained on TAP + paromomycin.

Testing of phenotypes

To test the various phenotypes, *TF1-5* and *TFU1* transformants were grown under nitrogen-replete conditions, while *TFD2* and *TFD3* were grown under nitrogen-deprived conditions. The rationale for this was the assumption that expression of the TF when it is normally repressed would give the most obvious phenotype. Testing over-expression of the TF when it is normally expressed (i.e. under nitrogen-deprived conditions for an up-regulated TF) was not considered, as it was assumed that any phenotype change would be minor at best, and difficult to detect. *TF1-5* were tested for lipid analysis, as described below, while *TFU1*, *TFD2* and *TFD3* were tested for a variety of phenotypes.

Transformants were grown in 10-mL TAP cultures. Samples for growth rate counts were taken every 12 hours and counted via Z2 Coulter cell counter. After the appropriate nitrogen

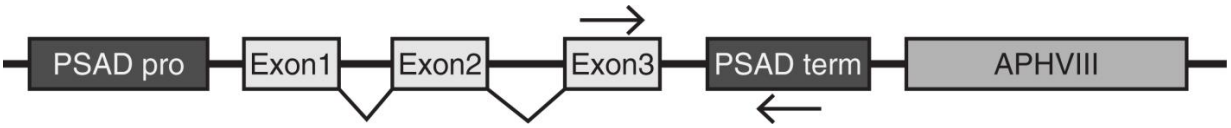


Figure 4-1. Construction of TF expression vectors. Expression vectors were constructed using a modified pJR38 plasmid. The *PSAD* promoter and terminators were used to drive expression. The inserted gene contains the full-length coding sequence, along with one or more native introns. The antibiotic resistance gene *APHVIII* was used to select and maintain transformants using paromomycin in the medium. The arrows indicate the location of the primers used to amplify cDNA to test for expression of the mRNA.

Table 4-3. Primers used in this experiment.

Name		5' Sequence 3'
TF1	F	GCGCGCACTAGTATGTCGAGTTGCGTCGTG
	R	GCGCGCGAATTCTTAGCACTCAGCGTCCAG
TF2	F	GCGCGCACTAGTATGCCGCCCCCGGGCAAC
	R	GCGCGCGAATTCCTACCGCCTCCCGGCCGC
TF3	F	GCGCGCACTAGTATGACGGAGACCGACCAC
	R	GCGCGCGAATTCTCAGCCAAGCAGGATGCG
TF4	F	GCGCGCACTAGTATGCCCCGCGTGGCAGGT
	R	GCGCGCGAATTCCTACACCTTGGACGGCGC
TF5	F	GCGCGCACTAGTATGAACGGAACTCAGCCG
	R	GCGCGCGAATTCTTACGATTCGGTGGGGCC
TFU1	F	GCGCGCACTAGTATGAACGAGGCGCTGGAC
	R	GCGCGCACTAGTCTAGGCGCGAGGCCGCTT
TFD2	F	GCGCGCACTAGTATGGCCGATGAGGGACCG
	R	GCGCGCGAATTCCTAGGCGTAGTCGGGCACGTCGTAGGGGTACT TCCCTCCCGCCGT
TFD3	F	GCGCGCACTAGTATGCTCTTCTGCAATCAG
	R	GCGCGCGAATTCTCATGACGCCATACCAAA
PSAD terminator		CCCGTATCAATCAGCGAAAT
Malate synthase	F	CAAGTACCCTCTGGCTGCTC
	R	CGATGTGGTCATAGCACAGG
Isocitrate lyase	F	AGAAGAAGCTGAGCGACGAC
	R	GCTGAACATGCCGTAGTTGA
Rack1	F	GACCACCAACCCCATCATC
	R	AGACGGTCACGGTGTTGAC

treatment, samples were taken for lipid extraction, protein extraction and RNA extraction. The lipid pellets were extracted with 1:1 (v:v) chloroform:methanol, and then phase separated with 0.2M H_3PO_4 and 1M KCl. The organic layer was extracted and loaded on a TLC plate, where the neutral lipids were separated by 80:20:1 (v:v:v) petroleum ether:ether:acetic acid. The TAG bands were scraped and converted to fatty acid methyl esters (FAMEs), along with an equivalent amount of total lipid. The FAMEs were separated and counted via GD-FID, and the amount of TAG calculated by the ratio of TAG-derived FAMEs to FAMEs from the total lipid fracture. The amount of pheophytin-derived FAME was also calculated from the total lipid fraction, to give a proxy for the amount of chlorophyll in the cells.

To extract protein, 1 mL cell pellets were mixed with SDS buffer and boiled for 10 minutes. The proteins were loaded on a 12% polyacrylamide gel and separated. The proteins were transferred to blots, and hybridized with anti-MLDP antibody. The blot was developed and exposed to x-ray film to detect the signal from the protein.

For RNA extraction, Trizol was used on 1 mL of *Chlamydomonas* pellet. The total RNA resulting was converted to cDNA using Invitrogen SuperScriptIII reverse transcriptase. The expression of the transgenes was detected using the forward cloning primers and the *PSAD* terminator primer, listed in Table 4-3. The primers are in the gene sequence and in the *PSAD* terminator sequence, thus differentiating between the native and inserted genes. The amplified gene sequences were run on an agarose gel and visualized with ethidium bromide to confirm expression. For Q-RT-PCR, the total RNA was converted to cDNA, and reacted using SYBR Green PCR mix. *Rack1* was used as a control gene for normalizing expression levels. The primer sequences used are listed in Table 4-3.

During the various screening processes, if any transformant showed a significant difference compared to the empty vector, that line was regrown in triplicate and retested to confirm the phenotype.

Results

For *TF1-5*, identified from the 454 sequencing dataset, the ratio of TAG to total fatty acid was analyzed in ~20 transformants for each construct. When analyzed, none of the transformants showed a significant difference in lipid level, compared to the empty vector control lines (Table 4-4). Because of these negative results, and their lack of significant difference in expression in the Illumina dataset, no further experiments were performed using these constructs.

For *TFU1*, *TFD2* and *TFD3*, identified in the Illumina dataset, ~50 transformants from each construct, along with empty vector control lines, were tested for the various phenotypes described earlier. Roughly half the transformants showed expression of the transgene mRNA (Figure 4-2). Some lines showed a significant difference during initial testing, and were tested again in triplicate. However, after the second round of testing, no significant differences were observed, compared to the empty vector lines (Table 4-4). The lines were maintained under selection between testings, to minimize the risk of silencing.

Discussion

These experiments show the difficulty of taking a reverse genetics approach to identifying regulators for a specific pathway. Genomic data is sufficient to identify putative transcription factors, but cannot tell anything about their role in cellular regulation. By determining which transcription factors were differentially regulated under TAG-inducing conditions, the number of targets could be narrowed, but this approach has two major drawbacks. The first is that multiple physiological changes occur during N-deprivation, not just TAG accumulation, and the induced

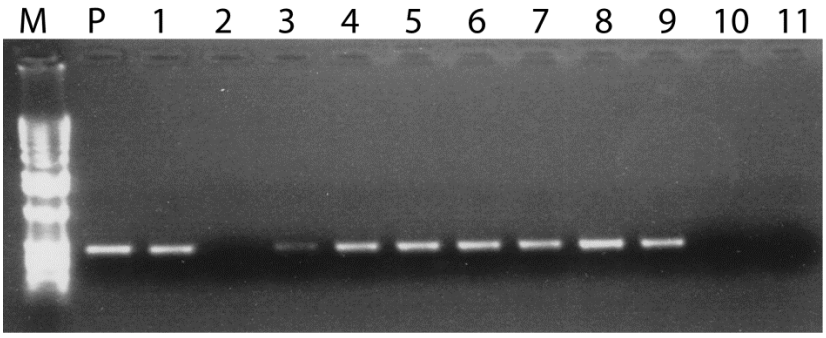


Figure 4-2. mRNA expression of TF overexpression constructs. This figure shows a representative agarose gel of RT-PCR results for *TFU1* expression in *Chlamydomonas*. M indicates the marker; P indicates the pJR-TFU1 plasmid, used as a positive control; 1-11 indicate the transformed lines tested. The band indicates the presence of the transgene mRNA.

Table 4-4. Results of testing TF over-expression constructs

Gene	mRNA	TAG	MLDP	pheophytin	Malate synthase	Isocitrate lyase	growth	protein
TF1	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A
TF2	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A
TF3	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A
TF4	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A
TF5	Yes	No	N/A	N/A	N/A	N/A	N/A	N/A
TFU1	Yes	No	No	No	No	No	No	No
TFD2	Yes	No	No	No	No	No	No	No
TFD3	Yes	No	No	No	No	No	No	No

"Yes" indicates that a positive and repeatable result was seen, "No" indicates that a positive result was not seen or was not repeatable. Shaded cells indicate that the construct was not tested for that particular phenotype

transcription factors may be related to any of them. Additionally, it is possible that the key transcription factor is not induced transcriptionally, but rather post-transcriptionally or post-translationally, and thus would not be among the list of candidate genes. Another factor is that the list of candidates is based on the quality of the transcript profiling data. The results from the 454-based dataset were unreliable, and therefore the candidate genes identified and tested apparently did not have significant changes in transcript level. The Illumina-based dataset was more reliable in terms of expression level, but both datasets suffer from being based on 48 hours of N-deprivation, whereas it is likely that transcription factors would be induced earlier. Their expression level may still be altered after 48 hours, as is the case with *NIT2*, but this is not certain. As mentioned in the previous chapter, the number of time points that could be sequenced was limited, preventing us from taking earlier samples for more detailed examinations. Therefore, it was known that this set of experiments had a significant possibility of failure to identify genes of interest. However, the potential benefits of identifying a regulator similar to WRI1 were considered worth the effort.

Because the function of the putative transcription factors could not be determined from sequence data alone, several phenotypes were examined. TAG content was the main phenotype examined, via TLC and GC. A second lipid-related phenotype examined was the level of MLDP protein, a lipid body associated protein, determined via Western blot (Moellering, E. R. and Benning, C. 2010). Chlorophyll content was estimated by measuring pheophytin levels via GC. The RNA levels of malate synthase and isocitrate lyase, two enzymes in the glyoxylate cycle that are strongly down-regulated under nitrogen deprivation, were measured via Q-RT-PCR. The growth rates of the cells were measured by cell counting at various time points. Finally, the

expression of the transcription factor candidates was tested by measuring the mRNA levels of the transgenes by RT-PCR.

All of these phenotypes were tested in roughly 50 transformants of each construct, along with appropriate wild-type controls. For expression of the transgene, as measured by RT-PCR, positive results were seen in roughly half of the transformants, and none of the wild-type samples. For the other phenotypes, no significant differences from the wild-type samples were detected. When outliers were detected, the experiments were repeated in triplicate, but with no positive results. Therefore, the three transcription factor candidates tested appear to have no effect on the phenotypes tested. This may be due to them operating in other pathways not tested. It could be that over-expression of the candidate genes is insufficient, and other genes in the regulatory pathway are the limiting step in activation of the particular pathway. Alternatively, it may be that the proteins were never properly expressed in *Chlamydomonas*. Transgene expression in *Chlamydomonas* is notoriously difficult, with silencing of inserted genes common. To mitigate the effect of this, I used proven tools, with an effective promoter and terminator sequence, and a strain mutated and selected for improved transgene expression (Fischer, N. and Rochaix, J. D. 2001, Neupert, J. et al. 2009). In addition, all constructs contained at least one native intron, which has been shown to improve expression (Lumbreras, V. et al. 1998). The positive results from the RT-PCR reactions indicate that expression at the mRNA level is occurring. However, although tagged constructs were used, protein expression was never detected using Western blots. This could indicate that silencing is occurring post-transcriptionally, although this is uncommon in *Chlamydomonas*. Alternatively, the proteins could be being degraded rapidly, or simply not being properly detected by the antibody. As the

lines were maintained on selective media, it is unlikely that the transgenes were lost between testing.

BIBLIOGRAPHY

BIBLIOGRAPHY

Baud,S., Mendoza,M.S., To,A., Harscoet,E., Lepiniec,L., and Dubreucq,B. (2007a) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant.J **50**:825-838.

Baud,S., Wuilleme,S., Dubreucq,B., de,A.A., Vuagnat,C., Lepiniec,L., Miquel,M., and Rochat,C. (2007b) Function of plastidial pyruvate kinases in seeds of *Arabidopsis thaliana*. Plant.J **52**:405-419.

Camargo,A., Llamas,A., Schnell,R.A., Higuera,J.J., Gonzalez-Ballester,D., Lefebvre,P.A., Fernandez,E., and Galvan,A. (2007) Nitrate signaling by the regulatory gene *NIT2* in *Chlamydomonas*. Plant Cell **19**:3491-3503.

Cernac,A. and Benning,C. (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. Plant.J **40**:575-585.

Fischer,N. and Rochaix,J.D. (2001) The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*. Mol.Genet.Genomics **265**:888-894.

Focks,N. and Benning,C. (1998) wrinkled1: A novel, low-seed-oil mutant of Arabidopsis with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant.Physiol **118**:91-101.

Goodenough,U., Lin,H., and Lee,J.H. (2007) Sex determination in *Chlamydomonas*. Semin.Cell.Dev.Biol **18**:350-361.

Harris,E.H. (1989) *The Chlamydomonas Source Book*. San Diego: Academic.

Lotan,T., Ohto,M., Yee,K.M., West,M.A., Lo,R., Kwong,R.W., Yamagishi,K., Fischer,R.L., Goldberg,R.B., and Harada,J.J. (1998) Arabidopsis LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93**:1195-1205.

Lumbreras,V., Stevens,D.R., and Purton,S. (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant.J **14**:441-447.

Moellering,E.R. and Benning,C. (2010) RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. Eukaryot.Cell **9**:97-106.

Mussgnug,J.H., Wobbe,L., Elles,I., Claus,C., Hamilton,M., Fink,A., Kahmann,U., Kapazoglou,A., Mullineaux,C.W., Hippler,M., Nickelsen,J., Nixon,P.J., and Kruse,O. (2005) NAB1 is an RNA binding protein involved in the light-regulated differential expression of the light-harvesting antenna of *Chlamydomonas reinhardtii*. Plant Cell **17**:3409-3421.

Neupert,J., Karcher,D., and Bock,R. (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant.J **57**:1140-1150.

Riano-Pachon,D.M., Correa,L.G., Trejos-Espinosa,R., and Mueller-Roeber,B. (2008) Green transcription factors: a *Chlamydomonas* overview. Genetics **179**:31-39.

Riano-Pachon,D.M., Ruzicic,S., Dreyer,I., and Mueller-Roeber,B. (2007) PlnTFDB: an integrative plant transcription factor database. Bmc Bioinformatics **8**.

Santos,M.M., Dubreucq,B., Miquel,M., Caboche,M., and Lepiniec,L. (2005) LEAFY COTYLEDON 2 activation is sufficient to trigger the accumulation of oil and seed specific mRNAs in Arabidopsis leaves. FEBS.Lett **579**:4666-4670.

Yang,Y., Yu,X., Song,L., and An,C. (2011) ABI4 activates DGAT1 expression in Arabidopsis seedlings during nitrogen deficiency. Plant Physiol **156**:873-883.

CHAPTER 5: CONCLUSIONS

The goal of my thesis project was to examine some of the basic biology behind TAG synthesis in *Chlamydomonas reinhardtii*. To accomplish this goal, I focused on two main areas of TAG synthesis: the final step in TAG synthesis, and the overall regulation of TAG synthesis.

For the first area, I attempted to identify and characterize the enzymes responsible for the final step in TAG synthesis, the addition of a fatty acyl group to DAG. There are three main types of enzymes that can carry out this step, including DGATs, PDATs and DGTAs, with DGATs appearing to play the major role in most organisms. Searching with BLAST revealed five putative type 2 DGAT candidates, which I then focused my research on. Of these five, four appear to be actively expressed genes, and three of the four show DGAT activity in yeast. The last DGAT, *DGTT1*, was not cloned by myself, but has recently been confirmed by others to also have DGAT activity (Boyle, N. R. et al. 2012). When expressed in Arabidopsis, DGTT2 continued to show DGAT activity, as well as the apparent ability to incorporate long-chain fatty acids, modifying the composition of surface lipids and sphingolipids (Sanjaya et al. 2013). The attempts to assay the function of DGTTs in *Chlamydomonas*, via overexpression or knockdown techniques, was not successful, likely due to the difficulty of expressing transgenes in *Chlamydomonas*, as well as the functional redundancy provided by the remaining DGTTs, along with the other TAG-synthesis enzymes discovered. Other labs have been able to perform these experiments, and their results indicate that DGTT1 and DGTT3 play partial roles in TAG synthesis in *Chlamydomonas* (Deng, X. D. et al. 2012, La Russa M. et al. 2012).

For the second area, I used transcript profiling on *Chlamydomonas* cultures under nitrogen deprivation to give an overview of the changes occurring that may contribute to TAG synthesis. The goal was to both identify large-scale changes in metabolic pathways, such as

carbon metabolism, that affect the flow of carbon into TAG, and to identify specific genes of interest for further study, including potential regulators of TAG synthesis. This experiment was successful, generating a large amount of useful data that has revealed both expected and unexpected changes in the expression of various genes (Miller, R. et al. 2010). Two methods of high-throughput sequencing were used, 454 and Illumina. Both datasets gave the same overall trends in metabolic pathways, but the Illumina dataset has much more complete and accurate data, due to a larger number of reads, and was used as the primary data source for the analyses performed in this thesis. The dataset is continuing to be useful for further research, as both a source for genes to study and as additional information about genes discovered by other methods. Due to the single time point used, the data is somewhat limited, but still provides a valuable resource for information about both specific genes and general physiological trends.

As mentioned, one of the goals for the transcript profiling experiment was to identify potential regulators. Both the 454 and the Illumina datasets contained putative transcription factors with a significant change in expression, five from the 454 set and 83 from the Illumina set. Because the function of a TF cannot be determined from its sequence, I selected several of the putative TFs to clone and over-expression *Chlamydomonas*, in order to test their effect on various metabolic pathways. The five transgenic lines expressing TFs from the 454 dataset, *TF1-5*, were all tested for changes in TAG level, and all were negative. Because their changes in expression were not significant in the Illumina dataset, they were dropped from further experiments. Three putative TFs from the Illumina set, *TFU1*, *TFD2* and *TFD3*, were successfully cloned and tested for their effect on various pathways. Although expression of the mRNA could be confirmed, no consistent change in phenotype was detected in the transgenic

lines, suggesting that the cloned TFs do not play a role in TAG metabolism or the other pathways tested.

There is still much future work that can be done in the areas I researched. With the DGTTs, there are many options for future experiments. One obvious step, already being done by my collaborator Dr. Sanjaya, is to express *DGTT3* and *DGTT4* in Arabidopsis, and to see whether they behave like *DGTT2*. Another is to repeat the competition assay with DGTT3 and DGTT4, and see if the results from the earlier assays hold (i.e., if DGTT3 has similar preferences to DGTT2, and if DGTT4 has different preferences). More straight-forward feeding assays, using more fatty acid species, would also be informative and help distinguish between actual competition and inhibition, but this approach is somewhat limited by the availability, or rather lack thereof, of less common radiolabeled acyl-CoAs.

Repeating the attempts to assay DGTT function in *Chlamydomonas* is also a potential future plan. There is still much that needs to be learned about how the various DGTTs function and interact during TAG synthesis. For example, why does *Chlamydomonas* have four functional DGAT genes, when most higher organisms have only one or two? The preliminary results suggest that the genes have different substrate preferences. Combined with the different expression patterns, it is possible that the multiple DGATs are used to determine both TAG amount and composition, in response to different conditions. Another possible explanation is that the DGATs are targeted to different cellular compartments. Although most TAG synthesis is thought to take place in the ER, recent experiments indicate that *Chlamydomonas* produces TAG in the chloroplast as well (Fan, J. et al. 2011, Goodson, C. et al. 2011). Overexpression or amiRNA knockdown of specific DGAT genes, and subsequent analysis of the TAG amount, composition and location, may provide further evidence of the role of the DGATs in

Chlamydomonas, such as which TAG species are contributed by each DGAT, which genes contribute the most to TAG synthesis under specific conditions, and which genes are active at the ER, chloroplast or both. At the very least, this would allow confirmation of the *in vitro* results from earlier experiments. One potential obstacle, especially with the knockdown approach, is the possibility of compensation by the untested DGATs, making a TAG phenotype difficult to detect.

An additional area of research is the other TAG-synthesizing enzymes discovered, a PDAT and a type 1 DGAT (Boyle, N. R. et al. 2012). Both of these enzymes are capable of significant contributions to TAG accumulation, and the PDAT may also play a role in membrane remodeling and recycling, as it takes fatty acyls from membrane lipids. The activity of the *Chlamydomonas* PDAT has been confirmed, and its specific function is currently being tested by a fellow graduate student in the Benning lab, Bensheng Liu. The *DGAT1* has not been studied, and experiments with it would reveal its role in TAG synthesis, and possibly give some insight into why green algae have only 0-1 *DGAT1*s, and multiple *DGAT2*s.

The transcript profiling experiment was a success, and will still be useful in the future for other researchers in the lab, and thus little more needs to be done with it. The logical next step would be to repeat the experiment with more time points, earlier during the N-deprivation period, but this has already been done recently by the Merchant lab (Boyle, N. R. et al. 2012), and would thus be redundant. One possibility is to look at N-resupply, post deprivation, to see what genes are affected during TAG breakdown. Controlling TAG degradation is an important factor in maximizing oil yields for biofuels purposes.

There is still much further work that can be done with the identified transcription factors from the transcript profiling experiments. To date, only three of the 83 regulated putative TFs

have been tested. It is entirely possible that one or more of the remaining TFs are involved in the regulation of TAG synthesis. However, as my work has shown, trying to determine the function of a putative transcription factor via reverse genetics is difficult, given both the large amount of processes possibly regulated, and the general difficulties of transgene expression in *Chlamydomonas*. This is especially true given the large number of putative transcription factors identified, which makes testing them one by one impractical. One possible alternative is to try to identify related transcription factors by using other lipid-related genes that are regulated by N-deprivation, such as *DGTT1*. Using the promoter of these genes for yeast-1-hybrid would help to identify regulatory factors involved in control of their expression patterns. Another approach would be to use the transcript profiling dataset to select knockout mutants from a collection, and use them for testing. This speeds up the screening process and avoids the limitations of transgene expression in *Chlamydomonas*, and has been done successfully by others (Boyle, N. R. et al. 2012).

One conclusion reached from working on these projects is the limitation of *Chlamydomonas* as a model organism. Although it is still the most studied and best characterized microalgae, it suffers in comparison to other model organisms, such as Arabidopsis. The main difficulty with *Chlamydomonas* is the difficulty of expressing transgenes and non-native constructs. This is something many *Chlamydomonas* researchers have struggled with. The likely cause is the extensive gene silencing mechanism present in *Chlamydomonas*. Several attempts have been made to improve the expression of transgenes, including incorporation of native introns, which have been shown to improve transcription, and the use of mutant strains selected for improved expression (Lumbreras, V. et al. 1998, Neupert, J. et al. 2009). However, successful expression of transgenes is still uncertain, with only some proteins being expressed at

detectable levels. One way around this issue is to use mutagenesis to disrupt a gene of interest, but the lack of homologous recombination, and no existing organized library of mutants equivalent to the Arabidopsis Salk lines, makes this approach difficult and time-consuming when looking for specific genes to be disrupted, rather than phenotypes of interest.

Another potential problem with *Chlamydomonas* is its inappropriateness for commercial biofuels production. Most algal biofuels systems are based around the use of marine algae, and require higher yields than *Chlamydomonas* produces. Although model organisms don't have to be commercially viable to be useful and informative (see Arabidopsis for example), the question still exists about how much the knowledge gained in *Chlamydomonas* can be applied to potential biofuels production strains. This is especially true for microalgae, given the large genetic diversity of this group of organisms. Algae being considered for production range from *Chlorella*, a green alga related to *Chlamydomonas*, to *Nannochloropsis*, a member of the eustigmatophyte lineage. Others in the lab are currently experimenting with *Nannochloropsis*, both to develop genetic tools and to compare its TAG synthesis to *Chlamydomonas*. These experiments will help determine how universal the *Chlamydomonas* TAG synthesis mechanism is.

BIBLIOGRAPHY

BIBLIOGRAPHY

Boyle,N.R., Page,M.D., Liu,B., Blaby,I.K., Casero,D., Kropat,J., Cokus,S., Hong-Hermesdorf,A., Shaw,J., Karpowicz,S.J., Gallaher,S., Johnson,S., Benning,C., Pellegrini,M., Grossman,A., and Merchant,S.S. (2012) Three acyltransferases and a nitrogen responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. J.Biol.Chem. **287**:15811-15825.

Deng,X.D., Gu,B., Li,Y.J., Hu,X.W., Guo,J.C., and Fei,X.W. (2012) The Roles of acyl-CoA: Diacylglycerol Acyltransferase 2 Genes in the Biosynthesis of Triacylglycerols by the Green Algae *Chlamydomonas reinhardtii*. Mol.Plant **5**:945-947.

Fan,J., Andre,C., and Xu,C. (2011) A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. FEBS.Lett **585**:1985-1991.

Goodson,C., Roth,R., Wang,Z.T., and Goodenough,U. (2011) Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost. Eukaryot.Cell **10**:1592-1606.

La Russa M., Bogen,C., Uhmeyer,A., Doebbe,A., Filippone,E., Kruse,O., and Mussgnug,J.H. (2012) Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. J.Biotechnol **162**:13-20.

Lumbreras,V., Stevens,D.R., and Purton,S. (1998) Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron. Plant.J **14**:441-447.

Miller,R., Wu,G., Deshpande,R.R., Vieler,A., Gartner,K., Li,X., Moellering,E.R., Zauner,S., Cornish,A.J., Liu,B., Bullard,B., Sears,B.B., Kuo,M.H., Hegg,E.L., Shachar-Hill,Y., Shiu,S.H., and Benning,C. (2010) Changes in transcript abundance in *Chlamydomonas reinhardtii* following nitrogen deprivation predict diversion of metabolism. Plant.Physiol **154**:1737-1752.

Neupert,J., Karcher,D., and Bock,R. (2009) Generation of *Chlamydomonas* strains that efficiently express nuclear transgenes. Plant.J **57**:1140-1150.

Sanjaya, Miller,R., Durrett,T.P., Kosma,D.K., Lydic,T.A., Muthan,B., Koo,A.J., Bukhman,Y.V., Reid,G.E., Howe,G.A., Ohlrogge,J., and Benning,C. (2013) Altered lipid composition and enhanced nutritional value of Arabidopsis leaves following introduction of an algal diacylglycerol acyltransferase 2. Plant Cell **25**:677-693.