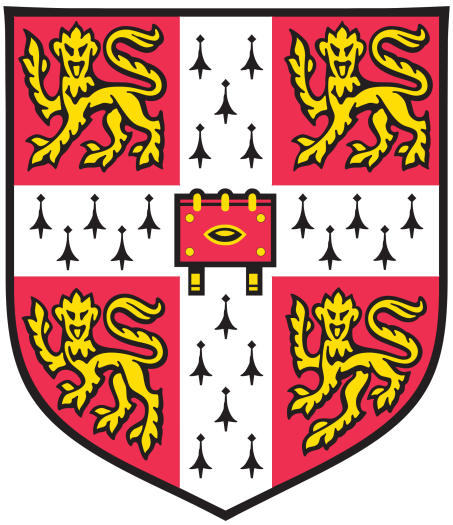
A thesis submitted for the degree of Doctor of Philosophy

Investigating mechanisms of triacylglycerol synthesis induced by nitrogen depletion in the diatom *Phaeodactylum tricornutum* and characterising its diacylglycerol acyltransferases and phospholipid:diacylglycerol acyltransferase



Iain David Bower

Hughes Hall
Department of Plant Sciences
University of Cambridge

Supervised by Professor Alison Smith

14th May 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text

It does not exceed the prescribed word limit for the relevant Degree Committee.

Acknowledgements

Professionally, thanks first go to my supervisor Professor Alison Smith for her guidance and support. I have learnt numerous things through her example and her help has made this thesis possible. To my examiners, thank you for your time and effort, I recognise that the marking process is not a trivial one and I do appreciate it. To the authors and lab members whose efforts I have made use of, thank you, particularly to Dr Alex Litvinenko for his previous work on DGATs and PDAT in *Phaeodactylum tricornutun* and the overexpressing lines he generated, to Dr Katrin Geisler for her generous assistance in the lab and to Dr Mary Hamilton and Dr Richard Haslam at Rothamsted Research for their teaching on experimental techniques and mass spectrometry measurements respectively. Finally, I would like to thank the BBSRC for their funding and the opportunities this has granted me.

This process would also not have been possible without the support of some key people in my life. I would like to thank John and particularly Martha Bower for their support throughout my education, especially in the last few months. Thanks to my brothers and sisters for creating our family-community and the happiness this brings. And thank you to my friends Kevin Heffernan and Lulu Shi, who have injected some joy into the process of thesis writing.

Investigating mechanisms of triacylglycerol synthesis induced by nitrogen depletion in the diatom *Phaeodactylum tricornutum* and characterising its diacylglycerol acyltransferases and phospholipid:diacylglycerol acyltransferase

Iain David Bower

Summary

The ability of the diatom *Phaeodactylum tricornutum* (*P. tricornutum*) to accumulate triacylglycerol (TAG) and synthesise the omega-3 fatty acids eicosapentaenoic acid (EPA) and docosahexaenoic (DHA) has led to interest in their development for biofuels and human nutrition. Nitrogen depletion is a well-studied experimental system for producing high levels of TAG in *P. tricornutum*. Identification of the mechanisms involved in TAG synthesis holds promise for increasing TAG yields. Numerous studies of nitrogen depletion have been carried out that combine physiological and transcriptomic or proteomic measurements. These studies provide insight into global and local metabolic responses but experimental setups varied. To identify a consistent response, experimental conditions and physiological changes were compiled and transcripts and proteins were functionally grouped in a systematic review. Comprehensive analysis of lipid metabolism and autophagy genes was undertaken. Analysis of regulated genes supports involvement of central pathways and lipid metabolism in TAG accumulation: acetyl-CoA synthesis may increase while FA synthesis, the Kennedy pathway, malonyl-CoA/ACP transacylase, plastidial desaturase and FA transport enzymes were upregulated. Desaturases, elongases, Lands cycle enzymes and long chain acyl-CoA synthases are targets for overexpression to increase EPA production during nitrogen depletion.

Several lipidomic experiments have highlighted the potential for TAG synthesis from chloroplast lipids, particularly monogalactosyl diacylglycerol (MGDG), during disassembly of thylakoid membranes under N-depletion. Identifying MGDG derived TAG accumulation may yield enzymes that target EPA to TAG and thereby enable use of developed extraction processes. Some archaeplastida also synthesise TAG from MGDG in nitrogen depletion and freezing stress. A similar response in diatoms would demonstrate this metabolic feature in diverse photosynthetic lineages. In this thesis,

the fatty acid synthase inhibitor cerulenin was used to probe TAG accumulation during N-depletion. Neutral lipid staining data support the possibility of an MGDG to TAG route but were not conclusive. Concurrently, candidate enzymes were identified based on conserved sequence domains and experimentally characterised genes from *Chlamydomonas reinhardtii* (*C. reinhardtii*) and *Arabidopsis thaliana* (*A. thaliana*). *P. tricornutum* was subsequently transformed with acyl hydrolase:YFP constructs. Confocal microscopy supported locations of two acyl hydrolases: in the endoplasmic reticulum (ER) and chloroplast ER for Phatr3_J44028 and in the ER or cytoplasm for Phatr3_J41624.

P. tricornutum encodes multiple isoforms of the TAG synthesis enzymes diacylglycerol acyltransferase (DGAT) and phospholipid:diacylglycerol acyltransferase (PDAT). Previous studies supported functionality of multiple PtDGATs and differential regulation and predicted locations suggest non-redundancy. To investigate their role, overexpression of multiple DGATs and an N-terminal truncated PDAT was induced under nitrogen depletion. Total lipid extraction was carried out and species and FA composition of TAG was measured using tandem mass spectrometry. Total TAG was not altered. Overexpression of DGAT2B decreased the proportion of the 48:2 (carbon atoms:double bonds) TAG species and increased 50:2 and 50:3 while palmitic and oleic acid formed an increased and decreased proportion of TAG respectively. Overexpression of PDAT decreased the 48:1 and 48:2 TAG species, decreased the proportion of TAG composed of palmitic acid and increased the amount composed of EPA. Knowledge of typical TAG species composition in *P. tricornutum* supported respective preferences for incorporation of 18 carbon acyl-CoAs and EPA by DGAT2B and PDAT and their use of C16/C16 diacylglycerols.

Contents

List of Figures

List of Tables

List of Abbreviations

ΔF Δfluorescence 182

A. thaliana *Arabidopsis thaliana* iv, 30, 32, 35, 38, 166, 169, 209, 210, 213, 214, 222

C. reinhardtii *Chlamydomonas reinhardtii* iv, 6, 30, 31, 35, 36, 38, 47, 50, 90, 162, 166, 167, 169, 173, 189, 194, 198, 210–212, 221

E. coli *Escherichia coli* vi, 43, 45, 46, 49, 50, 104

N. gaditana *Nannochloropsis gaditana* 8

N. oceanica *Nannochloropsis oceanica* 35, 161

P. tricornutum *Phaeodactylum tricornutum* iii, iv, vi, viii–xiv, xvi, 12–33, 38–40, 43, 45–51, 53, 56, 61, 63, 66, 67, 71, 72, 77, 78, 88, 90, 98–100, 102, 104, 106, 110, 113–116, 118, 121–124, 127, 128, 131–135, 137–139, 141, 142, 148, 150, 153, 155, 159–162, 164, 166, 167, 170, 171, 173–187, 189, 194, 196, 198, 200, 202–215, 217–224, 226, 279

S. cerevisiae *Saccharomyces cerevisiae* xii, 90, 118–121, 131, 147, 153, 161

T. pseudonana *Thalassiosira pseudonana* 8, 13, 14, 27

ACC acetyl-CoA carboxylase 23, 109, 168

ACP acyl carrier protein 23, 168

AMT1 ammonium transporter 1 196, 197, 200

AtSFR2 *A. thaliana* sensitive to freezing 2 32, 171, 189, 279

AWAT acyl-CoA:wax alcohol acyltransferase 34

BCAA branched chain amino acids 96, 108, 109, 111

BLAST basic local alignment tool 169

CDP cytosine diphosphate 24

CDS coding sequence 196

CoA coenzyme A 188

CrPDAT *C. reinhardtii* phospholipid:diacyglycerol acyl transferase 31, 37, 171, 189, 194, 198, 279

CrPGD1 *C. reinhardtii* plastid galactolipid degradation 1 31, 171, 189, 194, 279

DAG diacylglycerol 18, 27, 41, 99, 188

DGAT Diacylglycerol acyltransferase 27, 99

DGDG digalactosyl diacylglycerol 18, 26, 69, 138

DGTA diacylglycerol hydroxymethyltrimethyl β-alanine 18, 27, 138

DGTS diacylglycerol trimethylhomoserine 18

DHA docosahexaenoic acid 14, 20, 27, 69, 73

DMF dimethyl formamide 43

DMSO dimethyl sulphoxide 43, 46, 179–181, 185

DXR 1-deoxy-D-xylulose 5-phosphate reductoisomerase 198, 200, 201

EPA eicosapentaenoic acid 14, 26, 27, 69, 70, 73, 138

ER endoplasmic reticulum 27, 28, 193

ESI-TQMS electrospray ionisation triple quadrupole mass spectrometry 53, 116, 128

FA fatty acid 5, 16, 21, 29, 41, 73, 167, 168

FA synthesis fatty acid synthesis x, xiv, 167, 177, 179, 180, 184

FabB Ketoacyl-ACP synthase I 167

FabF ketoacyl-ACP synthase II 167

FabI Enoyl-ACP reductase I 167

FAD fatty acid desaturase 26, 27, 104

FAME fatty acid methyl ester 121, 128

FAPES fatty acid phytyl ester synthase 214

FAS fatty acid synthase 23, 109, 167, 184

FBA flux balance analysis 221

FC $\log_2$ fold change 76

FcpA fucoxanthin-chlorophyll a-c binding protein A 196, 197, 200

FcpC fucoxanthin-chlorophyll a-c binding protein C 197, 200

FDR false discovery rate 51, 74, 76, 77

FFA free fatty acid 105

FPKM fragment per kilobase per milion mapped reads 60

G3P glycerol 3-phosphate 24

GC-FID gas chromatography flame ionisation detection 52, 119, 211

GGGT galactolipid:galactolipid galactosyl transferase 32

GO Gene Ontology 170

GPAT glycerol 3-phosphate acyltransferase 21, 24, 99

HPLC high pressure liquid chromatography 61

LACS long chain acyl-CoA synthetase 105

LC-PUFA long chain polyunsaturated fatty acid 3, 14

LCAT lecithin:cholesterol acyl transferase 37

LPA lysophosphatidic acid 24

LPAT lysophosphatidylglycerol acyltransferase 21, 24, 26, 99

LPC lysophosphatidylcholine 138

LPCAT lysophosphatidylcholine acyltransferase 100, 105

LPLAT lysophospholipid acyltransferase 26

MAG monoacylglycerol 18, 99

MEP 2-C-methyl-D-erythritol 4-phosphate 198

METE B12 independent methionine synthase 196, 197, 200, 201

MGAT acyl-CoA:monoacylglycerol acyltransferase 34

MGDG monogalactosyl diacylglycerol ix, xiii, xiv, 18, 26, 69, 73, 99, 138, 168, 169, 172, 178, 188, 194, 198, 209–211, 279

NDE non-differentially expressed 74, 110

PA phosphatidic acid 24

PAP phosphatidic acid phosphatase 24, 27, 99

PC phosphatidylcholine 18, 27, 37, 38, 69, 138

PCR polymerase chain reaction 201, 202

PDAT phospholipid diacyglycerol acyl transferase 25, 27, 36, 192, 211

PDC pyruvate dehydrogenase complex 97

PDCT phosphatidylcholine:diacylglycerol cholinephosphotransferase 27

PDK Pyruvate dehydrogenase kinase 97

PE phosphatidylethanolamine 18, 27, 37, 99, 138

PES phytyl ester synthase 34, 214

PG phosphatidylglycerol 18, 26, 27, 69, 73, 99, 138, 211

PI phosphatidyl inositol 18, 27, 138

PLA phospholipase A 220

PLA2 phospholipase A2 26, 37, 100, 105

PLC phospholipase C 99

PPP pentose phosphate pathway 85

PSAT phospholipid:sterol acyl transferase 37

PtPDAT *P. tricornutum* phospholipid:diacyglycerol acyl transferase 200, 201

RNAseq RNA sequencing 58

ROS reactive oxygen species 166, 209, 218

RPKM reads per kilobase per milion mapped reads 60

SDS-PAGE sodium dodecyl sulphate polyacrylamide gel electrophoresis 119

SNP single nucleotide polymorphism 201

SQDG sulfoquinovosyldiacylglycerol 18, 26, 99

TAG triacylglycerol vi, ix, xiii, xiv, 2, 14–16, 21, 27, 41, 53, 69, 73, 99, 102, 105, 166–169, 172–188, 192, 210, 211, 221

UTR untranslated region 196, 197

VLC-PUFA very long chain polyunsaturated fatty acid 100, 101, 105

WS/DGAT wax ester synthase/DGAT 34

YFP yellow fluorescent protein x, 194, 196, 198, 200

If the sight of the blue skies fills you with joy, if a blade of grass springing up in the fields has power to move you, if the simple things of nature have a message that you understand, rejoice, for your soul is alive.

Eleonora Duse

1

Introduction

1.1 Towards genetically modifying lipid metabolism in microalgae for biotechnological use

There is now substantial interest in microalgae from the perspectives of the environmentalist, the earth scientist, the biologist and the industrialist. Together with other phytoplankton, microalgae play prominent roles in carbon, sulphur, iron and silica cycling (Alcolombri et al., 2015; Falkowski et al., 1998; Tréguer and De La Rocha, 2013; Field et al., 1998) and contribute substantially to net primary productivity, thereby underpinning trophic interactions (Tynan, 1998; Walsh, 1981; Field et al., 1998). Microalgae can sequester heavy metals and grow on industrial waste and are therefore of potential use in bioremediation (Thomas et al., 2016; Suresh Kumar et al., 2015). It is perhaps in the industrial product setting however, that microalgae are generating the most excitement, whether as producers of human food, animal feed, pharmaceuticals, pigments, fuels or biochemical feedstocks. Currently, commercially cultivated microalgal biomass is predominantly used as food and feed (Eppink et al., 2017). The ability to produce the carotenoid pigments and health claims surrounding algae as food underpins this. However, the characteristics of microalgae foreshadow great potential for more widespread use in the production of commodities. The ability to grow on non-arable land, with non-traditional nutrient sources and the ability for some species to grow in saline conditions, support cultivation that will not compete with arable crops for resources. High protein, lipid and carbohydrate contents make possible the use of microalgae as bulk foods while their projected high areal productivity holds promise for providing the required quantities of food and fuel for the growing human population (Eppink et al., 2017; Draaisma et al., 2013; Wijffels et al., 2010).

Features of lipid metabolism of microalgae provide potential at both ends of the product value scale, from accumulation of TAG for low-value biofuel production to synthesis of high-value LC-PUFAs, important in human health. The global market for liquid fuels used in transport currently stands at around 50 million barrels per day (b d^{-1}). Crude oil currently accounts for 94% of transport fuel (BP,

2018). The demand for transport fuel is expected to grow by 8 Mbd^{-1} by 2040. Awareness of the risks posed by human greenhouse gas (GHG) emissions has led to increasing efforts to mitigate global warming, most recently culminating in the Paris Agreement (UN, 2015). Given the significance of transport greenhouse gas emissions, many countries include the use of less carbon intensive transport fuels among their mitigation strategies (EU, 2011, 2001; ITF, 2018). Algal produced TAGs can be transesterified to produce biodiesel and, with little modification, biodiesel can be used in current infrastructure and engines (Fazal et al., 2011; Haseeb et al., 2011). GHG emission from the use of microalgal produced biodiesel can be lower than those of fossil fuels and first generation biofuels, providing appropriate inputs and cultivation techniques are used (Carneiro et al., 2017; Campbell et al., 2011; Clarens et al., 2010). Significant challenges do remain if liquid fossil fuels are to be replaced with algal biodiesel, including expensive production, unsatisfactory energy balance and a number of competing technologies such as alternative sources of biofuels, electric vehicles and hydrogen fuel cells. However, unlike first generation biofuels, algal biofuels need not compete with agriculture for land (Carneiro et al., 2017) while competitor technologies face an uncertain future of their own (Robinson, 2018; Ball and Weeda, 2015). Furthermore, strategies for gradually changing the fuel mix over time have been suggested (EU, 2001) while the final composition this mix will take remains unclear. Despite the challenges therefore, a very large market and comparative environmental benefits make the commercial production of algal biodiesel an attractive long-term goal.

Omega-3 long chain polyunsaturated fatty acid (LC-PUFA) are fatty acids of more than twenty carbon atoms in length and incorporate three or more double bonds. Omega-3 signifies that the first double bond from the methyl end occurs after the third carbon atom. These FAs are valuable due to their importance in cardiovascular and neural health and the poor ability of humans to synthesis them (Khozin-Goldberg et al., 2011). They have found large markets in aquaculture feed, infant milk formula and supplements. The majority of omega-3 oils are derived from harvested fish (Finco et al., 2017). This is damaging to wild fish stocks and is insufficient to cater for the current 1.3 Mty^{-1} demand projected from recommended daily consumption (Finco et al., 2017). The development of other sources is therefore a continuing effort.

Heterotrophic cultivation of fungi and algae has found commercial success but still caters for only a small proportion of the demand (Finco et al., 2017; Terezono, 2017). Recent efforts to accumulate omega-3 LC-PUFAs in plants has also met with success and field trials have been undertaken (Napier et al., 2015). Phototrophically cultured microalgae are another promising platform for omega-3 LC-PUFA production as they do not require a fixed carbon and energy source and many algae natively produce these FAs. Compared to engineered plants, photosynthetic microalgae may be less likely to compete for agricultural inputs, such as land and fresh water, provide higher yields and avoid some of the barriers of perception that face GM crops.

Algal biotechnology is still a nascent field and there are multiple challenges to further development of the industry. Techno-economic analyses have highlighted the high cost of cultivation, harvesting and product extraction (Ruiz et al., 2016). As a result, currently only production of high value compounds are thought to be commercially viable, in particular, pigments and LC-PUFAs. This has led to suggestions that algal biotechnology can advance by focusing on these high value products in the short term and moving to mid and low value products over time. The development of knowledge to first aid commercial production of high value LC-PUFAs and then the lower value fuels is therefore important in realising the advantages that microalgal production of these lipids offer.

The non-biological aspects of algal production contribute a significant part of the associated costs and these vary depending on the type of cultivation system, harvesting technique and extraction procedure used (Ruiz et al., 2016). These in turn are dependent upon the product and end use. Capital expenditure, culture temperature control, cleaning and wastewater treatment and operational energy usage are prominent costs variously associated with open or closed cultivation systems. Centrifugation and bead milling, prominent techniques for biomass harvesting and cell disruption respectively are also energy intensive while membranes and solvents used in product extraction can be costly. Much work is still to be done but promising R&D technologies and operational and conceptual advances are expected to lower these costs. One concept that has received significant funding from the EU Seventh Framework programme is that of the algal biorefinery (Eppink et al., 2017): in an effort to maximise revenue, an integrated process is envisaged, with the aim of

exploiting multiple components of the cell. However, advances in development of algal strains are also anticipated to play a major role in minimising cost per unit of product.

Important biological improvements relate to the general productivity and resilience of algae. Increased photosynthetic efficiency has been identified as a major factor in reducing production cost per unit biomass (Ruiz et al., 2016) and efforts to achieve this through truncation of antenna complexes has previously led to greater permeation of light through cultures and biomass productivity (Melis, 2009). Work to improve the tolerance of strains for culture conditions has also been undertaken. Temperature tolerant strains for example, offer potential for reduced energy consumption in cooling when cultures are grown in closed bioreactors (Ruiz et al., 2016).

Another important way that genetic engineering can increase the economic viability of microalgal production of lipids and also improve associated industrial processes is through direct manipulation of lipid metabolism. Microalgae do not yield enough TAG to cover the cost of biofuel production while for some applications it is desirable to accumulate omega-3 LC-PUFAs in TAG to aid their extraction (Khozin-Goldberg et al., 2011). TAG accumulation is primarily achieved through subjecting microalgae to physiological stress such as nitrogen deprivation (Sharma et al., 2012). While stress conditions cause TAG content to increase, they also lead to a slowing or cessation of cell division and so TAG yield drops (Sheehan et al., 1998). However, these conditions do provide experimental systems for the study of the mechanisms involved in TAG accumulation. These mechanisms may also provide insight into the movement of LC-PUFAs to TAG. Further to stress conditions, key TAG synthesis enzymes called diacylglycerol acyltransferases are found across organisms and their manipulation can alter TAG accumulation. This thesis therefore aims to improve knowledge of microalgal lipid metabolism so as to move towards the goals of improved TAG yield and fatty acid (FA) composition.

1.2 An introduction to the biology of microalgae and diatoms

1.2.1 Microalgal diversity

Algae are a polyphyletic grouping of photosynthetic organisms that are physiologically, metabolically and morphologically diverse. Some researchers include both eukaryotic and prokaryotic photosynthetic organisms in the term algae although typically the term refers to eukaryotes. Eukaryotic species can be unicellular microalgae, multicellular macroalgae or colonial. Algae and algal relatives that have lost the ability to photosynthesise can be found in marine, freshwater and terrestrial habitats, can be fee living, symbiotic or parasitic and demonstrate various nutritional modes including autotrophy, heterotrophy and mixotrophthy. Throughout this thesis I shall refer to eukaryotic microalgae simply as algae or microalgae and other types shall be explicitly identified where necessary.

Overall eukaryotic diversity is classified into the super clusters SAR (stramenopiles, alveolates, rhizaria), unikonta, excavata, archaeplastida, hacrobia and a number of smaller groups (Adl et al., 2012; Keeling, 2013) and this is illustrated in Figure 1.1, in which photosynthetic lineages are marked. Algae are thought to have first originated around 1.5 billion years ago (Yoon et al., 2004) with the uptake and subsequent domestication of a β-cyanobacterium by a eukaryote, generating the archaeplastida (Keeling, 2010). The conversion of a cyanobacterium to the plastid of a eukaryote that previously lacked one is termed primary endosymbiosis. A second primary endosymbiosis has also been identified, giving rise to *Paulinella*, but this lineage is currently not ecologically significant. Within the archaeplastida are the red algae, glaucophytes and green algae, from which the land plants derive. Primary green algal endosymbiotic species of note to this thesis are the chlorophyte *C. reinhardtii* and land plants. Lineages derived from the green and red algae form the vast majority of current algal diversity and include secondary, tertiary and serial endosymbionts (Keeling, 2013). Secondary endosymbioses are formed by eukaryotes integrating primary endosymbionts to form a plastid. Tertiary endosymbiosis involves the incorporation of these secondary endosymbionts. Serial endosymbiosis occurs

when an organism that has previously contained a plastid obtains a new one, losing the first.

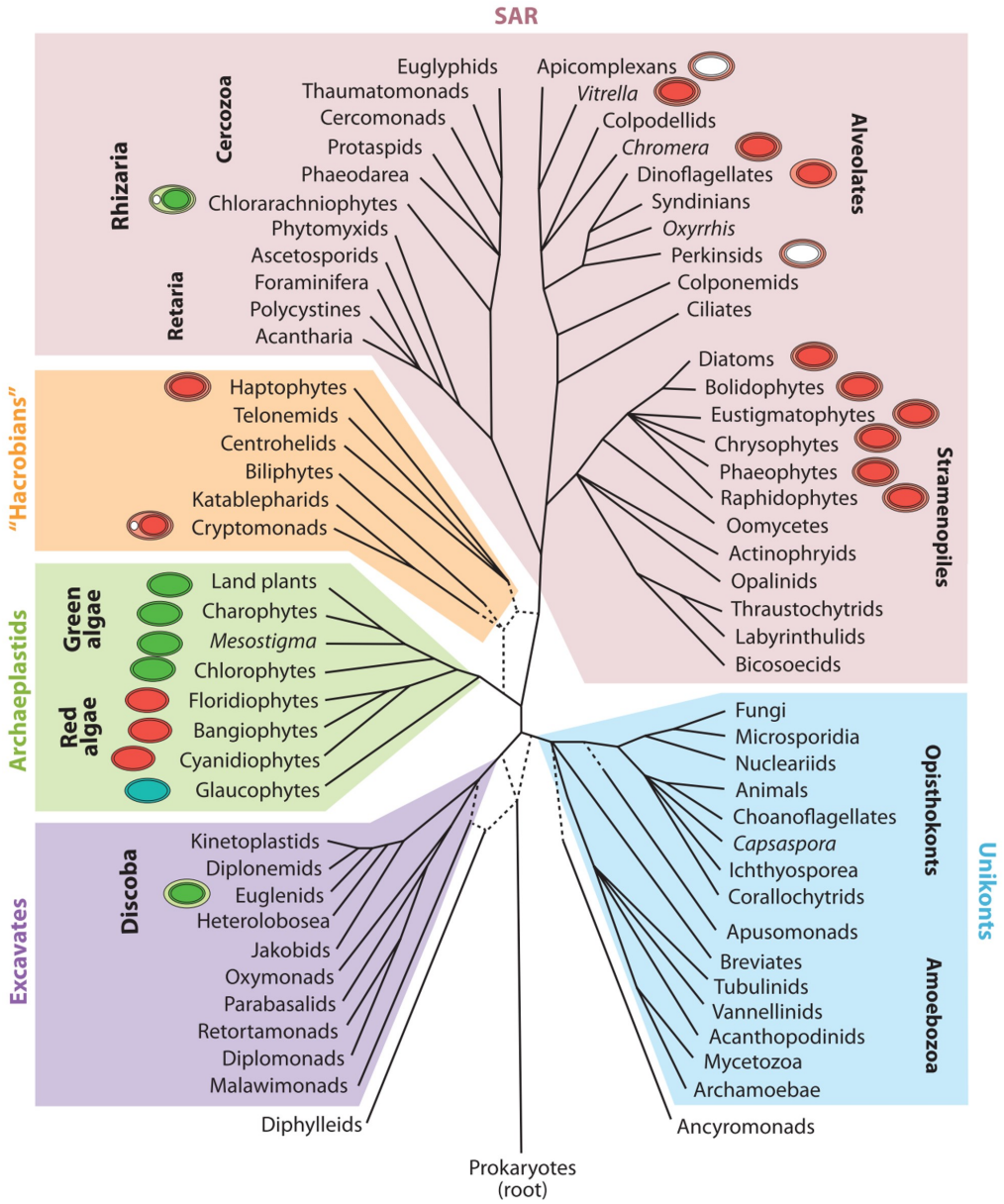


Figure 1.1 – A phylogeny of eukaryotes highlighting photosynthetic lineages. Plastids contained in lineages are coloured based on their red, green or glaucophyte (cyan) algal origins. Disputed relationships are shown with dashed lines and multiple dashed lines show alterntive relationships. *Paulinella*, which originates from a second primary endosymbiosis, is not included. Figure reproduced from Keeling (2013).

Multiple eukaryotic hosts and endosymbionts have contributed to the current diversity of microalgae, as depicted in Figure 1.2 (Keeling, 2013). Glaucophytes are not thought to have formed higher level endosymbioses. In contrast, higher order symbioses of the red algae have generated huge diversity and a number of

ecologically important groupings (Keeling, 2010). Secondary endosymbionts of red algae may have originated in a single endosymbiotic event, generating the ancestor of the stramenopiles, alveolates, rhizaria and CCTH clade (cryptomonads, centrohelids, telonemids and haptophytes). Support for a single endosymbiosis is complicated by the occurrence of many non-photosynthetic lineages within these groups (Burki et al., 2009; Okamoto et al., 2009; Rice and Palmer, 2006). Additionally, the existence of hacrobia as a monophyletic grouping is not settled: some phylogenies group the haptophytes with SAR and the cryptomonads with archaeplastida (Baurain et al., 2010; Burki et al., 2012). Secondary red endosymbionts of note to this thesis are the photosynthetic stramenopiles, termed the ochrophytes (Dorrell et al., 2017). Diatoms, such as *Phaeodactylum tricornutum* (*P. tricornutum*) and *Thalassiosira pseudonana* (*T. pseudonana*) and eustigmatophytes, such as *Nannochloropsis gaditana* (*N. gaditana*), are ochrophytes. The dinoflagellates are another significant group of secondary red algal endosymbionts and in the alveolate group. Green algae have been incorporated in at least three secondary endosymbioses, generating the euglenids, chlorachniophytes and green algal containing dinoflagellates (Keeling, 2013). These dinoflagellates are proposed to have previously contained a red algal plastid that was replaced through serial endosymbiosis (Keeling, 2013).

The limited number of lineages that provided the eukaryotic hosts and plastids of algae suggests that a knowledge of the phylogenetic relationships between them would provide clues about physiology and metabolism of related groups. The archaeplastida classes differ in a number of features including their photsynthetic apparatus. All oxygenic photosynthetic lineages contain chlorophyll a. In addition, green algae contain chlorophyll b and zeaxanthin epoxidase enzymes required for development of the xanthophyll cycle (Dautermann and Lohr, 2017). Starch is one of the main storage molecules in green algae and accumulates in their plastids. Red algae, in contrast, contain phycoerythrin as an accessory pigment (Gould et al., 2008) as well as the zeaxanthin epoxidase enzymes (Dautermann and Lohr, 2017). In contrast to green algae, red algae contain phycobilisomes, unstacked thylakoids and store starch in their cytoplasm (Panaud et al., 2013). Glaucophytes are the third group within the archaeplastida and are thought to have have diverged before the other lineages split, however, they have not contributed to any known higher order endosymbioses.

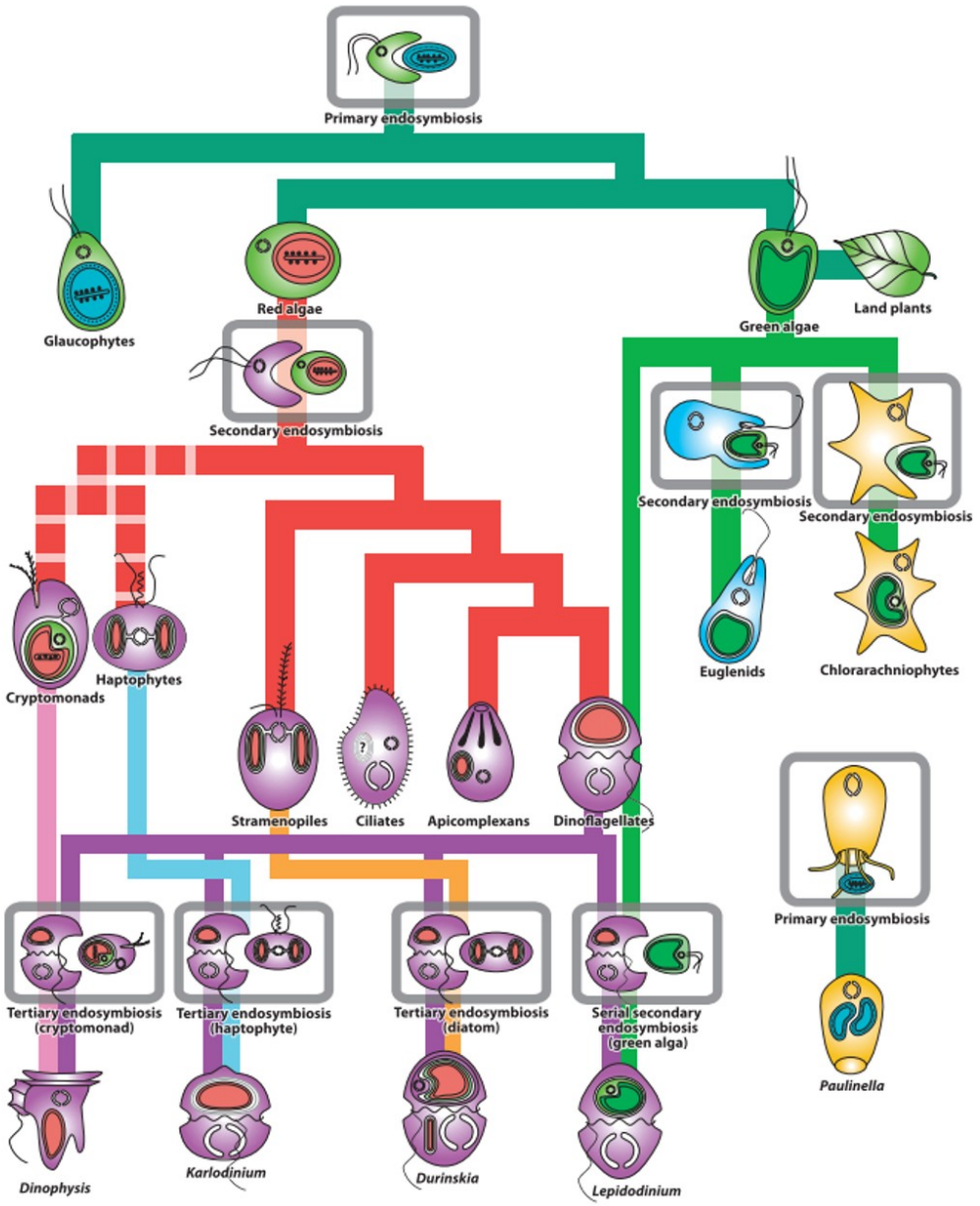


Figure 1.2 – The endosymbioses that produced eukaryotic algae. The primary endosymbiosis of a β-cyanobacterium produced the red and green algae and glaucophytes. Two secondary endosymbioses of green algae formed the euglenids and the chlorachniophytes. A secondary endosymbiosis of a red alga may have given rise to cryptomonads, haptophytes, stramenopiles and alveolates (dinoflagellates and the non-photosynthetic ciliates and apicomplexans). However, cryptomonads and haptophytes may have originated in a discrete endosymbiosis, illustrated by the broken line. Dinoflagellates have formed tertiary endosymbioses with cryptomonads, haptophytes and diatoms and a serial secondary endosymbiosis with green algae. A more recent primary endosymbiosis has also been documented and gave rise to *Paulinella*. Figure reproduced from Keeling (2013).

Knowledge of the primary endosymbiont lineages therefore provides insight into some of the metabolic features that may be found in the metabolism of higher level endosymbionts. The picture is less clear when considering related groups arising from the same eukaryotic host. For example, metabolic commonalities across the SAR species do not appear to have been identified (Grattepanche et al., 2018). Several factors could lead to this situation such as extremely diverse evolutionary histories resulting in conservation of only essential metabolism, a lack of metabolic comparisons or limited data from species across the grouping.

Each endosymbiotic event brought together novel combinations of genes and metabolic elements. During the domestication of a plastid, many genes are transferred from the endosymbiont genome(s) to the host nuclear genome and systems to target proteins to the plastid are established (Keeling, 2010, 2013). The number of membranes surrounding the plastid differs between groups. The archaeplastida have two surrounding membranes while secondary endosymbionts typically have three or four. Most secondary endosymbionts are thought to have been taken up by endocytosis and hence the outer plastid membrane corresponds to the host endomembrane derived vacuole membrane. The outer membrane of the plastid is contiguous with the host ER in stramenopiles, cryptomonads and haptophytes (Gibbs, 1981). The other three membranes are thought to correspond, from the outside in, to the endosymbiont cytoplasmic membrane and the two membranes of the endosymbiont plastid. Archaeplastida utilise the translocon inner-membrane chloroplast (TIC)/translocon outer-membrane chloroplast (TOC) system to recognise a transit peptide at the N-terminus of plastid targeted proteins (Jarvis and Soll, 2001). Secondary endosymbionts use a bipartite N-terminal signal (Bolte et al., 2011; Gruber et al., 2007). This is composed of a signal peptide involved in transport across the endoplasmic reticulum membrane followed by the chloroplast transit peptide. Differences in fully established plastids can be notable. For example, both chlorachniophytes and dinoflagellates have three plastidial membranes. This is despite a presumably endocytotic origin for the dinoflagellate plastid and the four membranes that would be expected to result. Additionally, chlorachniophytes and cryptomonads retain nucleomorphs within the periplastidial compartment of the plastid that represents the original nucleus of the endosymbiont. Genome sequencing

shows considerable loss of genes but some are still retained (Keeling, 2013). The transfer of genes to the nucleus provides substantial substrate for diverse evolutionary histories to shape and is thought to play an important part in the success of some groups.

1.2.2 Diatoms

Diatoms are secondary endosymbionts in the stramenopile group and due to their silaceous walls and high productivity play central roles in biogeochemical cycles of silicon and carbon (Tréguer and De La Rocha, 2013; Field et al., 1998). They are important to humans as the basis of highly productive fisheries (Sarthou et al., 2005), sources of diatomite and fossil fuels, as environmental markers (Smol and Stoermer, 2010) and into the future, of nanotechnological (Kröger and Poulsen, 2008) and biotechnological importance. They mainly exist as single cells but can also form chains. They are the most diverse phytoplankton group with an estimated 200,000 species but despite this have a relatively recent origin with molecular evidence suggesting an earliest origin of 250Myr and fossils from 190Myr. Within this time, substantial adaptations to the ancestral form has enabled diatoms to adapt to changes in atmospheric chemistry and novel niches (Armbrust, 2009). Thought to have originated in the coastal benthos, changes in morphology, motility and metabolic capability enabled the development of planktonic forms and dominance of the open ocean. Diatoms can be split into radial centrics, bipolar and multipolar centrics, araphid pennates and raphid pennates and the major division of centric and pennate are illustrated in Figure 1.3. These forms also correspond to significant occurrences in geochemistry and diatom biology over time (Armbrust, 2009). Radial centrics emerged around 100Myr as diatoms became dominant in the ocean carbon cycle, fossils of araphid pennates are found from around the time of the K-T extinction, 65Myr, while raphid pennates originated around 30Myr, after atmospheric oxygen and carbon dioxide were around contemporary levels.

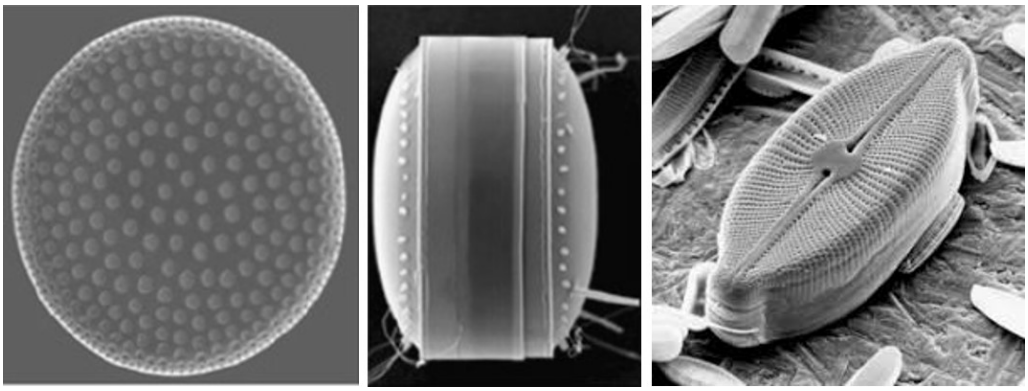


Figure 1.3 – Centric and pennate diatoms. Electron microscope images of a centric diatom from the top (left) and side (middle) and a pennate diatom (right). Images from Brayner et al. (2011) and Mann (2008).

A number of interesting biological and ecological differences exist between the radial and pennate datoms. The radial diatoms are circular when viewed from the top and rectangular from the side while pennate cells are elongated ovals. Other morphologies, such as triradiate are seen in some species, such as *P. tricornutum*. Almost all diatoms are encapsulated in silaceous shells, known as frustules. These frustules are formed from the precipitation of silicate onto an organic matrix, are genetically determined and are usually conserved in species, such that they are used for descriptive systematics (De Martino et al., 2007). A suite of enzymes have been identified as important in frustule formation, with silaffins, long-chain polyamines and silacidins that are embedded in the wall playing roles in silicic acid precipitation (Armbrust, 2009; Kröger and Poulsen, 2008). The silica wall can be considered to be like a petri dish, split into two halves or valves, with one slightly smaller than the other, the hypotheca and so slotting into the larger, the epitheca.

The nature of the frustule plays an important role in the life cycle of the diatoms. Diatoms are diplontic, with the majority of reproduction carried out vegetatively in the diploid state (von Dassow and Montresor, 2011). Each valve is used as the epitheca of the daughter cells and hence those forming from the hypotheca of the mother cell are reduced in size. A species dependent lower size limit exists at which the cell is primed for sexual reproduction, although environmental and chemical signalling can also play a role. Unencapsulated sperm, flagellated in the centrics but not in the pennates, fuses with eggs that remain in the frustule. An auxospore is then formed and grows to restore the size lost through continuous division. The old valves

are then shed and new ones grown. The size requirement for sexual reproduction has been proposed to act as a 'sex clock' (von Dassow and Montresor, 2011), determining how frequently sexual reproduction occurs, and for some species the whole cycle can take decades (Smol and Stoermer, 2010). There is great diversity in the life cycle of diatoms. A number of diatoms, including *P. tricornutum* and *T. pseudonana* do not undergo sexual reproduction nor shrink in size and loss of sexual redroduction or its facultative use is not uncommon in the diatoms. Centric diatoms are homophallic, having gametes with one mating type, while pennate cells are heterophallic (von Dassow and Montresor, 2011).

Frustule properties have significant uses in diatom ecology. The raphid pennates are of great ecological importance and contain a slit along their frustule, called a raphe, that enables movement through secretion of mucilage (Armbrust, 2009). Movement is also important in enabling diatoms to acquire nutrients and energy. Cell shape and processions called setae are important in increasing the surface area to volume ratio and thereby maintaining diatoms near the surface. Some large diatoms are able to alter their buoyancy, becoming quiescent and dense when nitrogen is depleted. Sinking enables nitrogen to be taken up at which point they increase their buoyancy and return to the surface (Suresh Kumar et al., 2015). Optical and proton buffering properties of frustules have been implicated in increased light capture and CO_2 uptake (Kröger and Poulsen, 2008).

Diatoms can be found in freshwater, marine and terrestrial habitats and photoautotrophic and facultative heterotrophic forms exist. The metabolic capabilities that facilitate the success of diatoms result from the secondary endosymbiotic origin of diatoms and subsequent gene transfer events, which have facilitated the attainment of ecological dominance in the water column and open ocean blooms, despite their benthic origins (Bowler et al., 2008; Armbrust, 2009). The genomes of seven diatoms have now been sequenced (Rastogi et al., 2018). This has enabled investigation of the origin of genes in species and comparison of diversity between groups. A number of unexpected metabolic abilities have been identified within diatoms, including a urea cycle (Allen et al., 2011), an Entner-Doudoroff glycolytic pathway (Fabris et al., 2012), and the ability to transfer electrons from the chloroplast to the mitochondria during stress (Bowler et al., 2008). The urea cycle is not found in any other pho-

tosynthetic lineage and its presence is thought to enable rapid recovery from long periods of nitrogen depletion. Even within the diatoms, great diversity is observed. Bipolar centric diatoms have altered photosynthetic apparatus to reduce iron use and have replaced iron dependent electron transfer chain proteins with ones that use copper (Armbrust, 2009). Raphid pennate diatoms in contrast survive iron shortages by downregulating iron dependent enzymes and use ferritin to store iron. Despite having diverged relatively recently at 90Myr, *P. tricornutum* only shares 10% of its genes with *T. pseudonana* (Rastogi et al., 2018).

1.2.3 *Phaeodactylum tricornutum*, a model diatom of biotechnological interest

P. tricornutum is one of the main model diatoms. A raphid pennate, it was first described in the late 19th Century and has become a focus for understanding diatom genetics and metabolism and for development of their biotechnological potential. *P. tricornutum* accumulates TAG under certain growth conditions, reaching levels of up to 60% of dry weight (Song et al., 2013) with an average overall productivity comparable to other oleaginous algae of interest (Sheehan et al., 1998). It also produces valuable and nutritionally essential LC-PUFAs, with eicosapentaenoic acid (EPA) at around 30 mol% of FAs and docosahexaenoic acid (DHA) in smaller amounts, up to 5mol% in the WT (Hamilton et al., 2014). A number of valuable pigments have been identified in *P. tricornutum* including β-carotene, a precursor to vitamin A and fucoxanthin, which is of interest for anti-cancer and anti-malarial properties (Fu et al., 2015; Borowitzka, 2013). *P. tricornutum* is easy to culture and grows rapidly, dividing once per day under optimal conditions. These characteristics have led to ongoing efforts to develop the organism for increased oil productivity and more favourable LC-PUFA composition (Dinamarca et al., 2017; Gong et al., 2011; Hamilton et al., 2014). The work presented in this thesis has focused on characterising lipid metabolism in line with these efforts.

In the wild, *P. tricornutum* is found in coastal areas such as estuaries and rock pools and is not considered of great ecological importance (De Martino et al., 2011). Like other stramenopiles, the plastid of *P. tricornutum* has four membranes. *P. tricornutum* is distinct among diatoms, having a pleomorphic nature that is facilitated

by an organic cell wall, contrasting with the silicified walls of other diatoms. Sexual reproduction has not been observed, nor a decrease in cell size with division. Three morphologies are recognised: fusiform, triradiate and oval as displayed in Figure 1.4. These morphologies are interchangeable with the oval morphology suggested as the intermediate between the pelagic fusiform and triradiate morphologies and a fourth, round form, proposed to be a quiescent form (De Martino et al., 2011). At least ten isolates of *P. tricornutum* exist, denoted Pt1 to Pt10 and these have been categorised into four genotypes (De Martino et al., 2007). Under constant conditions a dominant morphology is identified for each strain. There appears to be little correlation between morphotype, genotype, nor location of isolation however. Alterations in culture condition such as temperature and salinity can alter the dominant form. Unfavourable conditions such as low temperature and salinity stimulate production of the oval form. The oval form sediments readily and the presence in this morphotype of a silicified valve with a single raphe facilitates gliding. It is therefore proposed that the unfavourable conditions stimulate development of the oval, benthic form, that may further develop into the quiescent round morphotype, while favourable conditions for growth lead to the pelagic triradiate or fusiform states (De Martino et al., 2011). Variable levels of accumulated TAG (Abida et al., 2015) and optimal growth conditions (De Martino et al., 2007) indicate substantial differences between strains. However, conserved rRNA internal transcribed spacer 2 sequences suggests they do belong to the same species, and similar composition of TAG species in Pt1 and Pt4 supports substantial metabolic conservation (Yu et al., 2009; Bower, MRes-unpublished).

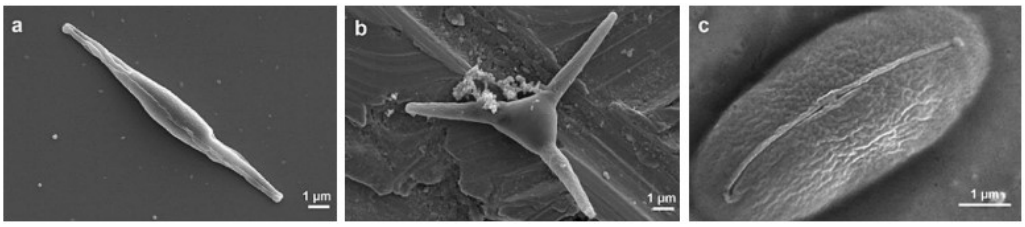


Figure 1.4 – Morphologies of *P. tricornutum*. Electron microscopy images of the three *P. tricornutum* morphotypes. a, fusiform; b, triradiate; c, oval. Images reproduced from Francius et al. (2008).

The biochemical composition and physiology of *P. tricornutum* has been the focus of numerous studies, for example Zhou and Wangersky (1985), Terry et al.

(1985) and Gatenby et al. (2003). Culture conditions of particular interest have been the irradiance level, temperature and variation in micronutrient source and quantity. Measured parameters of focus have included the proportions of protein, carbohydrate and lipid, pigment content, FA composition, photosynthetic efficiency, and growth rate. Chrysolaminarin and TAG are both diurnally regulated (Caballero et al., 2016), accumulating during the light and decreasing in the dark. Chlorophyll a content and cell size also increase during the light while division occurs in the dark (Ragni and D'Alcalà, 2007). As for other diatoms and microalgae, the composition of *P. tricornutum* is dependent upon growth conditions. Pigment content and chloroplast size increase in low light (Fawley, 1984). Low temperatures lead to an increase in the proportion of FAs that are composed of PUFAs (Sayanova et al., 2017). The particular nitrogen source that is used alters photosynthetic efficiency, lipid productivity and macromolecule composition (Huete-Ortega et al., 2018).

The genomes of the chloroplast, mitochondria and nucleus of *P. tricornutum* have been sequenced and annotated with gene models produced using RNAseq and expressed sequence tag data sets combined with protein sequences from multiple organisms. These gene models have been annotated with sequence, structure, function and location information and are easily accessible through the Ensembl Protist database (Ensembl Protist; Kersey et al., 2018; Kinsella et al., 2011; Rastogi et al., 2018). These genetic studies also identified features important in enabling the high diveristy and ecological success of diatoms. The genome of *P. tricornutum* is diverse in origin with 47% having originated within the stramenopiles and 26% of the genes unique to *P. tricornutum*. Around 600, 560 and 2000 genes were identified as originating from prokaryotes, red alga and green algae respectively (Rastogi et al., 2018). Significant amounts of alternative splicing was identified, underpinning high transcript diversity. High levels of transposable elements and diatom specific expansions of gene groups such as cyclins, heat shock transcription factors, histidine kinases and polyamine related genes are seen, which are proposed to have led to the rapid diversification of diatoms (Bowler et al., 2008).

Patterns of global metabolic regulation have also been studied in *P. tricornutum* using transcriptomic, proteomic and metabolomic technologies in numerous growth conditions. These include optimal conditions, high light, dark and nitrogen,

phosphorus and iron depletion. Several studies have boosted understanding of the epigenetic code of histone post-translational modifications and DNA methylation that partly underpin observed regulation (Veluchamy et al., 2015; Rastogi et al., 2018). The transcription factor complement in *P. tricornutum* has been compiled and functional studies have begun (Matthijs et al., 2016, 2017). Furthermore, non-coding RNAs in *P. tricornutum* have been identified and a number are potential regulators of expression (Cruz de Carvalho et al., 2016; Rogato et al., 2014; Huang et al., 2011, 2018). Metabolic models have been developed for *P. tricornutum* to integrate the available genetic and biochemical information to enable further characterisation of metabolism and the impact of metabolic alterations (Kim et al., 2016; Levering et al., 2016). A tool for visualising omic changes on metabolic pathways has also been created (Fabris et al., 2012).

The focus on *P. tricornutum* as a model organism and biotechnological host has led to the development of genetic manipulation tools and resources. *P. tricornutum* can be transformed using biolistics or electroporation and DNA introduced in this manner is randomly integrated into the genome (Niu et al., 2012; Miyahara et al., 2013; Falciatore et al., 1999). Bacterial conjugation of episomes has been developed (Karas et al., 2015; Diner et al., 2016) and the sequences that enable their maintenance have been characterised (Diner et al., 2017). Techniques for targeted genetic modification have also been developed (Kroth et al., 2018). Insertion can be carried out through use of meganucleases and TALENs (Daboussi et al., 2014). RNA-mediated interference has been developed for silencing in *P. tricornutum* and meganucleases, TALENs and CRISPR/Cas9 have been developed for mutagenesis and knock out studies (Nymark et al., 2016; Daboussi et al., 2014; De Riso et al., 2009). Optimisation of CRISPR/Cas9 has been undertaken to overcome implementation challenges, develop DNA free use and simultaneously knock out multiple genes (Serif et al., 2018; Sharma et al., 2018; Stukenberg et al., 2018).

Genetic engineering in *P. tricornutum* is also aided by a number of characterised or regularly used genetic parts such as promoters, terminators, untranslated regions, selectable markers (Kadono et al., 2015; Zaslavskaia et al., 2000; Sakaue et al., 2008) and the development of screening protocols and the broader efforts to improve cloning, increase modularity and reduce redundancy of effort, illustrated for example, by the

Golden Gate cloning system and the plant Modular Cloning Toolbox (Engler et al., 2014a; Weber et al., 2011).

1.2.4 Lipid metabolism in *P. tricornutum*

Lipids are hydrophobic or amphiphilic molecules including galactolipids, phospholipids, sphingolipids, betaine lipids, mono-, di- and tri- acylglycerols, sterols, isoprenoids and prenols (Fahy et al., 2008). Within this thesis I focus on lipids based on a glycerol backbone except where stated and so the terms lipids, glycolipids or phospholipids refer to glycerolipids, glycoglycerolipids and phosphoglycerolipids.

Understanding of lipid metabolism in *P. tricornutum* is inferred from analogy to lipid metabolism in green algae and plants and homology of genes from these organisms and other groups (Rastogi et al., 2018; Zulu et al., 2018; Dolch and Maréchal, 2015). However, several experimental approaches have been used as well, including measurements of the lipidome (Abida et al., 2015; Yu et al., 2009; Yongmanitchai and Ward, 1991; Remmers et al., 2017; Alonso et al., 2000), radiolabelling experiments (Arao and Yamada, 1994) and the perturbations that are caused by stress conditions (Levitan et al., 2015; Matthijs et al., 2016; Longworth et al., 2016; Mus et al., 2013; Alipanah et al., 2018) and genetic alterations (Xue et al., 2015; Ma et al., 2014; Hao et al., 2018). This has confirmed the general picture and knowledge from these areas is detailed in the next paragraphs.

A variety of glycerolipid classes have been detected in *P. tricornutum* and are shown in Table 1.1. These include the mainly plastidial lipids sulfoquinovosyldiacylglycerol (SQDG), acyl-SQDG, MGDG, digalactosyl diacylglycerol (DGDG) and phosphatidylglycerol (PG), the extraplastidial lipids phosphatidyl inositol (PI), phosphatidylcholine (PC), phosphatidylethanolamine (PE), lyso-PC, diacylglycerol hydroxymethyltrimethyl β-alanine (DGTA) and members of the Kennedy pathway monoacylglycerol (MAG), diacylglycerol (DAG) and TAG (Abida et al., 2015; Alonso et al., 2000). The betaine lipid diacylglycerol trimethylhomoserine (DGTS) has previously been identified in *P. tricornutum* but this can be caused by an inability to discern the lipid from DGTA (Abida et al., 2015). In order of abundance, MGDG, SQDG and PC are the most common classes followed by DGDG and PG. By analogy to land plants, PC is thought to be the main extraplastidial lipid, MGDG, DGDG

Table 1.1 – The glycerolipid and fatty acid content of *P. tricornutum*. The major glycerolipids and fatty acid detected in *P. tricornutum* are shown. The proportion of total glycerolipids that each of these glycerolipds and fatty acid contribute derives from Abida et al. (2015). Minor fatty acids detected in other studies are included (Alonso et al., 2000; Bower et al., 2015; Chrismadha and Borowitzka, 1994; Jiang and Gao, 2004). Fatty acid nomenclature is detailed in Table 1.2.

Major glycerolipids	% of glycerolipids	Subcellular location	Major fatty acids	% of glycerolipids	Minor fatty acids
Monogalactosyl diacylglycerol (MGDG)	36	Plastid	14:0	5	18:4
Digalactosyl diacylglycerol (DGDG)	8		16:0	13	20:1
Sulfoquinovosyl diacylglycerol (SQDG)	18		16:1	20	20:0
Acylsulfoquinovosyl diacylglycerol (ASQ)	1		16:2	5	20:2
			16:3	13	22:0
Phosphatidyl glycerol (PG)	8	Plastid/ Mitochondria/Plasma membrane	18:0	3	22:1
			18:1	1	
			18:2	2	
Phosphatidyl choline (PC)	15	Endomembrane/Plasma membrane	18:3	1	
Phosphatidyl inositol (PI)	2		20:4	1	
Phosphatidyl ethanolamine (PE)	1		20:5	32	
Diacylglycerol hydroxymehthyl trimethyl β-alanine (DGTA)	4		22:5	1	
			22:6	3	
Diacylglycerol (DAG)	2	Plastid/ Extraplastidial membranes			
Triacylglycerol (TAG)	1				
Free fatty acids	3	NA			

and SQDG are almost solely present in the plastid while the majority of PG is plastidial but with substantial proportions in the mitochondria and plasma membrane (Kobayashi and Wada, 2016).

Numerous fatty acids have also been detected in *P. tricornutum* and are detailed in Table 1.1, along with their abundances. Fatty acids can be named using the systematic IUPAC nomenclature, a trivial name or shorthand notation. Table 1.2 contains the names of the major FAs in *P. tricornutum*. I make heavy use of the shorthand, in which the number before the colon signifies the number of carbons in the fatty acid and the number after signifies the number of double bonds. Figure 1.5 depicts the numbering procedure for placement of double bonds. Double bonds are numbered from the carboxyl end, labelled Δ, or the methyl end, labelled n or ω. For example, a Δ9 double bond is located between the 9th and 10th carbon from the carboxylic acid group. Fatty acids of 14, 16, 18, 20, 22 and 24 carbons in length are present in the cell. Those fatty acids of 18 and 24 carbons are minor, as are 20 carbon species except for 20:5, also known as EPA. 20:5 constitutes a third of total FAs, 16:1 a fifth, 16:0 and 16:3 more than a tenth each and 14:0 and 16:2 each around a twentieth. 22:6, known as DHA, is present at around 3mol% (Abida et al., 2015; Alonso et al., 1994).

Table 1.2 – FA naming. a:bΔc notation - a is the length of the acyl chain, b is the number of double bonds. The double bond is placed c carbons from the carboxylic acid group, denoted Δ. Trivial names are given where available

Systematic name	Notation (C atoms/ desaturation)	Trivial name
Tetradecanoic acid	14:0	Myristic acid
Hexadecanoic acid	16:0	Palmitic acid
Hexadec- 9- enoic acid	16:1Δ9	Palmitoleic acid
Hexadec- 9,12- dienoic acid	16:2Δ9,12	
Hexadec- 6,9- dienoic acid	16:2Δ6,9	
Hexadec- 6,9,12- propenoic acid	16:3Δ6,9,12	
Hexadec- 6,9,12,15- butenoic acid	16:4Δ6,9,12,15	
Octadecanoic acid	18:0	Stearic acid
Octadec- 9 -enoic acid	18:1Δ9	Oleic acid
Octadec- 9,12 -dienoic acid	18:2Δ9,12	Linoleic acid
Eicosapentaenoic acid	20:5Δ5,8,11,14,17	
Docosahexaenoic acid	22:6Δ4,10,13,16,19	

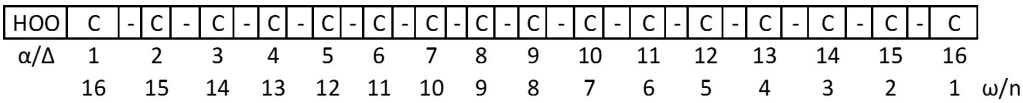


Figure 1.5 – Fatty acid nomenclature. Positions of double bonds in fatty acids are numbered from the carboxylic acid end, labelled α or Δ, or from the methyl end, named ω or n. For example, a Δ9 double bond is located between the 9th and 10th carbons from the Δ end.

In *P. tricornutum*, the FA profile of individual glycerolipid classes varies. MGDG and PC are composed of around 35mol% 20:5 and contain the majority of that FA in the cell with MGDG containing twice the amount in PC and PC containing around twice the amount in DGDG. MGDG contains the majority of the polyenoic C16 FAs, with 16:3 and 16:4 each contributing less than 5mol% of FAs in the other lipids. In TAG, the main FAs present are 16:0 and 16:1. These are present at roughly equal amounts in nutrient replete conditions but 16:1 is present at around twice the amount of 16:0 in nitrogen depletion (Abida et al., 2015; Bower, 2015). 14:0 and 20:5 are also present in TAG.

Several *P. tricornutum* lipid metabolism genes have been experimentally investigated in *P. tricornutum*, including FA desaturases, a FA elongase, a thioesterase and an acyl-ACP thioesterase acting in fatty acid metabolism, several diacylglycerol acyltransferases, a glycerol 3-phosphate acyltransferase (GPAT) and an lysophosphatidylglycerol acyltransferase (LPAT) in the Kennedy Pathway and TAG lipases (Domergue et al., 2003; Zhu et al., 2017; Xue et al., 2014; Domergue et al., 2002; Jiang et al., 2014; Li et al., 2018; Barka et al., 2016; Dinamarca et al., 2017; Cui et al., 2018; Gong et al., 2013; Guihéneuf et al., 2011; Niu et al., 2013; Radakovits et al., 2011; Niu et al., 2016; Wang et al., 2018a; Gong et al., 2011; Hao et al., 2018).

It should be noted that as a secondary endosymbiont of a red alga, there are significant metabolic differences between *P. tricornutum* and green algae and land plants (Gruber and Kroth, 2017). Nevertheless, a model that uses information from these organisms and the above knowledge of *P. tricornutum* has been developed that describes the major features of lipid synthesis in *P. tricornutum* (Figure 1.6).

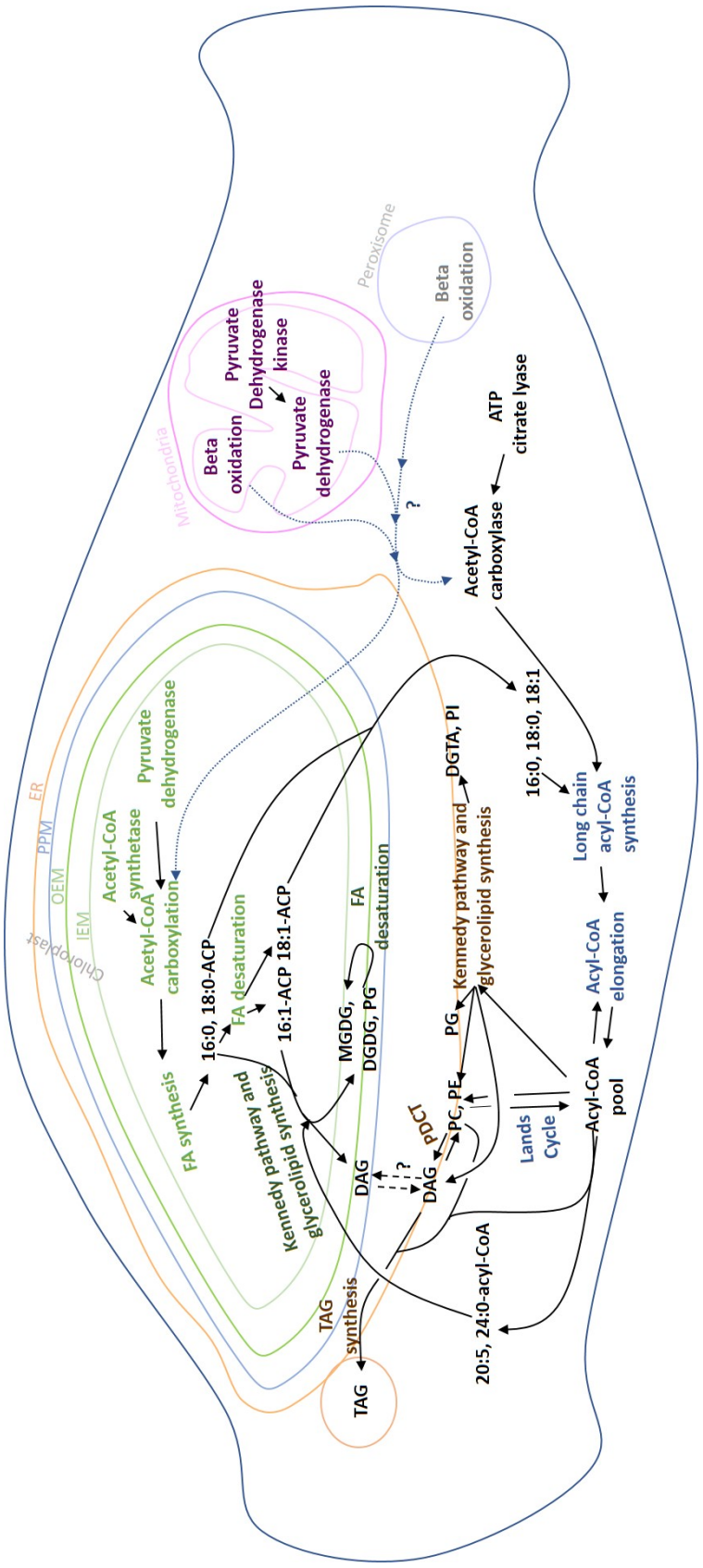


Figure 1.6 – Lipid metabolism in *P. tricornutum*.

Figure 1.6 – Lipid metabolism in *P. tricornutum*. Pyruvate dehydrogenase in the plastid and mitochondria, plastidial acetyl-CoA synthetase and mitochondrial and peroxisomal β-oxidation contribute to acetyl-CoA production. The contribution of each mechanism to the acetyl-CoA pool is uncertain, as are the transport mechanisms involved. Acetyl-CoA is carboxylated and transesterified to ACP in the plastid. Malonyl-ACP is used for FA synthesis in the plastid and FAs of 16 and 18 carbons in length are synthesised. C16 FAs are incorporated into DAG via a plastidial Kennedy Pathway and subsequently into MGDG, DGDG and PG, also in the plastid. Desaturation occurs up to 16:4 on MGDG. 16:0, 18:0 and 18:1 FAs are exported from the plastid by an unidentified mechanism and incorporated into acyl-CoA. Acyl-CoA is used to acylate glycerol 3-phosphate by the ER Kennedy Pathway and the DAG produced is subsequently used in the synthesis of PC, PE, PG, DGTA and PI on the ER membrane. FA acids are further desaturated on glycerolipids and elongated as acyl-CoA, efficiently producing 20:5. The Lands Cycle facilitates the exchange of FAs between the sn-2 position of PC and PE and the acyl-CoA pool that is required for LC-PUFA synthesis. PDCT produces DAG from PC, providing an additional source to the Kennedy Pathway. An exchange of DAG between the plastid and ER may occur but is not certain. DAG can be used to produce TAG in a canonically ER based process using DAG and FAs from either a glycerolipid or acyl-CoA. Across organisms, the donor glycerolipid is typically a phospholipid but can be MGDG. TAG sythesis may also occur in the plastid. 20:5 and 24:0 are reimported into the plastid by an unidentified mechanism. 20:5 is incorporated into all plastid lipids and 24:0 into SQDG. Information compiled from Fabris et al. (2012), Zulu et al. (2018), Abida et al. (2015) and Flori et al. (2016). IEM- inner envelope membrane, OEM- outer envelope membrane, PPM- periplastidial membrane, ER- endoplasmic reticulum, ACP-acyl carrier protein, FA- fatty acid, DAG - diacylglycerol, TAG- triacylglycerol, MGDG- monogalactosyldiacylglycerol, DGDG- digalactosyldiacylglycerol, PG-phosphatidylglycerol, PC- phosphatidylcholine, PE- phosphatidylethanolamine, PG-phosphatidylglycerol, DGTA- diacylglycerol trimethyl β-alanine, PI- phosphatidylinositol, PDCT- Phospholipid:DAG cholinephosphotransferase. FA nomenclature described in Table 1.2

Acetyl-CoA along with glycerol 3-phosphate and the head group molecules are the starting points for glycerolipid synthesis. Based on sequence similarity, sources of acetyl-CoA in *P. tricornutum* include a plastidial acetyl-CoA synthetase, a cytoplasmic ATP citrate lyase/synthase, a plastidial and a mitochondrial pyruvate dehydrogenase and beta oxidation located in both the mitochondria and peroxisome. Acetyl-CoA is then converted to malonyl-CoA by acetyl-CoA carboxylase (ACC) (Zulu et al., 2018), in what is a committed step for lipid metabolism, followed by transacylation to acyl carrier protein (ACP) in the plastid. ACC is thought to be located in both the plastid and cytoplasm in *P. tricornutum*. Enzymes of the plastidial Type II fatty acid synthase (FAS) are predicted to be located in the plastid (Ryall et al., 2003). Four monomeric enzymes constitute the Type II FAS system,

catalysing a cyclical sequence of condensation, reduction, dehydration and a further reduction, in which two carbons from malonyl-ACP are transferred onto the growing acyl chain with each round. In *P. tricornutum*, FA synthesis produces acyl chain lengths of 14, 16 and potentially 18 carbons that are then either incorporated into glycerolipids within the plastid or exported.

The Kennedy pathway encompasses the sequential addition of acyl chains to glycerol 3-phosphate (G3P) and as such is the origin of glycerolipids. Schematics of these reactions and others involved in DAG and TAG synthesis are shown in Figure 1.7. Gene homologues encoding putative prokaryotic and eukaryotic Kennedy pathways have been identified in *P. tricornutum*. The prokaryotic pathway is named due to its location in the plastid, while the eukaryotic pathway is located in the cytosol and endomembranes. Further evidence comes from FA content of lipids. Plastid lipids contain C16 FAs at the sn-2 position while extraplastidial lipids contain C16, C18 and C20 (Abida et al., 2015). As a result, it is not possible to identify transfer of 'eukaryotic' glycerolipids into the plastid as in plants, although this mechanism is not excluded. In the Kennedy pathway, GPAT and LPAT catalyse the addition of acyl chains at the sn-1 and sn-2 positions, producing lysophosphatidic acid (LPA) and phosphatidic acid (PA) respectively. By analogy with other organisms, PA can be dephosphorylated by phosphatidic acid phosphatase (PAP) to produce DAG or is combined with cytosine triphosphate by cytosine diphosphate (CDP)-DAG synthase to yield CDP-DAG. DAG and CDP-DAG can then be utilised in the synthesis of phospholipids and glycolipids.

C16 and small amounts of C18 FAs are incorporated into lipids in the prokaryotic pathway. Plastidial FAs can be desaturated (Dolch and Maréchal, 2015). A soluble Δ9 desaturase has been identified bioinformatically. This desaturase putatively catalyses the production of C16:1$^{\Delta 9}$ and to a much lesser extent, C18:1$^{\Delta 9}$. Further desaturation occurs on FAs incorporated into a glycerolipid. Plastid FAs contain up to four double bonds. The detected species are 16:2$^{\Delta 9,12}$, 16:2$^{\Delta 6,9}$, 16:3$^{\Delta 6,9,12}$ and very small amounts of 16:4$^{\Delta 6,9,12,15}$. A Δ12 desaturase catalyses conversion of C16:1$^{\Delta 9}$ to 16:2$^{\Delta 9,12}$ (Domergue et al., 2003). A Δ6 desaturase was identified by homology and is thought to catalyse the production of 16:2$^{\Delta 6,9}$. Either the Δ12 or Δ6 desaturase could then produce 16:3$^{\Delta 6,9,12}$, although it is not clear what each enzyme

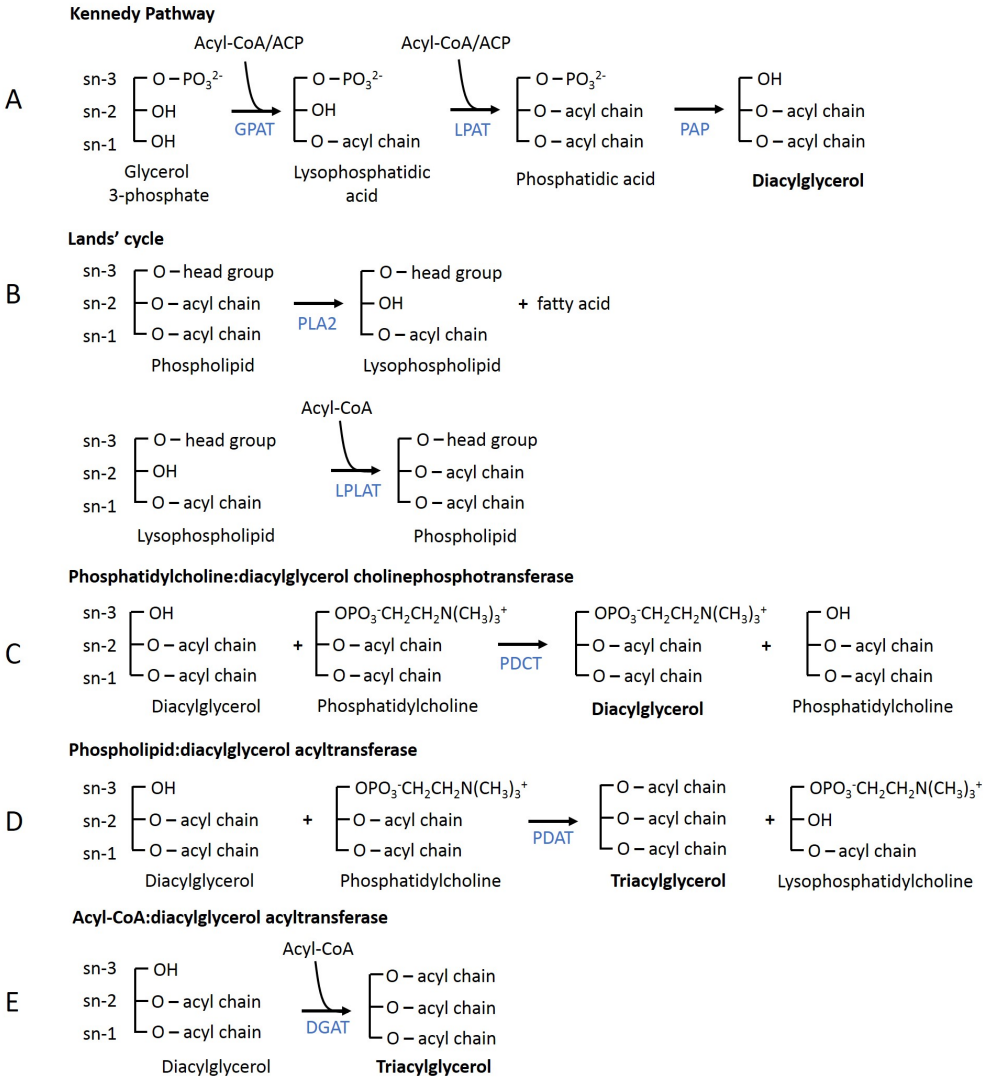


Figure 1.7 – Diacylglycerol and triacylglycerol synthesis reactions in *P. tricornutum*. Diacylglycerol (DAG) is synthesised via the Kennedy Pathway by the sequential action of glycerol 3-phosphate and lysophosphatidic acid acyltransferases (GPAT, LPAT) that acylate glycerol 3-phosphate in the plastid and ER membranes, followed by removal of the phosphate group by phosphatidic acid phosphatase (PAP). The Lands' cycle remodels extraplastidial lipids. Phospholipase A2 (PLA2) first removes a fatty acid from the sn-2 position and lysophospholipid acyltransferase (LPLAT) replaces this with one from the acyl-CoA pool in the cytoplasm. Phosphatidylcholine:diacylglycerol choline phosphotransferase (PDCT) transfers the phospocholine head group of phosphatidylcholine (PC) onto diacylglycerol, producing DAG with an altered fatty acid (FA) composition. TAG can be synthesised by the transfer of a FA from the sn-2 position of a lipid to the sn-3 position of DAG by phospholipid diacyglycerol acyl transferase (PDAT) or the addition of a fatty acid to DAG by diacylglycerol acyltransferase (DGAT). PDAT and DGAT are typically thought to transfer FAs from phospholipids and acyl-CoA respectively.

contributes to this activity. Finally, a bioinformatically identified ω-3 desaturase putatively catalyses synthesis of 16:4$^{\Delta 6,9,12,15}$. MGDG, DGDG, SQDG and part of the PG content are thought to be synthesised in the chloroplast (Abida et al., 2015) and several MGDG, DGDG and SQDG synthases have been identified by homology (Abida et al., 2015).

The mechanisms of import and export of fatty acids from the plastid are not well defined. An acyl-ACP/CoA thioesterase with substrate preference for 18:0-CoA over 18:1-CoA and 16:0-CoA and 18:1-ACP over 18:0-ACP may be involved in transport in or out of the plastid. This would occur through hydrolysis of the thioester bond to release an acyl chain that subsequently diffuses or is transported through membranes (Hao et al., 2018; Gong et al., 2011). Export of glycerolipids may also occur, as supported by the presence of C16 at the sn-2 position in extraplastidial lipids. However, this could also be produced by an LPAT with a promiscuous acyl chain length specificity and hence rather provide glycerolipids for import into the plastid. Indeed, import of EPA is required given its expected synthesis in the extraplastidial compartments. The position of EPA at the sn-1 position in MGDG has been proposed to arise from the import of free EPA rather than through the import of eukaryotic glycerolipids, however, experimental verification is still needed (Abida et al., 2015).

Exported free fatty acids are activated to acyl-CoA in the cytosol. Subsequently they are further desaturated and elongated. Recurrent transfer between the acyl-CoA and glycerolipid pools is required as most of the desaturation occurs on a glycerolipid backbone while elongation proceeds on acyl-CoA (Hamilton et al., 2014). The enzymes of the eukaryotic Kennedy pathway incorporate acyl chains into glycerolipids while enzymes of the Lands Cycle both incorporate and release them. The Lands cycle consists of phospholipase A2 (PLA2) and lysophospholipid acyltransferase (LPLAT) and respectively removes and adds acyl chains from/to the sn-2 position of phospholipids. The Lands Cycle reactions are illustrated in Figure 1.7B. Given the low levels of C18 FAs in *P. tricornutum* lipids, it has been suggested that EPA production is very efficient (Hamilton et al., 2014). A soluble Δ9 desaturase has also been identified in the cytoplasm and is another way in which 18:1$^{\Delta 9}$ could be sythesised. Synthesis of 18:2$^{\Delta 9,12}$ is catalysed by fatty acid

desaturase (FAD) 2 (Domergue et al., 2003). After this point synthesis of EPA and DHA can occur by various sequences of desaturation and elongation. However, based on radiolabelling experiments, desaturation is thought to mainly proceed to $18{:}4^{\Delta 6,9,12,15}$ (Dolch and Maréchal, 2015), followed by elongation to $20{:}4^{\Delta 8,11,14,17}$ and desaturation to $20{:}5^{\Delta 5,8,11,14,17}$. A longer FA 24:0, also constitutes part of SQDG (Abida et al., 2015). A Δ6 FAD, Δ5 FAD and Δ5 elongase have been experimentally characterised (Domergue et al., 2002; Jiang et al., 2014). The Lands Cycle enables the incorporation of these LC-PUFAs into glycerolipids.

PC, DGTA, some PG, PI and PE are all thought to be synthesised in the extraplastidial membranes via the eukaryotic pathway (Abida et al., 2015), several genes involved in their synthesis have been identified via homology (Levitan et al., 2015; Fabris et al., 2012). Further to the activities of PAP, DAG can be synthesised by phosphatidylcholine:diacylglycerol cholinephosphotransferase (PDCT). PDCT transfers the phosphocholine headgroup from a PC onto a DAG and as a result, DAG with an altered FA composition can be produced. The PDCT reaction is illustratd in Figure 1.7C. Diacylglycerol acyltransferase (DGAT) synthesises DAG and acyl-CoA while PDAT transfers an acyl chain to DAG, typically from PC, as outlined in Figure 1.7D and E. Seven DGATs have been identified in *P. tricornutum* and a number have been published in current literature (Dinamarca et al., 2017; Cui et al., 2018; Gong et al., 2013; Guihéneuf et al., 2011; Niu et al., 2013). A single PDAT is known and during the course of this work, I identified another gene annotated as such.

In many plants, TAG is synthesised on the endoplasmic reticulum (ER) membranes and accumulated as lipid bodies within the cytoplasm (Cagliari et al., 2011). Cytoplasmic lipid bodies are also observed in *P. tricornutum*. In addition, imaging of lipid droplets in 3D supports the presence of lipid bodies within the chloroplast or closely associated with it (Cavonius et al., 2015). In *T. pseudonana*, support was found for localisation of a DGAT in the chloroplast during exponential growth, with a movement to the ER during stationary phase (Manandhar-Shrestha and Hildebrand, 2015). In *P. tricornutum*, increased TAG accumulation resulted from overexpression of plastidial GPATs and an LPAT (Wang et al., 2018a; Niu et al., 2016) and support was found for the localisation of a TAG lipase in the periplastidial membrane. There are multiple isoforms of DGAT in *P. tricornutum* (Chen and

Smith, 2012). One is predicted to be localised to the chloroplast, another in the mitochondria and the remainder in the ER. Collectively this evidence indicates that TAG synthesis in diatoms may be different to that of other organisms, located within both the chloroplast and the ER.

Lipid catabolism has received little attention in diatoms except for a number of studies of TAG lipases and a putative membrane lipid lipase (Li et al., 2018; Barka et al., 2016; Trentacoste et al., 2013). The lipid accumulation that resulted from knock-down or knock-out of these genes highlights the importance of understanding the breakdown of lipids in diatoms.

1.3 Investigating the metabolism of TAG accumulation in *P. tricornutum*

1.3.1 Nitrogen depletion as a model of triacylglycerol accumulation in *P. tricornutum*

This thesis aims to improve knowledge of lipid metabolism in *P. tricornutum* with a particular focus on the metabolism of TAG accumulation so that TAG yield and composition may be improved. Algae cultured in adverse environmental conditions offer experimental systems for investigation of TAG metabolism. It has long been known that certain stress conditions cause microalgae to accumulate substantial amounts of TAG (Sheehan et al., 1998). Of these, nutrient deprivation generates some of the highest levels of accumulation. In *P. tricornutum*, both nitrogen and phosphorus depletion are commonly used systems in the investigation of TAG accumulation. Nitrogen depletion causes a substantially more rapid response and more literature into its effect on *P. tricornutum* was available at the beginning of this study. Nitrogen depletion was therefore chosen as the experimental system in which to investigate mechanisms of TAG accumulation.

Nitrogen depletion causes system-wide changes in *P. tricornutum* that lead to a slowing and the cessation of cell division. This reduction in biomass limits culture TAG yield and so is not considered a suitable strategy for industrial production of TAG. For this reason, it is desirable to identify mechanisms that can enable TAG

accumulation without substantially reducing growth. Numerous studies have been undertaken to pursue this aim and the study of *P. tricornutum* at a systems level has been particularly important in this.

The combination of transcriptomic, proteomic, lipidomic and physiological measurements has revealed the complexity of the process of acclimatisation to nitrogen depletion (Alipanah et al., 2015; Levitan et al., 2015; Ge et al., 2014; Yang et al., 2013; Valenzuela et al., 2012; Abida et al., 2015). Changes occur that ultimately lead to the development of a quiescent cell that has drastically reduced its ability to capture light and fix carbon, has accumulated large stores of TAG and is primed for nitrogen resupply. Specifically, reduced amounts of light harvesting proteins, chlorophyll a and photosynthetic efficiency underpin decreased light capture. The structure of the chloroplast is also altered and the quantities of MGDG and PG in the cell are decreased. Reduced assimilation of carbon is supported by downregulation of Calvin Cycle transcripts. Upregulation of nitrate and urea transporters, nitrate reductase and the glutamine synthetase/glutamine 2-oxoglutarate aminotransferase (GOGAT) transcripts supports increased nitrogen uptake and assimilation.

Although the current transcriptomic and proteomic literature has provided insight into the process of acclimatisation to nitrogen depletion, a greater focus on changes in expression that impact lipid metabolism would be useful. During nitrogen depletion, TAG accumulation itself is majorly driven by *de novo* synthesis of FAs and transcriptomic and proteomic studies have highlighted the likely role of alterations to central metabolism in providing carbon for these FAs. This carbon is partly provided by catabolism of branched chain amino acids (Ge et al., 2014; Pan et al., 2017) and partly by redistribution of carbon flux away from protein synthesis (Levitan et al., 2015; Valenzuela et al., 2012; Wagner et al., 2017). However, the same transcriptomic and proteomic studies also found changes in expression of lipid metabolism genes. For example, some studies found increased expression of FA synthesis genes and Kennedy Pathway genes (Longworth et al., 2016; Alipanah et al., 2015). Investigations to understand the contribution of lipid metabolism alterations to TAG accumulation are therefore important. However, several aspects of the current transcriptomic and proteomic studies limit the interpretations that can be drawn from their data. These include the basis of the studies on one experiment, the dynamic nature of the

acclimatisation response through time coupled with the limited number of sampled time points and the varied conditions cells were cultured under. These limitations make outlining a consensus response of expression to acclimatisation difficult and therefore these limitations need to be addressed.

1.3.2 Lipid recycling dependent synthesis of TAG

Lipid remodelling is important in acclimatisation to multiple adverse conditions in photosynthetic organisms although the exact response can differ depending upon the stress and between organisms. TAG accumulation is common to stress caused by unfavourable temperatures and salinities, high light, dehydration and deprivation of numerous nutrients including nitrogen, phosphorus, iron, zinc, silicon and sulphur (Du and Benning, 2016). Polar lipids are also altered. Specifically regarding acclimatisation to nitrogen depletion, degradation of MGDG coupled with synthesis of TAG has been identified as part of the response in land plants (Gaude et al., 2007), green algae (Martin et al., 2014; Yoon et al., 2012; Du et al., 2018; Li et al., 2012), diatoms (Abida et al., 2015) and eustigmatophytes (Simionato et al., 2013; Jia et al., 2015; Martin et al., 2014). Across the species the lipid changes are more variable but decreases in MGDG and increases in TAG are consistent features and are often accompanied by a decrease in PG. In addition, DGDG is commonly increased in *C. reinhardtii* and *A. thaliana* and decreases in SQDG and PC have been observed in *Nannochloropsis sp.* (Jia et al., 2015; Simionato et al., 2013; Martin et al., 2014) and *Chlorella sp.* (Martin et al., 2014). In addition to nitrogen depletion, a decrease in MGDG and increase in TAG has been observed in freezing and ozone stress in multiple plant species (Du and Benning, 2016).

In *P. tricornutum*, TAG is present in small amounts in the cell during replete growth. The quantity follows a diurnal cycle, accumulating during the day and decreasing at night. In contrast, during nitrogen depletion, TAG accumulates dramatically and can account for more than 50% mol/mol of lipids and 10% w/w of cellular carbon (Abida et al., 2015; Levitan et al., 2015). In this condition, the majority of TAG originates from de novo synthesis. This is supported by the high content of C16 FAs within TAG and the increased total lipid content under nitrogen depletion. 16:0 and 16:1 make the main contribution to this increase in lipids, but

the content of 20:5 (EPA) also increases. In addition to increased TAG quantity, MGDG and PG decrease during nitrogen depletion while other lipids may undergo minor alterations as noted by Alonso et al. (2000); Abida et al. (2015) and in my MRes work (for which data are presented in Figures A.1 and A.2). It is worth noting that Popko et al. (2016) identified a major decrease in DGTS and suggested this contributes to TAG accumulation in the Pt4 strain. However, such large quantities of DGTS have not been detected in other studies of *P. tricornutum* and so verification of this is needed. Calculations based on my MRes work and studies by Levitan et al. (2015) and Abida et al. (2015) indicate that recycled constituents of membranes could contribute 12-20% of accumulated TAG. The decrease in quantity of MGDG is much greater than that of PG, reflecting its abundance in the cell. As mentioned, MGDG also contains the majority of the EPA in the cell. TAG in *P. tricornutum* contains substantial amounts of EPA (Arao et al., 1987) and under nitrogen depletion, the increased quantity of EPA in TAG was observed to correspond to the decrease in MGDG (Abida et al., 2015). The potential for transfer of fatty acids from chloroplast lipids to TAG in *P. tricornutum*, particularly from MGDG, is therefore supported by these lipidomic observations.

The movement of MGDG constituents to TAG has also been observed in other photosynthetic organisms and under a number of stress conditions. Two enzymes, *C. reinhardtii* plastid galactolipid degradation 1 (CrPGD1) (Li et al., 2012; Du et al., 2018) and *C. reinhardtii* phospholipid:diacyglycerol acyl transferase (CrPDAT), have demonstrated this activity under nitrogen depletion. CrPGD1 hydrolysed *C. reinhardtii* derived MGDG *in vitro*, releasing 16:0 and 18:1Δ9, probably from the sn-1 position. Relative to other chloroplast lipids, CrPGD1 mutants contained higher proportions of MGDG and also had a 50% reduction in overall TAG content during nitrogen depletion. Pulse-chase labelling indicated the mutation prevented transfer of radiolabel from MGDG to TAG while 16:1Δ7 and 18:1Δ9 were enriched in MGDG but depleted in TAG in the mutant. CrPGD1 is upregulated during nitrogen depletion. The second enzyme, CrPDAT is located in the chloroplast and *in vitro* transfers acyl chains from a particularly wide variety of substrates onto DAG. These substrates include MGDG but not DGDG or SQDG. Knockdown of CrPDAT reduced TAG accumulation early during nitrogen depletion, with no effect

later in the response. Knock-downs demonstrated increased levels of the plastid lipids MGDG, SQDG and PG under nitrogen replete conditions (Yoon et al., 2012).

The *A. thaliana* sensitive to freezing 2 (AtSFR2) gene has been characterised as a galactolipid:galactolipid galactosyl transferase (GGGT) and implicated in production of oligogalactolipids during freezing stress, transferring the galactosyl moiety to other galactolipids (Moellering et al., 2010). TAG accumulation is impaired in AtSFR2 knock-out lines. DAG with 16:3 at the sn-2, characteristic of MGDG, was released in an assay of the activity of AtSFR2 and 16:3 was found at the sn-2 position of TAG in freeze treated *A. thaliana*. AtSFR2 is located in the outer envelope of the chloroplast and is therefore able to access the galactolipid substrates. Collectively this suggests DAG is released from MGDG by AtSFR2 and is further incorporated into the synthesis of TAG.

Nitrogen depletion in *P. tricornutum* is thought to cause an imbalance between the intake of carbon and energy and their cellular requirement (Wagner et al., 2017). This likely results from reduced and eventually halted growth although a contribution from altered macromolecule composition has also been proposed based on the varying numbers of electrons required to make a unit of carbohydrate, protein or lipid (Wagner et al., 2017; Kroon et al., 2006). A decreased requirement for reducing power results in light absorption exceeding the capacity of the electron transport chain due to decreased anabolic processes. As a result, excess light absorption causes damage to thylakoid components through mechanisms such as the Mehler reaction and photinhibition (Badger et al., 2000; Du and Benning, 2016). To prevent damage and balance the ATP/NADPH ratio, the cell can decrease light capture, dissipate captured light, direct electrons into alternate electron pathways and alter the macromolecular composition of the cell (Wagner et al., 2017; Curien et al., 2016). It would therefore be expected that nitrogen depletion would lead to processes that decrease light capture and electron excitation.

MGDG is located in the chloroplast of photosynthetic organisms although the distribution across the membranes of *P. tricornutum* is yet to be determined (Petroutsos et al., 2014). In land plants, MGDG plays roles in coordinating and stabilising interactions between photosystem (PS) II monomers, light harvesting complex (LHC) II components and between PSII and LHCII (Nakamura and Li-Beisson, 2016). The

lipid is involved in electron transport within PSII and between the two photosystems and its presence is necessary for a well ordered thylakoid. Knock-down of MGDG synthase led to a 50% reduction in cellular chlorophyll content. Therefore, in addition to the decreased quantities of light harvesting complex proteins and chlorophyll, decreased MGDG content during nitrogen depletion is also likely to reduce light capture and potentially the efficiency of the electron transport chain.

Alternative reasons for movement of chloroplast lipid constituents to TAG have also been suggested. These include the storage of chloroplast lipid FAs in TAGs to enable their rapid remobilisation (Du and Benning, 2016). MGDG is important in photosynthesis and EPA constitutes a third of its FAs in *P. tricornutum*. The synthesis of LC-PUFAs require a large number of electrons. Therefore, rapid remobilisation of these FAs upon return of favourable conditions may be advantageous for growth. Additionally, the decreased anabolic need for reducing equivalents evidenced by reduced growth and protein synthesis (Levitan et al., 2015) may indicate that storage of FAs is more beneficial to the cell than their direction to β-oxidation.

In conclusion, there are multiple strands of support for a transfer of MGDG constituents to TAG in *P. tricornutum*. These include both direct observations, inferences drawn from shared acclimatisation responses and postulated metabolic advantages. Decreases in MGDG and PG have been observed in *P. tricornutum* combined with increased EPA in TAG. Similar responses occur across diverse photosynthetic lineages and observations of transfers of MGDG constituents to TAG have been made in some. This can be explained by the shared need of photosynthetic organisms to avoid photodamage under adverse conditions or an advantage from storing MGDG constituents. Therefore, investigating the transfer of MGDG constituents to TAG during nitrogen depletion in *P. tricornutum* may identify mechanisms to increase TAG accumulation, produce oils of a desired composition and better understand the physiology of this diatom.

1.3.3 The triacylglycerol synthesis enzymes

Three types of enzyme activity are known that carry out TAG synthesis: acyl-CoA:diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase and DAG:DAG acyltransferase (DAGTA). The reactions of DGATs and PDATs are

outlined in Figure 1.7D and E. Enzymes demonstrating these activities make use of separate substrate pools. DGATs (sometimes referred to as DAGAT and DGTT) incorporate an acyl chain from acyl-CoA onto the sn-3 position of DAG. PDATs are typically thought to transfer an acyl chain from the sn-2 position of phospholipids to the sn-3 of DAGs. DAGTA transfers the acyl chain of one DAG to another. DAGTA activity has been observed in organisms (Buhman et al., 2002; Lehner, 1993; Stobart et al., 1997), however a specific DAGTA has not been identified and so this enzyme class was not pursued for this thesis.

1.3.3.1 Diacylglycerol acyltransferases

DGATs are widespread across lineages and belong to divergent classes. Prokaryotes synthesise TAGs using bifunctional wax ester synthase/DGAT (WS/DGAT) enzymes (Rottig and Steinbuchel, 2013; Arabolaza et al., 2008; Barney et al., 2012) and DGATs are thought to be ubiquitous in eukaryotes, synthesising TAG for carbon, energy and FA storage (Yen et al., 2008). Based on conserved domains, the eukaryotic DGATs are divided into four groups named Type 1, 2, 3 and WS/DGAT. The DGAT classes do not share sequence similarity and there are substantial structural differences between them. DGAT1 genes are longer, contain more exons and transmembrane domains than DGAT2s and the proteins are structurally related to acyl-CoA:cholesterol acyl transferase. DGAT2s in contrast are structurally related to acyl-CoA wax alcohol acyltransferases and acyl-CoA:monoacylglycerol acyltransferases. DGATs 1 and 2 are membrane bound, DGAT3 enzymes are soluble while WS/DGATs seem to vary between isoform (Yen et al., 2008; Liu et al., 2012; Cao et al., 2013; Turchetto-Zolet et al., 2016). Given their shared function and divergent sequences, the DGAT classes are proposed to be an example of convergent evolution (Liu et al., 2012). Additionally, a number of enzymes demonstrate DGAT activity in combination with others, for example mammalian acyl-CoA:monoacylglycerol acyltransferase (MGAT), acyl-CoA:wax alcohol acyltransferase (AWAT) and plant phytyl ester synthase (PES) (Lippold et al., 2012; Liu et al., 2012).

The complement of DGAT classes can vary between organisms, as can their physiological roles. DGAT1 and 2 are present in animals, fungi, land plants and microalgae (Chen and Smith, 2012; Yen et al., 2008; Liu et al., 2012). In contrast,

DGAT3s seem to be absent from animals but they have been identified in land plants, green algae and yeast (Turchetto-Zolet et al., 2016; Bagnato et al., 2017; Yen et al., 2008). The role of the DGAT classes can differ between organisms and tissues. In *A. thaliana* and soybean, DGAT1 plays the major role in TAG synthesis in seeds and demonstrates higher expression in this tissue than in others (Zou et al., 1999; Li et al., 2010). However, in tung tree, castor bean and ironwood, species that accumulate unusual FAs in TAG, this is reversed with DGAT2 playing the major role in TAG synthesis in seeds and demonstrating higher expression than DGAT1 (Shockey, 2006; Kroon et al., 2006; Li et al., 2010). This functional differentiation is supported by DGAT1 and 2 operating in distinct ER subdomains in the tung tree cells (Shockey, 2006). In mammals and yeast, DGAT2 seems to make the larger contribution to TAG synthesis (Yen et al., 2008; Xu et al., 2018). DGAT2 knockouts in mice were lethal but those of DGAT1 were not (Smith et al., 2002) while DGAT2 also demonstrated higher substrate affinity (Cases et al., 2001). DGAT2 is missing from a number of yeast species including *S. cerevisiae* but is likely the primary TAG synthesis enzyme in *Yarrowia lipolytica* despite containing DGAT1 (Turchetto-Zolet et al., 2011). In some cases, the DGATs have specific substrate specificities that aid their roles. Whereas some DGATs use a broad range of acyl-CoA substrates, others demonstrate substrate specificity. Such specificity has been noted in DGAT2s of organisms that incorporate unusual FAs into TAG (Jeennor et al., 2017; Shockey, 2006; Liu et al., 2012).

Although less well studied, DGAT3 and WS/DGAT can also function in TAG synthesis and may carry out complementary and differentiated roles to each other and DGAT1 and DGAT2. DGAT3 from *C. reinhardtii* produces TAG when heterologously expressed and is predicted to be targeted to the chloroplast (Bagnato et al., 2017). DGAT3 was found to have similar patterns of expression to DGAT1 and 2 in soybean seeds and is present in more tissues and developmental stages than WS/DGAT in *A. thaliana* and soybean (Turchetto-Zolet et al., 2016).

In addition to the multiple DGAT classes and their differentiated roles in organisms, DGATs also vary in gene copy number. DGAT1 is found as a single copy in the majority of studied microalgae. *Nannochloropsis oceanica* (*N. oceanica*) contains two however, while a number of chlorophytes may lack a DGAT1 gene entirely (Wang

et al., 2014; Wagner et al., 2010; Chi et al., 2014). DGAT3 has been identified in green algae (Bagnato et al., 2017) and is present as a single copy. DGAT2, however, is found at increased copy numbers in the majority of microalgae. This contrasts with the typically single copy in fungi, plants and animals (Chen and Smith, 2012; Chi et al., 2014).

Two studies have been undertaken to investigate the phylogenies of DGAT2 enzymes (Chen, in preparation; Bagnato et al., 2017). Bagnato et al. (2017) included the DGAT2s of 79 eukaryotes. and identified four clades: Clade I- all animal and fungal; Clade II- land plant and some algal DGAT2s; Clade III- SAR supergroup and haptophyte DGAT2s; Clade IV- SAR, green algae and excavata DGAT2s. Chen et al. carried out a larger study and included 578 DGAT2s from 298 eukaryotes and identified five clades: Clade I, animal and fungal DGAT2s; Clade II, some bryophyte and rhodophyte DGAT2s; Clade III, some chlorarachniophyte, haptophyte and rhodophyte DGAT2s; clade IV - all other land plant DGAT2s; Clade V - some chlorarachniophyte, haptophyte and rhodophyte DGAT2s. Chen et al. found that heterokont and chlorophyte DGAT2s were present across all clades while charophyte DGAT2s were absent from clade I. They also suggested that this diversity of DGAT2s is explained by DGAT2 gene duplication in early eukaryotic history, prior to subsequent gene loss in some descendant lineages.

The large number of DGAT2s in microalgae are suggested to play differentiated roles. This is underpinned by differential expression of DGAT2s under conditions that induce TAG accumulation (Levitan et al., 2015; Msanne et al., 2012; Alboresi et al., 2016; Li et al., 2015), predicted targeting of DGATs to multiple cellular compartments (Zienkiewicz et al., 2016) and studies demonstrating multiple functional DGATs within one microalga. In support of this, DGAT2s in *C. reinhardtii* were found to have different substrate specificities (Liu et al., 2016b).

1.3.3.2 Phospholipid:diacylglycerol acyltransferases

PDAT activity was discovered at the turn of the millenium in sunflower (*Helianthus annus*), castor bean (*Ricinus communis*) and *Crepis palaestina* along with the PDAT gene from *S. cerevisiae* (Dahlqvist et al., 2000). Subsequently, PDATs have been identified in many land plant, fungi and algal species but not animals (Pan

et al., 2015). PDAT activity has also been observed in the soil dwelling bacterium *Streptomyces coelicolor* (Arabolaza et al., 2008).

PDATs share sequence and structure similarity with lecithin:cholesterol acyl transferase (LCAT) and PLA2 and two phylogenetic studies of LCAT-like genes segregated LCATs, PDATs, PLA2s and phospholipid:sterol acyl transferase (PSAT) into four clades (Yoon et al., 2012; Pan et al., 2015). A further phylogeny of green algal and land plant PDATs divided them into seven clades, with green algal PDATs segregated from those of land plants (Pan et al., 2015). All plant PDATs contain LCAT and α/β-hydrolase fold motifs (Pan et al., 2015). Four LCAT motifs are present and their distribution relative to each other is conserved. Green algal PDATs vary with regard to the presence of transmembrane domains whereas those investigated in land plants tend to have one. Green algae and monocots typically have a single copy of PDAT while dicots contain multiple copies that are thought to originate from gene duplication at the base of the phylum and subsequent segment duplication in certain species.

PDATs of different organisms have demonstrated specificities for varied glycerolipid substrates and acyl chains. PDATs are characteristically thought to utilise PC or PE as substrates (Dahlqvist et al., 2000). However CrPDAT was observed to utilise phospholipids, DAG and MGDG as acyl donors in TAG accumulation (Yoon et al., 2012). Variation in PDAT gene composition is higher in green algae than land plants (Pan et al., 2015) and these non-typical functionalities may be specific to green algae. However, it is not clear whether land plant PDATs may also use varied substrates. PDATs have also been noted for their ability to incorporate unusual FAs into TAG. Castor PDAT incorporates the epoxy-FA vernoloyl acid and hydroxy-FAs such as ricinoleoyl acid. *Crepis palaestina* PDAT can also use vernoloyl acid while flax PDAT can integrate α-linolenic acid (Kim et al., 2011; van Erp et al., 2011; Pan et al., 2013). In phylogenies, these PDATs cluster separately to others (Xu et al., 2018).

The contribution that PDATs make to TAG accumulation is species, tissue, life stage and isoform dependent and therefore the physiological role of PDAT can vary. PDAT contributes the major part of TAG synthesis in exponentially growing yeast (Oelkers et al., 2002). Support was found for PDAT1 making the major

contribution to TAG synthesis in young leaves of *A. thaliana* but a contribution to seed oil production was only identified in a DGAT1 knock-down background (Zhang et al., 2009). In contrast, mutation, overexpression and downreglation of PDAT1 in *Camelina sativa* altered seed oil content and TAG fatty acid composition (Aznar-Moreno and Durrett, 2017; Marmon et al., 2017). PDAT1 does not appear to play an substantial role in seed oil production in oleic acid rich seed oil plants (Troncoso-Ponce et al., 2011; Woodfield et al., 2018) and a substantial impact on TAG accumulation of another isoform, PDAT2, has not been observed in flax, castor bean or *A. thaliana* (van Erp et al., 2011; Kim et al., 2011). Further to seed oil production, PDAT has been implicated in stress responses in land plants (Mueller et al., 2017) and green algae (Yoon et al., 2012; Liu et al., 2016b). PDAT from *C. reinhardtii* uses FAs from MGDG to synthesise TAG during nitrogen depletion as mentioned above. A similar role has been suggested for PDAT from *Myrmecia incisa* with PC acting as the acyl donor (Liu et al., 2016b). Therefore, despite the variation, PDATs can play an important role in TAG accumulation.

1.3.3.3 Diacylglycerol acyltransferases and phospholipid:diacylglycerol acyltransferases in *P. tricornutum*

Six DGATs and a WS/DGAT have been identified in *P. tricornutum* and named based on conserved domains, phylogenetic analysis and functional investigation. These are detailed in Table 1.3 and are a DGAT1, five DGAT2s, named A-E and a WS/DGAT (Guihéneuf et al., 2011; Gong et al., 2013; Cui et al., 2013; Niu et al., 2013; Chen, 2013; Litvinenko, 2016; Dinamarca et al., 2017; Cui et al., 2018). The WS/DGAT was originally identified as a DGAT3 but has been re-annotated (Cui et al., 2013, 2018). The DGAT2s have been structurally analysed and found to contain the 6 motifs that are thought to be conserved across DGAT2s (Cao et al., 2011; Gong et al., 2013; Litvinenko, 2016). PtDGAT2A, B and C are in Clade V from the phylogeny produced by Chen outlined above. PtDGAT2D is in Clade III and DGAT2E is in Clade I. During the course of this PhD, another potential PDAT was noted with the identifier Phatr3_J49702 and is also included in Table 1.3. This PDAT was not investigated in this study and so reference to PDAT in *P. tricornutum* refers to Phatr3_J8860 unless otherwise stated.

Table 1.3 – The DGATs and PDAT of *P. tricornutum*.

Gene Name	Gene Identifiers	Identified	Functionality confirmed
DGAT1	Phatr3_J9794		Guiheneuf et al., 2011
DGAT2A	Phatr3_J49462		Niu et al., 2013
DGAT2B	Phatr3_J49544		Gong et al., 2013
DGAT2C	Phatr3_J31662	Chen and Smith, 2012	
DGAT2D	Phatr3_J43469		Dinamarca et al., 2017
DGAT2E	Phatr3_EG00369	Chen and Smith, 2012	
WS/DGAT	Phatr3_J49708		Cui et al., 2013; Cui et al., 2018
PDAT	Phatr3_J8860		Litvinenko, 2016
PDAT (putative)	Phatr3_J49702	This study	

A number of studies have been carried out to investigate the function of PtDGATs and have demonstrated activity of all except DGAT2C and DGAT2E. DGAT1, DGAT2B, DGAT2D and the WS/DGAT were found to reinstate TAG synthesis in the TAG synthesis deficient H1246 *S. cerevisiae* strain (Cui et al., 2018; Niu et al., 2013; Litvinenko, 2016; Guihéneuf et al., 2011; Gong et al., 2013). Overexpression of DGAT2A, DGAT2D and WS/DGAT in *P. tricornutum* have been reported to increase TAG synthesis under replete growth conditions (Cui et al., 2018; Niu et al., 2013; Dinamarca et al., 2017) with little impact upon rate of division. WS/DGAT overexpression also increased TAG synthesis under nitrogen depletion and has been demonstrated to be mutifunctional, producing wax esters in yeast and *P. tricornutum* (Cui et al., 2018).

Some information on substrate preference is also available. DGAT1 appears to demonstrate a greater preference for saturated FAs than monounsaturated FAs, preferentially incorporating 16:0 and 18:0 and lower proportions of 16:1 and 18:1 than *S. cerevisiae* DGAT2D (Guihéneuf et al., 2011). Work by a former member of the Plant Metabolism Group, Dr Alex Litvinenko demonstrated that expression of DGAT2B in *S. cerevisiae* H1246 produced greater proportions of TAG that had an increased mass, likely due to containing C18 rather than C16 FAs (Litvinenko, 2016). In another study, heterologous expression of DGAT2B increased the proportion of 18:1 and 18:2 in the TAG of *P. pastoris* while that of 16:0 and 18:0 decreased (Gong et al., 2013). PtDGAT2B did not demonstrate a greater preference for EPA than *S. cerevisiae* DGAT2. DGAT2B overexpression in *P. tricornutum* increased the proportion of TAG species with 50 carbons in their combined acyl chains and decreased those species with 46 and 48 (Litvinenko, 2016). Considering WS/DGAT,

its overexpression increased the proportion of TAG composed of saturated FAs and reduced the proportion of unsaturated FAs (Cui et al., 2018).

A PDAT has been identified in *P. tricornutum* and its characterisation was begun by Dr Litvinenko. The PDAT contains an α/β-hydrolase fold and LCAT domain (Litvinenko, 2016). The seven structural motifs that are characteristic of LCAT-like enzymes, includind PDATs, were also identified in PtPDAT (Litvinenko, 2016; Yoon et al., 2012; Pan et al., 2015). Overexpression in *P. tricornutum* during nitrogen depletion was found to alter the TAG species composition, decreasing the proportion that contained acyl chains with a total of 48 carbon atoms and increasing the proportion collectively composed of 52, 54 and 56 carbon atoms (Litvinenko, 2016). Greater detail on the current characterisation of the DGATs and PDAT is provided in the introduction to Chapter 5.

The multiplicity of functional TAG synthesis enzymes in *P. tricornutum* raises questions about their physiological purpose. As characterisation in plants and other organisms has demonstrated, these enzymes may play differentiated roles, acting at different times towards separate ends. This differentiation is not only interesting in understanding the basic biology of TAG synthesis in *P. tricornutum* but also offers a potential resource for biotechnological exploitation. Knowledge of which of the enzymes demonstrates the greatest activity is important for efforts to engineer TAG accumulation while information on substrate specificities may enable tailoring of oil fatty acid composition.

1.4 Thesis Aims

With the above in mind, I pursued a number of aims in this thesis:

1. Produce a comparison of current studies to identify consensus regulation of physiology, transcripts and proteins during nitrogen depletion: Chapter 3.

2. Use this consensus of the regulatory response to nitrogen depletion to investigate lipid metabolism, with a particular focus on mechanisms of TAG accumulation: Chapter 3.

3. Continue characterisation of *P. tricornutum* DGATs and PDATs: Chapter 4.

4. Investigate the sources of DAG and FAs used in TAG synthesis during nitrogen depletion: Chapter 5.

5. Identify lipid catabolism genes putatively involved in degradation of chloroplast membrane lipids and characterise their functions: Chapter 5.

2

Materials and Methods

2.1 Chemical reagents

Analytical grade dimethyl formamide (DMF) and HPLC grade heptane, chloroform and methanol were used and obtained from Fisher Scientific. Stearic acid (>98.5%), palmitic acid (>99%), trimyristate (> 99%), Nile Red, and analytical grade dimethyl sulphoxide (DMSO) were bought from Sigma-Aldrich while cerulenin (>98%) was bought from Cayman Chemical. All restriction endonucleases and Taq DNA ligase were provided by NEB. Nuclease free water was obtained from Thermo Fisher Scientific. Oligonucleotide primers were synthesised by Sigma Aldrich. Hyperladder 100bp and 1kb molecular weight markers and deoxynucleotide triphosphates were provided by Bioline. Spectinomycin, carbenicillin and kanamycin were supplied by Thermo Fisher Scientific and Zeocin was supplied by Invitrogen. Plasmids used within this study are presented in Table 2.1

2.2 General Techniques

E. coli cell concentration was measured via optical density of cultures at 600 nm (OD600) blanked against appropriate media, in a UV1 UV-Vis spectrophotometer from Thermo Spectronic, using a 1 cm path length. *P. tricornutum* cell density was measured using a Z2 Coulter Counter from Beckman Coulter, with a particle diameter between 3 to 9μm and with cells suspended in Coulter Isoton II Dilute solution from Beckman Coulter. Dry mass of *P. tricornutum* was measured by first washing cells with 0.5M ammonium formate solution and then filtering onto pre-combusted, dessicated GF/F Whatman filter papers of a known weight. Cells were dried at 50°C until a constant mass was obtained. Centrifugation of *P. tricornutum* was carried out at 1500g for 10minutes. F/2 media was filter sterilised through filter paper with 0.2μm pores. LB media was autoclaved.

Chlorophyll a in *P. tricornutum* was quantified spectrophotometrically. 1ml of DMF was added to a cell pellet, typically derived from 1ml of culture. The mixture was vigorously mixed for 15min then centrifuged at 10,000rpm for 2 minutes. The supernatant was transferred to a BRAND UV-cuvette and diluted as necessary to ensure an optical density below 1 was obtained. Absorbance was measured at 647

Table 2.1 – Plasmids used in this study. Description is of the form plasmid + inserted sequences. pICH and pAGM plasmids were obtained from the MoClo kit (Engler et al., 2014a). DGAT containing plasmids were produced by Alex Litvinenko. UTR - untranslated region.

Plasmid	Description	MoClo Level
pIB112.Egst-2	pICH41308 + Phatr3_EG01099 with stop codon	0
pIB120.Pt54st-1	pICH41308 + Phatr3_J54569 with stop codon	0
pIB117.DXR_b2-2	pAGM1276 + DXR targeting peptide for position B2 in level 1 assembly	0
pIB114.Egno-1	pICH41258 + Phatr3_EG01099 without stop codon	0
pIB128.Pt49no-7	pICH41258 containing Phatr3_J49702 without a stop codon	0
pIB124.Pt54no-1	pICH41258 containing Phatr3_J54569 without a stop codon	0
pIB039.lip1sp	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_41624 signal peptide:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pIB038.lip1	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_41624:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pIB040.lip2sp	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_J31492 signal peptide:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pIB090.lip2	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_J31492:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pIB093.lip3sp	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_44028 signal peptide:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pIB096.lip3	pAGM4723 + FcpA promoter and 5' UTR + Phatr3_44028:mVenus + FcpC 3' UTR and terminator + Ble resistance cassette	2
pPT1+PtDGAT1fi	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J9794 + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT2Afi	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J49462 + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT2Bi	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J49544 + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT2Ci	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J31662 + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT2Di	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J43469 + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT2Ei	pPT1 + AMT1 promoter and 5'UTR + DGAT2Ei + AMT1 3' UTR and terminator	NA
pPT1+PtDGAT3i	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J49708 + AMT1 3' UTR and terminator	NA
pPT1+PtPDATfi	pPT1 + AMT1 promoter and 5'UTR + Phatr3_J8860 + AMT1 3' UTR and terminator	NA

and 664.5nm in a UV1 UV-Vis spectrophotometer blanked with DMF. Chlorophyll a content was calculated using the following equation: Chl a (μg) = 12.70$A_{664.5}$-2.79A_{647}, where A is the absorbance at the respective wavelength (Wellburn, 1994).

2.3 Biological materials and their growth

2.3.1 Biological materials

P. tricornutum Pt1 (CCAP 1055/1) from the Culture Collection of Algae and Protozoa (Oban, Scotland) was used for algal work throughout. Strains DH5α and NEB5α of *E. coli* were used. DH5α was obtained from the Plant Metabolism group and NEB5α was obtained from New England Biolabs.

2.3.2 Growth for diacylglycerol acyltransferase overexpression analysis

Starter cultures were grown from stationary phase stocks in conical flasks on an orbital shaking platform at 120rpm, 80μmol photons m^{-2} s^{-1}, 16h:8h light:dark cycle and 18°C for 10 days in F/2 media (Guillard and Ryther, 1962), until they reached logarithmic phase. Cells were then inoculated at 2.5x10^5 cells ml^{-1} in 60ml of F/2 in 100ml flasks and grown for 10 days under the same conditions. Cells were then washed by two rounds of resuspension in F/2 without nitrogen (-N) and recultured in F/2 -N and grown for 8 days under the same conditions. 30ml of culture was taken for dry mass measurements and 5ml for lipid extraction.

2.3.3 Growth for fatty acid synthesis inhibition experiments

P. tricornutum was acclimatised to constant irradiance of 100μmol photons m^{-2} s^{-1}, temperature of 18°C and shaking of 120rpm in F/2 with 10 times the usual nitrate and phosphate (10N10P) and kept in exponential phase. At the beginning of the experiment, the cells were washed with F/2 0N10P (without nitrogen) media and resuspended in 50ml of either F/2 0N10P or F/2 10N10P at 1x10^6 cells ml^{-1} in 100ml conical flasks. Cell density, pigment quantification and Nile Red staining were carried

out at the reported time points. Cerulenin, triclosan, stearic acid and palmitic acid were dissolved in DMSO and added as described for specific experiments.

2.3.4 *E. coli* growth

Regrowth from glycerol stocks of *E. coli* carried out by streaking stocks on Luria-Bertani (LB) media plates containing 1.5% w/w agar. For miniprep plasmid extraction and glycerol stock production, *E. coli* was cultured from a single colony in LB media at 37°C for 16-20h.

2.4 Molecular Methods

2.4.1 DNA extraction from *E. coli*

During molecular cloning, plasmids were extracted from *E. coli* using the Monarch Plasmid Miniprep Kit available from New England Biolabs. The Qiagen Midiprep kit was used to prepare plasmids for transformation into *P. tricornutum*. The manufacturer's instructions for *E. coli* growth and extraction procedures were followed.

2.4.2 DNA extraction from *P. tricornutum*

40ml of *P. tricornutum* culture was centrifuged at 1500g for 10 min and resuspended in 1500μl H_2O. 1500μl of phenol:chloroform:isoamyl alcohol, at a ratio of 25:24:1, were added and thoroughly mixed. The mixture was then centrifuged at room temperature for 5 minutes at 16,000g and the aqueous phase transferred to a fresh tube. One tenth of the mixture volume of 3M sodium acetate were added followed by addition of two and a half times the mixture volume of ice cold 100% ethanol. The solution was mixed by inversion and incubated at -20℃ for 30 minutes. The mixture was then centrifuged at 4℃ for 10 minutes at 16,000g to pellet the DNA. Supernatant was removed without disturbing the pellet. The pellet was washed with 1 ml 70% (vol/vol) ethanol and then air dried before resuspension in H_2O.

2.4.3 RNA extraction and cDNA synthesis

RNA was extracted from *P. tricornutum* and *C. reinhardtii* using the RNeasy Plant Mini Kit from Qiagen. Genomic DNA contamination was removed using Ambion TURBO DNA-free DNase Treatment from Life Technologies. cDNA was then synthesised using the SuperScript III First-Strand Synthesis System for RT-PCR also from Life Technologies. All processes were carried out as described in the kit instructions.

2.4.4 PCR

PCR was carried out using iProof High-Fidelity DNA-polymerase from Bio-Rad, following their recommended setup instructions.

2.4.5 Gel electrophoresis and band recovery

Gel with an agar content ranging between 0.5% (w/vol) and 2% (w/vol) was produced and submerged in 1xTAE buffer. Voltages of between 90 and 120V were applied using a Consort E835 gel bath. DNA was recovered from gel bands using the Gel Band Purification kit from GE Healthcare.

2.4.6 Golden Gate Cloning

Golden Gate cloning was used to incorporate PCR products and generate expression cassettes. The syntax described by Engler et al. (2014a) was used along with the vectors made available and as described in the section 5.3.2.3. Primers were designed to amplify gDNA and cDNA and incorporate BpiI or BsaI restriction sites. Level 0 and 1 reactions were carried out with 75ng of empty vector or plasmids containing desired parts with a 2:1 molar ration of insert:backbone. 0.5 μL of BpiI was added to Level 0 and Level 2 reactions while 0.5 μL BsaI were added to Level 1. 0.5 μL of T4 DNA ligase, 1.5 μL T4 buffer, 1.5 μL BSA and water made up the solution to 20μL. Restriction-ligation was carried out as follows for Level 0 and 1: 20 seconds at 37°C followed by 26 rounds of 3min at 37°C and 4min at 16°C after which it is held at 5min at 50°C then 5min at 80°C. For Level 2, 45 rounds of 2 min of 37°C and 5min at 16°C are followed by 5min at 50 °C and 10min at 80°C for Level 2. Amplicon

and plasmid sequences were confirmed by Sanger sequencing carried out by Source Biosciences.

2.4.7 Electroporation based transformation of *P. tricornutum*

P. tricornutum cultures were grown in F/2 -Si to early exponential phase and a density of around 2-4×10^6 cells ml^{-1}. The volume required to produce a suspension of 3.57x10^8 cells per ml was calculated. Cells were spun down and washed in 2ml of a mixture of 0.77 M mannitol and F/2 -Si at a volume/volume ratio of 97.5:2.5 and subsequently suspended in the correct volume of the same solution. 5μL of linearised plasmid at 0.5μg μL^{-1} was added to 35μL of cell suspension. The mixture was transferred to a 0.2cm gap NEPA21 electroporation cuvette. The NEPA21 electroporator was set with the following values: poring pulse of 300 V, 5 ms length, 50 ms interval, 10% decay rate, positive polarity; for transfer pulse use 20 V, 50 ms length, 50 ms interval, 5 pulses, 40% decay rate, positive/negative polarity. The impedance of the mixture was measured. Three poring pulses were used for an impedance between 0.4 and 0.5kΩ,four pulses between 0.5 and 0.58 kΩ, 5 between 0.58 and 0.63kΩ and 6 between 0.63 and 0.68kΩ as determined by Alex Litvinenko, a previous member of the Plant Metabolism Group. Suspensions were diluted if impedance was more than 0.68kΩ. The process was repeated with fresh cells if the poring pulse energy was not between 2.4 and 3J. The electroporated mixture was added to 5ml F/2 + Si solution and left in the darke to recover for two hours. Cultures were then transferred to a light incubator with an irradiance of 40μM photons $m^{-2}s^{-1}$ for 16-20 hr at 18°C without shaking. The following day, after resuspending cells in 0.6ml F/2 +Si, 0.15ml of suspension was plated onto F/2 +1%weight/volume agar containing 75 µg μL^{-1} of zeocin and incubated at 16-20 hr at 18°C without shaking until colonies appear after 3 weeks. Colonies were picked and subcultured twice in F/2 + 75 µg μL^{-1} of zeocin. Phire PCR and confocal screening was then carried out.

2.4.8 Preparation of chemically competent *E. coli*

200ml of LB was inoculated with 2ml of *E. coli* culture and grown at 37°C with shaking until an OD of 0.6-0.8 was reached. The culture was placed on ice for 10min and then centrifuged at 2000rpm for 10min at 4°C. The supernatant was discarded and the pellet resuspended in 40ml of ice cold 10% aqueous glycerol solution with 0.1M $CaCl_2$. The culture was incubated on ice for 15min after which it was centrifuged for 10min at 2000rpm and 4°C. The supernatant was discarded and the pellet resuspended in 40ml of ice cold 10% aqueous glycerol solution with 0.1M $CaCl_2$. The solution was portioned into 50μL aliquots and frozen in liquid nitrogen before being stored at -80°C.

2.4.9 Transformation of *E. coli*

Chemically competent *E. coli* were thawed on ice. Either 5μL of the solution of a Golden Gate restriction-ligation or 0.1ng of previously constructed plasmid were mixed with the cells and the mixture was left on ice for 20min. The mixture was then heat shocked at 42°C for 30-45s and incubated on ice for 2min. 450μL of LB was added to the cells and the culture was incubated at 37°C for 60min. The culture was then spread on LB-agar plates with the appropriate antibiotic.

2.4.10 Screening plasmids with colony PCR of *E. coli*

Small amounts of each colony were mixed in 20μL of water. PCR solutions containing 10μL REDTaq PCR ReadyMix, 2μL of each primer and 6μL of the colony water were mixed. Amplification was carried out and the product screened using gel electrophoresis.

2.4.11 Direct PCR measurement of transgene presence in *P. tricornutum*

The presence of transgenes in *P. tricornutum* was assayed using the Phire Plant Direct PCR kit, from Thermo Fisher Scientific. Dense cultures of *P. tricornutum* were diluted 1:10 using Phire dilution buffer. PCR was carried out in 96 well plates

following the instructions in the kit.

2.5 Selective markers

For growth and selection of *E. coli*, LB media and LB-agar plates were prepared with spectinomycin, carbenicillin and kanamycin at 100μg ml^{-1}. xgal was added to spectinomycin and carbenicillin plates at 40μg ml^{-1}. Selection of *P. tricornutum* was carried out in F/2 media + 75μg ml^{-1} and F/2-1% agar plates + 75μg ml^{-1}.

2.6 Bioinformatics, plotting and software use

2.6.1 Identifying candidate genes

BLAST and protein domains were used to search for candidate lipase, galactosyl hydrolase and acyltransferase candidate genes. Using TBLASTN, implemented on the Ensemble Protist webpage , the genomic sequence of *P. tricornutum* (assembly GCA_000150955.2) was queried with protein sequences of the characterised genes detailed in Section 5.3.1. An E value of 1^{-1}was used.

Interpro, Gene3D and Panther domains present in CrPDAT, CrPGD1 and AtSFR2 were identified using InterProScan (Jones et al., 2014). InterPro, Gene Ontology Biological Process identifiers were manually collected based on relevance to acylhydrolase and acyltransferase activity. These identifiers, and the term "lipase" were used to query the gene models of the Phatr3 gene annotation (Rastogi et al., 2018) that are stored in Ensemble Protist.

Interpro identifiers were assigned to candidate genes using InterProScan (Jones et al., 2014). ASAFind (Gruber et al., 2015) in conjunction with SignalP 3.0 (Dyrløv Bendtsen et al., 2004) and separately, HECTAR (Gschloessl et al., 2008) were used to predict *P. tricornutum* protein location. Location of *C. reinhardtii* genes was predicted with TargetP (Emanuelsson et al., 2007) and PredAlgo (Tardif et al., 2012). Secondary structures were predicted using an implementation of the Chou-Fasman algorithm (Kumar, 2013) and transmembrane domain predictions were made using TMHMM2.0 (Krogh et al., 2001).

2.6.2 Comparison of transcriptomic and proteomic datasets

Processed transcriptomic and proteomic datasets for a number of studies are available as detailed in their publications (Levitan et al., 2015; Yang et al., 2013; Matthijs et al., 2016; Longworth et al., 2016). Raw microarray data from Alipanah et al. (2015) are available on the Gene Expression Omnibus website with accession GSE58946 and were processed to produce $\log_2$ fold change and Benjamini and Hochberg false discovery rate (FDR) values (Benjamini and Yosef, 1995) using GEO2R. Data from Huete-Ortega et al. are under preparation for publication. Uniprot protein identifiers were converted to Phatr2 identifiers from the first annotation of gene models (Bowler et al., 2008) using Biomart on Ensemble Protist. Phatr2 identifiers were semi-manually converted to Phatr3 identifiers with reference to Rastogi et al. (2018).

Compiled data were assigned to genes sets for analysis. The gene sets within Levitan et al. (2015) were supplemented with groups from other sources. Gene sets for the ribosome, amino acid synthesis and tRNA charging were obtained from KEGG, reference T01075. Those of autophagy and the proteasome were obtained from the reannotation of *P. tricornutum* gene models (Rastogi et al., 2018). Chrysolaminarin synthesis genes were collected from DiatomCyc (Fabris et al., 2012). The proteasome and autophagy groups were manually curated to remove unrelated genes. Data were assigned to a manually curated set of lipid metabolism genes for further investigation, with genes in the set originating from Zulu et al. (2018), the dataset from Levitan et al. (2015) and DiatomCyc (Fabris et al., 2012).

2.6.3 Software

Primer and construct design, sequence analysis and restriction site identification were carried out using SnapGene software (from GSL Biotech and available at snapgene.com). Data analysis and plotting were carried out using R studio version 0.99.891 and aided by the ggplot2, dplyr and ggpubr packages. Microsoft Excel and Python 2.7 were used for data handling and preparation.

2.7 Lipid measurements

2.7.1 Nile Red staining of neutral lipids

Cells were resuspended at 3x10^6 cells ml^{-1} in media as described in Chapter 5. 100μL of either cell suspension, F/2 media (with or without nitrogen) or a triolein standard were added to the well of a black Greiner F-bottom Fluotrac 96 well plate. 100μL of 1μg ml^{-1} Nile Red in 1:1 (vol/vol) DMSO:F/2 media (with or without nitrogen). The plate was then incubated for 15min at 37°C. The plate was then placed in a BMG Fluorstar Omega plate reader, shaken and each well excited at 525-535 nm. Fluorescence was detected at 570-580nm and the gain was set at 2500.

2.7.2 Total lipid extraction

5ml of each culture was centrifuged at 1500rpm for 10 minutes at room temperature. The supernatant was removed and the pellets were frozen in liquid nitrogen and stored at -80C. 200μL of 1mg ml^{-1} glycerol trimyristate standard was added to each pellet. Each pellet was resuspended in 10ml 2:1 (vol/vol) of chloroform:methanol and sonicated for 30 min in water cooled with ice. Suspensions were then placed on ice for 30min. 5ml of 2:1 (vol/vol) of chloroform:methanol and 5ml of distilled water were added to each suspension. Suspensions were mixed vigorously for 2min and centrifuged for 3min at 4°C and 972g. The methanol and water forming the top layer was removed prior to transfer of the chloroform layer, that contained the lipids, to a clean vessel. The chloroform solution was evaporated at below 45°C in a Genevac EZ-2 evaporating centrifuge. Lipids were resuspended in 200μL heptane. 120μL was used for measurement on gas chromatography flame ionisation detection (GC-FID) while 80μL were evaporated at below 45°C in a Genevac EZ-2 evaporating centrifuge and resuspended in 300μL of chloroform for measurement by tandem mass spectroscopy.

2.7.3 Gas chromatography-flame ionisation detector measurement of TAGs

1 μL of total lipid extract was injected into a gas chromatograph (Trace GC Ultra; Thermo Scientific) fitted with an Agilent Select Biodiesel for Glycerides metal capillary GC column (10 m, Ø 0.32 mm, film thickness 0.1 m; Agilent Technologies; Edinburgh, Scotland) and quantified by FID. Injector and detector temperatures were set to 300 °C and 380 °C respectively. He(g) was used as the carrier at a constant flow rate of 1.5 ml/min. H_2(g) was used for FID. An 8:1 split ratio was used. The temperature program was: 50°C for 1min, followed by rising to 150°C at 15°C min^{-1}, followed by rising to 210°C at 7 °C min^{-1}, followed by rising to 310°C at 15°C min^{-1}, followed by rising to 370°C at 7°C min^{-1} followed by a final hold for 5min. Retention times of TAG species were previously identified by Alex Litvinenko. Glyceryl tripalmitate standards (0 to 1 mg ml^{-1}) were used to generate calibration curves to quantify TAG. Glyceryl trimyristate in the sample was quantified to account for sample loss during extraction.

2.7.4 Electrospray ionisation triple quadrupole mass spectrometry of total lipid extract

Dr Richard Haslam at Rothamsted Research carried out measurement of TAGs by electrospray ionisation triple quadrupole mass spectrometry (ESI-TQMS) from total lipid extracts I prepared. The TAG fraction within the total lipid extract was analysed by API 4000 QTRAP, ESI-TQMS from Applied Biosystems, using the method described by Hamilton et al. (2014). Neutral loss profiling was used to quantify TAG species and FA composition.

2.8 Confocal microscopy

P. tricornutum was grown to exponential phase in F/2 media and transferred to slides for viewing. Cells were excited at 514nm using the argon laser of a Leica SP5 fluorescent confocal microscope. mVenus fluorescence was detected between 520 and 550nm and chlorophyll autofluorescence was detected between 680 and 700nm.

Images were processed using the LAS X software.

2.9 Statistics

Statistical analysis was carried out using R studio version 0.99.891. Tests were core R functions or obtained from the onewaystats, car or rstatix packages.

3

Systematic review of nitrogen depletion literature in *P. tricornutum*

3.1 Introduction

As outlined in Chapter 1, the system-wide nature of acclimatisation to nitrogen depletion has led to the use of techniques that can measure large proportions of various classes of cellular constituents. The desire to identify the factors that cause the acclimatisation process and for some researchers, the mechanism of TAG accumulation, has resulted in numerous studies that employ techniques to measure the transcripts and proteins in the cell. These studies have provided overall narratives of the physiological changes that occur during nitrogen depletion and have led to the identification of a number of important aspects of the process. However, interpretation of the results is limited due to the nature of these studies as a single experiment. In addition, such large amounts of data are generated that a single study is unlikely to fully utilise the information available. As a result, the available transcriptomic and proteomic studies of nitrogen depletion in *P. tricornutum* were compared and analysed and this is presented in this chapter. Below I first introduce the types of omic technologies and how biological insight is drawn from the data generated. I then introduce the current nitrogen depletion literature, the authors' key findings and the rationale for a review of their work.

3.1.1 Measurements of the transcriptome and proteome

The '-omic' suffix is applied to a number of procedures that focus on collecting information on a large proportion of a given class of molecules within a sample. These molecules typically include transcripts, proteins, lipids and metabolites and techniques to measure them are respectively referred to as transcriptomics, proteomics, metabolomics and lipidomics, with lipidomics forming a sub-part of metabolomics. For each of these areas, multiple measurement techniques can be used. Generally, measurements of the transcriptome can be carried out using microarrays or RNA sequencing while proteomic measurements typically use mass spectrometry or gel based techniques. The measurement techniques relevant to the studies detailed in this chapter are outlined below. This chapter makes use of data on differential expression of transcripts and so the methods to produce and analyse these data are also introduced.

3.1.1.1 Microarrays

Microarrays are an early technology for high throughput measurement of mRNAs that enable relative quantification of transcripts between two samples (Seidel, 2008). Microarrays are formed by binding, or synthesising *in situ*, specific oligonucleotide probe sequences at points on glass slides. Each point contains many copies of a unique sequence and the location of the point therefore corresponds to an oligonucleotide sequence. Probes are designed so that their sequence is present in a single gene. Probes can originate from several sources but the prevalence of whole genome sequencing enables synthesis of oligonucleotides for known genes in many cases. To measure expression, cDNA is first produced from the RNA. The cDNA is either labelled with dyes or biotin which can subsequently be bound by fluorescntly labelled antibodies. The final amount of cDNA is proportional to the initial quantity of mRNA and the cDNA is then hybridised with the corresponding oligonucleotides on the slide. The contained dye is excited using lasers and the location and intensity of fluorescence provides information on the expression of genes that contain the probe oligonucleotide sequence (Seidel, 2008).

Typically, comparisons of expression are carried out between samples, for example, from control and experimental conditions or cells. The generation of the fluorescence data for these samples and how these data are processed differ depending on the type of microarray used. Ultimately however, data is normalised to enable comparison between the samples. Data collection varies due to the use of two main types of chips, Affymetrix and spotted microarrays. For analysis, the primary difference arises from the use of a single Affymetrix chip for each sample, with multiple probes per gene. In contrast, for spotted microarrays, two samples that are labelled with differently coloured dyes are hybridised to a single spotted chip and the chip only contains one probe per gene. Relative expression is therefore calculated from multiple signals on separate chips for Affymetrix microarrays. For spotted arrays in contrast, the ratio of the dyes present at a spot dictate its colour and hence this is an indication of the relative abundance of each transcript (Seidel, 2008). Normalisation is needed to account for non-biologically produced differences in samples and numerous methods have been develped for both types of microarray. For spotted microarrays, such differences include variations in dye decay or fluorescence. For these microarrays,

normalisation is carried out on the data from the two samples to enable comparison to be made. Affymetrix data processing involves first calculating a value for gene expression from the multiple probes and then scaling the measurements for each chip to take account of variations between the two. Once the normalised and scaled values have been produced, these are typically log transformed and the ratio between the samples calculated (Seidel, 2008).

3.1.1.2 RNA sequencing

Although microarray technology is still extensively used in transcriptomics, RNA sequencing is now favoured (Illumina, 2016) as it has numerous benefits over microarray RNA measurements. These include a wider dynamic range that allows detection of strongly altered transcripts, higher sensitivity that enables the detection of low abundance transcripts, high specificity and the ability to detect genetic variance, differential splicing and unknown transcripts (Zhao et al., 2014; Marioni et al., 2008; Nookaew et al., 2012).

RNA sequencing (RNAseq) enables the abundance of transcripts to be determined through random sequencing of the transcriptome. Numerous sequencing technologies exist and although there are substantial differences in the platforms, sample preparation and some processing steps, the principle of their operation and quantification of the transcriptome remains the same (Wolf, 2013; Hrdlickova et al., 2017). As such, a general RNAseq workflow can be outlined and one is shown in Figure 3.1. RNAseq determines the abundance of transcripts by randomly sequencing fragments of transcripts. The abundance of sequences of a given transcript are therefore proportional to the quantity of that transcript in the transcriptome. It should be mentioned that the more recently developed platform of Oxford Nanopore is capable of directly sequencing complete mRNAs but the application of this technology to RNA sequencing is still emerging and will not be outlined here. With the other technologies, sequencing of the fragments is typically carried out by extracting RNA from the sample, enriching for mRNA, fragmenting the transcripts and synthesising cDNA from these fragments (Wolf, 2013). These cDNA fragments frequently consist of a few hundred base pairs but can be up to 1kbp in length. The cDNAs are ligated to oligonucleotide adaptors to produce a library of fragments. The adaptors anneal

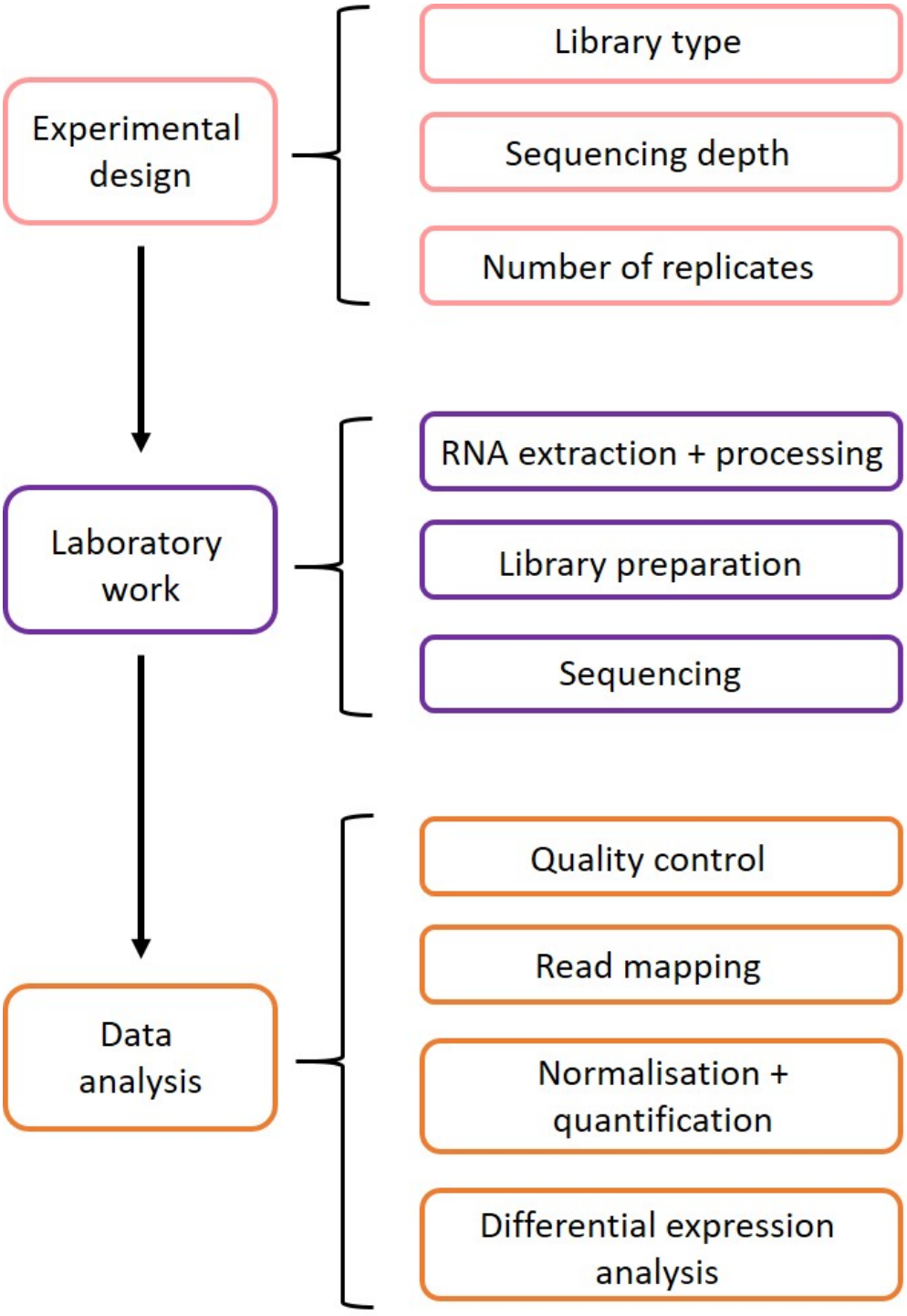


Figure 3.1 – Generalised RNAseq workflow.

to either a slide or bead, depending on the technology (EMBL-EBI, 2019). The cDNAs are amplified by PCR and also anneal to either the same location on the slide or the same bead. When beads are used, they are placed in individual wells. In this way, a slide or well location is associated with a specific sequence. Sequence information is collected through DNA replication: as each base is added, a signal is detected. This signal is either fluorescence, where the colour corresponds to the added base, or a flash of light or change in pH as a known base is added. The spatial origin of the signal combined with the sequence of signals is used to build up the sequences of the fragments. A certain number of sequencing reactions, referred to as read depth, are carried out on each sample.

Aspects of quantifying transcripts from the sequenced fragments are dependent upon the the sequencing platform used and the investigated organism. However, the principles of processing, quantifying and calculating differential expression for fragment sequences are the same (Wolf, 2013; Hrdlickova et al., 2017). Fragment sequences first undergo quality control to minimise errors. The adaptor sequences are removed from those of the fragments and duplicate sequences, which are not expected at typical read depths, are adjusted for. Additionally, quality scores for the data can be summarised by various tools and spot checking of individual fragments is carried out to observe aspects such as coverage distribution and effects of CG biases. For organisms for which the transcriptome is known, processed reads are mapped to transcript models. Programs are available to deal with ambiguous mapping and mapping of spliced reads.

After mapping, quantification is carried out by adding the total reads that map to a transcript. These quantities are then normalised in three ways. Higher numbers of sequencing reactions will generate higher numbers of reads for a given transcript. Similarly, longer transcripts will also generate higher numbers of reads. The number of reads detected for a transcript are therefore normalised by the total number of mapped reads and transcript length. This produces values known as either fragment per kilobase per milion mapped reads (FPKM) or reads per kilobase per milion mapped reads (RPKM). An additional normalisation is carried out to account for carry-over effects (Wolf, 2013). These effects are the expected decrease in the number of reads of a transcript that results from increased expression of a different transcript,

rather than decreased expression of the first. Frequently, it is desirable to know how the abundance of a transcript differs in two samples and this is carried out through differential expression analysis. Software packages are employed to do this and produce statistical distributions to represent the data. These distributions allow significance testing of changes in expression to be carried out and also return fold change data for a transcript, typically normalised to $\log_2$ values. Quantifying isoforms poses a challenge given their overlapping sequences. Software has been developed to do this however, making use of differences in reads mapped to exons shared between isoforms and those mapped to unique exons.

3.1.1.3 Mass spectrometry based proteomics

Several platforms are used for proteomics, including gel electrophoresis and liquid chromatography (Guo et al., 2007), but the proteomic studies in *P. tricornutum* made use of mass spectrometry and so that is the focus here. Mass spectrometry (MS) based proteomics enables identification and relative or absolute quantification of proteins in samples. It has a number of applications including the investigation of protein interactions and post-translational modification but one of its main uses is in identifying changes in expression of proteins (Aebersold and Mann, 2003).

There are multiple strategies and variations in the type of mass spectrometer or scan available for MS based proteomics but a general workflow has now been established (Aebersold and Mann, 2003; Hosp and Mann, 2017). Proteins are first separated from the sample and then fragmented to form peptides. These peptides are separated using high pressure liquid chromatography (HPLC) and introduced into a mass spectrometer, typically using electrospray ionisation (ESI). Peptides are separated by mass or may undergo fragmentation followed by selection in a tandem mass spectrometer. The signal intensity at each mass to charge ratio is recorded and the resulting spectra are compared to ion databases to identify the detected peptide. Unambiguous assignment of spectra to peptides is not a simple task and multiple methods that utilise statistical methods have been developed to aid differentiation between true and false spectral assignment. This process is assisted if the proteome of the organism is known as assignment is made to a limited number of peptides. The identified peptides are then mapped to proteins using software tools.

Relative quantification of peptides is possible by labelling proteins or peptides from each sample with stable isotopes or isobaric tags, such as iTRAQ reagents (Ross et al., 2004; Aebersold and Mann, 2003). This enables ions to be differentiated by the mass/charge ratio or a molecule that is released during ion fragmentation that is characteristic of the tag. For relative quantification, differentially labelled sampled are mixes at a 1:1 ratio and the intensity of peptide peaks are averaged for each protein (Hosp and Mann, 2017). Absolute quantification can be achieved by comparing peak intensity with protein standards that incorporate an isobaric tag. One limitation of MS-based proteomics is that often only a small proportion of the proteins present are detected. This results from the detection of the more common proteins over the less abundant. This situation is improved by using separation techniques on the sample prior to MS but is difficult to solve entirely.

3.1.1.4 Generation of biological insight from differential expression data

Large quantities of data are produced for each omics study, typically consisting of thousands of differential expression values. Data analysis tools are required to sort and visualise this information. Furthermore, sequence based functional annotation is often required for the majority of organisms for which functional characterisation of genes is mostly lacking. Two commonly used methods to highlight the differentially regulated genes are heat maps and enrichment analysis. Heat maps are useful in visualising changes in expression between samples and can be coupled with clustering to aid this. Enrichment analysis makes use of functional annotation and evaluates whether these functional groupings are altered more than would be expected by chance. Commonly used functional annotation includes Gene Ontology terms (Ashburner et al., 2000), that provide a structured ontology for biological data, or assignment to pathways, provided by KEGG for example (Kanehisa and Goto, 2000).

To derive biological insight from information on differential expression of transcripts and proteins between samples, it is assumed that the amount of transcript correlates to the amount of its respective protein(s) and that the amount of protein correlates with its activity. Aspects of a number of processes alter these relationships.

Among other things, the transcript-protein relationship is altered by translation rate, protein synthesis delay and protein half-life (Haider and Pal, 2013; Liu et al., 2016b) while the protein-activity relationship is impacted by post-translational modifications, protein interactions and protein location. Nonetheless, transcript and protein level are commonly used as proxies for biological activity and this has been useful in generating biological insights.

3.1.1.5 Literature of nitrogen depletion studies in *P. trictornutum*

To identify the mechanisms that constitute the nitrogen depletion acclimatisation response in *P. tricornutum*, a number of transcriptomic and proteomic studies have been carried out (Longworth et al., 2016; Ge et al., 2014; Yang et al., 2013; Alipanah et al., 2015; Levitan et al., 2015; Matthijs et al., 2016). These studies were selected for comparison in this chapter as their experimental designs enable investigation of the specific response to nitrogen depletion. In addition, an unpublished study carried out by a former member of the Plant Metabolism Group, Dr Maria Huete-Ortega, was included. The data produced by this study is available on the accompanying USB stick. Despite collecting omic data, other published studies were not considered as their results were confounded with regard to investigating nitrogen depletion, (e.g Valenzuela et al., 2012).

Further to transcriptomic and proteomic studies, a couple of studies have been conducted to investigate changes to the lipidome of *P. tricornutum* during nitrogen depletion. One of these, the study of Abida et al. (2015) is published. I conducted the other as part of my master's degree and the lipidomic data from this is included in Figures A.1 and A.2, along with information on experimental design.

The design of the gene expression studies enabled physiological alterations to be associated with those of either the transcriptome and proteome. The authors also used a variety of techniques to gain insight into the implications of changes in expression. The impacts of nitrogen depletion were investigated by comparing cellular characteristics under nitrogen replete and nitrogen deplete conditions. This comparison was primarily carried out by reculturing exponentially growing cells in media containing or lacking nitrogen although Ge et al. (2014) took measurements of a culture before and after TAG accumulated, using TAG accumulation as a marker

that nitrogen had been depleted from the media. Physiological and in some cases, metabolite data were collected alongside transcriptomic or proteomic measurements. This information included cell density, chlorophyll content and macromolecular composition.

Transcripts and proteins were first annotated with functional information and alterations in expression investigated. This enabled major alterations in metabolism to be identified. Gene ontology (GO) annotation and assignment to enzymatic functions underpinned enrichment analysis, pathway analysis and expression clustering. From these analyses, the authors outlined some of the processes that follow nitrogen depletion (Levitan et al., 2015; Alipanah et al., 2015; Longworth et al., 2016; Matthijs et al., 2016). The studies observed prominent changes in nitrogen metabolism, with increased expression of genes involved in uptake and assimilation of nitrogen but also its recycling from proteins and amino acids. Photosynthesis was also prominently reduced with decreased expression of genes in the Calvin Cycle and chlorophyll synthesis reflected in decreased photosynthetic efficiency and reduced chlorophyll content. Upregulation of tricarboxylic acid cycle genes, pyruvate metabolism and glycolysis were also highlighted as likely to play a role in recycling carbon and reductants to lipid metabolism.

3.1.1.6 Purpose of the systematic review of nitrogen depletion literature in *P. tricornutum*

Transcriptomic and proteomic studies are often carried out to generate hypotheses about the functioning of organisms. Indeed the previously conducted nitrogen depletion studies enabled a number of biological insights to be gained and further investigated in some cases, including an outline of the response, identification of a transcription factor, a role for branched chain amino acid degradation in lipid accumulation and a suggested negative feedback mechanism by which nitrogen assimilation regulates carbon fixation and lipid metabolism. However, given the complex nature of the nitrogen depletion response it was reasoned that there are likely other observations to be made. In particular, it was thought that further omic investigation of two questions in lipid metabolism would be useful. The first regards the processes that lead to the remodeling of the chloroplast membrane lipids:

MGDG may have decreased due to increased catabolism or decreased synthesis. A second question, regards whether *de novo* biosynthesis of TAG occurs as a result of regulation of lipid metabolism or due to changes that increase carbon flux into lipid metabolism, at least relative to the need of the cell for lipids for growth. Gene expression data provides a way to investigate both of these questions but they were not addressed in the previous studies.

Given the existing data sets, it was questionable that further measurements, particularly a transcriptomic investigation, would warrant the effort and although the existing data sets were subject to several limitations, comparing across the studies mitigates these. Overall, there is a need for robust data that reflects a defined cellular state. This is important as interpretation of omic data is complicated by the quantity of data and each alteration in expression cannot be considered and accounted for individually. It is therefore useful to remove ambiguities that may alter interpretation of the data. Regarding the limitations specifically, the first point to make is that, as is typical given the effort and resources involved in undertaking omic studies, experiments were not repeated. Accordingly, the results of each study may be influenced by factors unrelated to nitrogen depletion and are therefore less robust. Secondly, the regulation, physiology and metabolism of the cell alter through time during nitrogen depletion. As a result, making conclusions based on individual or few time points, as was the case in the existing literature, does not provide a complete outline of the process and at worst is misleading. Finally, as microalgae are single celled organisms, their physiology is dependent upon their environment. Differences in experimental conditions may therefore alter the response to nitrogen depletion and lead to different conclusions being drawn from the multiple studies. A comparison of the existing transcriptomic and proteomic data was therefore considered a useful way to both provide further biological insight into lipid metabolism and aid researchers in making use of the existing studies.

Further to a review of the transcriptomic and proteomic data, it was also decided to compile the physiological and biochemical responses to nitrogen depletion. This was undertaken to produce an outline of the acclimatisation process that could be used to compare between the gene expression data of different studies and to identify the sequence of physiological changes. The changing acclimatisation response through

time and the varying time taken for this response to unfold in different conditions require a method to identify at which point in nitrogen depletion gene expression data was measured. To achieve this, the physiological data that was measured in the studies was compared. It was assumed that differing experimental conditions between the studies would impact the timings of the processes involved in nitrogen depletion but the constituent processes themselves would be unaltered. That is to say, the same sequence of processes would be observed in each study. It was further assumed that patterns of gene expression would also be temporally altered by experimental conditions but that the same patterns would occur. In this way, expression data can be compared to each other relative to the physiological observations. Furthermore, a physiological outline of the acclimatisation response was necessary for investigating contributions of chloroplast lipid remodeling to TAG accumulation as undertaken in Chapter 5.

3.2 Results

3.2.1 Summary and comparison of the experimental conditions used in the nitrogen depletion literature

The cellular response to nitrogen depletion in *P. tricornutum* has been investigated in a number of studies that variously measured cell composition, biological processes, the transcriptome and the proteome. These studies therefore provide insight into the coordinated processes that acclimatise the cell to a lack of nitrogen. Table 3.1 details the experimental conditions of eight studies that undertook relevant measurements. Six of the studies have been published elsewhere (Alipanah et al., 2015; Levitan et al., 2015; Yang et al., 2013; Ge et al., 2014; Longworth et al., 2016; Abida et al., 2015), another I accessed through a collaborator and subsequent member of the Plant Metabolism lab, Maria Huete-Ortega. The final study I carried out myself in my MRes (Bower et al., 2015).

The experimental set-up of the studies differed in a number of important ways as summarised in Table 3.1. Several *P. tricornutum* isolates were individually used, namely Pt1, Pt4 and FACHB-863. As outlined in Section 1.2.3, these display

Table 3.1 – A summary of experimental conditions, procedures and strains used in studies of nitrogen depletion in *P. tricornutum*. The effects of nitrogen depletion were investigated by comparing between either concurrently grown nitrogen deplete and replete cultures (Alipanah et al., 2015; Levitan et al,. 2015; Yang et al,. 2013; Longworth et al,. 2016; Abida et al,. 2015; Huete-Ortega et al,. unpublished; Bower et al,. unpublished) or before and after nitrogen was depleted from the media of a culture (Ge et al., 2014).

	Alipanah	Levitan	Yang	Huete-Ortega	Ge	Longworth	Abida	Bower	Matthijs
Strain	Pt1	Pt1	FACHB-863	Pt1	Pt1	Pt1	Pt1	Pt4	Pt1
Temperature (°C)	15	18	21+/-0.05	Room temp	22	25	19	20	21
Light (µmol photons $m^{-2}s^{-1}$)	Continuous, 60	Continuous, 120-150	12hr light/12hr dark, 200	Continuous, 250	Continuous, 100 (says mmol but unlikely)	Continuous, 200±50 along tube length	12hr light/12hr dark, 40 (says mmol but unlikely)	60	Continuous, 100
Media	F/2 made with filtered seawater	F/2 made with artificial seawater	F/2 (-Si) made with filtered seawater	Marine F/2	F/2 made with artificial seawater	F/2 made with artificial seawater	ESAW with 10xN and 10xP (when present)	F/2 made with artificial seawater	ESAW
Acclimation	Grown at exponential phase in above conditions for three weeks	Grown to exponential phase in above conditions for two days prior	Grown to late exponential phase over 5 days prior	Acclimated for several days until exponential, same conditions except light = 200	Cultures grown to mid logarithmic phase, unclear what conditions	Cultured for 48h to a certain density beforehand	Not clear	Cultured in same conditons until late exponential phase	Cultures kept in exponential phase
Experimental procedure	Culture resuspended in N replete or deplete media	Culture washed and resuspended in N replete or deplete media	Culture washed and resuspended in N replete or deplete media	Culture resuspended in N replete or deplete media	Cells innoculated in F/2 with 500mM nitrate. Nitrate was used up by 48h	Culture resuspended in N replete or deplete media	Culture resuspended in 10xN replete or deplete, 10xP media	Culture resuspended in 10xN replete or deplete, 10xP media	Culture resuspended in nitrogen replete or deplete ESAW
N concentration in deplete media	1% f/2	0	0	0	NA	0	0	0	0
Density at resuspension (cells/ml)	$5x10^4$	NA	NA	NA	$4x10^5$	NA	0.5-1.1x10^6	5x10^6	NA
Aeration	NA	Bubbled air	NA	Stirred, 300 rpm + 1% CO_2 litre per min	Bubbled air	Air bubbled at 2.4L/min	100 rpm shaking in air	70rpm shaking in air	120rpm shaking in air

physiological differences and so may respond differently to experimental conditions. Three culture conditions that have substantial effects on cell physiology also varied between the studies: temperature ranged from 15 to 25°C, irradiance varied from 40 to 250μmol photons $m^{-2}s^{-1}$ and from continuous irradiance to diurnal cycles while gas exchange was either facilitated by bubbling, stirring, shaking, provided by supplementation with CO_2 or was unaided. Cultures were generally acclimated to the culture conditions through growth to exponential phase, although the length of these acclimation periods varied from three weeks to several days. The acclimation conditions were the same as experimental conditions in most of the studies, were unclear in two of the studies (Abida et al., 2015; Ge et al., 2014), and had reduced irradiance in that of Huete-Ortega. Most of the studies used F/2 media (Guillard and Ryther, 1962), made either with filtered seawater or artificial seawater and with or without silicon (F/2 -Si). Abida et al. (2015) used ESAW media (Falciatore et al., 2012) supplemented with 10 fold the usual nitrogen and phosphorus content to ensure these nutrients remained in excess throughout the culture period. ESAW media is a less defined media than F/2, containing soil extract.

In all of the studies but one, nitrogen replete control cultures, containing NO_3^- at the concentration specified in their recipes, were grown in parallel to deplete ones, enabling comparisons at the same sampling time. Ge et al. (2014) in contrast, grew a single culture that exhausted the original supply of nitrogen after several days and comparisons between deplete and replete conditions were therefore made between the culture from before and after this point. The experimental cultures themselves did not contain nitrogen, except in the study of Alipanah et al. (2015) where 1% of the usual F/2 recipe was included.

3.2.2 Comparison of changes in cellular composition and physiology in the selected nitrogen depletion literature

Differing growth conditions between the studies may result in varying cell composition at the start of nitrogen depletion and rates of growth during the experiment. The timings of changes to cell composition and physiology were therefore outlined relative

to each other within a study prior to comparison between studies. Tables 3.2 and 3.3 summarise the changes in composition and cell function reported by the previously introduced literature. These changes are shown for nitrogen deplete samples relative to replete ones. Table 3.2 contains information on changes that occurred between 0 and 38h while Table 3.3 shows changes between 38 and 120h. The time points at which measurements were taken are summarised in Table 3.3.

Carbohydrate levels increased within four hours of the onset of nitrogen depletion (Longworth et al., 2016), before returning to a similar level to that seen in replete conditions at around 48h, as observed by Huete-Ortega. The C:N ratio was observed to have increased early in nitrogen depletion (Huete-Ortega) and was also increased, along with P:N ratios, at later stages in nitrogen depletion (Alipanah et al., 2015). The early increases in carbohydrate content and C:N ratio occurred prior to the first observations of changes to chlorophyll a or TAG content. Other observed features of nitrogen depletion were TAG accumulation and alterations to photosynthesis. These changes to photosynthesis were in the form of a decreased pigment content or reduced photosynthetic efficiency (Abida et al., 2015; Alipanah et al., 2015; Yang et al., 2013). In the studies of Longworth et al. (2016) and Huete-Ortega, neutral lipids accumulated in nitrogen depleted cells immediately after chlorophyll a degradation. Cell division also consistently decreased and subsequently ceased during nitrogen depletion. This slowing of division occurred after the decrease in chlorophyll a content (Huete-Ortega) and photosynthetic efficiency (Alipanah et al., 2015) and initiation of increase in TAG (Ge et al. 2014, Huete-Ortega).

Changes to membrane lipids and fatty acid content were also observed during nitrogen depletion. The cellular content of the predominantly chloroplast lipids MGDG and PG were observed to decrease by Abida et al. (2015), Levitan et al. (2015) and Bower (unpublished), while Levitan et al. (2015) also observed a decrease in PC and DGDG. The changes in fatty acid content were less uniform between the studies. Abida et al. (2015) found the cellular content of 14:0, 16:0, 16:1, EPA and DHA were increased while Yang et al. (2013) and Bower (unpublished) also noted an increase in the percentage of total fatty acids consisting of 16:0 and 16:1 but a decrease in the contribution of EPA. Abida et al. (2015) and Bower (unpublished) both observed an increase in the amount of 14:0, 16:0 and 16:1 in TAG while

individually increases in 18:2 (Bower, unpublished) and EPA (Abida et al., 2015)

Table 3.2 – Physiological changes of *P. tricornutum* observed between 0 and 38h after nitrogen depletion. Changes are recorded at the first time point they became apparent. The described changes refer to observations of nitrogen deplete cultures relative to replete ones. Alipanah et al., 2015; Levitan et al,. 2015; Yang et al,. 2013; Longworth et al,. 2016; Abida et al,. 2015; Huete-Ortega et al,. unpublished; Bower et al,. unpublished; Ge et al,. 2014. See Table 3.1 for experimental conditions.

	Time after nitrogen depletion (h)						
	0	4	12	18	20	24	38
Alipanah							
Levitan						Cell division ceased	
Yang							
Huete- Ortega		Carbohydrates ↑ C:N ratio ↑	Chlorophyll ↓ Total lipids ↑	Rate of cell division ↓	Cell division ceased		
Ge			Nile red fluorescence ↑				Cell division ceased
Longworth		Carbohydrates ↑	Chlorophyll A ↓ Neutral lipids ↑				
Abida						Cell division ceased	
Bower						TAG ↑ MGDG ↓	

Table 3.3 – Physiological changes of *P. tricornutum* observed between 48 and 120h after depletion of nitrogen. Changes are recorded at the first time point they became apparent. The described changes refer to observations of nitrogen deplete cultures relative to replete ones. The time points at which samples were collected are recorded. Alipanah et al., 2015; Levitan et al,. 2015; Yang et al,. 2013; Longworth et al,. 2016; Abida et al,. 2015; Huete-Ortega et al,. unpublished; Bower et al,. unpublished; Ge et al,. 2014. See Table 3.1 for experimental conditions.

	Time after nitrogen depletion (h)				
	48	72	96	120	Sampling time points (h)
Alipanah	Maximum quantum yield ↓, C:N ratio ↑	Rate of cell division ↓			Growth, Fv/Fm: 0, 24, 48 and 72.
	N:P ratio ↓, Chlorophyll a ↓ fucoxanthin ↓, TAG levels ↑				Others: 48 and 72
Levitan	As a % of cell carbon: carbohydrate↑, TAG ↑, protein ↓				48
	MGDG, DGDG, PG and PC ↓				
	Amount per cell: Polar lipids except SQDG, ↓ TAG per cell ↑				
Yang	TAG ↑				Photosynthesis, fatty acids: 48. TAG: 48 and 72
	Fv/Fm, Fv/Fo, Fo, and Fm ↓				
	As a % of FA: saturated and monounsaturated FAs ↑ PUFAs ↓				
	Cell division ceased, Oil body size ↑				
Huete-Ortega	Carbohydrates equal to control				0, 4, 12, 18, 24, 48, 72 and 96
Ge					0, 12, 24, 36, 28 and 60 after N depletion
Longworth					4, 12, 18, 24, 36, 48, 72
Abida	Fv/Fm ↓		Neutral lipids ↑	MGDG ↓	Cell density: 24, 48, 72 and 96. TAGs: 96 and 120
				PG ↓	Polar lipids: 120h
Bower	PG ↓				0, 24, 48 and 72

were also noted. Increases were seen in the proportion of FAs in TAG composed of 16:1 (Levitan et al., 2015; Bower, unpublished) and EPA (Levitan et al., 2015), while decreases were noted in the proportion of 16:0 (Levitan et al., 2015), 18:0, 18:1, 18:2 and DHA (Bower, unpublished).

In summary, excluding inconsistent observations, there was support from one study or more for a role in acclimatisation to nitrogen depletion of the following: an increase in the cellular content of carbohydrate and ratio of C:N in cells; reduced photosynthetic efficiency and content of chlorophyll a, MGDG and PG; an increased cell content of 16:0 and 16:1 FAs; accumulation of TAG that contains a larger quantity of 14:0, 16:0 and 16:1 and is composed of a higher proportion of 16:1; and cessation of cell division. Comparisons of the relative timings of these changes support initiation of increases in carbohydrate and the C:N ratio preceding reduced chlorophyll a content and photosynthetic efficiency. TAG subsequently begins to accumulate, followed by a slowing and halting of cell division.

Cell number was commonly measured at a temporal resolution that enabled comparison between studies. Timing of cessation of cell division was altered in a manner consistent with changes in experimental conditions. In a number of studies, cell division ceased within 24hrs of the onset of depletion (Matthijs et al., 2016; Levitan et al., 2015; Huete-Ortega et al., unpublished; Abida et al., 2015). These studies made use of temperatures over 18°C. Three of these studies used continuous irradiance of 100 µmol photons $m^{-2}s^{-1}$ while the fourth used 40 µmol photons $m^{-2}s^{-1}$ with a 12h light/dark cycle. These findings support acclimatisation to absolute nitrogen depletion occurring rapidly under appropriate growth conditions.

Division halted at later time points in three of the studies. In Ge et al. (2014), division halted 36h after nitrogen depletion. Nitrogen deplete conditions were preceded by a gradual decline in media nitrate content, in comparison to the immediate reculturing of the other studies. Other growth conditions were comparable with studies that demonstrated a rapid response to nitrogen depletion and the nature of depletion therefore seems to have played a role in the delayed response. Division halted 48h after onset of depletion in the study of Yang et al. (2013). Cells had ceased dividing prior to reculture in experimental media and therefore the response that followed was likely confounded by factors unrelated to nitrogen depletion, despite

optimal growth conditions after reculture. Alipanah et al. (2015) observed that cell division had not ceased 72h after reculture in deplete conditions. Prominent differences with more rapidly acclimatising studies were a growth temperature of 15°C, an irradiance of 60 µmol photons $m^{-2}s^{-1}$ combined with the use of deplete media that contained 1% of the nitrate concentration found in f/2 media.

3.2.3 Overview of the changes in gene expression observed in nitrogen depletion studies for selected metabolic and regulatory groupings

3.2.3.1 Description of the data sets and their processing

Of the nitrogen depletion studies previously introduced, five included transcriptome measurements. Four of these studies have been published (Levitan et al., 2015; Yang et al., 2013; Matthijs et al., 2016; Alipanah et al., 2015) and the data from the fifth was collected by Maria Huete-Ortega. Two studies collected proteomic data (Longworth et al., 2016; Ge et al., 2014) but the data set from Ge et al. (2014) was not available. Table 3.4 contains information on the measurement technique used, statistical analyses carried out to determine differential expression and a summary of the number of detected, non-differentially expressed (NDE), upregulated and downregulated transcripts or proteins. Of the transcriptomic studies, Alipanah et al. (2015) used a microarray while the rest used RNA-seq.

Between the studies, differential expression was calculated using several methods and resulted in a variety of outputs. All studies calculated $\log_2$ normalised fold change. Yang et al. (2013), Alipanah et al. (2015), Levitan et al. (2015) and Huete-Ortega et al. (unpublished) calculated FDR alongside calculations of differential expression. It is unclear whether Matthijs et al. (2016) provided FDR or, as with Longworth et al. (2016), unadjusted p values. All transcriptomic studies apart from that of Huete-Ortega et al. (unpublished) mapped reads to Phatr2 gene models from the Joint Genome Institute (JGI, 2008). Peptides were mapped to the reference proteome UP000000759. Huete-Ortega et al,. (unpublished) mapped reads to Phatr3 gene models (Rastogi et al., 2018), available from the Ensembl Protist database (Kersey et al., 2018).

Table 3.4 – Transcriptomic and proteomic metadata. Significance thresholds were taken from the literature and used to determine differential expression. The studies are denoted by a number referring to the hour of the time point and a letter referring to the author: H, Huete-Ortega; M, Matthijs; A, Alipanah; L48, Levitan; L24, Longworth; Y, Yang. Fold change (FC) values refer to nitrogen deplete measurements relative to replete as outlined in Table 3.1. No. common refers to the number of genes that had data in both experimental and control conditions. DE - differentially expressed, NDE - not differentially expressed, FC - fold change, no. - number.

Study /Timepoint	Technology used	DE statistical analysis	DE statistic	FC values	FC cutoff	DE statistic cutoff	Transcript \protein models	No. transcript /protein models	No. common transcripts /proteins	No. NDE	No. Upregulated	No. Down-regulated
H4	RNA-seq	baySeq	FDR	log2	NA	0.01	Phatr3	12089	11481	10828	253	400
H12									11481	8984	1270	1227
H18									11481	8305	1649	1527
M4	RNA-seq	CuffDiff	NA	log2	NA	0.05	Phatr2	10402	9999	6789	1621	1589
M8									9999	7451	1157	1391
M20									9986	5885	2003	2098
A48	Microarray	Limma package/ GEO2R	FDR	log2	NA	0.05	Phatr2	10402	10292	7885	1092	1315
A72									10291	5976	2258	2057
L48	RNA-seq	Deseq R/Bioconductor	FDR	log2	NA	0.05	Phatr2	10402	10133	4518	2728	2887
Y48	RNA-seq	NA	FDR	log2	NA	0.01	Phatr2	10402	8376	2636	1213	4527
L24	iTRAQ-MS	t test	unadjusted	log2	<-1, >1	0.05	UP000000759	10465	1759	859	429	471

Inter-study comparisons of the timing of physiological responses characteristic to nitrogen depletion and growth conditions, summarised above, provide insight into the relative progression of acclimatisation to nitrogen depletion. These factors were used to classify the studies as a way of grouping those exhibiting the most similar responses to nitrogen depletion. As outlined in Section 3.2.2, increases in carbohydrate and neutral lipid content, decreases in chlorophyll a content and cessation of cell division are consistent markers of this progress. Reference to Tables 3.2 and 3.3 indicates that transcriptomic measurements were taken prior to cessation of cell division in the studies by Huete-Ortega et al. (unpublished) and Alipanah et al. (2015). As recorded in Table 3.2, changes in carbohydrate, chlorophyll a and neutral lipids occurred at similar time points in the studies by Longworth et al. (2016) and Huete-Ortega et al. (unpublished) and therefore the proteomic measurement of 24h in Longworth et al. (2016) is likely to have occurred a short time after division ceased. Physiological and cell number data is not available for the study by Matthijs et al. (2016) but the author states that within the 20h of depletion, cells had bleached, began accumulating lipids and halted their cell cycle. As growth conditions were similar to those of Huete-Ortega et al. (unpublished), it is assumed cell division ceased shortly before 20h. The measurement of Levitan et al. (2015) occurred 24h after cell division had ceased. Taking cessation of cell division as a marker between 'early' and 'late' nitrogen depletion, the majority of the transcriptomic measurements fall into the early category. The 20h time point of Matthijs et al. (2016) and the proteomic measurement were taken at the beginning of the late period, shortly after cell division ceased and the measurement taken by Levitan et al. (2015) occurred substantially later. This relative ordering is used to present and assist interpretation of the comparisons of differential expression that follow.

The summaries of change in expression in Table 3.4 come from the numbers calculated in the studies themselves, prior to reprocessing the data sets for this review. Significant regulation in the studies was defined as a FDR or p value of less than 0.05 for Matthijs et al. (2016), Alipanah et al. (2015) and Levitan et al. (2015), $p<0.05$ combined with $\log_2$ fold change (FC)>1 or <-1 for Longworth et al. (2016) or FDR<0.01 for Yang et al. (2013) and Huete-Orteg et al. (unpublished). Increasing numbers of significantly regulated genes with time were seen in the study

by Huete-Ortega et al. (unpublished) and Alipanah et al. (2015). In contrast, ordering time points from Matthijs et al. (2016) based on the number of regulated genes at each results in 8, 4, 20h. As summarised in Table 3.4, at fourfold higher, the proportion of downregulated to upregulated genes in the study by Yang et al. (2013) was much larger than in the other studies. Assignment of fold change data to functional groups also demonstrated the same pattern (data not shown). Cell division data indicated that cells may have already been in stationary phase at the beginning of nitrogen deprivation and thus the study was deemed not sufficiently comparable to the others. The data set was therefore excluded from further analysis.

The data sets were processed to provide the required information for analysis. As the whole data set was not available from the study by Alipanah et al. (2015), the raw data set was reprocessed to provide $\log_2$ and FDR values using GEO2R. For the genes with processed data available, the $\log_2$ fold change values were the same. The recalculated FDR values were found to be more stringent than those provided by Alipanah et al. (2015). Some $\log_2$ fold changes for Matthijs et al. (2016) were not consistent with calculations that were carried out as a check and hence these genes were excluded from the data sets. Data from 166, 177 and 3750 genes were excluded from the 4, 8 and 20h data sets respectively. The large number of genes removed at the 20h period had an important effect, removing many of the most strongly regulated transcripts.

In order to analyse the data, Phatr2 identifiers for *P. tricornutum* genes were mapped to Phatr3 models (Rastogi et al., 2018), available from the Ensemble Protist database. A number of proteins could not be mapped to gene identifiers. Additionally, chloroplast transcripts detected by two of the studies (Matthijs et al., 2016; Alipanah et al., 2015) could not be mapped to Phatr2 or Phatr3 gene models and so were excluded from the data sets. Collectively, these factors resulted in the removal of 160, 88 and 210 genes with data from the Alipanah et al. (2015), Yang et al. (2013) and Longworth et al. (2016) studies respectively.

Table 3.5 summarises the expression data after the data sets were processed. Cutoffs for differentially expressed genes were set as $\log_2 > 1$ or < -1, FDR < 0.01 for the Huete-Ortega et al. (unpublished) data set and FDR or $p < 0.05$ for the other studies. Protein quantification data was used for proteins identified through

Table 3.5 – Summary of differential expression of transcriptomic and proteomic data sets based on significance thresholds used in this comparison. The studies are denoted by a number referring to the time point and a letter referring to the author: H, Huete-Ortega; M, Matthijs; A, Alipanah; L48, Levitan; L24, Longworth; Y, Yang. Fold change (FC) values refer to nitrogen deplete measurements relative to replete as summarised in Figure 3.1. DE: differentially expressed, NDE: non-differentially expressed, FC: fold change.

Study /Timepoint	FC cutoff	DE statistic cutoff	No. genes with data	No. genes without data	No. NDE	No. upregulated	No. down-regulated
H4	<-1, >1	0.01	11481	0	11042	150	289
H12			11481	0	10179	602	700
H18			11481	0	10107	668	706
M4	<-1, >1	0.05	9664	166	8378	594	692
M8			9656	177	8209	529	918
M20			6755	3750	6633	44	78
A48	<-1, >1	0.05	10128	160	8250	775	1103
A72			10128	160	7166	1467	1495
L48	<-1, >1	0.05	10133	0	6408	1668	2057
L24	<-1, >1	0.05	1549	210	770	390	389

annotation with at least one unique peptide. The number of downregulated genes in each study was larger than the number upregulated but by less than two fold. The proportion of regulated genes increased at subsequent time points within studies and reached a maximum of 36% differentially expressed genes in Levitan et al. (2015). The genes within these groups were mapped to Phatr3 IDs to enable matching with expression data.

Previous changes in expression observed across the studies were collated. Genes were grouped into 22 categories that were selected based on previous observations of regulatory and metabolic changes. These categories are displayed in Figure 3.2. Data from the 20h time point of Matthijs et al. (2016) was not included as the large number of removed genes would provide a misleading comparison. The data set of Longworth et al. (2016) is presented after the transcriptomic data sets in Figure 3.2 as its data is not directly comparable to that of the transcriptomic studies. The pre-curated gene groups created by Levitan et al. (2015) formed the majority and were supplemented by genes annotated to autophagy and the proteasome from the reannotation of *P. tricornutum* gene models (Rastogi et al., 2018), ribosome, amino acid synthesis and tRNA charging genes from KEGG and chrysolaminarin synthesis

genes from DiatomCyc (Fabris et al., 2012). The proteasome and autophagy groups were curated to remove unrelated genes.

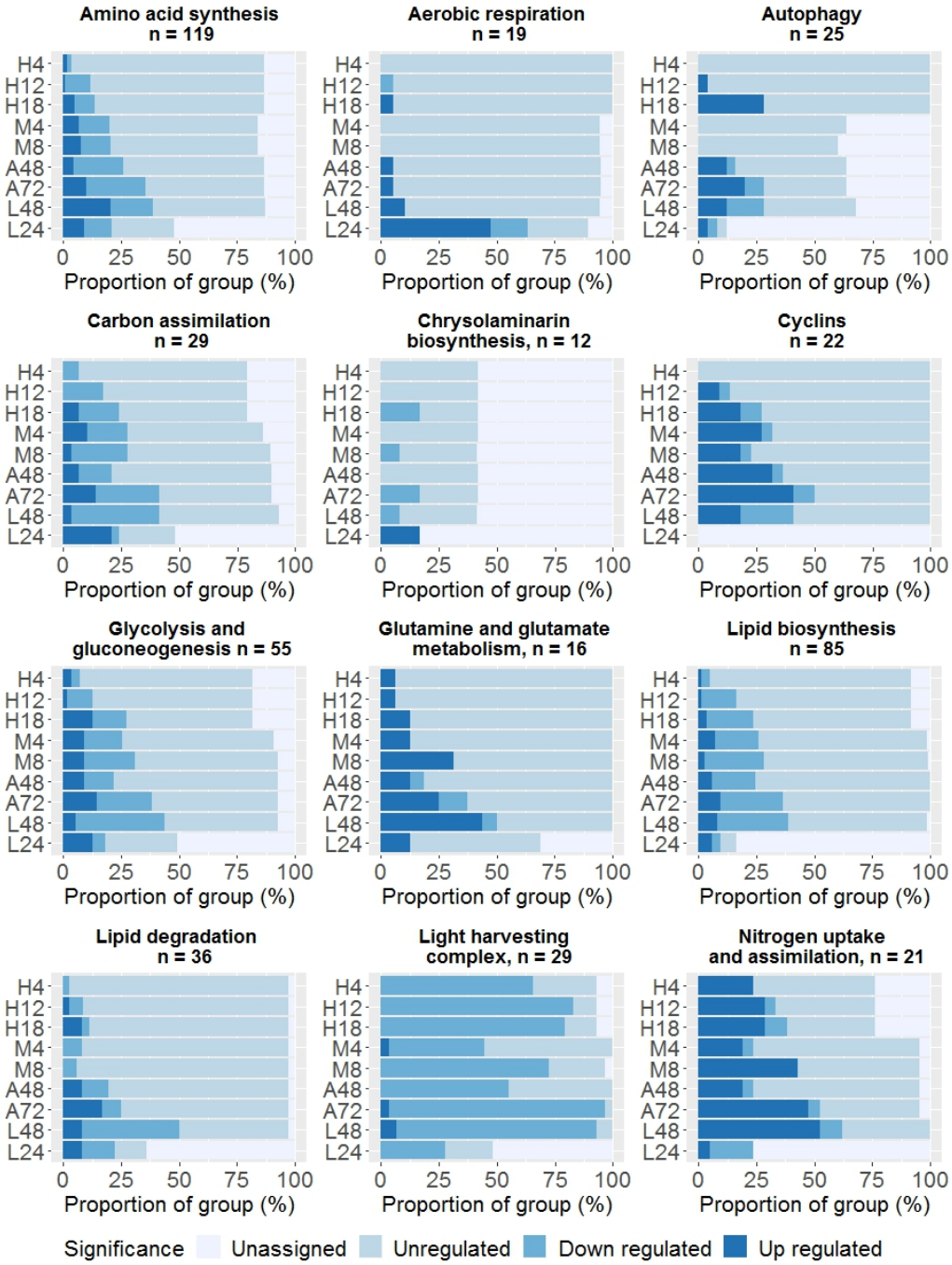


Figure 3.2 – Summary of the transcriptomic and proteomic response of selected metabolic and regulatory groups. The percentage of genes within a group that are significantly differentially upregulated or downregulated, insignificantly regulated (unregulated) or were not assigned data are shown. The number of genes in each group is shown. The studies are denoted by a number referring to the time point and a letter referring to the author: H, Huete-Ortega; M, Matthijs; A, Alipanah; L48, Levitan; L24, Longworth. Differential expression was calculated for nitrogen deplete measurements relative to replete. Significance thresholds were $\log_2$ FC < -1 or > 1 and p or FDR < 0.01 for Huete-Ortega and 0.05 for the other studies.

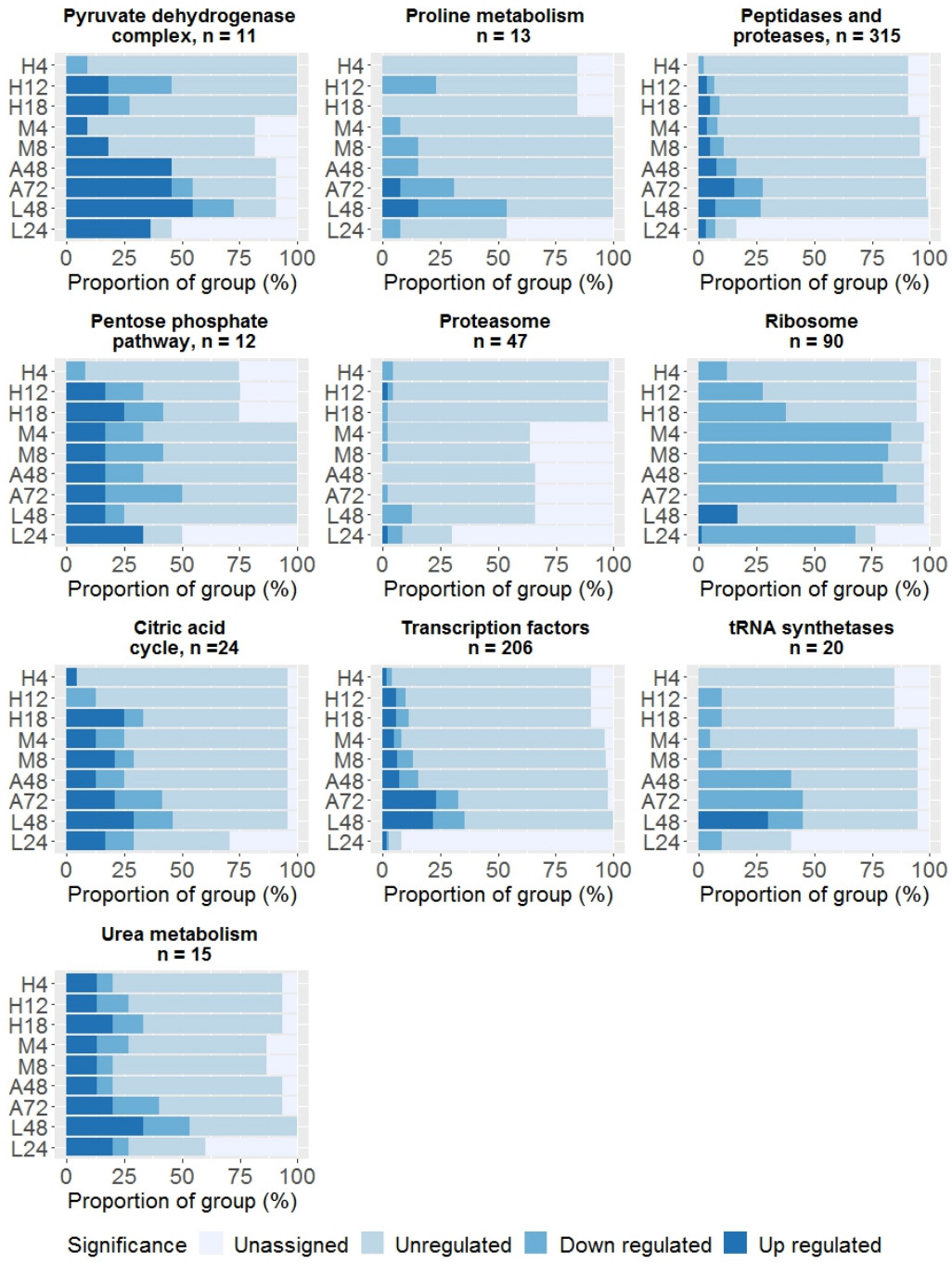


Figure 3.2 – Continued.

3.2.3.2 Comparison of regulation of gene groups before and after cessation of cell division

The data assigned to metabolic and functional groups was classified as upregulated, downregulated, unregulated or unassigned and is displayed in Figure 3.2. The requirements to be counted as differentially expressed are outlined above. The unassigned category contained genes for which data was not available. The number

of differentially expressed genes varied from each study due to differences in how the expression data was processed, therefore, the ratio of upregulated to downregulated genes was used to assess overall similarity between the depletion response between different studies.

The early/late nitrogen depletion distinction previously outlined was assessed by comparing the pattern of differential expression between studies. Alipanah et al. (2015) was classified as an early study as samples were taken prior to cessation of cell division. This classification was supported by this data set's similar ratio of upregulated to downregulated genes compared to the Matthijs et al. (2016) and Huete-Ortega et al,. (unpublished) data sets for the following gene groups: amino acid synthesis, cyclins, peptidases and proteases, the pentose phosphate pathway, the ribosome, tRNA synthetases, the proteasome, carbon assimilation and chrysolaminarin biosynthesis. In addition transcription factors and glutamine and glutamate metabolism were similarly regulated between the Alipanah and Matthijs studies, while lipid degradation was similarly regulated between those of Alipanah and Huete-Ortega (unpublished). Further supporting the difference in response following cessation of cell division is the unique regulation of amino acid synthesis, proline metabolism, the ribosome, tRNA synthetases, autophagy, cyclins, lipid degradation and peptidases and proteases in the study by Levitan at al (2015). Collectively, these changes indicate that changes in expression observed after cessation of cell division differ substantially to those in the studies that took measurements beforehand. Only a small number of discrepancies contradict this, with differences in the regulation of gene groups in early studies occurring for only two or three groups. Additionally, greater similarity of regulation of gene groups between early and late studies as opposed to between early studies was similarly noted for only two or three groups.

In contrast to the differences between data from before or after cell division ceased, some gene groups were commonly regulated across all studies. These include the light harvesting complex, nitrogen uptake and assimilation, lipid biosynthesis and urea metabolism. The pyruvate dehydrogenase complex and proline metabolism showed variable regulation both between studies and over time within studies, indicating that comparisons within a finer time scale or on an individual pathway or gene level is required.

3.2.3.3 Comparison of transcript and protein regulation

Due to the lower numbers of measured proteins, assignment of protein expression data to gene groups was much less extensive than for transcript data. Despite this, some differences and similarities with the other studies could still be seen, either as a result of better assignment to a particular group or regulation that obviously contrasted with that of transcripts. No cyclin and very few transcription factor proteins were detected but 75% of proteins in the ribosome group were. These were mostly downregulated, suggesting the same pattern of differential expression as the transcriptomic studies that were measured prior to halting of the cell cycle. Interestingly, a larger number of proteins were upregulated in the aerobic respiration, carbon assimilation and chrysolaminarin biosynthesis groups than transcripts as illustrated in Figure 3.2. This is the case despite the low number of proteins detected in the latter two groups. The response is particularly striking in aerobic respiration where 9 of the 19 genes in the group are upregulated compared to 2 in the transcriptomic data set that has the most upregulated transcripts. Furthermore, three genes are downregulated at the proteomic level, with downregulation only seen in one transcriptomic study, of 1 transcript. Additionally, despite detecting only 25% of the proteins in the group, four proteins assigned to nitrogen uptake and assimilation were downregulated compared to a maximum of 2 in transcript studies.

3.2.3.4 Regulation of protein metabolism during nitrogen depletion

The proteasome and autophagy were included as two catabolic groups. Low numbers of the genes assigned to the proteasome were differentially regulated, with slightly higher numbers seen at the proteomic level and later in the transcriptional response Levitan et al. (2015). Regulation of autophagy genes differed between the data sets. Differential expression increased over time in two of the transcriptomic studies (Alipanah et al,. 2015; Huete-Ortega et al. unpublished). Some downregulated transcripts were seen in the data of Alipanah et al. (2015) but upregulated transcripts were the majority in both studies. However, later in the depletion response this changed, with downregulated transcripts in the majority in the study of Levitan et al. (2015). Overall, more than 25% of the autophagy genes were differentially expressed once the response was established after around 18h, although few autophagy proteins

were detected.

A number of gene groups relating to protein synthesis were included. These were amino acid synthesis, proline metabolism, ribosome and tRNA synthetases. For amino acid synthesis, across the majority of the transcriptomic studies, differentially expressed genes made up around 12 to 40% of those in the group. More than three quarters of these differentially expressed genes in each study were downregulated apart from that of Levitan et al. (2015), where there were many more differentially expressed genes. The proteomic data also showed a mixture of up and downregulated genes. The majority of differentially expressed genes in the proline metabolism group were downregulated but expression did alter between study and time point. Some upregulated transcripts were seen in the Alipanah et al. (2015) and Levitan et al. (2015) data sets.

Genes in the ribosome and tRNA synthetase groups were regulated in a similar manner with differentially expressed genes majorly or completely downregulated in all studies apart from that of Levitan et al. (2015). The proportion of genes downregulated in the ribosome group was close to or more than 75% in three studies (Longworth et al., 2016; Matthijs et al., 2016; Alipanah et al., 2015) while the proportion in the tRNA synthetases group ranged between 5 and 45%. In contrast, more than 15% and 25% of the ribosome and tRNA synthetase genes respectively were upregulated in the study by Levitan et al. (2015). In addition, 15% of the synthetase genes were downregulated in the same study.

Considering the above regulation, a number of aspects of protein metabolism during nitrogen depletion can be summarised. Protein synthesis is likely to be much reduced prior to cell division due to extensive downregulation of amino acid synthesis, ribosome components and tRNA synthetases. Surprisingly, this situation may reverse after cell division with increased upregulation seen in these groups. The effect of protein degradation on the cellular protein content is less clear however. Proteasome genes demonstrate little differential expression, while peptidases and proteases are up and downregulated in equal amounts prior to cell division ceasing but are largely downregulated afterwards, further supporting an increase in protein synthesis at that time. There are also signs that autophagy plays a role in the response to depletion, particularly at early stages. When differentially expressed, autophagy

genes are majorly upregulated prior to cell division halting, but regulation is evenly split between up and downregulation subsequently.

3.2.3.5 Regulation of lipid metabolism during nitrogen depletion

Lipid metabolism clearly plays an important role in the response to nitrogen depletion. Apart from the earliest time point of the data set from Huete-Ortega (unpublished), of the 85 genes grouped to lipid biosynthesis, between 20 and 35% of transcripts were differentially expressed across the studies. Around 80% of these genes had reduced expression. This supports decreased lipid synthesis during nitrogen depletion. Of the 36 genes annotated to lipid degradation, the situation was more variable. Around 5 to 10% of genes across the data sets taken prior to cessation of cell division (Huete-Ortega et al., unpublished; Alipanah et al., 2015; Matthijs et al., 2016) were downregulated. Upregulated genes were also noted by Huete-Ortega et al. (unpublished) and Alipanah et al. (2015), and increased to constitute a larger proportion than the downregulated at 18 and 72h. In the Levitan data set, differential expression reached 50% of the genes in the group, the majority of which were downregulated. Nuanced regulation of lipid degradation is therefore supported, with increasing degradation prior to cell division but with lipids mostly conserved afterwards. The protein data in these two groups showed similar regulation to the transcripts, but the majority of proteins in the group did not have data assigned. To provide more insight, investigation of the regulation of specific genes is needed and this is reported in Section 3.2.5.

3.2.3.6 Regulation of photosynthesis during nitrogen depletion

Genes involved in photosynthesis are contained in the light harvesting complex and carbon assimilation gene groups. The carbon assimilation group contains genes from the Calvin cycle. Consistently, across both trancriptomic and proteomic studies, light harvesting complex gene expression is majorly decreased. This is seen from very early after nitrogen depletion and continues to the latest stage measured in these studies. Several of the studies contain genes which are upregulated (Matthijs et al., 2016; Alipanah et al., 2015; Levitan et al., 2015), but this remains less than 5% of the group. Within carbon assimilation, the expression of the majority of genes

remain unchanged but a substantial number of genes are differentially expressed, the majority of which are downregulated. Each study also noted upregulated transcripts and the majority of differentially expressed proteins were upregulated. While these responses indicate that light capture is decreased from an early stage in nitrogen depletion, alterations to carbon assimilation are ambiguous and need more refined investigation with a focus on the regulation of particular reactions in the pathway.

3.2.3.7 Regulation of central carbon and energy metabolism during nitrogen depletion

Carbohydrate metabolism is represented by the glycolysis and gluconeogenesis, pentose phosphate pathway (PPP) and chrysolaminarin biosynthesis gene groups. A mixture of up and down differential expression of transcripts were observed in the glycolysis and gluconeogenesis and PPP groups, accounting for more than 25% of the genes in most studies. Less than 50% of chrysolaminarin biosynthesis genes had data assigned, but of those that did, downregulation increased within studies over time. Regulation of protein expression differed for the PPP and chrysolaminarin synthesis, with a higher proportion of upregulated expression for PPP genes. This is despite detection of only 50% of the proteins in the PPP group.

Differential expression of genes of the pyruvate dehydrogenase complex was observed in each of the studies. Upregulation of transcripts is an important component of this, with up to 50% (6) of the genes upregulated. Downregulation of genes was also observed, with up to 3 regulated in this manner. Downregulation was less consistent across the data sets. These genes are investigated in more detail in Section 3.2.5.

Expression of the citric acid cycle was consistently regulated across the studies with up to 50% of the 24 genes in the group differentially expressed. Upregulation was slightly predominant across the studies apart from in the first two time points of the data set from Huete-Ortega (unpublished).

The aerobic respiration group, containing 19 genes from the mitochondrial electron transfer chain, was unregulated at the transcript level apart from 1 or 2 genes. In contrast, almost 50% of the genes are upregulated at the protein level, while several are downregulated. This supports regulation of aerobic respiration occurring mainly

at the protein level during nitrogen depletion. Collectively, the enzymes of central carbon metabolism and aerobic respiration require investigation of the regulation of individual enzymes to better conclude on their role during nitrogen depletion.

3.2.3.8 Regulation of nitrogen uptake and assimilation

Nitrogen uptake and assimilation, urea metabolism and glutamine and glutamate metabolism gene groups were included to investigate regulation of genes involved in the sourcing and incorporation of nitrogen. The regulation of differentially expressed transcripts across these groups was similar, with genes predominantly upregulated across all studies. Downregulated transcripts also made up a sizeable proportion in urea metabolism. In total, differentially regulated transcripts accounted for between 5 and 50% of glutamine and glutamate metabolism and 20 and 50% of nitrogen uptake and assimilation and urea metabolism. Differential protein expression in these groups was similar for glutamine and glutamate metabolism and urea metabolism but downregulated proteins made up a larger proportion of the nitrogen uptake and assimilation group, at around 20%.

3.2.3.9 Regulation of regulatory genes

Two regulatory groups were included in the analysis, namely cyclins and transcription factors. The transcriptional response of the cyclins was similar across the transcriptomic data sets measured prior to the cessation of cell division, except from the 4h time point of the Huete-Ortega data set, with between 15 and 50% of the transcripts differentially expressed, the vast majority of which were upregulated. In contrast, after cell division ceases in the study of Levitan et al. (2015), although around 40% of the transcripts were differentially expressed, slightly more than 50% of these were downregulated. A number of transcription factors were also differentially expressed in nitrogen depletion. There was a roughly equal split between up and downregulated genes in the data sets from Huete-Ortega et al., Matthijs et al. (2016) and 48h time point of Alipanah et al. (2015). Upregulation predominated at the 72h timepoint and in the Levitan et al. (2015) study however. Protein expression data for these two groups was very limited. Collectively these changes highlight the substantial reprogramming of the cell that is undertaken during nitrogen depletion.

3.2.3.10 Consensus of changes in gene expression during nitrogen depletion

In summary, the results described above support substantial commonalities in the response to nitrogen depletion between studies but with some distinctions between the regulation seen before and after cessation of cell division, something that was not noted in the studies this work made use of. Of general note is that many genes were not differentially expressed. Apart from this, support was seen for a decrease in the ability to harvest light, assimilate carbon and synthesise lipids and an increase in the ability to recycle nitrogen through the urea cycle from early in the response and subsequent continuation of these processes throughout. Although alterations to central carbon metabolism were consistent throughout nitrogen depletion, without further investigation, the impact these changes will have on the cell are less certain due to their ambiguous regulation.

A number of processes were regulated differently before and after cell division ceased. Early in the response, changes to expression were seen that would lead to decreased protein synthesis and increased degradation. However, afterwards this altered with expression supporting renewed protein synthesis or preparation for protein synthesis upon the return of nitrogen. Early changes in expression of cyclins and transcription factors were noted, providing evidence of regulatory activity that brings about the acclimatisation response and cessation of division. As might be expected from the cell adopting a quiescent state during late acclimatisation, the expression of these groups was different after cell division ceased. Similar patterns of changed expression were also seen for autophagy and lipid degradation groups, suggesting that the activity of these proteins increased prior to cessation of cell division but decreased afterwards. A decrease in expression of peptidases and proteases is also coherent with this decreased degradation post cessation of cell division.

Another notable difference between the studies is the differentiated response of the transcripts and proteins of some groups. The different expression of proteins and transcripts in aerobic respiration, carbon assimilation, chrysolaminarin biosynthesis and nitrogen uptake and assimilation highlight the difficulties of drawing conclusions on these groups from transcript data.

3.2.4 Regulation of autophagy genes during nitrogen depletion studies

Autophagy is a process that is conserved across eukaryotes, in which intracellular components of a cell are engulfed by a double membrane-bound organelle termed the autophagosome that subsequently fuses with the lysosome or vacuole, leading to the degradation of the engulfed material. Autophagy plays an important role in acclimatisation to nutrient depletion and has previously been observed to play a role in chloroplast degradation in plants (Ishida et al., 2014; Izumi et al., 2017). It was therefore decided to investigate expression of autophagy genes in a more in-depth manner. The process by which the phagophore is formed, expansion into a autophagosome and the subsequent degradation of its contents involves autophagy related genes (ATGs).

A number of core autophagy genes were transcriptionally regulated during these studies and are presented in Table 3.6 along with annotation from Pfam (El-Gebali et al., 2019). Among the earliest regulated genes is Phatr3_EG00153 which is annotated as Atg101, the protein of which forms a part of the complex involved in initiation of phagophore formation, called the Ulk complex in mammals or the Atg1 complex in yeast. It is upregulated 18h after depletion in Huete-Ortega's unpublished study. Annotation as Atg101 suggests *P. tricornutum* uses the Ulk complex (Suzuki et al., 2015). Phatr3_44229, annotated as Atg11 is upregulated even earlier, at 8h and subsequently 18 h after nitrogen depletion. The Atg11 protein activates Atg1 in selective autophagy in yeast, and may function in a similar manner in the Ulk complex. Phatr3_44229 was also found to be upregulated at later stages of depletion, at 48 and 72h (Alipanah et al., 2015).

The Ulk or Atg1 complex recruits the phosphatidyl inositol 3-kinase nucleation complex which participates in membrane nucleation and consists of Atg6 and Atg14. A number of regulated genes were annotated as such in these studies: Phatr3_J43402 and Phatr3_46466 as Atg6 and Phatr3_J50209, Phatr3_J31423, Phatr3_J35961 and Phatr3_J39167 as Atg14. One Atg6 gene is upregulated at 18h and 48h and two of the Atg14 genes are upregulated at 48h in Levitan et al. However, the other Atg6 gene is downregulated at the timepoints after 20h and two of the Atg14 genes are

Table 3.6 – Significantly regulated autophagy genes. Autophagy genes with absolute log2 fold change values > 1 or < -1 and p < 0.01 (H2, H8, H18) or p < 0.05 in at least one nitrogen depletion study. Transcriptomic studies: H4, 8, 18 - Huete-Ortega 4h, 8h, 18h; M4, 8, 20 - Matthijs 4h, 8h, 20h; A48, 72 - Alipanah 48h, 72h; L48 - Levitan 48h. Proteomic study: L24 - Longworth 24h.

		Fold change (log_2)									
#IDs	Pfam annotation	H4	H8	H18	M4	M8	M20	A48	A72	L48	L24
Phatr3_J49352	Atg2 CAD motif			1.440				1.392	1.830	-1.006	
Phatr3_J16669	Atg2 CAD motif			1.484					1.620		
Phatr3_J43402	Atg6			1.251						1.095	
Phatr3_J46468	Atg6						-1.016	-1.148	-1.350	-1.187	
Phatr3_J50303	Atg8 ubiquitin like			1.026				2.353	2.438		
Phatr3_J44229	Atg11		1.138	1.300				1.388	2.010		
Phatr3_J50209	Atg14 subunit										
Phatr3_J31423	Atg14 subunit									1.167	
Phatr3_J35961	Atg14 subunit									1.365	
Phatr3_J39167	Atg14 subunit									-1.658	
Phatr3_J50005	Atg27								-1.128		
Phatr3_EG00153	Atg101			1.900							
Phatr3_J1143	Atg C terminal domain									-2.397	1.457
Phatr3_J47968	Atg C terminal domain										-2.524
Phatr3_J9342	ThiF family			1.076					1.147		

downregulated 48h after depletion.

The phosphatidyl inositol 3-phosphate complex is then recruited to the other complexes where it functions in phagophore expansion. It is composed of Atg2, Atg9 and Atg18. A requirement for Atg27 in Atg9 cycling between the phagophore and mitochondria has also been observed (Yen et al., 2007). Two genes, Phatr3_J49352 and Phatr3_J16669 have been annotated as Atg2. Both genes are upregulated at 18h and in the study by Alipanah. Phatr3_J49352 is also downregulated at 48h in the study by Levitan. A gene annotated as Atg27 was downregulated late in nitrogen depletion as detected by Alipanah et al..

Cargo recognition, phagophore expansion and closure to produce the autophagosome and subsequent targeting to its destination are achieved through the combined action of two ubiquitin-like conjugating complexes. The first is composed of Atg5, Atg7, Atg10, Atg12 and Atg16. The second is composed of Atg8, Atg3, Atg4, Atg5 and Atg7. Phatr3_J50303 is annotated as Atg8 and was upregulated at 18h, 48h and 72h. Phatr3_J9342 contains the ThiF domain, which is found in the ubiquitin activating E1 family. It is therefore a candidate for Atg7 (Delorme-Axford and Klionsky, 2018). Phatr3_J9342 is upregulated at 18h in the study by Huete-Ortega and 72h in that of Alipanah.

Transcriptional upregulation of Atg8, Atg7 or Atg3 has been identified as a marker

of autophagy activation in *C. reinhardtii* (Pérez-Pérez et al., 2017). Upregulation of Atg8 and Atg14 has been observed in *S. cerevisiae* under nutrient depletion, with the amount of Atg8 directly correlating with the size of the autophagosome (Delorme-Axford and Klionsky, 2018). The transcriptional upregulation of genes annotated as Atg8, a potential Atg7 and Atg14 genes in nitrogen depletion studies in *P. tricornutum* therefore supports activation of autophagy in *P. tricornutum* under these conditions. Upregulation of other core Atg genes, Atg101, Atg11 Atg2 and Atg6 adds further strength to this interpretation. With reference to Table 3.2, timings of physiological changes of the Huete-Ortega study can be compared to transcriptional ones. Interestingly, an autophagy activating gene, Atg11, is upregulated at 8h. Between 4 and 12h, the level of chlorophyll decreases and total lipids, acting as a proxy for TAG, increases. A suite of other core autophagy genes, including Atg101, Atg8 and Atg7 are upregulated at 18h. Activation of autophagy, degradation of chlorophyll and increase in TAG may therefore occur cotemporally.

3.2.5 Regulation of lipid metabolism genes during nitrogen depletion studies

To investigate the expression of lipid metabolism genes further, genes were assigned to metabolic pathways. These genes and their pathway and functional associations were compiled from a review paper (Zulu et al., 2018), the DiatomCyc pathway genome database (Fabris et al., 2012) and the study by Levitan et al. (2015) and partially overlapped with the lipid metabolism biosynthesis and degradation sets described in Section 3.2.3. The genes, assigned to pathways, are displayed in Table 3.7 alongside expression data, predicted function and predicted subcellular location. Where data was not available for expression data, NA was recorded. Protein location was predicted using ASAFind (Gruber et al., 2015) to assign genes to cytoplasm, endoplasmic reticulum/secretory system or chloroplast. TargetP (Emanuelsson et al., 2007) was then used to predict mitochondrial targeting. A mitochondrial location was accepted only for genes predicted to be targeted to the cytoplasm or ER/secretory system by ASAFind. For reference, Figure 1.6 provides a schematic of lipid biosynthesis and beta oxidation processes encompassed in Table 3.7.

Table 3.7 – Differential expression of transcripts and proteins of lipid metabolism during nitrogen depletion. The studies are denoted by a number referring to the hour of the time point and a letter referring to the author: H, Huete-Ortega; M, Matthijs; A, Alipanah; L48, Levitan; L24, Longworth. Fold change (FC) values refer to nitrogen deplete measurements relative to replete at different time points. Significance thresholds were $\log_2$ FC < -1 or > 1 and p or FDR < 0.01 for Huete-Ortega and 0.05 for the other studies. FC values were omitted if not significantly altered. Time points without data were denoted NA.

	Gene	Predicted function	Location	H4	H12	H18	M4	M8	M20	A48	A72	L48	L24
Beta oxidation	Phatr3_J20310	Acyl-CoA dehydrogenase	Mito							1	1.9	-1	NA
	Phatr3_J11014	Acyl-CoA dehydrogenase/oxidase	Mito										-2
	Phatr3_J11811	2-enoyl-CoA hydratase	Mito							1.9	1.7		NA
	Phatr3_J35240	2-enoyl-CoA hydratase	Mito										-1
	Phatr3_J40988	2-enoyl-CoA hydratase	Mito										
	Phatr3_J54494	2-enoyl-CoA hydratase	Mito								1.2		NA
	Phatr3_J54528	2-enoyl-CoA hydratase	Mito			2.2			NA	1.1			NA
	Phatr3_J55192	2-enoyl-CoA hydratase	Mito		-1							-2	NA
	Phatr3_J5527	2-enoyl-CoA hydratase	Mito										NA
	Phatr3_J39681	3-hydroxyacyl-CoA dehydrogenase.	Mito			1.4			NA			-2	NA
	Phatr3_J41886	3-hydroxyacyl-CoA degydrogenase	Mito		1.3	1.1					1.2	1.4	NA
	Phatr3_J28068	3-oxoacyl-CoA thiolase	Mito							-1			1.5
	Phatr3_J45947	3-oxoacyl-CoA thiolase	Mito							-1	-1		
	Phatr3_J42907	Acyl-CoA dehydrogenase/oxidase	Cyt										
	Phatr3_J19979	Precursor of oxidase acyl-CoA oxidase	Cyt						NA				
	Phatr3_J55069	Peroxisomal bifunctional enzyme. 2-enoyl-CoA hydratase or 3-hydroxyacyl-CoA degydrogenase	Cyt									-1	NA
	Phatr3_J41969	3-oxoacyl-CoA thiolase	Cyt										NA
Entry of unsaturated FAs into β-oxidation	Phatr3_J35240	3-hydroxyacyl-CoA dehydrogenase	Mito										-1
	Phatr3_J18064	delta-3,5-delta-2,4-dienoyl-CoA isomerase (DEI1)	Mito	-1			-2	-1	NA	-2	-2	-2	NA
	Phatr3_J31367	2,4-dienoyl-CoA reductase.	Mito				-2		NA	-2		-1	NA
	Phatr3_J11319	Peroxisomal 2,4-dienoyl-CoA reductase (DER1)	Cyt						NA		-1	-3	-2
	Phatr3_EG02333	delta-3,5-delta-2,4-dienoyl-CoA isomerase (DEI1)	Cyt				NA	NA	NA	NA	NA	NA	NA
	Phatr3_J55069	Peroxisomal bifunctional enzyme	Cyt									-1	NA

Table 3.7 – Table 3.7 continued

	Gene	Predicted function	Location	H4	H12	H18	M4	M8	M20	A48	A72	L48	L24
Glycerolipid degradation	Phatr3_J34489	Monoacylglycerol lipase	Mito						NA		1.7	-2	NA
	Phatr3_J46570	TAG lipase	Mito									1	NA
	Phatr3_J41254	Monoacylglycerol lipase	Cyt										NA
	Phatr3_J42683	Phospholipase C	Cyt									1.1	NA
	Phatr3_J49771	Phospholipase C	Cyt						NA			-1	1.6
	Phatr3_Jdraft1000	Phospholipase C	Cyt	NA	NA	NA	NA	NA	NA	NA	NA		
	Phatr3_Jdraft1611	Phospholipase C	Cyt	NA	NA	NA	NA	NA	NA	NA	NA		
	Phatr3_J12431	Phospholipase D	Cyt						NA		1.1		NA
ATP citrate lyase/synthase	Phatr3_J54477	ATP citrate lyase/synthase	Cyt									-3	
Acetyl-CoA synthetase	Phatr3_Jdraft645	Acetyl-CoA synthetase	ER/SS										
	Phatr3_EG02293	Acetyl-CoA synthetase	Cyt										
	Phatr3_J22974	Acetyl-CoA synthetase	Cyt									-5	
Pyruvate dehydrogenase	Phatr3_J20183	Precursor of E1 component beta subunit	Mito						NA			1.7	
	Phatr3_J9476	Pyruvate dehydrogenase E1 component	Cyt		2.7	5.1		1.2	NA	3.3	5.7	2.4	NA
	Phatr3_J55035	E1 component alpha subunit	Mito			1			NA			1.2	NA
	Phatr3_J17401	Dihydrolipoamide acetyltransferase E2 component	Mito						NA	1.2	1.2	1.2	1.2
	Phatr3_J40430	Dihydrolipoamide acetyltransferase E2 component	Mito									1.6	1.5
	Phatr3_J26432	Dihydrolipoyl dehydrogenase, E3 component	Cyt							1.3	1.2	1.3	1.3
	Phatr3_J23850	Dihydrolipoamide acetyltransferase E2 component	Chlp		-2		NA	NA	NA		-1	-5	NA
	Phatr3_Jdraft1828	Dihydrolipoamide acetyltransferase E2 component	Chlp	-3	-3	-1	NA	NA	NA	NA	NA	NA	NA
	Phatr3_J30113	Lipoamide dehydrogenase, E3 component	Chlp		-2							-3	1.2
Pyruvate dehydrogenase kinase	Phatr3_J13123	Pyruvate dehydrogenase kinase	Mito	1.6	1.3					-1	-1	1.8	NA
	Phatr3_J15081	Pyruvate dehydrogenase kinase	Mito								1		NA
	Phatr3_J50961	Pyruvate dehydrogenase kinase	Mito						NA				
	Phatr3_J46264	Pyruvate dehydrogenase kinase	Cyt		2.1	1							NA

Table 3.7 – Table 3.7 continued

	Gene	Predicted function	Location	H4	H12	H18	M4	M8	M20	A48	A72	L48	L24
Acetyl-CoA carboxylation	Phatr3_EG01955	Acetyl-CoA carboxylase	Chlp		-1		NA	NA	NA	NA	NA	NA	NA
	Phatr3_EG01956	Acetyl-CoA carboxylase	Chlp				NA	NA	NA	NA	NA	NA	NA
	Phatr3_J37652	Malonyl-CoA:ACP transacylase	Chlp		-4	-4	-1	-3	NA	-2	-5	-6	1.4
	Phatr3_J55209	Acetyl-CoA carboxylase	Cyt				-1		NA	-1		-2	NA
Fatty acid synthesis	Phatr3_J37367	3-oxoacyl-ACP synthase	Chlp		-3	-5	-2	-3	NA	-2	-5	-5	
	Phatr3_J52648	3-oxoacyl-ACP synthase	Chlp		-1	-1		-2	NA	-1	-3	-5	1.6
	Phatr3_J13073	3-oxoacyl-ACP reductase	Chlp		-2	-1		-1	NA			-4	1.7
	Phatr3_J55157	3-hydroxyacyl-ACP dehydratase	Chlp				NA	NA	NA	-1	-3	-3	NA
	Phatr3_J10068	enoyl-ACP reductase	Chlp	-2	-3	-2	-2	-3	NA	-2	-4	-5	
	Phatr3_J18940	3-oxoacyl-ACP synthase	Mito	1.3		1.2	1		NA		1		NA
	Phatr3_J21608	3-oxoacyl-ACP reductase	Cyt		-1								NA
	Phatr3_J10306	3-oxoacyl-ACP reductase	Cyt										-1
	Phatr3_J47536	enoyl-ACP reductase	Mito										NA
FA desaturation in the chloroplast	Phatr3_J9316	Soluble delta 9 desaturase. 16:1 or 18:1 production	Chlp			1.9	1.2		NA	1.7	2.2		
	Phatr3_J48423	FAD6. Generates 16:2n-4	Chlp			-2		-1	NA		-2	-3	NA
	Phatr3_EG02619	PtFAD6. 16:3n-4 production.	Chlp			-2	NA	NA	NA	NA	NA	NA	NA
	Phatr2_41301	FAD4, fatty acid desaturase	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Kennedy Pathway - chloroplast	Phatr3_J11916	GPAT	Chlp				-1	-2	NA	-1	-2		NA
	Phatr3_J3262	GPAT	Chlp									1.4	NA
	Phatr3_J11916	LPAT	Chlp				-1	-2	NA	-1	-2		NA
	Phatr3_J43469	DGAT2D	Chlp								1.1	1.4	NA
	Phatr3_J8663	Phosphatidyl glycerol phosphate synthase	Chlp										NA
	Phatr3_J54168	MGDG synthase	Chlp		-1	-1			NA		-1	-1	NA
	Phatr3_J14125	MGDG synthase	Cyt			-1			NA		-1		NA
	Phatr3_J9619	MGDG synthase	Cyt						NA				NA
	Phatr3_J50356	SQD2	Cyt		-2	-2		-1	NA	-1	-2	-2	NA

Table 3.7 – Table 3.7 continued

	Gene	Predicted function	Location	H4	H12	H18	M4	M8	M20	A48	A72	L48	L24
FA membrane transport	Phatr3_J10454	Fatty acid thioesterase	ER/SS						NA	1.2	1.5	-1	NA
	Phatr3_J33198	Fatty acid thioesterase	Cyt										NA
Fatty acid elongation in cytoplasm	Phatr3_J26714	3-oxoacyl-CoA synthase	Cyt										NA
	Phatr3_J20508	ELO6. 18:3 to 20:3 or 18:4 to 20:4	Cyt	-2	-3	-3	-2	-3	NA	-2	-4	-4	NA
	Phatr3_J22274	ELO6. 18:3 to 20:3 or 18:4 to 20:4	Cyt			-1							NA
	Phatr3_J34485	ELO5. 20:4 to 22:4 or 20:5 to 22:5	Cyt			-2	-1	-2	NA	-1	-2	-2	NA
	Phatr3_J37372	enoyl-CoA reductase	Cyt						NA			1.4	NA
	Phatr3_J10497	enoyl-CoA reductase	Mito										
Long chain acyl-CoA sythesis or Acyl-CoA-glycerolipid transacylation	Phatr3_J20143	Long chain acyl-CoA synthetase, ACS1	Chlp										NA
	Phatr3_J45510	Long chain acyl-CoA synthetase, ACS4.	Cyt									-2	-1
	Phatr3_J12420	Long chain acyl-CoA synthetase	Cyt									-3	NA
	Phatr3_J17720	Precursor of long-chain acyl-CoA ligase, ACL1.	Mito						NA			-2	NA
	Phatr3_J54151	Long chain acyl-CoA synthetase, ACS3	Mito				-1						-2
Fatty acid desaturation in ER	Phatr2_32224	Stearoyl-CoA desaturase, 18:0 to 18:1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Phatr3_J29488	PTD6. 18:2 to 18:3 or 18:3 to 18:4	Cyt										NA
	Phatr2_22469	Delta 17 desaturase. 18:4 or 20:4 synthesis	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	Phatr3_J46838	Delta 17 desaturase. 18:4 or 20:4 synthesis	Cyt				1.2					-1	NA
	Phatr3_J46830	Delta 5 desaturase. 20:3 to 20:4 or 20:4 to 20:5	Cyt						NA	1.2	1.8		NA
	Phatr3_J22510	Delta 19 desaturase. 22:4 to 22:5 or 22:5 to 22:6	Cyt						NA			1.9	NA
	Phatr3_J22459	Delta 5 desaturase. 20:3 to 20:4 or 20:4 to 20:5	Mito									1.1	NA
	Phatr3_J25769	FAD2, delta12 desaturase. 18:1 to 18:2	Mito					-1	NA				NA
	Phatr3_J41570	PTD15. 18:3 from 18:2 or 18:4 from 18:3	Chlp					-2	-1.6		-1		NA

Table 3.7 – Table 3.7 continued

	Gene	Predicted function	Location	H4	H12	H18	M4	M8	M20	A48	A72	L48	L24
Kennedy Pathway in ER	Phatr3_EG02461	GPAT	Cyt				NA	NA	NA	NA	NA	NA	NA
	Phatr3_J54709	GPAT	ER/SS										NA
	Phatr3_J43099	GPAT	Mito				-1	-2	NA	-1	-2		NA
	Phatr3_EG02461	LPAT	Cyt				NA	NA	NA	NA	NA	NA	NA
	Phatr3_J43099	LPAT	Mito				-1	-2	NA	-1	-2		NA
	Phatr3_J45551	LPAT	Mito			1.4			NA		1		NA
	Phatr3_J39949	PAP	Cyt										NA
	Phatr3_J40261	PAP	Cyt										NA
	Phatr3_J9794	DGAT1	Cyt										NA
	Phatr3_J49462	DGAT2A	Mito				-1	-2	NA		-2	-3	NA
	Phatr3_J49544	DGAT2B	Cyt									-1	NA
	Phatr3_J31662	DGAT2C	Cyt						NA				NA
	Phatr3_EG00369	DGAT2E	Cyt										
	Phatr3_J49708	DGAT3	ER/SS						NA			-2	NA
The Lands Cycle and lipid remodelling	Phatr3_J39425	PLA2	ChIp									-2	NA
	Phatr3_J46193	PLA2	Mito	-1	-1				NA				NA
	Phatr3_J49702	acyl-CoA:LPCAT	Cyt			-2							NA
	Phatr3_J20460	acyl-CoA:LPCAT	Cyt				-1	-2	NA		-2		NA
	Phatr3_J8860	PDAT	Mito						NA		1.3		NA
	Phatr3_J49702	PDAT	Cyt			-2							NA
Glycerolipid sythesis	Phatr3_J14202	DAG kinase	Cyt						NA				NA
	Phatr3_J47250	Phosphatidate cytidyl transferase	Cyt										NA
	Phatr3_J54756	Phosphatidate cytidyl transferase	Cyt										NA
	Phatr3_J7678	Phosphatidate cytidyl transferase	ER/SS										NA
	Phatr3_Jdraft559	Phosphatidate cytidyl transferase	Mito				NA	NA	NA	NA	NA	NA	NA
	Phatr3_J43010	Phosphatidylinositol synthase	Cyt										NA
	Phatr3_J47576	Phosphatidyl glycerol phosphate synthase (PGPS)	Mito										NA
	Phatr3_J39949	Phosphatidylgycerophosphatase	Cyt										NA
	Phatr3_J33864	Ethanolaminephosphotransferase or DAG cholinephosphotransferase	Mito				-1	-1	NA				NA

3.2.5.1 Regulation of acetyl-CoA production

Acetyl-CoA is required for initiation of and elongation in fatty acid synthesis. Four processes are predicted to be involved in acetyl-CoA production in diatoms (Zulu et al., 2018). These include pyruvate dehydrogenation, beta oxidation, ATP citrate lyase and acetyl-CoA synthetase genes. Most of the genes assigned to mitochondrial beta oxidation were regulated. Prior to cell division, the majority of the beta oxidation genes were transcriptionally upregulated or not differentially expressed. Six transcripts, including isoforms of acyl-CoA dehydrogenase, 2-enoyl-CoA hydratase and 3-hydroxyacyl-CoA dehydrogenases were upregulated while a further four were unregulated including an acyl-CoA dehydrogenase and three 2-enoyl-CoA hydratases. Three transcripts were downregulated, including a 2-enoyl-CoA hydratase and two 3-oxoacyl-CoA thiolases. An acyl-CoA dehydrogenase and a 2-enoyl-CoA hydratase that were transcriptionally not differentially expressed, were downregulated at the protein level while a 3-oxoacyl-CoA thiolase was upregulated, in contrast to the downregulation of its transcript. After cell division had ceased, the majority of transcripts were not differentially expressed. Three were downregulated however, including an acyl-CoA dehydrogenase and a 3-hydroxyacyl-CoA dehydrogenase that were previously upregulated.

The branched chain amino acids (BCAA), isoleucine and valine are partly broken down by similar enzymes to those used in beta oxidation and therefore the genes assigned to beta oxidation may be involved in amino acid degradation. Isozymes acting on amino acids are likely to be soluble (Houten et al., 2016) and only Phatr3_J41969 is predicted to contain a transmembrane helix, supporting the potential role in protein degradation of the other genes annotated to beta oxidation. Irrespective of the substrate of the predicted beta oxidation enzymes, their action leads to the production of acetyl-CoA, along with other short chain acyl-CoAs in the case of BCAA degradation.

In addition to mitochondrial beta oxidation genes, four enzymes with predicted cytoplasmic localisation have been identified. These are likely targeted to the peroxisome. These genes were not differentially expressed prior to cell division. However, downregulation of peroxisomal bifunctional enzyme was observed after cessation of cell division (Levitan et al., 2015). A number of genes were annotated

with functioning to prepare unsaturated fatty acids for entry into beta oxidation. Three are predicted to be localised to the mitochondria and three to the cytoplasm. All but one of the genes annotated to this function were downregulated at some point during depletion. Three genes were downregulated or otherwise not differentially expressed prior to halting of cell division, two of which were annotated to the mitochondria and the third to the cytoplasm. One of these was also down regulated at the proteomic level. These three genes were also downregulated after cell division ceased. The peroxisomal bifunctional enzyme, also annotated to this function, was not differentially expressed until after cessation of cell division. A 3-hydroxyacyl-CoA dehydrogenase gene with a mitochondrial location whose transcript was not differentially expressed, had reduced protein in nitrogen depletion.

Another major source of acetyl-CoA is pyruvate dehydrogenation. Genes annotated to the pyruvate dehydrogenase complex (PDC) were predicted to be localised to the chloroplast, mitochondria and cytoplasm. Only plastidial and mitochondrial complexes are expected and so genes annotated as cytoplasmic were grouped with mitochondrial genes. This is supported by similar regulation between the two groups. All of the mitochondrial genes annotated as PDC proteins are transcriptionally upregulated at some point during nitrogen depletion. Genes of the E1, E2 and E3 components were upregulated before and after cessation of cell division. Furthermore, proteins of the E2 and E3 component were also upregulated prior to cell division. Pyruvate dehydrogenase kinase (PDK) isoforms, involved in deactivation of PDC are also regulated throughout depletion, seemingly counter to the upregulation of PDC genes. Two of the genes are transcriptionally upregulated or non-differentially expressed prior to halting of cell division. A third was observed be upregulated in one study but downregulated in another prior to cessation of cell division and was upregulated afterwards. A fourth gene was not differentially expressed.

Conversely, plastidial PDC genes are transcriptionally downregulated during nitrogen depletion. An E3 component and an E2 component genes are downregulated before and after cessation of cell division. However, the same E3 component gene has increased protein shortly after cell division halts. Another E2 component gene is transcriptionally downregulated prior to cell division stopping. Genes involved in the other two sources of acetyl-CoA production, ATP citrate lyase and acetyl-CoA

synthetase were not differentially expressed except after the termination of cell division after which point some of the isoforms were downregulated (Levitan et al., 2015).

3.2.5.2 Regulation of fatty acid synthesis and plastidial desaturation

Malonyl-ACP is used in the elongation of fatty acids and its synthesis is catalysed by acetyl-CoA carboxylase and malonyl-CoA:ACP transacylase. Chloroplastidic acetyl-CoA carboxylase was downregulated or not differentially expressed early in depletion. The cytoplasmic form was downregulated at both early and late time points. Plastidial malonyl-CoA:ACP transacylase was transcriptomically downregulated before and after cell division halted but was upregulated at the protein level at 24h.

Fatty acid synthesis for use in the majority of glycerolipid synthesis is thought to occur in the chloroplast (Zulu et al., 2018). All chloroplastic genes were downregulated before and after cell division halted. Contrary to transcript regulation, a 3-oxoacyl-ACP synthases, catalysing the condensation of acyl-ACP with malonyl-ACP and a 3-oxoacyl-ACP reductase were upregulated at the protein level shortly after cell division ceased. As with malonyl:CoA ACP transacylase, this may indicate differential regulation of transcripts and proteins of some fatty acid enzymes. Contrary to chloroplast lipid synthesis, a mitochondrial 3-oxoacyl-ACP synthase was upregulated at the transcript level prior to the cessation of cell division. Another 3-oxoacyl-ACP reductase was downregulated or not differentially expressed while a third was increased at the protein level. An enoyl-ACP reductase was not differentially expressed. Four fatty acid desaturases are thought to be localised to the chloroplast in *P. tricornutum*. Of these, the soluble $\Delta 9$ desaturase, which acts mainly to desaturate 16:0-ACP to produce 16:1$\Delta 9$ is upregulated prior to the cessation of cell division. The following desaturase, involved in 16:2 production was downregulated before and after cell division ceased. The next desaturase, involved in 16:3 production was downregulated in one study prior to cell division halting (Huete-Ortega et al., unpublished).

3.2.5.3 Glycerolipid metabolism

Members of the Kennedy pathway, GPAT, LPAT and DGAT, catalyse lysophosphatidic acid, phosphatidic acid and triacylglycerol synthesis respectively. Of the genes thought to belong to the plastidial Kennedy pathway, none were observed to be upregulated early in depletion. One gene, for which it is unclear whether it acts as a GPAT or LPAT, Phatr3_J11916, was transcriptionally downregulated prior to termination of cell division. Another GPAT was upregulated after cell division halted (Levitan et al., 2015). DGAT2D was upregulated in one study before cell division ceased and was also upregulated afterwards. A plastidial PAP has not been identified to date.

Regarding galactolipid synthesis, three MGDG synthases have been detected in *P. tricornutum*, in addition to a SQDG synthase. Two of the MGDG synthase isoforms and the SQDG synthase are predicted to be localised to the cytoplasm. The location of MGDG synthesis in diatoms is still unclear (Kalisch et al., 2016). Despite this, the included galactolipid synthesis enzymes were all similarly downregulated at time points prior to the cessation of cell division in various studies. Of the remaining glycerolipid synthesis genes, four phosphatidate cytidyl transferases, that activate DAG for PE and cardiolpin synthesis and a phosphatidylglyerophosphatase involved in PG synthesis, were not differentially expressed in any of the studies. Neither were a phosphatidyl glycerol phosphate synthase, a phosphatidylinositol synthase and a DAG kinase. A gene annotated to the transfer of phosphoethanolamine or phosphocholine onto DAG was downregulated 4 and 8h (Matthijs et al., 2016) after the initiation of depletion. Glycerolipid degradation enzymes were transcriptionally unregulated early in depletion. Regulation varied at later timepoints however. Prior to the cessation of cell division, a mitochondrially targeted MAG lipase is not differentially expressed or upregulated, whereas afterwards it is downregulated. A TAG lipase, also putatively targeted to the mitochondria, was not differentially expressed prior to the halting of the cell cycle but was upregulated afterwards. A cytoplasmic MAG lipase was non-differentially expressed across all studies. Two phospholipase C (PLC) genes were transcriptionally oppositely regulated after cell division had ceased for 24h. In contrast to its transcript, the protein of one of these genes was upregulated shortly after division halted. A phospholipase D was also

observed to be transcriptionally upregulated in one early study but was otherwise non-differentially expressed.

3.2.5.4 Long chain polyunsaturated fatty acid synthesis

Two fatty acid thioesterases, expected to be involved in fatty acid transport over membranes, have been identified in *P. tricornutum*. Phatr3_J10454 was upregulated or non-differentially expressed early in depletion (Alipanah et al., 2015) but downregulated in the later stages. Phatr3_J33198 was upregulated after cell division ceased but was not differentially expressed in the early stages of depletion. Long chain acyl-CoA synthetase and PLA2, involved in acyl-CoA formation have been identified in *P. tricornutum*. Long chain acyl-CoA synthetases are predicted to be localised in the chloroplast, cytoplasm and mitochondria. The chloroplast isoform is non-differentially expressed at the transcript level. Those located in the cytoplasm are transcriptionally unregulated prior to cessation of cell division but Phatr3_J45510 was downregulated at the protein level at shortly afterwards and transcriptionally downregulated later still. Phatr3_J12420 is oppositely regulated at 48h in two studies. One of the mitochondrial genes was transcriptionally upregulated after division ceased but was otherwise non-differentially expressed. The other transcript was non-differentially expressed or downregulated prior to termination of division, as was its protein shortly afterwards.

In one half of the Lands cycle, hydrolysis of phospholipids liberates acyl chains into the cytoplasm which are subsequently incorporated into acyl-CoA by long chain acyl-CoA synthetases. In the other half, acyl-CoA:lysophospholipid acyltransferases incorporate acyl chains into lipids. Collectively, these processes enable the elongation of fatty acids and remodelling of membrane lipids. Two PLA2 isoforms have been identified, however, they are predicted to be localised to the plastid and mitochondria. The mitochondrial form was downregulated or non-differentially expressed prior to cessation of cell division, while the plastidial form was upregulated afterwards but non-differentially expressed beforehand. Two acyl-CoA:lysophosphatidylcholine acyltransferase (LPCAT) genes are downregulated prior to the halting of cell division. Synthesis of very long chain polyunsaturated fatty acid (VLC-PUFA) is thought to occur via the action of ER and plastid localised desaturases in combination with

cytoplasmic acyl-CoA elongases (Zulu et al., 2018). Four of the enzymes predicted to be involved in fatty elongation in the cytoplasm were either downregulated or unregulated. An ELO6 gene, thought to produce 20:3 and 20:4 and ELO5, expected to produce 22:4 and 22:5 were transcriptionally downregulated throughout depletion. Another ELO6 gene was downregulated at one early timepoint but was otherwise non-differentially expressed. A 3-oxoacyl-CoA synthase was non-differentially expressed, as was a putative mitochondrial enoyl-CoA reductase. A cytoplasmic enoyl-CoA reductase was non-differentially expressed early in depletion but was upregulated after division ceased. Domains within the 3-oxoacyl-CoA synthase and one of the mitochondrial enoyl-CoA reductase indicates a potential role in polyketide synthesis.

FAD2, thought to represent the start of VLC-PUFA synthesis, was downregulated or non-differentially expressed early in depletion. Unexpectedly, FAD2 is predicted to be localised to the mitochondria. PTD15 and PTD6 are thought to desaturate 18:2 and 18:3 to 18:3 and 18:4 respectively. PTD15 has a plastid targeting peptide and is downregulated prior to cessation of cell division. PTD6, is predicted to be localised to the cytoplasm and hence the ER membrane and is non-differentially expressed. A Δ17 desaturase, thought to produce 18:4 or 20:4, is upregulated or non-differentially expressed early in depletion but was downregulated after division halted. Two Δ5 desaturases, thought to catalyse 20:4 or 20:5 synthesis are both upregulated later in depletion, at 48h for Phatr3_J22459 and Phatr3_J46830, but were unregulated at earlier timepoints. Again unexpectedly, Phatr3_J22459 is predicted to be targeted to the mitochondria. The Δ19 desaturase, thought to catalyse 22:5 or 22:6 synthesis, is also unregulated early in depletion but is upregulated at 48h.

3.3 Discussion

3.3.0.1 Impacts of experimental design on acclimatisation to nitrogen depletion

Across the nitrogen depletion studies, the same physiological change occurred at varying time points. Cells acclimatised rapidly in a number of studies and more slowly in others. Varied experimental conditions can explain these observations, among them culture temperature and irradiance. Alterations in temperature change

the metabolic rate of organisms as described by the Boltzmann-Arrhenius function (Marañón et al., 2018) while different light levels result in varied energy capture and carbon fixation in photosynthetic organisms. However, in the compared studies with slower acclimatisation, temperature, at 15 and 22 °C, was within the optimal range for division of *P. tricornutum* (Bojko et al., 2013). Irradiance did vary over a range that would alter growth rate (Geider et al., 1985), however, a rapid acclimatisation response is seen in one study with low light (Abida et al., 2015) but a growth temperature of 19°C while a much slower response is seen in a second study with slightly higher light but a growth temperature of 15°C (Alipanah et al., 2015). Light levels and temperature have been observed to interact to alter growth rate in diatoms (Spilling et al., 2015) and may explain the discrepancies between these two studies. Differences in the manner of the transition between replete and deplete conditions, whether this is gradual or instantaneous, also differ between the more slowly acclimatising studies and the others. Studies with slower acclimatisation entered nitrogen depletion more gradually (Alipanah et al., 2015; Ge et al., 2014). However, similar duration of continued cell division has been observed with multiple initial nitrate concentrations (Yodsuwan et al., 2017).

It therefore appears that differences in temperature, irradiance or onset of depletion led to slower acclimatisation. Further investigation would be needed and may prove worthwhile to standardise experimental growth conditions, outline ways to modulate the timing of acclimatisation for experimentation and provide insight into diatom ecology. Furthermore, certain growth conditions impact the ultimate composition of the cell, as may varying initial cell states. For example, temperature and light alter the lipid content (Minhas et al., 2016) observed during nitrogen depletion. Interactions between physical conditions and nutrient depletion may cause changes in cell composition in microalgae. Whether this affects the quantity and quality of TAG accumulation is of biotechnological interest and therefore warrants further investigation.

3.3.0.2 Limitations of the systematic review of transcript and protein regulation

A number of limitations affect interpretation of the results of the systematic comparison. The studies made use of various methods to measure and process the raw data. The numbers of identified and differentially expressed transcripts and proteins may therefore have differed between studies. To minimise the impact that this has upon comparisons between studies, the proportions of up to down regulated genes were used. This assumes the proportion is representative of the actual biological changes. However, sampling time, significance threshold and number of identified transcripts and proteins may impact this ratio. Reprocessing the raw data and repeating the statistical analysis using the same methods would provide more certainty in the results. However, there would still be differences in the number of proteins and transcripts detected due to the various measurement techniques. Use of absolute numbers of regulated genes would only be accurate for studies that used the same technology.

Dividing the transcriptomic and proteomic responses with reference to the cessation of cell division is supported by the current studies. However, only one study took measurements a substantial time after cell division ceased. This division would therefore be better supported with further measurements carried out after cell division had halted. Additionally, given that only one proteomic study was included, inclusion of additional studies in the comparison would provide further information on the correspondence between transcript and protein levels and how protein levels alter throughout the acclimatisation process.

3.3.0.3 The role of fatty acid desaturation and elongation during nitrogen depletion

The plastidial $\Delta 9$ desaturase is transcriptionally upregulated during nitrogen depletion. This enzyme is thought to work preferentially on C16:0 rather than C18:0 (Dolch and Maréchal, 2015) to synthesise C16:1$\Delta 9$. The enzymes that catalyse the subsequent steps in plastidial desaturation are downregulated and collectively this supports increased accumulation of 16:1. Recent work on a thioesterase, Phatr3_J33198, supports a role for it in transport of fatty acids in or out of the chloroplast (Hao

et al., 2018). The authors favour a role for it in import of FAs but previous work illustrated the enzymes ability to act on acyl-ACP (Gong et al., 2011) with a preference for 18:1-ACP over 18:0-ACP and no observable activity on 16:1-ACP in *E. coli*. Were the thioesterase to act on acyl-ACPs then this may indicate that the fatty acids primarily transported from the chloroplast are 18:0 and 18:1. The majority of 16:1 is incorporated into chloroplast lipids (Abida et al., 2015) while evidence of activity converting 16:1 into 18:1 has not been reported in *P. tricornutum*. Increasing production of 16:1 may therefore reduce the amount of 18:0 synthesised and the quantity of fatty acid available for export by the thioesterase.

16:1 is the largest constituent of TAG during nitrogen depletion (Abida et al., 2015; Bower, unpublished). This, and the potentially reduced export of 16:1, would suggest that TAG synthesis is majorly dependant upon the chloroplast Kennedy pathway for DAG synthesis or that the final step of TAG synthesis is located in the chloroplast. The presence of C16 at the sn-2 position of TAG makes possible the use of prokaryotic DAG backbones (Abida et al., 2015). DGAT2D is the only DGAT observed to be transcriptionally upregulated at the transcript level (Levitan et al., 2015; Alipanah et al., 2015) during nitrogen depletion. Increased levels of its protein have not been observed however. Work in the Plant Metabolism group with fluorescently tagged DGAT2D indicates a plastidial location for the enzyme, as does its ASAFind targeting prediction. If DGAT2D is the major isoform utilised in depletion then increased retention of fatty acids in the chloroplast via conversion to 16:1 and decreased further desaturation could directly provide for its activity and the synthesis of TAG during nitrogen depletion, of which 16:1 is the major component (Abida et al., 2015). Synthesis of VLC-PUFAs requires export of fatty acids from the chloroplast, their subsequent elongation and desaturation and then import back into the chloroplast. In this scenario, an increased synthesis of 16:1 due to regulation of the $\Delta 9$ desaturase would serve to redirect fatty acids from VLC-PUFA synthesis into TAG accumulation.

Other mechanisms that impact VLC-PUFA synthesis also occur during nitrogen depletion. During the systematic review, a transcriptional decrease in expression of FAD2 was observed. FAD2 catalyses $\Delta 12$ desaturation of 18:1 (Domergue et al., 2003) that is thought to occur in the ER membrane as the first step of VLC-

PUFA synthesis (Zulu et al., 2018). Additionally, a Δ6 and a Δ5 elongase are transcriptionally downregulated across four nitrogen depletion studies while a further Δ6 elongase was downregulated in another. These enzymes catalyse the elongation of the fatty acid in VLC-PUFA synthesis. In order to provide the acyl-CoA substrate for elongation to occur, fatty acids are liberated from glycerolipids by PLA2 and incorporated into acyl-CoAs by long chain acyl-CoA synthetase (LACS). LACS may also play a role in incorporating free fatty acid (FFA) transported out of the chloroplast. Acyl-chains are then incorporated in glycerolipids via the action of LPCAT. The systematic review revealed that during nitrogen depletion, PLA2 and LPCAT were transcriptionally downregulated while the levels of several LACS transcripts, proteins or both were downregulated. These changes could underpin the decreased EPA synthesis required during nitrogen depletion and further decrease the pull of metabolites into VLC-PUFA synthesis from the chloroplast. It also highlights potential engineering strategies for increasing PUFA content in TAG, namely, increase the flux through the lands cycle, the rate of elongation and FAD2 and LACS activity. Increasing expression of the Δ5 elongase has already been successfully employed by Hamilton et al. (2014) to increase DHA content in TAG.

In summary, a collection of changes in gene expression observed during nitrogen depletion provide support for alterations to flux within fatty acid desaturation and elongation so that long chain unsaturated or monounsaturated fatty acid production is favoured over that of VLC-PUFAs. This would provide the fatty acids required for the increased proportion of 16:0 and 16:1 in the cell after acclimatisation to nitrogen depletion. Combined with knowledge of the fatty acid composition of TAG, location of DGAT2D and decreased MGDG content during nitrogen depletion, a cogent argument emerges for a significant role for the plastid in TAG accumulation during nitrogen depletion. Investigation of the roles of these enzymes is therefore warranted. Determination of the substrate preference of the plastidial Δ9 desaturase and further investigation of FA movement in the thioesterase knock out would be useful. More generally, the importance of research to develop an improved understanding of plastidial lipid import and export has been highlighted.

3.3.0.4 Insights into TAG accumulation in *P. tricornutum* during nitrogen depletion

Newly synthesised FAs play the major role in TAG accumulation, as noted in Section 1.3.2. Identifying what drives the incorporation of a large amount of these FAs into TAG is important in understanding acclimatisation to nitrogen depletion and for developing strategies for more productive TAG accumulation under optimal conditions. It is first important to consider whether an increase in FA synthesis occurs. The large accumulation of TAG and limited decrease in the amount of membrane acyl lipids would support increased FA synthesis as a redirection of FAs from membrane lipids would not be sufficient to support TAG accumulation. However, the cessation of cell division will reduce the need for synthesis of membrane glycerolipids and therefore reduce the overall demand for FAs. More *de novo* FAs will then be available for TAG synthesis and this could lead to TAG accumulation without increased FA synthesis. Studies of the flux into lipid synthesis have been carried out to investigate whether increased FA synthesis occurs. Two studies found that the amount of carbon in TAG and other lipids support at least modestly increased flux into FA synthesis (Levitan et al., 2015; Valenzuela et al., 2012). Guerra et al. (2013) instead identifies the greater decrease in flux into carbohydrate and protein synthesis as leading to the direction of an increased proportion of the cell's carbon to FA synthesis but the absolute flux into FA synthesis was reduced by half over three days of nitrogen depletion.

Together, consideration of whether FA synthesis increases and subsequently, what proportion of FAs are incorporated into TAGs compared to membrane lipids provides a framework to think about the mechanisms that lead to TAG accumulation during nitrogen depletion. This framework is depicted in Figure 3.3 and takes into account the halted cell division, accumulated TAG and moderately reduced membrane glycerolipid content. Briefly, if FA synthesis increases then the increased flux into lipid metabolism would not require substantially increased amounts of TAG synthesis enzymes to lead to TAG accumulation. Alternatively, if FA synthesis is decreased, TAG accumulation could be brought about by increased TAG synthesis and reduced membrane glycerolipid synthesis.

The overall flux of carbon into fatty acid synthesis is therefore important in

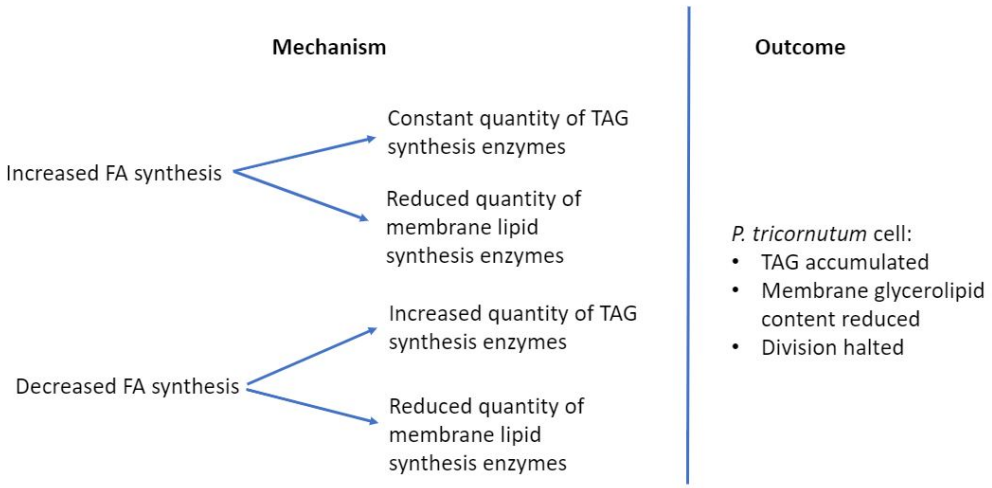


Figure 3.3 – Framework to aid understanding of TAG accumulation during nitrogen depletion. Nitrogen depletion causes cell division to halt, TAG to accumulate and the abundance of membrane glycerolipids to decrease moderately. Fatty acid synthesis may increase or decrease resulting in two suggested routes by which the changes to lipids in *P. tricornutum* could be brought about.

understanding TAG accumulation. Consideration of Guerra et al. (2013), Levitan et al. (2015) and Valenzuela et al. (2012) show it is still uncertain how flux alters. However, this information is important in differentiating between which of the two scenarios shown in Figure 3.3 is likely to be operating during nitrogen depletion. Considering an increase in flux into FA synthesis, generally, this could derive from 'push' or 'pull' factors as defined in metabolic control analysis and described by Cortassa et al. (2009). This means that the factors responsible for increasing flux into lipid metabolism would act upstream to push metabolites in or downstream to pull them in. Multiple mechanisms could be involved on the the push side.

Levitan et al. (2015) argues for flux of carbon into lipid metabolism resulting from degradation of proteins and redirection of flux from photosynthesis and competing pathways. This does find broader support across the transcriptomic studies from the review with the majority of differentially expressed genes downregulated in amino acid synthesis, ribosome and tRNA synthetase groups. However, gene groups involved in central and intermediate metabolism require more refined investigation to assess their contribution. Burrows et al. (2012) reports that around 40% of TAG accumulation is not derived *de novo*, that is to say, the carbon is supplied from internal stores as opposed to from photosynthesis and Ge et al. (2014) observed

that a knockdown of a 3-methylcrotonyl-CoA carboxylase subunit (MCC2) led to decreased TAG accumulation during nitrogen depletion. MCC2 is involved in leucine catabolism, mitochondrially located and its activity provides precursors for acetyl-CoA production. Recycling of carbon from other cell components into TAG is therefore well supported.

The upregulation of mitochondrial pyruvate dehydrogenase and mitochondrial β-oxidation identified in the systematic review could lead to increased acetyl-CoA. However, increased pyruvate dehydrogenase kinase and decreased plastidial pyruvate dehydrogenase may counter this effect. Caveats also exist for the β-oxidation genes. It is unclear how protein levels correspond to transcripts for many of the β-oxidation enzymes. Additionally similar enzymes are used in β-oxidation and BCAA degradation and it is not possible to separate their functionalities based on annotation alone.

Pull factors in contrast, would be observed within lipid metabolism. Metabolic engineering of lipid metabolism, through the overexpression of DGAT2D, DGAT2A, WS/DGAT and silencing of a fatty acid thioesterase (Niu et al., 2013; Dinamarca et al., 2017; Cui et al., 2018; Hao et al., 2018), has led to accumulation of TAG in replete conditions. The systematic review revealed a number of lipid metabolism genes that were upregulated during nitrogen depletion. Two fatty acid synthase enzymes, a 3-oxoacyl-ACP synthase and a 3-oxoacyl-ACP reductase, were both upregulated at the protein level. This is counter to regulation of their transcripts. Several fatty acid synthase genes were also transcriptionally upregulated when DGAT2D was overexpressed (Dinamarca et al., 2017). These include 3-oxoacyl-ACP synthases, 3-hydroxyacyl-ACP dehydratases and an enoyl-ACP reductase. This highlights the potential for fatty acid synthase genes to cause TAG accumulation both during nitrogen depletion and under replete conditions more generally. Dinamarca et al. (2017) also observed that two LPATs and pyruvate dehydrogenase genes were upregulated. Introduction of the DGAT2D therefore altered the expression of a number of other enzymes with potential roles in TAG accumulation. Other enzymes that are candidates for increasing metabolic pull were also identified from the systematic review: a malonyl-CoA:ACP transacylase, Phatr3_J37652 that was upregulated at the protein level but transcriptionally downregulated, while the

plastidial Δ9 desaturase, Phatr3_J9316, a fatty acid thioesterase, Phatr3_J10454, an LPAT, Phatr3_J45551 and a PDAT, Phatr3_J8860 were transcriptionally upregulated. However, similar to previous observations of the lack of impact on TAG levels of overexpressing ACC in diatoms (Courchesne et al., 2009), evidence of upregulation of ACC was not seen. There is therefore support for enzymes within lipid metabolism increasing the flux of carbon into into this area of metabolism by controlling the *de novo* synthesis of FAs, production of DAG backbones and final TAG synthesis step required for TAG accumulation.

The systematic review has therefore led to some valuable insight: increased production of acetyl-CoA from branched chain amino acid degradation, fatty acid degradation and pyruvate dehydrogenase could provide a push factor while an increased activity of FAS enzymes, an LPAT, DGAT2D, a PDAT or a Δ9 desaturase could provide pull factors for increasing flux through lipid metabolism. Further work to quantify the β-oxidation proteins during nitrogen depletion combined with measurement of flux into acetyl-CoA would provide insight into the contribution of mechanisms that generate acetyl-CoA to TAG accumulation during nitrogen depletion. Quantification of the rate of FA synthesis would supplement this, providing evidence for the connection between flux into acetyl-CoA synthesis and FA synthesis during nitrogen depletion. The LPAT, DGAT2D, PDAT and Δ9 desaturase are also highlighted as genes of interest with regard to TAG accumulation. Additionally, given the known importance of BCAA degradation in the nitrogen depletion response and the role these enzymes may play in degrading fatty acids released from potentially catabolised chloroplast membrane lipids, it would be useful to differentiate their functions.

The framework outlined in Figure 3.3 also suggests the relative direction of fatty acids into either TAG or membrane lipids is important in TAG accumulation, regardless of how flux into metabolism is increased. Indeed, the regulation of TAG synthesis enzymes and those involved in membrane lipid biosynthesis (Figure 3.2) demonstrated regulation that could lead to redirection of FAs between the two pools. Although many of the TAG synthesis enzymes are unregulated or downregulated, both PDAT and DGAT2D are upregulated in some of the studies. Of the membrane lipid synthesis enzymes, the MGDG and SQDG synthase enzymes are downregulated across

the studies. Other enzymes such as phosphatidate cytidyl transferase that synthesises CDP-DAG for PI and PG synthesis, and several membrane lipid synthetases are unregulated. However, it would be useful to try to identify more of these enzymes to ensure the analysis is comprehensive. Another way to favour TAG synthesis over membrane lipids is to prevent fatty acids from reaching the enzymes of membrane lipid synthesis. Indeed, ways in which this might be brought about were suggested in Section 3.3.0.4: increased synthesis of 16:1 may prevent fatty acid export from the chloroplast. Altering fatty acids such that they do not fit the substrate specificity of membrane lipid synthesis enzymes could also decrease membrane lipid synthesis. Again, increased 16:1 content, of which the majority is found in chloroplast lipids (Abida et al., 2015), may not suit the specificity of extraplastidial enzymes. As an engineering strategy, decreasing the demand of membrane lipids for fatty acids as a mechanism to increase TAG accumulation while maintaining optimal growth is likely to be challenging but is highlighted here as an option.

3.3.0.5 Regarding subcellular location of proteins

A number of genes in lipid metabolism had unexpected predictions for subcellular location. These include long chain acyl-CoA elongation, a PDAT, a DGAT and desaturases in the mitochondria and phospholipase A2s in the chloroplast and mitochondria. Knowledge of the location of these genes and others is essential for understanding lipid metabolism within diatoms. Currently, protein localisation is determined on a gene by gene basis. Were the plastid proteome to be determined, this would provide a useful resource in furthering understanding of lipid metabolism in *P. tricornutum*.

3.3.0.6 β-oxidation during nitrogen depletion

The systematic review of nitrogen depletion studies revealed that two sets of beta oxidation enzymes, with predicted locations of peroxisome and mitochondria were regulated in different ways. Many of the mitochondrial enzymes had increased transcript levels and differential expression of proteins was also seen. Fewer beta oxidation enzymes were identified in the peroxisome but these were almost entirely NDE. Algal beta oxidation is thought to occur in the mitochondria or peroxisome

dependent upon phylogenetic origin (Kong et al., 2018). It is not known where diatom beta oxidation is located but it has been observed to be located in the mitochondria of an ochrophyte, *Bumilleriopsis* (Winkler et al., 1988). Upregulation of beta oxidation enzymes in the mitochondria could indicate a role for FA catabolism in underpinning decreased chloroplast lipid quantity during nitrogen depletion. This would support a mitochondrial location of beta oxidation in diatoms. However, some of these enzymes may be involved in BCAA degradation and so it will be necessary to separate these processes before conclusions can be drawn.

3.3.0.7 Autophagy during nitrogen depletion

Nitrogen depletion also offers an interesting model for investigating the role of autophagy in lipid catabolism. Autophagy has been observed to play a role in TAG catabolism and it has been hypothesised that it may play a role in TAG accumulation from degradation of membrane lipids under nitrogen depletion in *Chlorella zofingiensis* (Zhang et al., 2018). Autophagy of chloroplasts has been observed under a variety of conditions in plants and implicated in nitrogen remobilisation (Ishida et al., 2014; Izumi et al., 2017). Transcriptional regulation of autophagy genes identified during the systematic review suggest upregulation of autophagy during nitrogen depletion, at a similar time to which TAG begins to accumulate, highlighting the potential role for autophagy in recycling of chloroplast membranes to TAG and acclimatisation to nitrogen depletion more widely.

4

Investigating the functions of the DGATs and PDAT of *Phaeodactylum tricornutum*

4.1 Introduction

DGATs and PDATs are key TAG synthesis enzymes, catalysing the addition of the final acyl chain to DAG. Seven DGATs and a PDAT have previously been identified in *P. tricornutum* and these are detailed in Table 1.3. Functionality has been demonstrated for multiple of these DGATs. During this study I also identified another gene with a PDAT/LCAT domain as described in Chapter 1.3.3. How these enzymes are functionally differentiated and the roles of the multiple TAG synthesis isoforms in *P. tricornutum* is not known. Characterised DGAT and PDAT isoforms have demonstrated differentiated functions and substrate specificities within and between species, noted in Section 1.3.3. Therefore, understanding the differentiation of the isoforms in *P. tricornutum* may provide insight into the physiology of this diatom. Of particular focus in this regard are the DGAT2s that occur in multiple copies in alga but individually in other groups (Chen and Smith, 2012). Additionally, DGAT or PDAT isoforms with desirable properties may be of use in biotechnology, for example, in developing strains that produce oils with a tailored FA content. Investigating the roles of these enzymes is therefore an important step in understanding TAG accumulation in *P. tricornutum*.

To investigate the function of these enzymes it was necessary to measure the lipids in *P. tricornutum*. TAG, FA and DAG molecules are therefore mentioned frequently in this Chapter and notation is needed to discuss them efficiently. A TAG species is denoted by a T, followed by the total number of carbons in its three constituent acyl chains. The TAG species mentioned in this thesis are T44, T46, T48, T50, T52, T54 and T56. The number of double bonds in a TAG species is denoted by the number after the colon e.g T48:1. The notation and naming of FAs was introduced in Chapter 1.2 and I primarily use the shorthand e.g 20:5. This resembles the TAG notation but is distinguishable in that the number of carbons never exceeds 24 e.g 24:0. The FA constituents of TAG and DAG species are introduced using the FA/FA/FA notation for TAGs and FA/FA for DAGs. FA refers to the FA present. When the number of double bonds is not specified then this refers to all FAs of that length. The sn position of the acyl chain within the molecules is not indicated by this notation unless explicitly stated. In DAGs FAs are located at the sn-1 and sn-2

positions, shown in Figure 1.7.

There is a broad diversity of TAG molecules in *P. tricornutum* that are composed of different combinations of acyl chains. Yu et al. (2009) previously measured the TAG molecules in *P. tricornutum* during nitrogen depletion and identified the FA composition of the major species. These are detailed in Table 4.1 and provide a resource to deduce the likely FA composition of TAG molecules measured in the current study. All except three of the TAG species compositions were derived from the study. I have suggested the likely FA composition of these three species, T50:1, T50:2 and T50:3, based on potential DAG species present in major TAGs and overall chain length and desaturation restrictions.

Table 4.1 – FA composition of major TAG species in *P. tricornutum* during nitrogen depletion. All species except T50:1, T50:2 and T50:3 were compiled from Yu et al. (2009). Compositions of T50:1, T50:2 and T50:3 were deduced based on potential DAG species in major TAG species and total carbon and saturation limitations inherent to the species.

TAG	FA composition	TAG	FA composition	TAG	FA composition
46:1	14:0/16:0/16:1	50:1	16:1/16:0/18:0[h]	50:5	16:1/16:1/18:3
46:2	14:0/16:1/16:1	50:1	16:0/16:0/18:1[h]	50:6	14:0/16:1/20:5
48:0	16:0/16:0/16:0	50:2	16:1/16:1/18:0[h]	50:6	16:1/16:1/18:4
48:1	16:0/16:0/16:1	50:2	16:1/16:0/18:1[h]	52:5	16:0/16:0/20:5
48:2	16:0/16:1/16:1	50:3	16:1/16:1/18:1[h]	52:5	16:0/16:1/20:4
48:3	16:1/16:1/16:1	50:4	16:0/16:0/18:4	52:6	16:1/16:0/20:5
48:4	16:0/16:1/16:3	50:4	16:0/16:1/18:3	52:7	16:1/16:1/20:5
48:4	14:0/16:0/18:4	50:4	16:1/16:1/18:2		
48:4	14:0/16:1/18:3	50:5	16:0/16:1/18:4		
48:4	16:1/16:1/16:2	50:5	14:0/16:0/20:5		

h = probable composition

Several lipid measurement methods were used in the work described in this chapter and TAG measurements made up the majority of these. One of the methods used to measure TAG was gas chromatography-flame ionisation detection (GC-FID) of total lipid extracts. Gas chromatography is commonly used to separate analytes based upon their retention time in a column. The column contains a carrier gas in which the analyte travels surrounded a liquid stationary phase that coats the surfaces within the column. The time taken for analytes to move through the column depends upon the proportion of time they partition into the carrier gas compared

to the stationary phase. The proportion of time spent in each phase depends upon the chemical characteristics of the analyte and stationary phase and the length of the column. The time spent in the stationary phase can be reduced by increasing the column temperature. Increasing temperatures are therefore used to elute more strongly retained analytes. If mobile phase flow rate, column length and temperature are kept constant then specific chemicals will elute at known times, called retention times. The analyte represented by a specific retention time can be identified by measurement of known standards.

Various detectors can be used depending upon the analyte to be detected. The work in this chapter made use of the flame ionisation detector (FID). The GC column is directly coupled to the FID such that eluted analytes are burnt in the flame of the detector. The burning produces ions in a manner proportional to the number of carbons in the molecule. This is known as the unit carbon response. The number of ions generated are therefore proportional to the quantity of analyte. These ions collide with a detector and produce a current that is also proportional to the quantity of analyte. An analyte is eluted over a discrete period of time, generating a signal trace as it is detected. Integrating the area under this trace provides the total signal of eluted analyte. A known amount of a standard analyte is also measured. Due to the unit carbon response, signal from this standard can be used to quantify the analyte. It should be noted that the presence of oxygen, nitrogen and sulphur reduce the unit carbon response and it is important to use appropriate standards when measuring analytes that contain these elements. GC-FID results presented in this chapter were quantified using a C15:0 standard when measured by Dr Chen and Dr Litvinenko and trimyristate (C14:0/C14:0/C14:0) in the measurements I made. The protocol used for total lipid extractions and GC-FID is described from Section 2.7.2 onwards. This method enabled the discrimination of TAG species with differing total numbers of carbon in their constituent acyl chains but does not enable discrimination based upon the desaturation of acyl chains. An example GC-FID trace of TAG from *P. tricornutum* is shown in Figure 4.1 and highlights the different lipids that are eluted. Five TAG species can be identified with acyl chains containing a total of 46, 48, 50, 52 or 54 carbons.

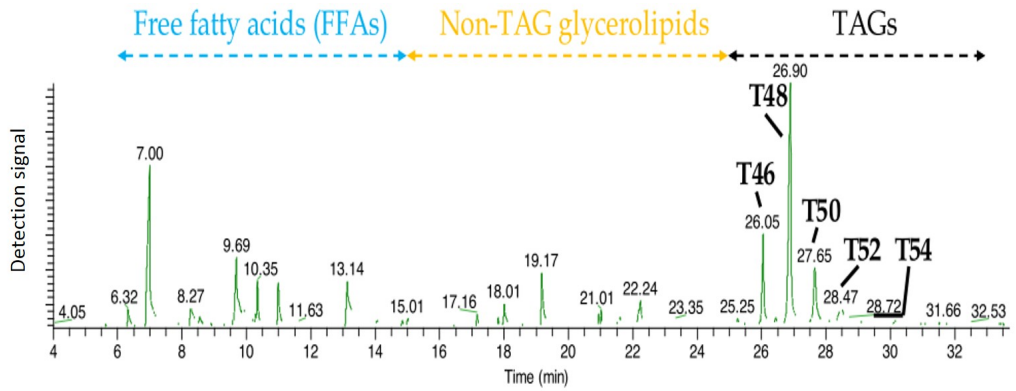


Figure 4.1 – Example GC-FID trace of total lipid extracts from *P. tricornutum*. Free FAs, non-TAG glycerolipids and TAGs were separated using a biodiesel column. Representative TAG peaks are labelled as T46-T54 whereby the number signifies the number of carbons in the three acyl chains. Figure reproduced from Litvinenko (2016).

The TAG in *P. tricornutum* cells is composed of a complex mixture of species that differ in acyl chain constituents and the position of these on the glycerol backbone. Depending on the technique used, mass spectrometry can identify the FA constituents of TAG. A number of TAG species have the same mass when ionised and are termed isobaric. This prevents their separation when measured by certain types of mass spectrometry. However, a type of tandem mass spectrometry called ESI-TQMS is able to measure TAG species and has improved segregation of isobaric species compared to use of a single mass spectrometer. This method was employed to measure TAG in total lipid extracts as described in Chapter 2. ESI-TQMS measurements were carried out by Dr Richard Haslam at Rothamsted Research, using total lipid samples that either myself or Dr Litvinenko prepared.

A schematic of ESI-TQMS is depicted in Figure 4.2. Three quadrupoles are connected in series, termed Q1, q2 and Q3. The first and third quadrupoles (Q1 and Q3 respectively) carry out separation of ions by mass/charge ratio while the second quadrupole (q2) fragments ions passed from Q1. A neutral loss scanning protocol was used in this work as described by Krank et al. (2007). This involved sequential selection of specific mass/charge ratios in Q1 that pass to q2. In q2, the TAG ion is fragmented to release a neutral acyl chain, ammonia and a diacylglycerol ion that pass to Q3. The acyl chain can be lost from any of the sn-positions although loss from sn-1 or sn-3 is more likely (Byrdwell and Neff, 2002). Q3 sequentially scans through mass/charge ratios, detecting DAG ions. By subtracting the mass

of the detected DAG ions from the parent TAG ion, the lost acyl chains can be identified, including the number of double bonds present. This information enables discrimination of isobaric species, providing information on the total number of carbon atoms and double bonds present in the TAG species. Furthermore, the proportional FA composition of the TAG fraction as a whole can be calculated. However, the protocol does not provide information on the position of the acyl chains on the glycerol backbone. A TAG internal standard was used to enable quantification of the TAG species. In a similar manner, polar lipids were also identified and quantified using ESI-TQMS. This was carried out by neutral loss or precursor scanning depending on the lipid class, as described by Shiva et al. (2013). The process of selection in Q1 followed by fragmentation in q2 and detection of ion fragments that pass through Q3 is conceptually similar to the process for TAG measurements. Lipid class, total acyl chain length and total number of carbon-carbon

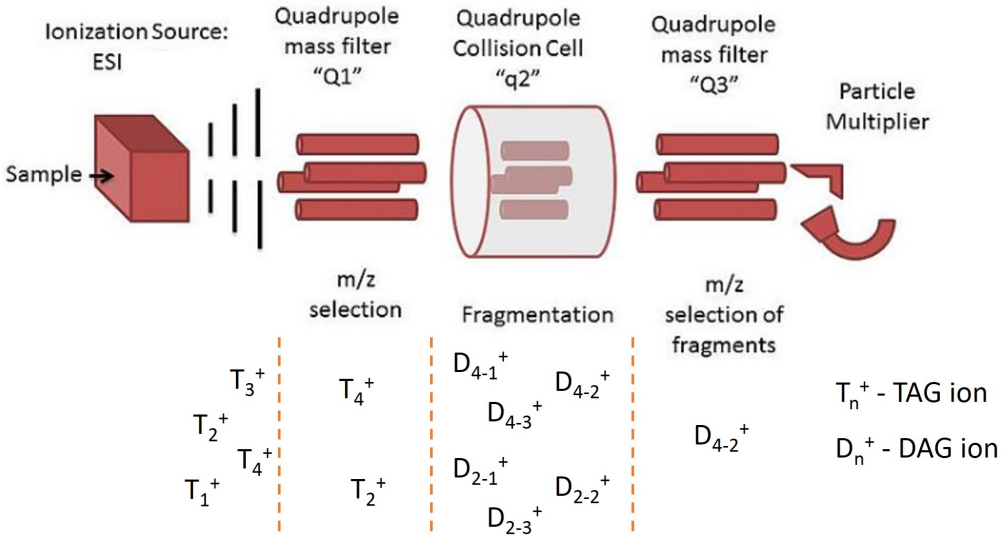


Figure 4.2 – Schematic of electrospray ionisation-triple quadrupole mass spectrometry (ESI-TQMS) illustrating the neutral loss scan protocol employed for TAG identification. The total lipid sample is ionised by electrospray ionisation giving rise to multiple TAG ions. TAG ions with a specific mass/charge (m/z) ratio are selected in the first mass spectrometer quadrupole, Q1 (for example, T_4^+ and T_2^+ have the same m/z ratio). These TAG species are fragmented in quadrupole q2 to generate DAG ions. DAG ions with a specific mass/charge ratio are selected in the second mass spectrometer, Q3 (for example, $D_{4\text{-}2}^+$), and is detected by a particle multiplier. Subtracting the mass of the DAG ion from the TAG ion selected in Q1 enables identification of the lost acyl chain. Q1 and Q2 scan through different mass/charge ratios to identify the DAG and acyl chain species that originate from each TAG ion.

double bonds can be identified in this way. However, unlike for TAGs, FA composition of the polar lipid classes is not available using this technique.

Investigation of the seven DGATs and the previously identified PDAT from *P. tricornutum* was previously begun by Dr Jit Ern Chen and Dr Alex Litvinenko, former PhD students of the Plant Metabolism Group (Chen, 2013; Litvinenko, 2016). Except where noted, the following work detailed in this introduction originates from their theses and is provided as background information.

To assess the functionality of the five PtDGAT2s, Dr Chen and Dr Litvinenko used the TAG synthesis deficient *S. cerevisiae* strain, H1246, in which all four TAG synthesis genes were deleted (Sandager et al., 2002). The five PtDGAT2s were cloned into the pYES2.1/V5-His-TOPO vector from the TOPO TA yeast expression system. The *S. cerevisiae* DGAT2 gene, commonly referred to as DGA1, was also cloned into the vector as a positive control. The GAL1 promoter was used in the constructs and drove expression of the genes. Thus, the presence of galactose induced expression in the absence of glucose (Johnston et al., 1994). The stop codons were removed from the genes to enable their expression from the plasmid as C-terminal V5-6×His tag fusion proteins. α-V5 or α-6×His antibodies could then be used in Western blots of these proteins. Six plasmids were thus cloned and transformed into *S. cerevisiae*, producing six strains that are detailed in Table 4.2. Furthermore, BMA64-1A, which is the parent strain of H1246 was also used in some experiments. BMA64-1A expressed ScDGAT2 from the native promoter.

Table 4.2 – DGAT2 yeast expression plasmids and resulting strains. ScDGAT2 and PtDGAT2A-E were incorporated into the pYES2.1/V5-His-TOPO vector and transformed into H1246. Figure reproduced from Litvinenko (2016).

Parental strain	Plasmid in derived strain	Name of strain
H1246	pYES2.1+ScDGAT2	Sc2
H1246	pYES2.1+PtDGAT2A	Pt2A
H1246	pYES2.1+PtDGAT2B	Pt2B
H1246	pYES2.1+PtDGAT2C	Pt2C
H1246	pYES2.1+PtDGAT2D	Pt2D
H1246	pYES2.1+PtDGAT2E	Pt2E

Appropriate conditions were established for induction of expression. Strains Sc2, Pt2A, Pt2B, Pt2C, Pt2D and Pt2E were grown under these conditions and their proteins were separated by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) and the DGAT2 protein bands were visualised using α-V5-HRP conjugate antibodies that bound to the V5 epitope tag. Western blot bands were detected for Sc2, Pt2A, Pt2B and Pt2D as depicted in Figure 4.3 and indicate that these proteins were expressed. The control strains BMA64-1A, Sc2 and the experimental strains Pt2A, Pt2B and Pt2D were cultured under the expression induction conditions for 6, 9 or 12 days to identify a time point with substantial TAG accumulation. TAG was extracted and quantified using GC-FID. The GC-FID measurements differentiate TAG species by differing total numbers of carbon atoms and this information is shown in Figure 4.4A. PtDGAT2B and PtDGAT2D were found to restore TAG synthesis to the TAG-less yeast, supporting their function as DGATs. PtDGAT2A did not restore TAG synthesis however.

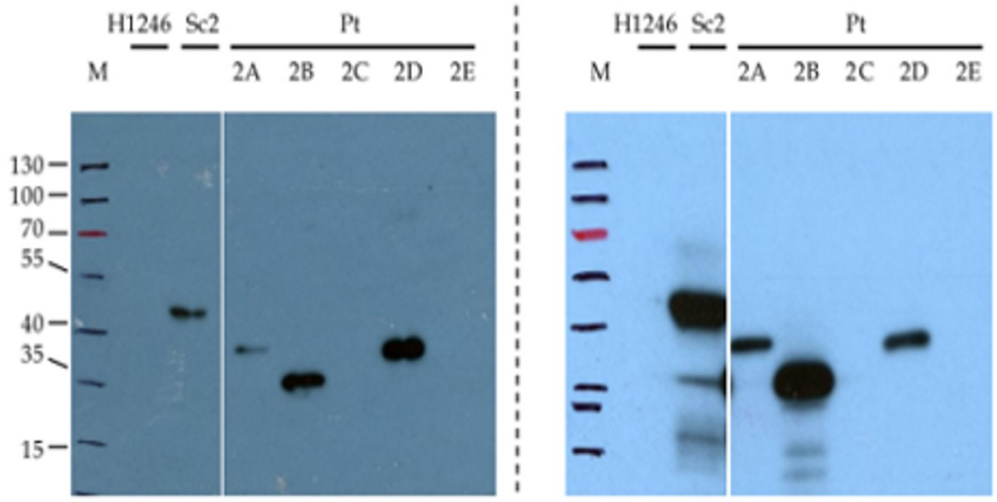


Figure 4.3 – Western blots of H1246 *S. cerevisiae* expressing ScDGAT2 and PtDGAT2A-E. Each figure originated from independent induction experiments. M- molecular weight ladder in kDa, H1246- TAG deficient *S. cerevisiae*, Sc2 - ScDGAT2, 2A to 2E - PtDGAT2A to PtDGAT2E. Proteins were bound by the α-V5-HRP conjugate antibody. Figure reproduced from Litvinenko (2016).

Strains Pt2B and Sc2 were recultured under the induction conditions and their lipids were extracted and quantified using GC-FID. The TAG species composition of PtDGAT2B expressing line was compared to ScDGAT2. These results are displayed in Figure 4.4B. T50 carbon constituted a lower proportion of TAG in PtDGAT2B

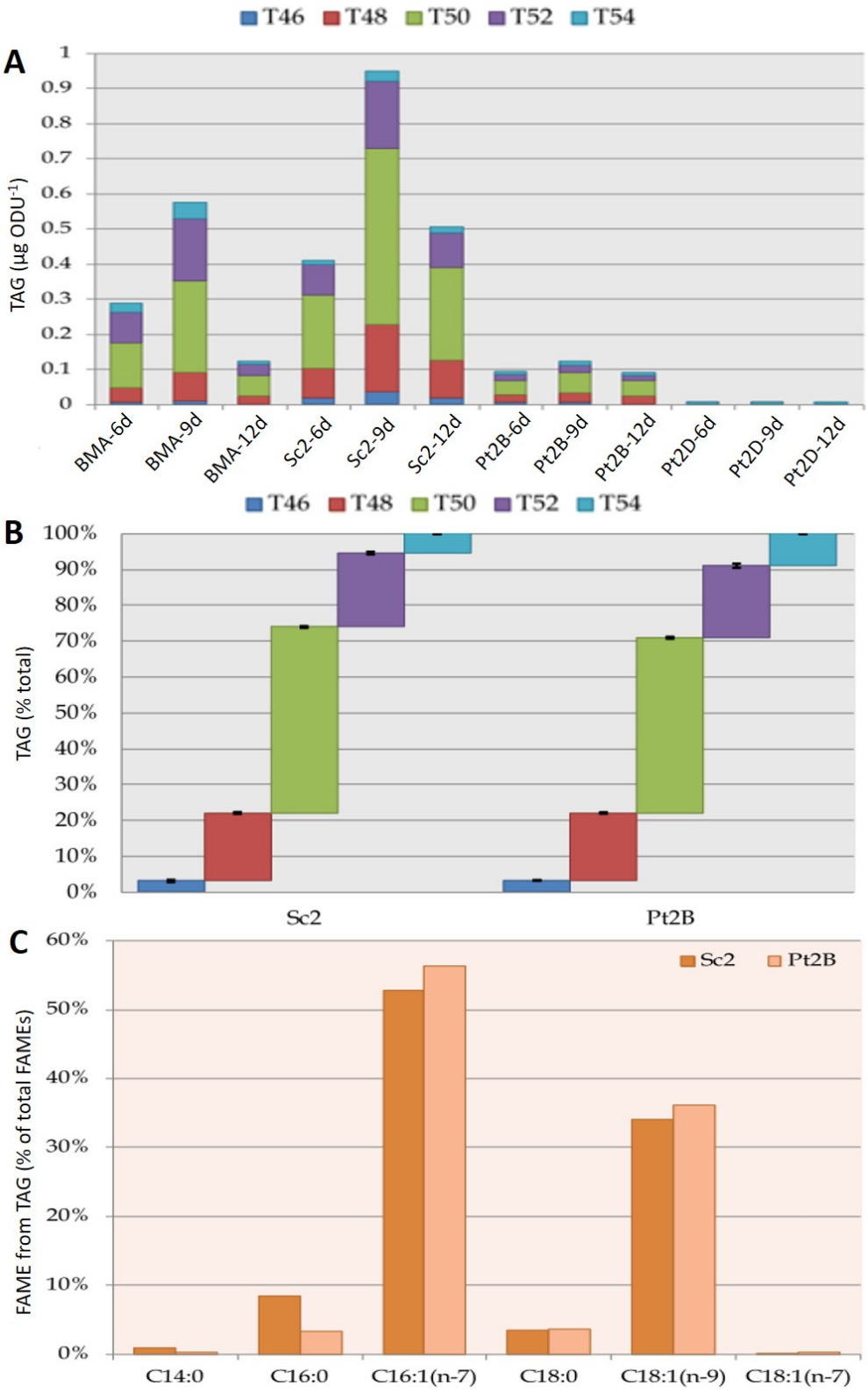


Figure 4.4 – Quantification of TAG in TAG synthesis deficient yeast complemented with PtDGATs. In independent experiments, (A) TAG was quantified by GC-FID at 6, 9 and 12 days after induction, (B) the proportional composition of TAG was calculated and (C) FAME composition of TAGs was calculated. BMA- wt *S. cerevisiae*, Sc2- ScDGAT2, Pt2B- PtDGAT2B, Pt2D- PtDGAT2D, ODU- standardised optical density unit. Reproduced from Litvinenko (2016)

expressing lines while T54 constituted a larger proportion. In *S. cerevisiae*, T50 is usually composed of 16/16/18 and T54 of 18/18/18. The results therefore support DGAT2B preferring to incorporate 18 carbon FAs from either 18-CoA or 18/18 DAGs compared to ScDGAT2. To measure the FA composition of the TAG fraction, thin layer chromatography was used to separate TAG from other lipid classes in lipid extracts of the Pt2B strain. The constituent FAs were derivatised to fatty acid methyl ester (FAME)s and quantified by GC-FID. The procedures used to carry out this analysis are detailed by Litvinenko (2016), and the results are displayed in Figure 4.4C. Compared to ScDGAT2B, the proportion of TAG composed of 14:0 and 16:0 was lower in PtDGAT2B while the proportion of 16:1 and 18:1 was higher. This supports PtDGAT2B preferring to incorporate longer, monounsaturated FAs compared to the preference of ScDGAT2 for shorter, saturated FAs.

Dr Litvinenko hypothesised that expression of DGAT2C and DGAT2E were not detected in Pt2C and Pt2E strains as codon usage of the genes was not optimised. He further suggested that incorrect localisation of DGAT2A or a substrate specificity not suited to the FA substrate pool of yeast may have prevented TAG synthesis in the Pt2A strain. To overcome these obstacles and gain information on the *P. tricornutum* TAG synthesis genes in the native environment, Dr Litvinenko transformed the PtDGAT2 genes into *P. tricornutum* to lead to increased enzyme expression. In addition, he also transformed *P. tricornutum* with PtDGAT1, PtWS/DGAT and one of two PtPDATs. Oligonucleotide primers were designed to amplify these genes from the Phatr2 gene models and are available in the thesis of Dr Livinenko (Litvinenko, 2016). Notably, the gene model used for DGAT1 was that described by Guihéneuf et al. (2011), which extends the Phatr2 gene model by 33bp to include a stop codon. Furthermore, the gene model used for PDAT lacks 25 amino acids from the N terminus of the Phatr3 protein model, Phatr3_J8860, as recorded in Table 4.3, which presents the produced strains and associated information. This resulted from the need to estimate the start of the coding region prior to the release of the Phatr3 gene models. All PDAT constructs in this Chapter made use of this truncated sequence. Bioinformatic analysis indicated that both the full gene model and the model lacking 25 amino acids contained a chloroplast transit peptide but not a recognised signal peptide. The subcellular targeting of PDAT is therefore ambiguous and as a result,

it is not clear whether the truncation impacts targeting of the protein. DGAT1 undergoes alternative splicing. Retention of its stop codon containing intron produces an enzyme without DGAT functionality (Guihéneuf et al., 2011). DGAT2A and PDAT also have introns that contain stop codons. To test whether the presence of introns resulted in functional impacts, alternative versions of these three genes were cloned into constructs, either containing or lacking the introns.

Table 4.3 – The DGAT and PDAT overexpression strains and the plasmids they contain.

Short hand	Strain name	Overexpressed gene	Promoter	Introns present	Number of clonal lines
pEV1	pEV1	NA			14
Pi*	PDATi	PDAT	inducible		14
Pfi*	PDATfi	PDAT	inducible	Yes	17
1i	DGAT1i	DGAT1	inducible		22
1fi	DGAT1fi	DGAT1	inducible	Yes	23
2Ai	DGAT2Ai	DGAT2A	inducible		24
2Afi	DGAT2Afi	DGAT2A	inducible	Yes	17
2Bc	DGAT2Bc	DGAT2B	constitutive		6
2Bn	DGAT2Bn	DGAT2B	native		12
2Bi	DGAT2Bi	DGAT2B	inducible		11
2Ci	DGAT2Ci	DGAT2C	inducible		45
2Di	DGAT2Di	DGAT2D	inducible		21
2Ei	DGAT2Ei	DGAT2E	inducible		22
Wi	WS/DGATi	WS/DGAT	inducible		14

i = inducible, c = constitutive, n = native

f = gene with introns

*Lacks the first 25 amino acids

Several sets of promoters and terminators were chosen to control expression of these TAG synthesising genes of interest. The promoters and terminators of fucoxanthin-chlorophyll a/c-binding protein (FCP) genes were chosen to provide constitutive expression. These promoters have previously been characterised and used for expression of transgenes in *P. tricornutum* (Siaut et al., 2007; Radakovits et al., 2011). The FCPB promoter and FCPE terminator were the versions that were used. It was noted that constitutive expression may lead to slow culture growth and thereby cause issues such as transgene loss. Therefore, an inducible promoter was also selected to drive expression. The ammonia transporter 1 (AMT1) promoter was selected due to its low expression under nitrogen replete conditions and high expression under nitrogen deplete conditions (Litvinenko, 2016). In this way, the

inducible promoter would lead to high expression of the genes during high levels of TAG accumulation, when substrate would be available. In addition to the constitutive and inducible promoters and terminators, the native DGAT2B versions were also used for the expression of this gene. This enabled evaluation of the level of expression from the native promoter compared to the constitutive and inducible promoter.

The genes were cloned into a custom vector. The vector contained the *Streptoalloteichus hindustanus* (Sh) ble selectable marker that provides resistanace to zeocin and phleomycin. Sh ble expression was controlled by the FCPC promoter and terminator. The vector backbone was derived from pMS2-RBCS2-Thi4-5'UTR-Ble, which is a plasmid used in the Plant Metabolism group. The finalised empty vector was termed pEV. The genes were cloned into this plasmid. For reference, this vector was called pPT1 in Dr Litvinenko's thesis and has been renamed here for clarity. In summary, constructs varied in whether they contained introns and on the type of promoter and terminator used in expression of the genes. Plasmid names were derived by adding a suffix to the gene name that indicated the promoter-terminator combination: i for inducible, c for constitutive and in the caase of DGAT2B, n for native. In addition, a suffix of f, for full gene, was given to indicate that the gene contained introns. For example PDAT with AMT1 promoter and terminator and introns is designated PDATfi. Shorthands will be used in this chapter, particularly in figures, and consist of the number for DGATs, a P for PDAT and an f or i as required e.g 2Bi for DGAT2Bi or Pfi for PDATfi. This is detailed in Table 4.3.

Constructed plasmids were assessed using restriction digestion and sequencing. A number of single base changes were noted in some of the gene coding regions. Multiple independent amplifications indicated that these were likely polymorphisms however and therefore amplicons containing them were still used in constructing the final constructs. The plasmids were then transformed into *P. tricornutum* Pt1 using electroporation as outlined in Section 2.4.7. Transformation occured through random integration into the genome. The presence of transgenes was tested using PCR after three rounds of subculturing to ensure stability. The numbers of stable transformants, their strain names and the plasmids they contain are detailed in Table 4.3. Each stable transformant was dubbed a line of a strain and is referred to by a number separated by a dash from the strain name, e.g DGAT2Bi-3. The use of the

strain name without a following number refers to the strain.

The DGAT2B overexpressing lines, including DGAT2Bn, DGAT2Bc and DGAT2Bi, were investigated first. The majority of experiments on all overexpressing lines involved culturing the cells without nitrogen. TAG was quantified 8 days after transfer to nitrogen deplete media. The methods used for cell culture, lipid extraction and quantification are presented in Section 2.7.2 and are the same that I used in the work presented in Section 4.2. TAG quantity was normalised to the amount present in 1ml of a culture that has an optical density of 1, referred to as an optical density unit (ODU). In one experiment shown in Figure 4.5A, lines were split and cultured with and without nitrogen. Measurements were only taken for one biological replicate per line. However, the results indicated that under replete conditions, overexpression of DGAT2B increased TAG accumulation relative to the controls. Despite the use of different promoters, the amount of TAG accumulated is similar between lines, except for DGAT2Bc-5. A similar increase in TAG content was not observed for nitrogen deplete cultures. The quantity of TAG was measured for deplete cultures in a second experiment with between 1 and 5 biological repeats and is displayed in Figure 4.5B. Although substantial variation in TAG quantity were observed, there were not significant differences in TAG quantity between control and overexpressing lines.

In addition to TAG quantity, the species composition of TAG was calculated. In *P. tricornutum* T52, T54 and T56 species typically contain EPA and T54 and T56 species are minor. For these reasons data on these species is presented collectively as T52-T56. Preliminary work on DGAT2B overexpressing lines indicated that those lines using the native and constitutive promoters did not contain a different TAG composition to control cultures, in contrast to those lines that used the inducible promoter. DGAT2Bi lines were grown under nitrogen deplete conditions and the TAG composition obtained from 3 or 4 biological replicates is shown in Figure 4.5C. All six of the measured lines contained an increased proportion of T50 in their TAGs. DGAT2Bi-3 contained a smaller proportion of T46 while DGAT2Bi-9 contained a smaller proportion of T48 and larger proportion of species from T52-56 species. One explanation for only registering differences for inducible constructs is that these may lead to higher expression of the DGAT or PDAT. Although overexpression of the

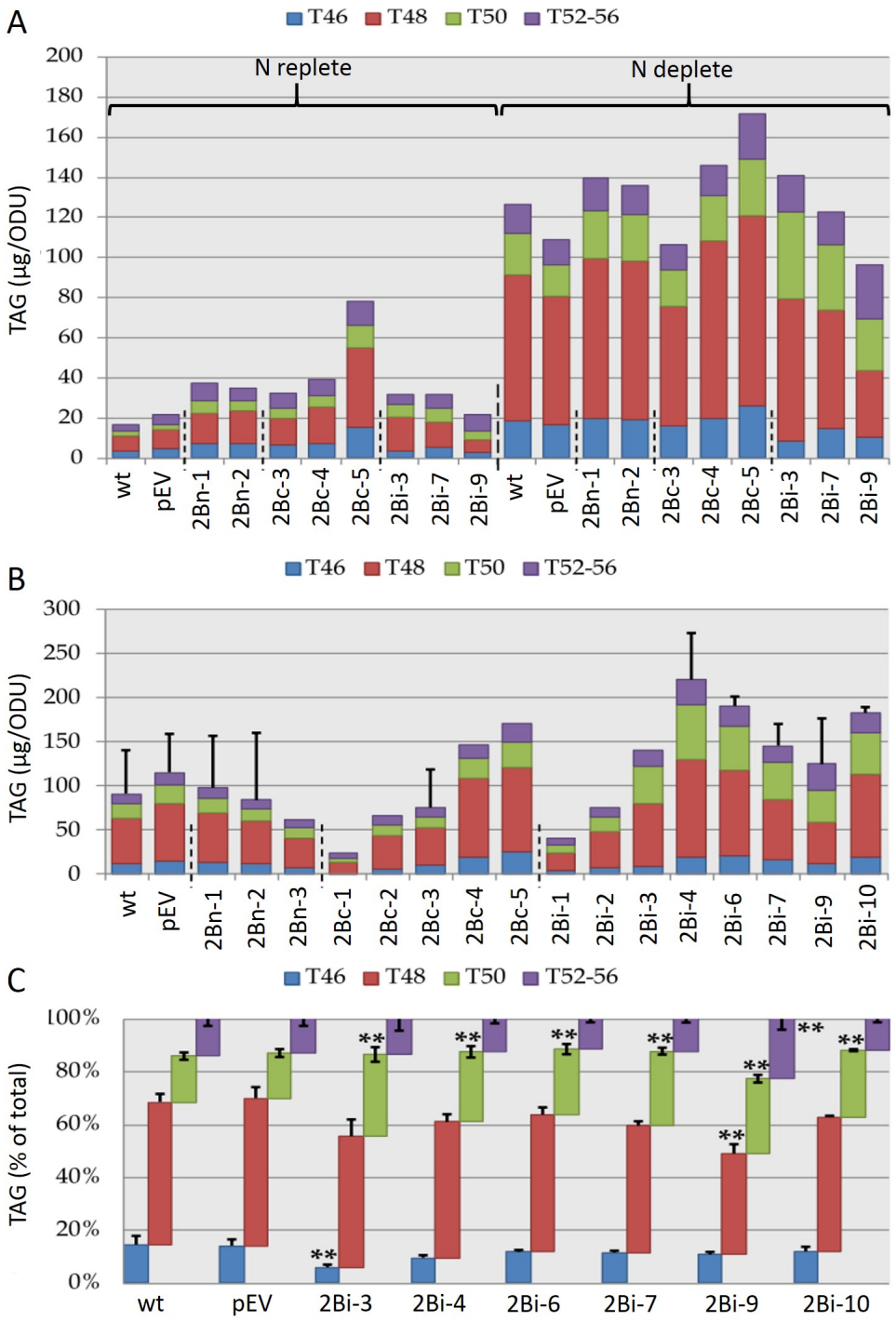


Figure 4.5 – TAG content of *P. tricornutum* lines overexpressing DGAT2B measured using GC-FID. A) TAG content measured from DGAT2 lines grown in F/2 (+N) and F/2 without nitrogen (-N). Data is from one biological replicate. B) TAG quantified from DGAT2 lines grown in F/2 without nitrogen. Results from 1 to 5 biological repeats. C) TAG composition of DGAT2Bi lines grown in F/2-N. Results from 3 to 4 biological repeats. Bars marked ** are significantly different to wt, $p<0.01$. Error bars are one standard deviation. ODU - standardised to optical density. i- AMT1 promoter used, c- FCPC promoter used, n- DGAT2B native promoter used. See Table 4.3 for strain descriptions. Figures reproduced from Litvinenko (2016).

gene did not lead to increased TAG accumulation in any strain, this may indicate that the TAG synthesis step is not limiting rather than providing information about expression of the gene.

The quantity of TAG and TAG composition were also measured for the lines of DGAT1, DGAT2A, DGAT2C, DGAT2D, DGAT2E and DGAT3. Given that DGAT2B only had an observable impact on TAG when expressed from the inducible promoter, it was decided to focus investigation of the remaining genes on the lines that also used the inducible promoter. Differences in the quantity of TAG accumulated and the proportional composition were observed, particularly for DGAT1i-7, DGAT1fi-9, DGAT2Ai-3, DGAT2Afi-6, DGAT2Bi-3, DGAT2Ci-8, DGAT2Di-2, DGAT2Ei-1, WS/DGAT-3, PDATi-5, PDATfi-6,. However, only one or two biological replicates were used and so significance of differences between lines could not be determined. As previously mentioned, constructs randomly integrate into the genome and so the mean of a number of lines must be taken into account when evaluating the impact of overexpression of the enzymes. The quantities and proportions of TAG measured for the DGAT2Bi lines shown in Figure 4.5B and C and those of the PDATi, PDATfi, DGAT1i, DGAT1fi, DGAT2Ai, DGAT2Afi, DGAT2Ci, DGAT2Di, DGAT2Ei and 3i lines in Figure 4.6A and B were averaged (mean). The resulting mean quantities for strains are shown in Figure 4.7A and the resulting TAG compositions are shown in 4.7B. Although some differences in TAG accumulation were seen, these were not significant. In contrast significant changes in the composition of TAG were observed. Alterations were seen for the DGAT2Bi strain, with an increased proportion of T50 at the expense of decreased proportions of T46 and T48. This is consistent with DGAT2B having a preference for longer FA substrates as indicated by its relative preference compared to yeast DGAT2, shown in Figure 4.4C. Additionally, alterations were seen in the TAG composition of the PDAT overexpressing line PDATi and PDATfi. The proportion of the T52-56 species increased across both lines while the proportion of T48 decreased in PDATfi.

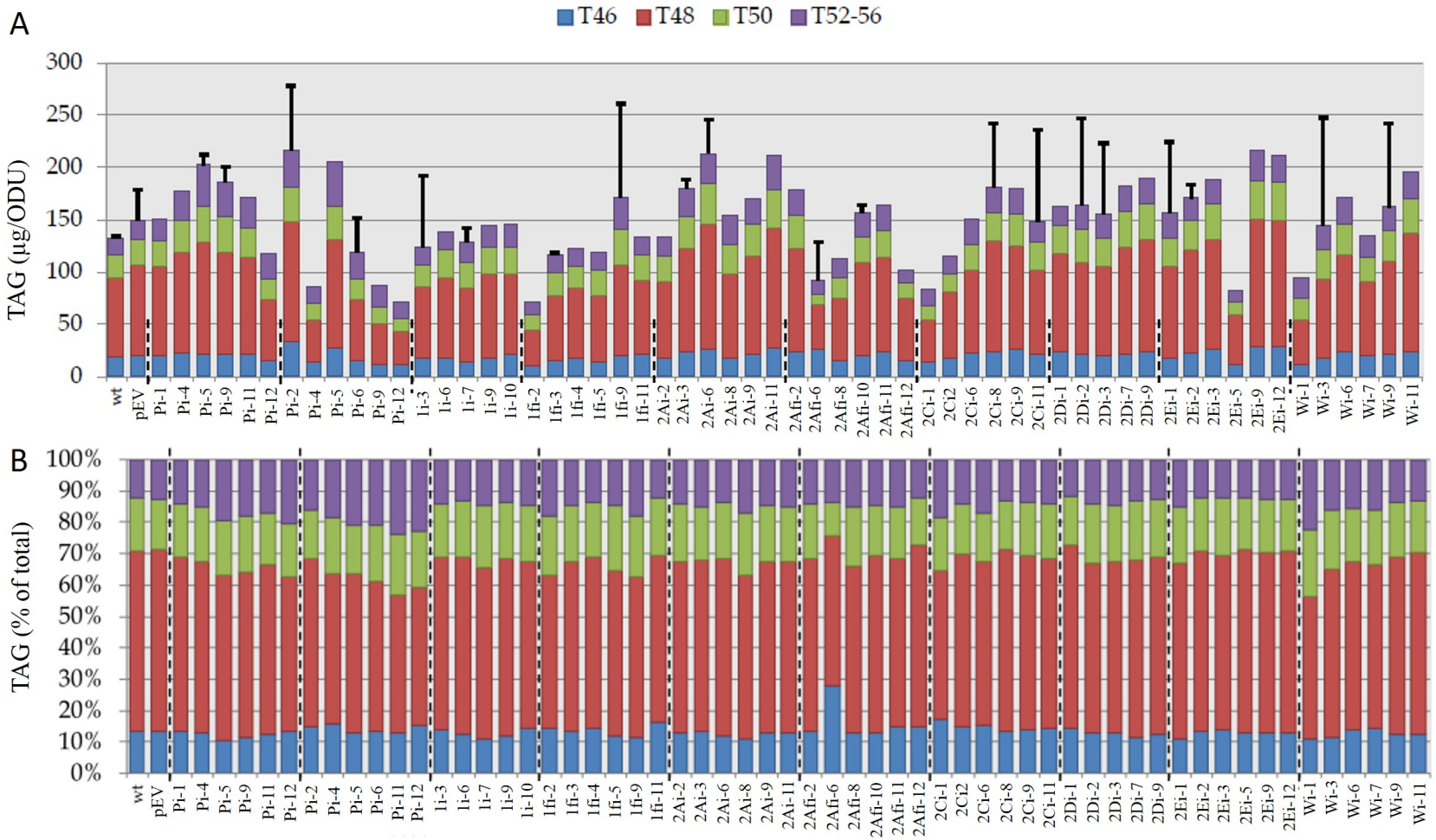


Figure 4.6 – TAG contents of *P. tricornutum* lines overexpressing PDAT and DGAT. TAG was extracted from wt *P. tricornutum* and overexpressing strains. A) TAG quantity and B) TAG composition. 1 to 2 biological replicates were measured. Nomenclature is describe in Table 4.3. f- gene containd introns, i- AMT1 promoter used. Error bars are one standard deviation. See Table 4.3 for strain descriptions. Figures reproduced from Litvinenko (2016).

To investigate the effect of DGAT2B overexpression further, the FA composition of TAG from one of the DGAT2Bi lines, DGAT2Bi-3, was measured. Thin layer chromatography was used to isolate the TAG fraction that was then derivatised to produce FAMEs. FAMEs were quantified using GC-FID and the resulting FA composition of TAG is shown in Figure 4.7C. Compared to the controls, 14:0 and 16:0 constituted a lower proportion of TAG while 18:1 constituted a large proportion. Although the results were only from one line, they were consistent with the previous observations for DGAT2B. To gain a similar insight into the FA composition of the other overexpressing lines, ESI-TQMS measurements were made on lipid extracts from the strains by Dr Richard Haslam and this is presented in Table 4.4. In addition to FA composition, these measurements provide more detailed information on TAG species composition. Due to limited time, measurements were summarised as a proportion of MS signal, which does not correlate directly with abundance of measured species. Time constraints also meant that only a single line from each strain was measured. Ion scans for mass to charge ratios corresponding to T44, T54, T56 were not conducted and hence these TAG species were not measured. The protocol also did not measure the C12:0 or hexadecapolyenoic (C16:2/3/4) FAs.

Overall seventeen TAG species were detected and nine FAs. In contrast to the GC-FID data, TAG species that contain acyl chains with an odd number of carbons were detected. These TAG species were T47:1 and T47:7 but individually contributed less that 1% of the MS signal. DGAT1i-7, DGAT1fi-9, DGAT2Ai-3, DGAT2Afi-6 , PDATi-5 and PDATfi-6, all contained T46:2 (typically 14:0/16:1/16:1) that was absent from the wt. This may explain why 16:0 decreased in these lines and 16:1 increased. Proportional alterations were seen for DGAT2Afi-6, DGAT2Bi-3, DGAT2Ei-1, PDATi-5 and PDATfi-6. Either individually or collectively, strains PDATi-5 and PDATfi-6 had higher proportions of T52:6 and T52:7 that in *P. tricornutum* typically contained 16:1/16:0/20:5 and 16:1/16:1/20:5 respectively. Proportions of T50:2, T50:3 and T50:4 were also higher. In contrast there were substantial decreases in T48:1 and T48:2 that typically consist of 16:0/16:0/16:1 and 16:0/16:1/16:1. The proportion of 16:0 in TAG decreased while 16:1, 18:1, 18:2 and 20:5 increased. This is consistent with the increased proportion of T52-56 measured by GC-FID but also indicates that T50 species may have been increased compared to wt. Furthermore,

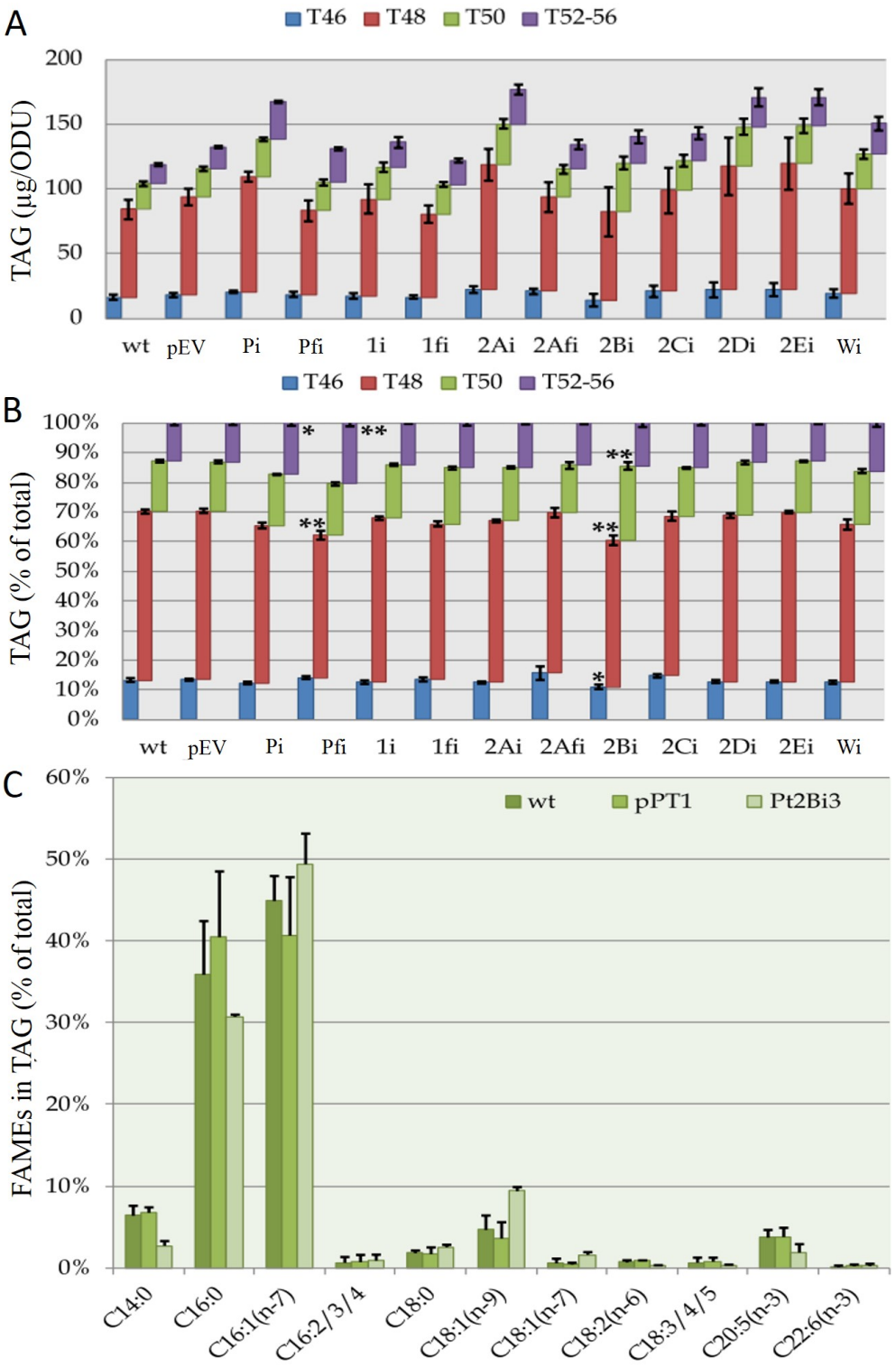


Figure 4.7 – TAG content, TAG and FA species composition of *P. tricornutum* lines overexpressing PDAT and DGAT after 8 days of nitrogen depletion. Absolute TAG quantity (A) and Relative TAG composition (B) of overexpressing strains. Results are the mean of the values for lines shown in Figures 4.5A+B and 4.6A+B. The FA composition of TAG from line DGAT2Bi-3 (C) resulting from two biological replicates. Error bars are one standard deviation. See Table 4.3 for strain descriptions. Figures reproduced from Litvinenko (2016)

Table 4.4 – TAG composition of DGAT and PDAT overexpressing lines. TAG species (A) and FA (B) composition of TAG measured by ESI-MS/MS in a single line for each strain. Results from two biological replicates. Error bars are one standard deviation. See Table 4.3 for strain descriptions. Tables reproduced from Litvinenko (2016).

A

TAG (% MS signal)	*P. tricornutum* strain analysed											
	wt	Pi	Pfi	1i	1fi	2Ai	2Afi	2Bi	2Ci	2Di	2Ei	Wi
T46:1	6.9 ± 0.8	4.5 ± 0.6	5.7 ± 1.6	7.4 ± 1.1	5.8 ± 1.5	6.0 ± 1.0	15.4 ± 1.7	2.7 ± 0.8	6.0 ± 0.6	7.3 ± 0.3	4.6 ± 0.4	6.4 ± 2.1
T46:2	0	2.4 ± 0.3	4.1 ± 0.7	4.0 ± 0.7	6.2 ± 3.0	5.0 ± 2.4	10.0 ± 0.2	0	0	0	0	0
T47:1	0.4 ± 0	0.3 ± 0	0.3 ± 0	0.4 ± 0	0.5 ± 0.2	0.3 ± 0	0.4 ± 0.1	0	0.3 ± 0.1	0.4 ± 0	0	0.3 ± 0
T47:7	0.6 ± 0	0.4 ± 0.1	0.3 ± 0.1	0.7 ± 0.2	0.4 ± 0.1	0.5 ± 0.1	0.9 ± 0.1	0	0.4 ± 0	0.6 ± 0.3	0	0
T48:0	2.8 ± 0.1	2.5 ± 1.0	1.6 ± 0.6	2.2 ± 0.1	1.7 ± 0.2	1.9 ± 0.7	1.4 ± 0	1.2 ± 0.3	2.0 ± 0.4	2.2 ± 1.5	1.5 ± 0	2.2 ± 0.2
T48:1	32.4 ± 0.4	29.0 ± 1.6	24.7 ± 0.3	29.2 ± 1.7	24.3 ± 0.8	29.3 ± 0.2	21.7 ± 0.7	20.2 ± 1.4	29.4 ± 2.4	30.5 ± 7.7	23.2 ± 3.9	29.5 ± 1.3
T48:2	37.9 ± 3.1	33.7 ± 2.9	37.3 ± 1.3	33.5 ± 0.3	34.3 ± 2.3	36.2 ± 0.4	30.2 ± 0.7	34.8 ± 0.9	40.8 ± 0.4	37.6 ± 7.1	42.7 ± 0.4	38.2 ± 3.3
T48:3	8.5 ± 0.2	9.6 ± 0.7	10.7 ± 0.4	8.5 ± 1.1	10.5 ± 0.1	8.8 ± 0.1	12.4 ± 0.4	19.4 ± 0.9	10.3 ± 1.0	8.1 ± 2.4	12.5 ± 0.7	9.3 ± 0.1
T48:4	0.6 ± 0.3	0.7 ± 0	0.7 ± 0.1	1.0 ± 0.1	0.5 ± 0.3	0.6 ± 0.1	1.2 ± 0	1.5 ± 0.3	0.6 ± 0.1	0.6 ± 0.3	0.6 ± 0.3	0.6 ± 0.4
T50:1	0.7 ± 0	1.9 ± 1.4	0.8 ± 0.1	0.8 ± 0	0.6 ± 0.1	0.7 ± 0.1	0.4 ± 0.3	0.9 ± 0.4	0.4 ± 0.4	0.6 ± 0.2	1.3 ± 0.9	0.8 ± 0.3
T50:2	2.8 ± 0.2	4.0 ± 0.2	3.0 ± 0.1	3.0 ± 0.4	3.5 ± 0.4	3.1 ± 0.5	0.9 ± 0.3	6.5 ± 0.7	2.8 ± 0.2	3.4 ± 0.4	3.1 ± 0.5	3.9 ± 0.8
T50:3	3.5 ± 0.5	5.2 ± 0.4	4.4 ± 0.2	3.5 ± 0.5	5.1 ± 0.4	4.1 ± 0.1	1.9 ± 0.2	12.2 ± 0.7	3.7 ± 0.1	3.9 ± 0.8	4.7 ± 0.6	5.0 ± 0
T50:4	0.6 ± 0.2	2.0 ± 0.1	0.7 ± 0	0.7 ± 0.1	2.5 ± 0.4	0.6 ± 0.2	0.3 ± 0	0	0.5 ± 0	1.2 ± 0.2	1.7 ± 1.1	0.5 ± 0
T50:5	0.3 ± 0	0.3 ± 0	0.4 ± 0.1	0.4 ± 0.1	0.6 ± 0.5	0.3 ± 0	0.3 ± 0	0	0.4 ± 0	0.5 ± 0.1	0.6 ± 0.1	0.4 ± 0
T50:6	0	0.3 ± 0	0.5 ± 0.2	0.6 ± 0.4	0	0	0.4 ± 0	0	0	0	0.4 ± 0.1	0
T52:6	1.3 ± 0.3	1.9 ± 0.1	2.5 ± 0.4	2.3 ± 0.8	2.0 ± 0.5	1.5 ± 0.2	1.0 ± 0.2	0.7 ± 0.1	1.4 ± 0.1	1.3 ± 0.1	1.5 ± 0.2	1.7 ± 0.3
T52:7	0.8 ± 0.4	1.3 ± 0.1	2.4 ± 0.2	1.9 ± 0.6	1.6 ± 0.3	1.0 ± 0.1	1.1 ± 0.2	0	1.0 ± 0.2	1.8 ± 1.0	1.5 ± 0.4	1.3 ± 0.2

B

TAG FA (% MS signal)	*P. tricornutum* strain analysed											
	wt	Pi	Pfi	1i	1fi	2Ai	2Afi	2Bi	2Ci	2Di	2Ei	Wi
C14:0	0	0	0	0	0	0	0.3 ± 0	0	0	0	0	0
C16:0	66.2 ± 6.4	48.7 ± 4.3	53.2 ± 4.7	56.8 ± 10	49.5 ± 4.8	57.0 ± 2.6	57.0 ± 2.6	41.9 ± 6.0	65.8 ± 5.6	70.3 ± 1.4	60.3 ± 1.5	62.4 ± 6.2
C16:1	24.5 ± 1.6	36.0 ± 3.2	34.4 ± 2.9	31.1 ± 7.1	37.0 ± 3.7	30.9 ± 2.7	38.3 ± 2.6	43.4 ± 3.6	26.4 ± 4.0	22.1 ± 0.5	30.7 ± 1.4	26.8 ± 3.6
C18:0	1.5 ± 0.5	2.2 ± 0.1	1.1 ± 0.1	1.6 ± 0.4	2.2 ± 0.5	1.6 ± 0.3	0.8 ± 0.2	2.1 ± 0.7	1.3 ± 0.2	1.2 ± 0.3	1.3 ± 0.1	1.7 ± 0.4
C18:1	5.2 ± 2.5	8.6 ± 0.7	6.3 ± 1.0	5.9 ± 1.1	5.9 ± 0.1	6.5 ± 0.3	2.1 ± 0	11.7 ± 1.4	4.6 ± 1.1	4.5 ± 1.3	4.5 ± 0.2	6.7 ± 1.2
C18:2	0.6 ± 0.3	1.5 ± 0.1	0.6 ± 0	0.7 ± 0.1	1.9 ± 0.1	0.7 ± 0	0.5 ± 0.1	0.7 ± 0.3	0.6 ± 0.1	0.5 ± 0.1	1.0 ± 0.2	0.7 ± 0.4
C18:3	0.4 ± 0.2	0.5 ± 0.1	0.4 ± 0	0.5 ± 0.1	0.7 ± 0.4	0.8 ± 0.4	0	0	0	0.6 ± 0.3	0.6 ± 0.2	0
C20:5	1.3 ± 1.0	2.2 ± 0.4	3.4 ± 1.0	2.9 ± 1.5	2.1 ± 0.5	1.7 ± 0.1	1.0 ± 0.3	0.4 ± 0	1.4 ± 0.2	0.8 ± 0.2	1.6 ± 0	1.7 ± 0.6
C22:6	0.4 ± 0.2	0.3 ± 0	0.5 ± 0.1	0.5 ± 0.1	0.7 ± 0.4	0.7 ± 0.4	0	0	0	0	0	0

these increases seem to come at the expense of T48:1 and T48:2. For DGAT2Bi-3, there were particularly large decreases in T46:1, T48:1 and T48:2 that are typically formed of 14:0/16:0/16:1, 16:0/16:0/16:1 and 16:0/16:1/16:1 respectively. There were also large increases of T48:3, T50:2 and T50:3. T48:3 is typically formed of 16:1/16:1/16:1, T50:2 is likely formed of either 16:1/16:0/18:1 or 16:1/16:1/18:0 and T50:3 is likely formed of 16:1/16:1/18:1, as shown in Table 4.1. The proportion of the FAs 16:0 and 20:5 were lower in TAG from DGAT2Bi-3 while 16:1, 18:0 and 18:1 were higher. This again supports DGAT2B preferentially incorporating C18 and monounsaturated FAs. DGAT2Afi-6 was another line in which marked differences with the control were noticed. It contained much higher amounts of T46:1 and T46:2 compared to other lines, typically composed of 14:0/16:0/16:1 and 14:0/16:1/16:1 respectively. The FA C14:0 was also detected in the TAG fraction of DGAT2Afi-6 but not in those of other lines. However, comparison to other DGAT2Afi lines in the GC-FID data presented in Figure 4.6B indicated that DGAT2Afi-6 had a unique response, indicating that these results may be specific to the line. The DGAT2E overexpressing line, DGAT2Ei-1, demonstrated decreased proportions of T48:0 and T48:1 and increased T48:2 and T48:3. The proportion of 16:0 in TAG decreased while that of 16:1 increased. Changes were not observed for DGAT2Ci-8, DGAT2Di-2 or WS/DGAT-8. In addition to ESI-MS/MS measurements, light microscope analysis of the strains indicated that DGAT2Ci-8 and DGAT2Ei-1 contained small spherical bodies not seen in the wt. These may be lipid droplets.

Collectively Dr Litvinenko's expression studies in *S. cerevisiae* and the preliminary ones in *P. tricornutum* indicated that DGAT2B and PDAT function in the synthesis of TAG. For DGAT2B, the quantity of TAG accumulated under nitrogen depletion did not increase when the enzyme was overexpressed in Pt. However, compositional changes were observed with the proportion of T50 species and T48:3 increased and T46 and the remaining T48 species decreased. Together this supports synthesis of T50 species and T48:3 at the expense of T46 and T48 species. Information was also gathered on the substrate preference of DGAT2B. Compared to ScDGAT2, PtDGAT2B accumulated TAG with higher proportions of 16:1 and 18:1 over 14:0 and 16:0. FAME analysis of TAGs from a *P. tricornutum* line that overexpressed DGAT2B contained lower proportions of 14:0 and 16:0 and higher proportions of

18:1. ESI-MS/MS measurements of a line of *P. tricornutum* overexpressing DGAT2B indicated that the proportion of 16:0 in TAG decreased while that of 16:1 and 18:1 increased. Typical FA compositions of T46, T48 and T50 species in *P. tricornutum* and their proportional changes correspond to these alterations in FA proportion. Collectively this supports a substrate preference of DGAT2B for incorporating acyl chains of 16:1 and 18:1 over 16:0 and 14:0. It should be borne in mind that as overexpression occurs against a background of native DGAT and PDAT expression, alterations to TAG profiles are relative to the native TAG content.

Regarding PDAT, overexpression in *P. tricornutum* during nitrogen depletion did not lead to higher levels of TAG accumulation. In GC-FID results the proportion of T52-56 species were increased while T48 was decreased. In ESI-MS/MS data, there was a decrease in the proportion of T48 species and increases in the proportion of T50 and T52 species in two lines overexpressing PDAT. These ESI-MS/MS measurements also identified a proportional increase in the amount of 16:1, 18:1, 18:2 and 20:5 in TAG and a decrease in 16:0. These changes in FA composition find support in the typical FA makeup of the altered TAGs T48, T50 and T52-56. In particular, 20:5 is uniquely found in TAG species with a total of 50 or more carbons in the acyl chains. Collectively, this indicates that PDAT may prefer to incorporate longer, more desaturated acyl chains into TAG.

In order to substantiate the observations made above, further experiments were necessary. In particular, ESI-TQMS measurements on multiple lines were needed to identify which observations are associated with enzyme overexpression rather than line specific variations.

4.2 Results

4.2.1 Comparison of TAG and glycerolipids between the DGAT and PDAT overexpressing strains during nitrogen depletion

In order to further investigate the functions of PtDGATs and PtPDAT and to validate the preliminary observations, lines of *P. tricornutum* overexpressing the

various enzymes detailed above were made use of. I carried out the work presented in this results section, except where stated. Lines were selected for experimentation based on at least one previously observed difference with wt or the empty vector pEV containing strains. The lines selected were as follows: DGAT1 (DGAT1fi-5, DGAT1fi-9, DGAT1fi-11), DGAT2A (DGAT2Afi-2, DGAT2Afi-6, DGAT2Afi-11), DGAT2B (DGAT2Bi-3, DGAT2Bi-4, DGAT2Bi-7), DGAT2D (DGAT2Di-2, DGAT2Di-3, DGAT2Di-9), DGAT2E (DGAT2Ei-1, DGAT2Ei-9, DGAT2Ei-12) and PDAT (PDATfi-2, PDATfi-5, PDATfi-6). These lines are detailed in Table 4.5 for reference. DGAT2C and WS/DGAT were not investigated as Dr Litvinenko did not observe changes in the lipid profile of strains that overexpressed them. To enable high amounts of TAG accumulation the overexpressing lines and controls were cultured in nitrogen deplete media for 8 days prior to total lipid extraction. The total lipid extracts were split to allow GC-FID and ESI-TQMS measurements of the same samples. The methods used for total lipid extraction, GC-FID and ESI-TQMS are detailed from Section 2.3.2 onwards. Three technical replicates were cultured and extracted for each *P. tricornutum* line. Unfortunately, wt control samples were lost during handling. However, comparison between the lines still provides insight into the effects of overexpressing these enzymes on lipids and information from this experiment is therefore detailed here as further support for subsequent experiments.

Table 4.5 – DGAT and PDAT independent transformants used in the comparison of TAG and glycerolipids. *the PDAT gene lacks the first 25aa in these transformant lines. See Table 4.3 for strain descriptions.

Strain	Gene	Lines
1fi	DGAT1	1fi-5, 1fi-9, 1fi-11
2Afi	DGAT2A	2Afi-2, 2Afi-6, 2Afi-11
2Bi	DGAT2B	2Bi-3, 2Bi-4, 2Bi-7
2Di	DGAT2D	2Di-2, 2Di-3, 2Di-9
2Ei	DGAT2E	2Ei-1, 2Ei-9, 2Ei-12
Pfi	PDAT*	Pfi-2, Pfi-5, Pfi-6

GC-FID measurements enabled differentiation of TAG species at the level of total acyl chain length. The proportion of TAG contributed by TAG species with the same total acyl chain length is displayed in Figure 4.8. Three cultures were

harvested to provide samples for each independent transformant. This data was then averaged (mean) for each *P. tricornutum* strain. The proportion of T46, T48, T50 and T52 are similar between DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei shown in Figure 4.8, with the exception of T46 in strain DGAT2Afi. The proportion of T46 in strain DGAT2Afi has high variation. In comparison to these strains, the species T46 and T48 form a lower proportion in DGAT2Bi, while T50 constitutes a higher proportion. TAG species in PDATfi are also altered compared to DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei. T48 forms a lower proportion while T52 forms a higher proportion. These results are consistent with the observations of Dr Litvinenko, as illustrated in Figure 4.7B.

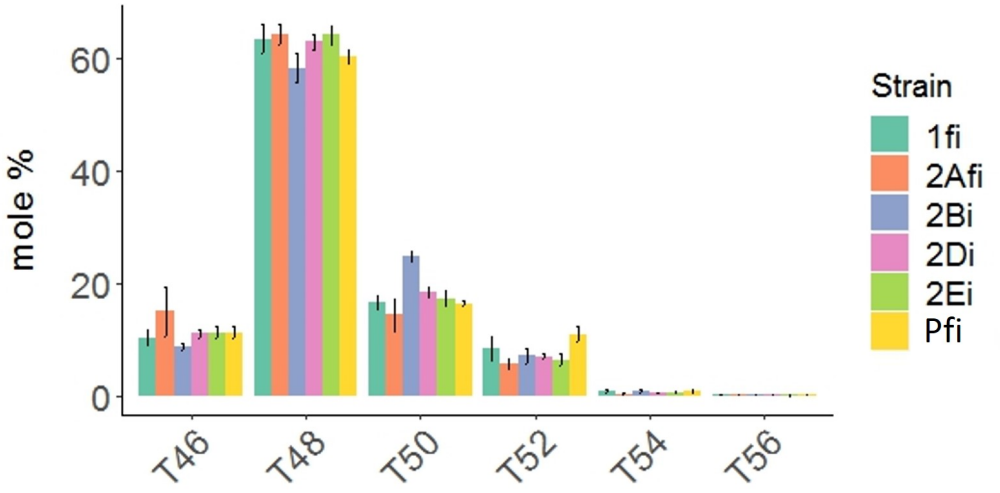


Figure 4.8 – TAG composition of *P. tricornutum* lines overexpressing DGATs and PDAT measured by GC-FID. Total lipid extracts were taken from strains during nitrogen depletion and TAGs quantified by GC-FID. The major detected TAG species are detailed by total acyl chain length: T46, T48, T50, T52, T54 and T56. The results are the mean of three independent transformants. Three biological repeats were measured for each transformant. 95% confidence intervals are shown. Significance of differences was not calculated. 1fi- DGAT1fi, 2Afi- DGAT2Afi, 2Bi- DGAT2Bi, 2Di- DGAT2Di, 2Ei- DGAT2Ei, Pfi- PDATfi.

Determining the total acyl chain length of TAGs with GC-FID loses information on the desaturation and FA content of TAG species. The measurements carried out by ESI-TQMS therefore complement GC-FID measurements by enabling the total saturation of TAG species and the FA makeup of the TAG fraction to be determined. The genetic constructs were expected to integrate randomly into the genome and independent transformants would therefore be expected to demonstrate variation in their response. To investigate the variability between lines, the proportion of TAG

species determined by ESI-TQMS was plotted for each independent transformant (Figure 4.9). A number of lines show large variation in the proportion of TAG species accumulated. DGAT2Afi-6 demonstrated particularly large differences in its TAG composition compared to two other DGAT2Afi lines, DGAT2Afi-2 and DGAT2Afi-11. The higher proportions of T46:0 and T46:2 in DGAT2Afi-6 likely underpinned the large variation in T46 observed in the GC-FID data in Figure 4.8. DGAT2Bi-3 also demonstrates substantial differences with other lines, namely DGAT2Bi-4 and DGAT2Bi-7, containing a lower proportion of T48:1 and higher proportion of T48:3. Variation was also seen between the DGAT1fi and DGAT2Di lines although this was substantially lower than that seen for DGAT2Afi and DGAT2Bi.

The TAG species data was averaged (mean) across independent transformants to provide data for the strains and this is shown in Figure 4.10A. The FA composition of the TAG fraction for strains was also averaged (mean) from independent lines and is displayed in Figure 4.10B. Several of the strains demonstrated TAG profiles that differed from the majority of the other strains. DGAT2Bi contained lower proportions of 46:0 and 48:2 and a higher proportion of 50:2. These support the GC-FID measurements and identify T46:0, T48:2 and T50:2 as the species that underpin the alterations in T46, T48 and T50 in DGAT2Bi. Considering the FA composition of TAG in DGAT2Bi, the proportion of 16:0 decreased and the proportion of 18:1 increased. These observations support the observations of Drs Chen and Litvinenko that DGAT2B has a preference for 18:1 over 16:0 as illustrated in Figures 4.4C and 4.6C. Yu et al. (2009) observed that the FA composition of T46:0 and T48:2 in *P. tricornutum* are 14:0/16:0/16:0 and 16:0/16:1/16:1 respectively. T50:2 is likely to have a FA composition of 16:1/16:1/18:0 or 16:1/16:0/18:1. Decreases in 46:0 and 48:2 could therefore lead to a decreased contribution of 16:0 to the TAG fraction. Likewise, an increase in T50:2 (16:1/16:0/18:1) could underpin the increased proportion of 18:1. There may also be alterations in the proportion of T48:1 and the FAs 18:0 and 18:2 for strain DGAT2Bi but these changes are less clear.

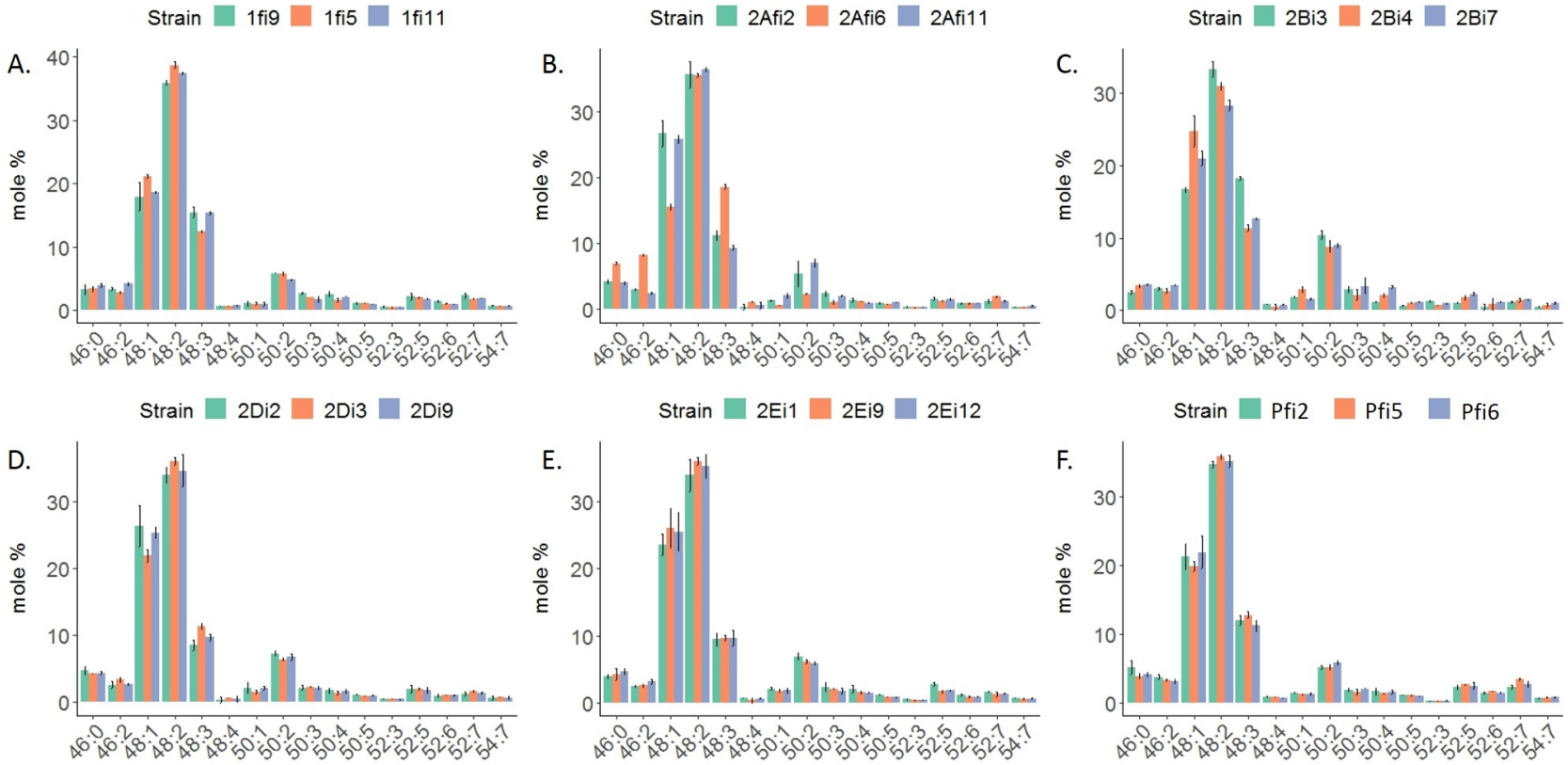


Figure 4.9 – TAG species composition of individual DGAT and PDAT overexpressing lines. TAG species were quantified by ESI-TQMS during nitrogen depletion and are arrayed on the x-axis. TAGs were measured from individual transformants overexpressing (A) DGAT1, (B) DGAT2A, (C) DGAT2B, (D) DGAT2D, (E) DGAT2E and (F) PDAT. Results from three biological repeats. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

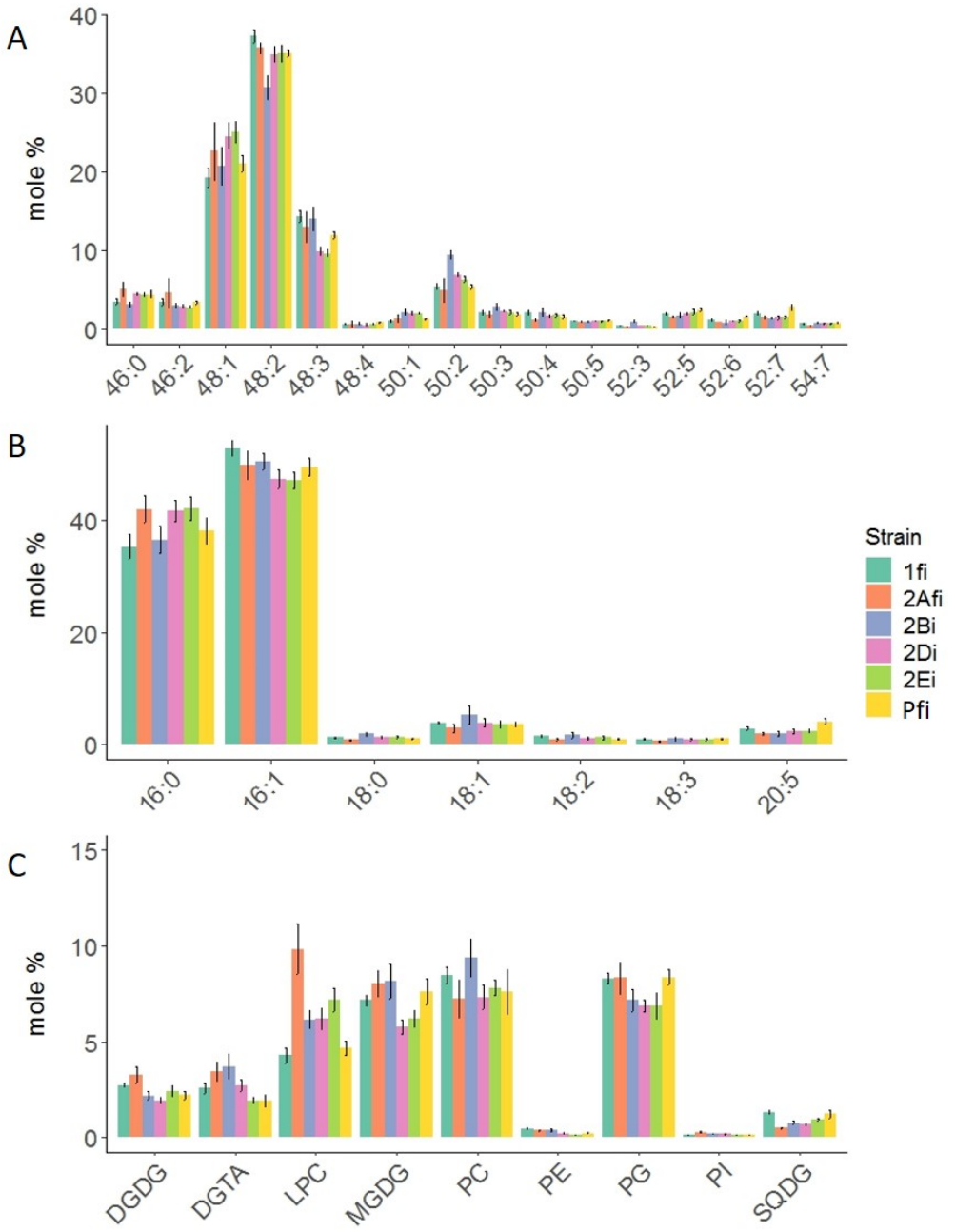


Figure 4.10 – TAG and membrane glycerolipid composition in DGAT and PDAT overexpressing *P. tricornutum* strains. TAG species, the FAs that composed TAG and membrane glycerolipids were quantified by ESI-TQMS after 8 days of nitrogen depletion. (A) Proportion of TAG constituted by major TAG species. (B) Proportion of TAG constituted by major FAs. (C) The proportion of measured glycerolipids constituted by membrane glycerolipid classes. Major species and FAs contributed more than 1% of TAG. DGDG- digalactosyldiacylglycerol, DGTA- diacylglycerol hydroxymethyltrimethyl β-alanine, LPC- lysophosphatidylcholine, MGDG - monogalactosyldiacylglycerol, PC- phosphatidylcholine, PE- phosphatidylethanolamine, PG- phosphatidylglycerol, PI- phosphatidylinositol, SQDG- sulphoquinovosyldiacylglycerol. The results are the mean of three independent transformants. Three biological repeats were measured for each transformant. 95% confidence intervals are shown are shown in a and b. Standard errors are shown in c. 1fi- DGAT1fi, 2Afi- DGAT2Afi, 2Bi- DGAT2Bi, 2Di- DGAT2Di, 2Ei- DGAT2Ei, Pfi- PDATfi.

Compared to the other strains, strain PDATfi demonstrates an increased proportion of 52:7 and a slight increase in 52:6, as shown in Figure 4.10. T52:7 and T52:6 are therefore supported as the species that caused alteration in T52 in the GC-FID data. However, a clear change corresponding to the decrease in T48 was not seen; Although T48:1 may decrease, variability in this species in the strains containing the other enzymes makes this uncertain. The ESI-TQMS measurements of the FA composition of the TAG fraction indicates the proportion of 20:5 increased. This supports a preference by PDAT for incorporation of the FA 20:5 (the omega-3 LC-PUFA EPA). This has not previously been observed for PtPDAT. In *P. tricornutum* during nitrogen depletion, the major T52:6 and T52:7 species contain 16:1/16:0/20:5 and 16:1/16/1/20:5 respectively. The increases in T52:6 and T52:7 therefore account for the increase in 20:5 in TAG.

Strains DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei also demonstrated changes in TAG composition, something that was not seen in the GC-FID results nor across multiple lines in the work of Dr Litvinenko. TAG from strain DGAT1fi contained a smaller proportion of T46:0, T48:1, a larger proportion of T48:2 and a lower proportion of the FA 16:0 compared to most other overexpressing strains. The TAG from strain DGAT2Afi contained an increased proportion of T46:0, T46:2 and 48:1 and TAG from strains DGAT2Di and DGAT2Ei contained a higher proportion of T48:1 and a lower proportion of T48:3. Previous observations by Dr Chen and Litvinenko support the observations reported here for strains DGAT2Bi and PDATfi.

In addition to impacting TAG content, overexpression of PDAT and DGAT may also alter other glycerolipids. PDAT typically hydrolyses an acyl chain from PC but has also been observed to act on MGDG. Overexpressing PDAT may therefore alter the amount of its acyl-donating substrate. Both DGAT and PDAT use DAG and so their overexpression could alter the levels available for synthesis of other lipids. To investigate this, non-TAG glycerolipid classes were measured using ESI-TQMS on the same total lipid extracts from which TAG was measured. The membrane glycerolipids included have been previously identified in *P. tricornutum* (see Table 1.1) and were MGDG, DGDG, PG, DGTA, PC, lysophosphatidylcholine (LPC), PE, PG and PI. The proportion of total lipids constituted by each non-TAG lipid class is shown in Figure 4.10C. Differences between the strains are seen for every membrane

glycerolipid class.

4.2.2 TAG content and composition of DGAT and PDAT overexpressing lines compared to WT and pEV lines during nitrogen depletion

4.2.2.1 Experimental outline

To further investigate the above observations and enable comparison with controls, *P. tricornutum* DGAT and PDAT overexpressing lines, wt and pEV empty vector control lines were cultured under nitrogen deplete conditions for 8 days and total lipid extractions were carried out as described in Chapter 2. The samples were filtered and internal standards required for quantification by ESI-TQMS were added. TAG within the lipid extracts was then measured by Dr Richard Haslam at Rothamsted Research using ESI-TQMS.

As mentioned, the observations detailed in Figure 4.10 demonstrated substantial variation between some of the measured lines that contained the same construct. This results from the formation of the lines by discrete random integration events. These extreme variations may impact interpretation of the results and I therefore aimed to ensure the lines selected for measurement were representative of the variation across all transformed lines. The data on TAG quantity and composition of different lines that is shown in Figure 4.6 was therefore reviewed. The lines with a representative TAG composition for their strain were determined. These selected lines are detailed in Table 4.6. Specifically, DGAT1fi-4 replace DGAT1fi-9, DGAT2Afi-8 and DGAT2Afi-10 replaced DGAT2Afi-2 and DGAT2Afi-6, DGAT2Bi-10 replaced DGAT2Bi-4, DGAT2Di-7 replaced DGAT2Di-9, DGAT2Ei-2 and DGAT2Ei-3 replaced DGAT2Ei-1 and DGAT2Ei and PDATfi-11 replaced PDATfi-2. Lipid extractions were carried out for three biological repeats of each line. Unfortunately, due to equipment failure, errors occurred during the ESI-TQMS measurements of a number of the samples. Still, at least one biological repeat of each of the lines was measured except for DGAT2Bi-3.

Table 4.6 – The DGAT and PDAT clonal lines grown in this study. *- the PDAT lacks the first 25 amino acids. See Table 4.3 for strain descriptions.

Strain	Gene	Lines
wt		
pEV		pEV-3, pEV-4, pEV-5
1fi	DGAT1	1fi-5, 1fi-4, 1fi-11
2Afi	DGAT2A	2Afi-8, 2Afi-10, 2Afi-11
2Bi	DGAT2B	2Bi-3, 2Bi-7, 2Bi-10
2Di	DGAT2D	2Di-3, 2Di-2, 2Di-7
2Ei	DGAT2E	2Ei-2, 2Ei-3, 2Ei-12
Pfi	PDAT*	Pfi-11, Pfi-6, Pfi-5

4.2.2.2 ESI-TQMS measurements of TAG quantity

ESI-TQMS measurement identified the TAG species in the lipid extracts of each independent transformant. These measurements provided information on the total acyl chain length and total desaturation of each species. To produce the TAG species proportional composition of each line, I calculated the proportion that each species constituted of the total moles in the TAG fraction. In addition, the quantity of TAG species was normalised to dry weight. The mole percentage and normalised amount of TAG species measured in each independent transformant are shown in Figure 4.11 and 4.12 and the number of biological repeats are detailed in the captions. There were differences in the proportions of TAG species between some lines transformed with the same construct. As seen in 4.11E, extractions from line DGAT1fi-11 had higher mole percentages of T48:1 and lower T48:3 than DGAT1fi-4 and DGAT1fi-5. Those from DGAT2Afi-8, shown in 4.11G did not contain T50:1 while DGAT2Afi-10 and DGAT2Afi-11 did. DGAT2Afi-8 extracts also had higher proportions of T52:7. Depicted in 4.12A, DGAT2Bi-7 TAGs contained T50:4 and T52:5, unlike those from DGAT2Bi-10. In 4.12E, DGAT2Ei-12 extract contained a higher proportion of T50:2 than DGAT2Ei-2. In contrast, consistent proportions of TAG species were seen across transformed lines for DGAT2Di, PDATfi and pEV. As mentioned, this is not unexpected as the constructs are randomly integrated into the genome during transformation.

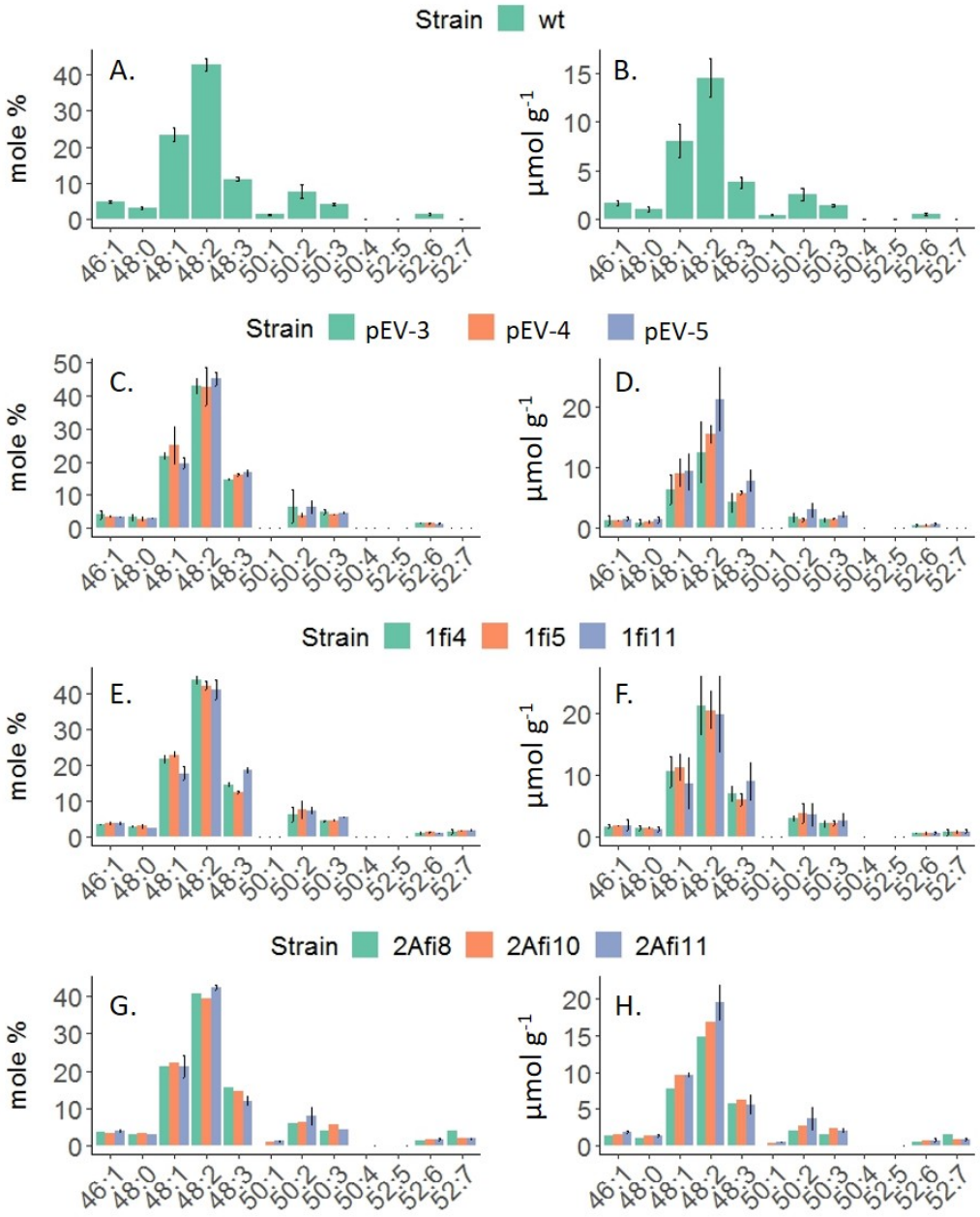


Figure 4.11 – TAG species content of *P. tricornutum* lines overexpressing DGAT1, DGAT2A. TAG was measured using ESI-TQMS from total lipids that were extracted from *P. tricornutum* lines after 8 days of nitrogen depletion. Figures on the left show proportional composition of TAGs and figures on the right show TAG species quantities. Lines and biological repeats: (A+B) wt, 6 repeats; (C+D) pEV-3, 2 repeats; pEV-4, 2 repeats; pEV-5, 3 repeats; (E+F) DGAT1fi-4, 2 repeats; DGAT1fi-5, 3 repeats; DGAT1fi-11, 2 repeats; (G+H) DGAT2Afi-8, 1 replicate; DGAT2Afi-10, 1 replicate; DGAT2Afi-11, 3 repeats. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

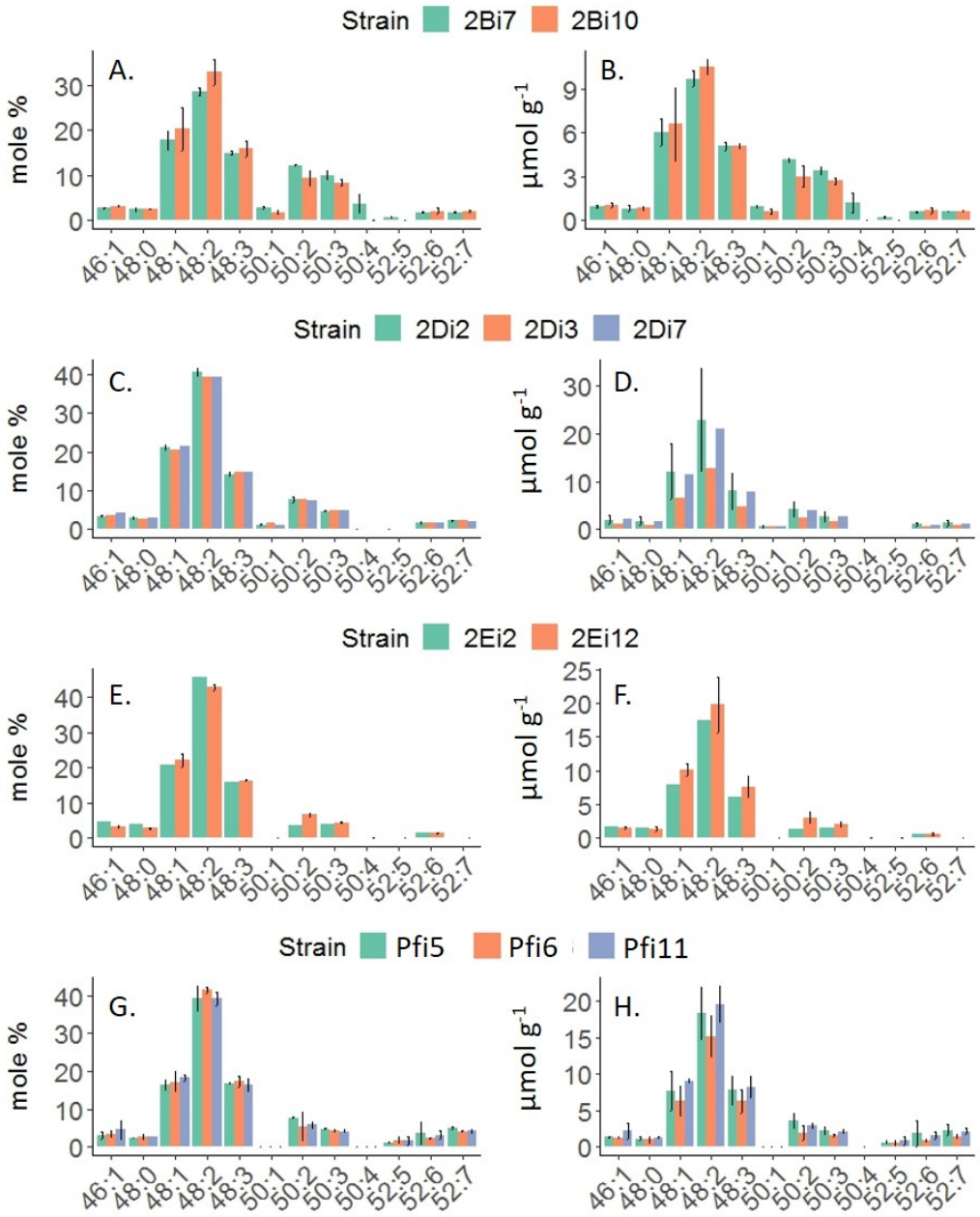


Figure 4.12 – TAG species content of *P. tricornutum* lines overexpressing DGAT2B, DGAT2D, DGAT2E and PDAT. TAG was measured using ESI-TQMS from total lipids that were extracted from *P. tricornutum* lines after 8 days of nitrogen depletion. Figures on the left show proportional composition of TAGs and figures on the right show TAG species quantities. Lines and biological repeats: (A+B) DGAT2Bi-7, 3 repeats; DGAT2Bi-10, 3 repeats; (C+D) DGAT2Di-2, 3 repeats; DGAT2Di-3, 1 replicate; DGAT2Di-7, 1 replicate; (E+F) DGAT2Ei-2, 1 replicate; DGAT2Ei-12, 2 repeats (G+H) PDATfi-5, 2 repeats, PDATfi-6, 2 repeats, PDATfi-11, 3 repeats. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

The TAG species that were detected during this experiment are presented on the x-axis of Figure 4.13A. Overall, fewer TAG species were detected than in the previous mass spectroscopy measurements detailed in Section 4.2.1 and previous studies (Yu et al., 2009; Bower- MRes, 2015). Specifically, T46:2, T48:4, T50:5, T50:6, T52:2, T52:3 and T54:8 were not detected in the present study. Furthermore, detection of T50:1, T50:4, T52:5 and T52:7 varies considerably between strains and these species were not detected in every strain of the current study. This varied detection does not appear to be the result of biological differences between the strains. Firstly, these species were consistently detected across lines in the relative comparison work presented in Section 4.2.1 and in measurements taken by Dr Litvineno displayed in Table 4.4. Secondly, there are differences in the detection of these species between independent transformants that contain the same construct: T50:1 was not detected in DGAT2Afi-8 while T50:4 and T52:5 were not detected in DGAT2Bi-10 as shown in Figures 4.11G and 4.12A.

Both the TAG species that are not detected and those that are variably detected are usually present at less than 5mol% of the TAG fraction as detailed in 4.10A and are the lowest abundance TAG species during nitrogen depletion. This suggests poor detection of low abundance species is the cause of the above observations and it is possible that the ESI-TQMS equipment problems contributed to this. To minimise the impact of variable detection of low abundance species on interpretation of the results, measurements of T50:1, T50:4, T52:5 and T52:7 were excluded from the following analysis. In addition, FAs that were present in the TAG fraction in only some strains, namely 18:0, 18:2, 18:3 and 20:4, were also excluded for the same reasoning. Given the small proportion of TAG constituted by these species, it is unlikely that their lack of detection and omission will have a large impact on interpretation of the data. However, potential impact is considered further in Section 4.3.

The TAG species mole percentage and normalised quantity data was averaged (mean) across lines to provide the data for the lines overexpressing each enzyme. These data are shown in Figures 4.13 and 4.14. As are those for the FA composition of TAG. The data from the control pEV and wt lines is included in both sets of figures. Surprisingly, there were differences in the content of TAG species between

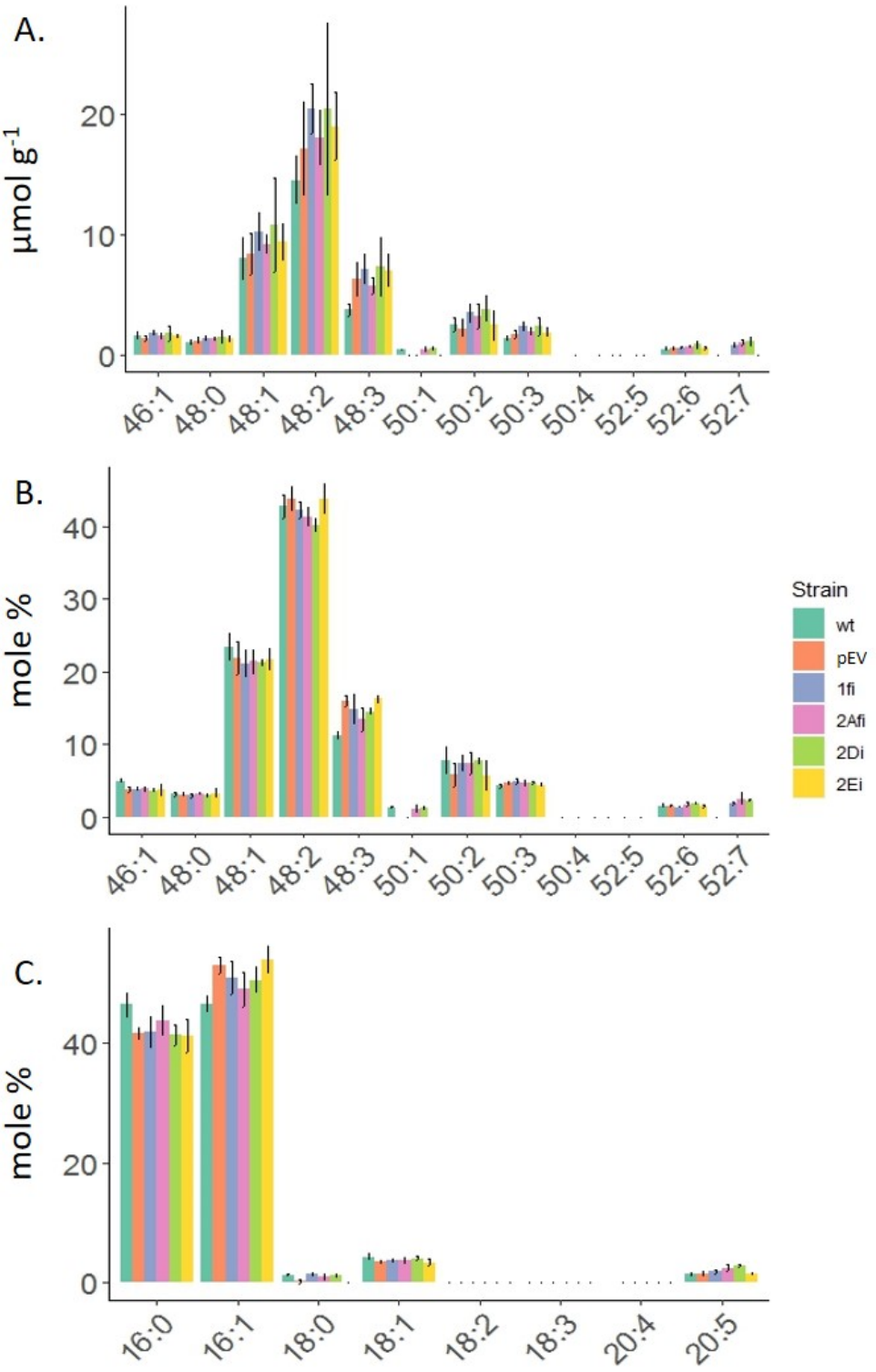


Figure 4.13 – TAG species and FA content of TAG extract from lines containing constructs DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei compared to controls. a. μmol of TAG normalised to dry cell mass. b. Mole percentage of TAG species. c. FA composition of TAG extract. wt - wild type, pEV- empty vector, 1fi- DGAT1fi, 2Afi- DGAT2Afi, 2Di- DGAT2Di, 2Ei- DGAT2Ei. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

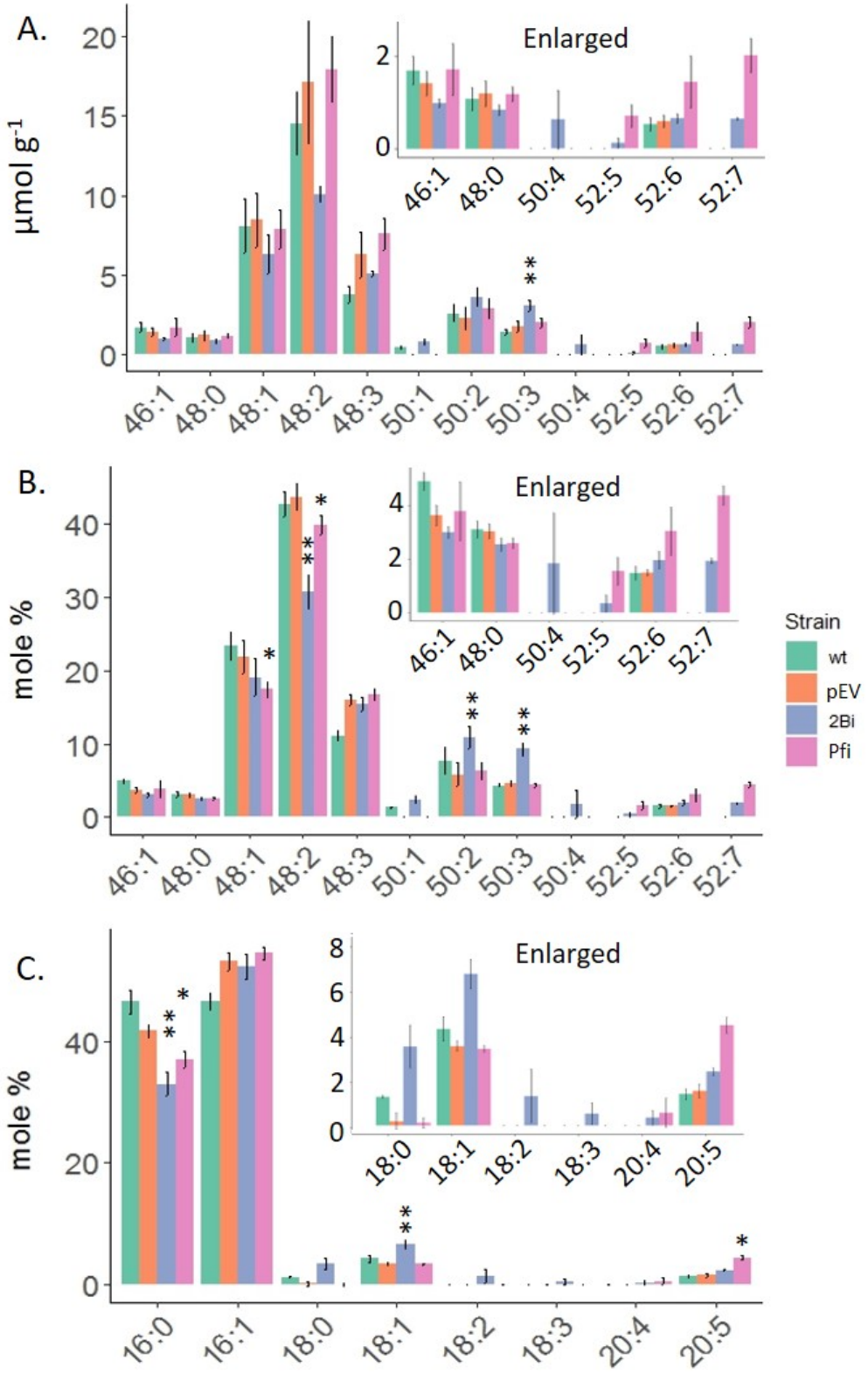


Figure 4.14 – TAG species and FA content of TAG extract from lines containing constructs DGAT2Bi and PDATfi compared to controls. a. μmol of TAG normalised to dry cell mass. b. Mole percentage of TAG species. c. FA composition of TAG extract. wt - wild type, pEV- empty vector, 2Bi- DGAT2Bi, Pfi- PDATfi. Significance was not tested for 50:1, 50:4, 52:5, 52:7, 18:0, 18:2, 18:3 and 20:4. Significance compared to pEV, ** - $p < 0.01$, * - $p < 0.05$. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

these two controls. pEV extracts contained a higher normalised amount of T48:3 than wt, as seen in Figure 4.13a and had a decreased proportion of T46:1 and increased proportion of T48:3, shown in Figure 4.13b. FA composition of TAG is shown in 4.1 and FA content was also altered, with a decreased proportion of 16:0 and 18:0 and an increased proportion of 16:1, depicted in Figure 4.13c. Dr Litvinenko did not observe variation in TAG composition between wt and pEV but only measured TAG in one pEV line. Three lines were measured in the current experiment, but the lines demonstrated similar TAG composition, as shown in 4.11C. It is therefore not clear why TAG content differed between the wt and pEV lines. Overall however, the choice of wt or pEV as the control has limited impact. Therefore, to simplify interpretation of the effect of candidate genes expression, it was decided to use pEV as the control.

Limited alterations to the TAG fraction were observed for a number of the overexpressing strains. The proportion of T48:2 and T48:3 decreased in the TAG of DGAT2Di as shown in 4.13b. However, significant changes in TAG species amount or composition were not observed for DGAT2Di nor strains DGAT1fi, DGAT2Afi, DGAT2Di or DGAT2Ei as depicted in Figures 4.13a and b. FA composition of TAG was also unaltered, as depicted in Figure 4.13C. A lack of change in the TAG fraction of these strains is supported by the GC-FID observations of Dr Litvinenko shown in 4.7B. It is however notable that using the wt as the control may alter the interpretation of changes in TAG composition, particularly for T46:1, T48:1 and T48:3 and the FAs 16:0 and 16:1.

Compared to pEV, there were clear differences in the amounts, proportions and composition of TAGs in DGAT2Bi as seen in Figure 4.14. Furthermore, where differences existed between TAG fractions of this line and pEV, they were also seen between wt and overexpressing lines. Normalised abundance of the TAG species T46:1, T48:0, T48:1, T48:2 and T48:3 decreased in DGAT2Bi extracts by 0.5, 0.5, 2.5, 7 and 1.5μmol g^{-1} respectively. However, these changes were insignificant (Tukey HSD test, $p > 0.05$). In contrast T50:3 was significantly increased by 1.5μmol g^{-1} (Tukey HSD test, $p < 0.01$). An increase of T50:2 by 1.5 μmol g^{-1} was also seen but this was insignificant. As seen in Figure 4.14b, these changes in quantity resulted in a significant decrease of 12.5% in the proportion of TAG composed of

48:2 of (Tukey HSD test, $p < 0.01$) and decreases in 46:1 and 48:1 of 0.5 and 3% although the latter changes were not significant. Significant increases in 50:2, and 50:3 (Games-Howell Test [non-parametric, equal variance not required], $p < 0.01$) of 4.5% each were observed. These changes in TAG species underpinned alterations in the FA composition of TAG as displayed in Figure 4.14c. A significant 9% decrease in the proportion of the FA 16:0 was observed (Tukey HSD test, $p < 0.01$), alongside a significant increase of 18:1 by 2.5% (Games-Howell test [non-parametric, equal variance not required], $p < 0.01$). Both significant and insignificant changes in TAG species are described above as these species are likely composed of FAs (Table 4.1 for which proportional contribution to TAG altered significantly, as illustrated by 16:0 in Figure 4.14c. It is therefore important to be aware of the TAG species that this may have caused.

The changes to TAG seen in the DGAT2Bi lines are consistent with previous measurements. A decrease in the proportion of TAG composed of the FA 16:0 and an increase of 18:1 were noted in the ESI-TQMS measurements of the relative comparison of the overexpression strains presented in Section 4.2.1. As with that previous study, there are hints that the proportion of other 18 carbon FAs were also increased in the TAG of DGAT2Bi strains but poor detection of low abundance TAG species obscures this information. GC-FID measurements of line DGAT2Bi-3 carried out by Dr Litvinenko also indicated that the proportion of 16:0 decreased while 18:1 increased as depicted in Figure 4.7. In addition, Dr Litvinenko noted a preference for PtDGAT2B for incorporation of 18:1 over 16:0 into TAG compared to the *S. cerevisiae* DGAT2 as shown in Figure 4.4. However, the current study did not provide support for a decreased incorporation of 14:0 into TAG in DGAT2Bi lines as Dr Litvinenko observed.

The alterations in TAG species seen for strain DGAT2Bi in the current study were also previously noted. A decreased proportion of T48:2 and increased proportions of T50:2 and T50:3 were observed in the ESI-TQMS measurements of the relative comparison of overexpressing lines that is presented in Section 4.2.1. This provided further support that T48:2, T50:2 and T50:3 were the altered species that underpinned the observed changes in T48 and T50 in the GC-FID data of DGAT2Bi lines presented in Section 4.2.1 and identified in the results of Dr Litvinenko presented in 4.7. There

were also indications in the current study and the relative comparison study that the minor species T50:1 and T50:4 may also be increased in TAG of strain DGAT2Bi.

Prominent alterations to the TAG fraction were also noted in strain PDATfi as shown in Figure 4.14. The species T52:6 increased by 1.5μmol g^{-1} compared to strain pEV but this was not significant. The species constitution of TAG also changed, with the proportion of T48:1 and T48:2 significantly decreased (Tukey HSD, $p < 0.05$) by 5% and 3.5% respectively. T48:0 decreased by 0.5% but this was not significant. An increase in the proportion of T52:6 was observed but this was not significant. Prominent increases in T52:5 and T52:7 were also noted, but the issues surrounding detection of low abundance species reduces the certainty of changes in these two species. Significant changes in the FA composition of TAG were observed in strain PDATfi. The proportion of 16:0 in TAGs decreased by around 4% (Tukey HSD, $p < 0.05$) and that of 20:5 increased by around 3% (Tukey HSD, $p < 0.05$).

The observations of changes to the TAG in strain PDATfi are supported by previous observations made in the relative comparison of overexpressing strains presented in Section 4.2.1 and the GC-FID results of Dr Litvinenko shown in Figure 4.7. The observations of decreased T48:1 in the current study was also observed in the ESI-TQMS measurements of the relative comparison. In addition to T48:1, T48:2 was also observed to be decreased in the current study. This supports a change in these two species in underpinning the decreased proportion of T48 observed in GC-FID results, both by Dr Litvinenko and those of the relative comparison shown in Figure 4.8. The increase in T52:6 observed in the current study was also noted in the relative comparison, as depicted in Figure 4.10a. As mentioned, it is possible that the minor species T52:5 and T52:6 were increased in the current study. An increase in T52:7 was previously noted in PDATfi in the relative comparison. T52:6 therefore contributes to the increase in the T52-56 species that was observed in the GC-FID results of Dr Litvinenko and in those of the relative comparison, as may T52:5 and T52:7. The increased proportion of the FA 20:5 in TAG was previously observed in the relative comparison as seen in Figure 4.10. In contrast, the decreased proportion of 16:0 observed in the current study has not been observed before. However, observation of a decreased proportion of 16:0 is consistent with decreases in T48 species that are typically composed of 16 carbon FAs in *P. tricornutum* (Yu

et al., 2009).

There was no difference between pEV and DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei in quantity or proportion of the majority of TAG species, as shown in Figure 4.13b. However, T50:1, T50:4, 52:5 and T52:7 were not detected in pEV extracts while in contrast, T50:1 was detected in DGAT2Afi, DGAT2Bi and DGAT2Di extracts, 52:5 was detected in extracts of DGAT2Bi and PDATfi and 52:7 was detected in those of DGAT1fi, DGAT2Afi, DGAT2Bi, DGAT2Di and DGAT2Ei, as illustrated in Figures 4.14 and 4.13. Based on the FA makeup of TAG species shown in Table 4.1, the lack of 50:1 may also be the cause of the drop in 18:0 FA in pEV extracts and the lack of 52:7 may cause a reduction in 20:5 in pEV. This could account for the increased amount of these species and FAs in the strains DGAT1fi, DGAT2Afi, DGAT2Di and DGAT2Ei (depicted in Figure 4.13) over that in pEV lines. These increases are likely artifacts as the variable detection of 50:1, 50:4, 52:5 and 52:7 across strains are most likely caused by poor detection of low abundance species, as mentioned above. The impact of this variable detection on changes to the TAGs of strains PDATfi and DGAT2Bi is dealt with in the Section 4.3 but does not alter interpretation of the TAG species and FAs that are significantly altered.

Figure 4.15 displays the mean total TAG content of the lines containing various constructs. Overall changes in mean TAG content were not observed for any of the constructs, however, there is large variation around the mean total TAG quantity of each. Interestingly, decreases in the combined normalised abundance of the TAG species T46:1, T48:0 and T48:2 in strain DGAT2Bi seem to be greater than the combined increased quantity of T50:2 and T50:3, as seen in Figure 4.14a. This would suggest that not all of the decreased species are compensated by a gain in others but contrasts with the lack of observed change in the total quantity of TAG. The cause of this discrepancy is unclear although the large variation in total TAG amount or poor detection of low abundance TAG species could be one reason.

The functional annotation of DGAT and PDAT presupposes their role in the addition of acyl chains to DAG, with the acyl chains originating from either acyl-CoA in the case of DGAT or a glycerolipid for PDAT. The overall quantity of TAG in the overexpressing lines did not change but the quantity and proportion of individual TAG species in DGAT2Bi and PDATfi lines did. I therefore hypothesised that TAG

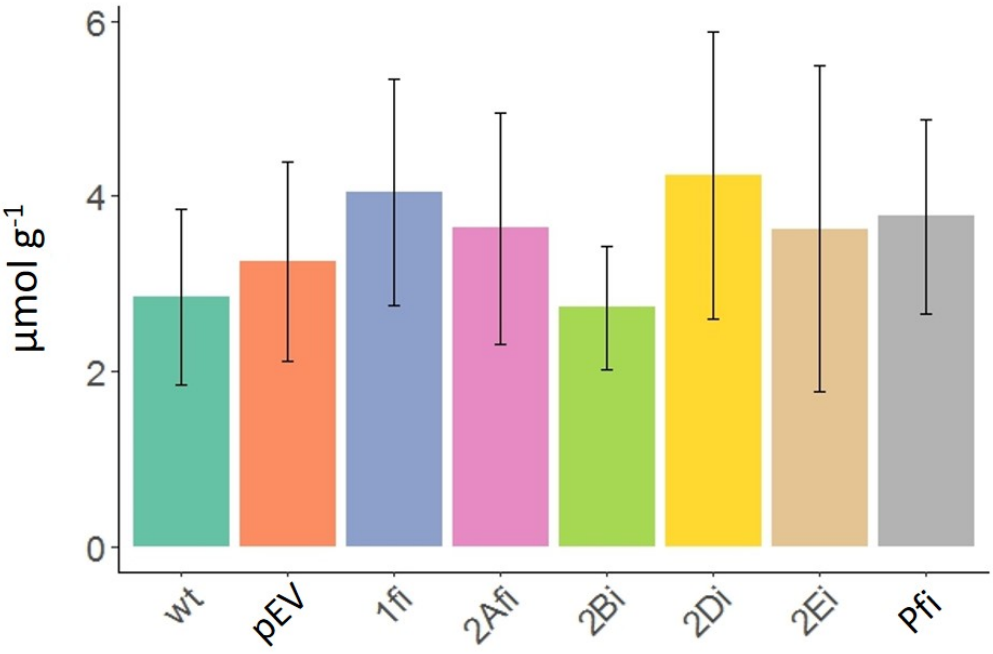


Figure 4.15 – Total TAG from *P. tricornutum* lines overexpressing DGATs and PDAT. μmol of TAG normalised to dry cell mass for lines containing wt - wild type, pEV- empty vector, DGAT1fi- DGAT1, DGAT2Afi- DGAT2A, DGAT2Bi- DGAT2B, DGAT2Di- DGAT2D, DGAT2Ei- DGAT2E, PDATfi- PDAT. 95% confidence intervals are shown. See Table 4.3 for strain descriptions.

synthesis activity was not limiting the amount of TAG accumulated and that the observed changes in TAG species were the result of substrate competition between the different overexpressed enzyme and the DGAT(s) and/or PDAT(s) natively active under nitrogen depletion. The observations of alterations in TAG in the PDATfi and DGAT2Bi strains have provided information on the acyl chains that are preferentially incorporated into TAG by PDAT and DGAT2B. In both lines, the proportion of TAG species with lower masses decreased while the proportion of those with higher masses increased. Additionally, the proportion of TAG constituted of 16:0 FA decreased in clones of both lines, the proportion of 20:5 increased in PDATfi lines while that of 18:1 increased in DGAT2Bi lines. Substrate preferences resulting in incorporation of 18:1 and 20:5 acyl chains into TAG by DGAT2B and PDAT respectively are therefore supported, disregarding whether the FAs originate from DAG, acyl-CoA or a glycerolipid. The next two Sections investigate this hypothesis and its implications by making use of previous data provided by Yu et al. (2009) on the FA composition of TAG during nitrogen depletion in *P. tricornutum*.

4.2.2.3 Interpretations of substrate preference of DGAT2B

As mentioned, overall TAG amount did not alter in transgene lines overexpressing DGAT2Bi and PDATfi but the TAG species composition did. In both strains, one or more TAG species with a larger mass increased. In DGAT2Bi, this increase cooccurred with a decrease in the quantity of T48:2. In PDATfi, the quantities of T48:1 and T48:2 were not decreased but the proportions that they contributed to TAG were. I therefore hypothesised that competition for substrates between the overexpressed enzyme and the native pool of DGAT(s) and/or PDAT(s) led to the alterations to TAG species seen in these strains. In this manner, substrates that are used to produce the lower mass TAG species in control cells are co-opted for synthesis of heavier TAG species in overexpressing strains. Under this hypothesis, there are two scenarios that could have produced the observed results. In the first scenario, the enzymes demonstrate differentiated preferences for the source of the acyl chain but share a common DAG preference with the native TAG synthesis enzymes. In the alternative scenario, DGAT2B and PDAT have preferences for particular DAG species but share with the native enzymes a preference for the same acyl chains. This section outlines the two scenarios for DGAT2B and the next section oultines them for PDAT.

Scenario 1: A acyl-CoA preference

Different combinations of DAG species and FAs may have been used to produce increased proportions of certain TAG species, or not used as a result of decreased proportions of other TAG species. To investigate the implications of these changes for the substrate specificity of DGAT2B, the TAG species that were altered by overexpression, along with the magnitude of the resulting changes, were compiled and this is shown in Table 4.7. In addition, the combinations of DAGs and FAs that could produce these TAGs were derived based on information from Yu et al. (2009) and are included for both scenarios. This enabled the identification of Those DAGs and FAs that, compared to pEV extracts, may have either been used or not used in TAG synthesis in the DGAT2Bi strain. Only the DAG species that may have contributed to the observed alterations in TAG, based on limitations of total carbon and saturation number, were included in the table.

Table 4.7 – Changes to TAG species and FA content in DGAT2Bi lines and likely FA content and DAG species used or released under the two substrate preference scenarios. Changes were calculated relative to pEV. TAG FA composition composition from Yu et al. (2009).

			Change			Acyl chain preference				Preference for C16/C18 DAGs			
						DAG species usage		FA usage		DAG species usage		FA usage	
	FA species	TAG species	µmol g^{-1}	%	FA composition	Decreased	Increased	Decreased	Increased	Decreased	Increased	Decreased	Increased
2Bi	16:0			-9									
	18:0			+2.5									
	18:1			+2.5									
		46:1[b]	-0.5	-0.5	14:0/16:0/16:1	16:0/16:1		14:0		16:0/16:1		14:0	
										14:0/16:0		16:1	
										14:0/16:1		16:0	
		48:0[b]	-0.5		16:0/16:0/16:0	16:0/16:0		16:0		16:0/16:0		16:0	
		48:1[b]	-1.5		16:0/16:0/16:1	16:0/16:1		16:0		16:0/16:1		16:0	
						16:0/16:0		16:1		16:0/16:0		16:1	
		48:2	-7	-12.5	16:0/16:1/16:1	16:0/16:1		16:1		16:0/16:1		16:1	
						16:1/16:1		16:0		16:1/16:1		16:0	
		48:3[b]	-1.5		16:1/16:1/16:1	16:1/16:1		16:1		16:1/16:1		16:1	
		50:1[c]		+1	16:1/16:0/18:0		16:1/16:0		18:0		16:1/18:0		16:0
											16:0/18:0		16:1
					16:0/16:0/18:1		16:0/16:0		18:1		16:0/18:1		16:0
		50:2	+1.5	+4.5	16:1/16:1/18:0		16:1/16:1		18:0		16:1/18:0		16:1
					16:1/16:0/18:1		16:1/16:0		18:1		16:1/18:1		16:0
											16:0/18:1		16:1
		50:3	+1.5	+4.5	16:1/16:1/18:1		16:1/16:1		18:1		16:1/18:1		16:1

[b]Change in quantity and proportion is insignificant

[c]Significance of change not tested

In the first scenario, DGAT2B would show a preference for the acyl chain that it incorporates onto DAG. In this scenario, it is competition with the natively expressed DGATs and PDATs for these DAG species that prevents an increased total accumulation of TAG. In other words, the amount of 16:1/16:0 and 16:1/16:1 DAG is limiting. It is assumed that the various DGATs have different acyl-CoA preferences. This is shown schematically in Figure 4.16. The Figure illustrates how the native combination of DGATs and PDATs in the wt would each use a proportion of the 16:1/16:0 and 16:1/16:1 DAG pools and attach acyl chains to these DAGs in line with their substrate specificities. Overexpression of DGAT2B in the DGAT2Bi strain results in DGAT2B utilising a higher proportion of the 16:1/16:0 and 16:1/16:1 DAG pools, thereby incorporating more 18:1 onto these backbones, at the expense of the acyl chains preferred by the native mix of DGATs and PDATs.

When DGAT2B is overexpressed, an increased proportion of 18:1 is detected in TAG and T50:2 and T50:3 increase. In contrast, the proportion of 16:0 and T48:2 decrease. Although these changes are the only statistically significant alterations to TAG composition in strain DGAT2Bi, changes in several other species were observed as well. The proportions of T46:1, T46:2, T48:0, T48:1 and T48:3 in TAG were decreased. Furthermore, despite the uncertainty of measurements of low abundance TAG species, the proportion of T50:1 seemed to be increased, similar to its likely constituent acyl chain, 18:0. These observations of T50:1 and 18:0 are weakly supported by the results presented in Figure 4.10a and b from the comparison of TAG between overexpressing lines. Furthermore, Dr Litvinenko previously observed preliminary results indicating that although PtDGAT2B clearly has a preference for incorporating 18:1 over 16:0 when overexpressed in *P. tricornutum* and *S. cerevisiae*, it also alters the incorporation of 14:0 and potentially 16:1, as depicted in Figures 4.7C and 4.4C. Several FA compositions of each of these TAG species have been observed in *P. tricornutum* (Yu et al., 2009). As a result, I included alterations in these FAs and TAG species in a consideration of the DAG and FA used by DGAT2B below.

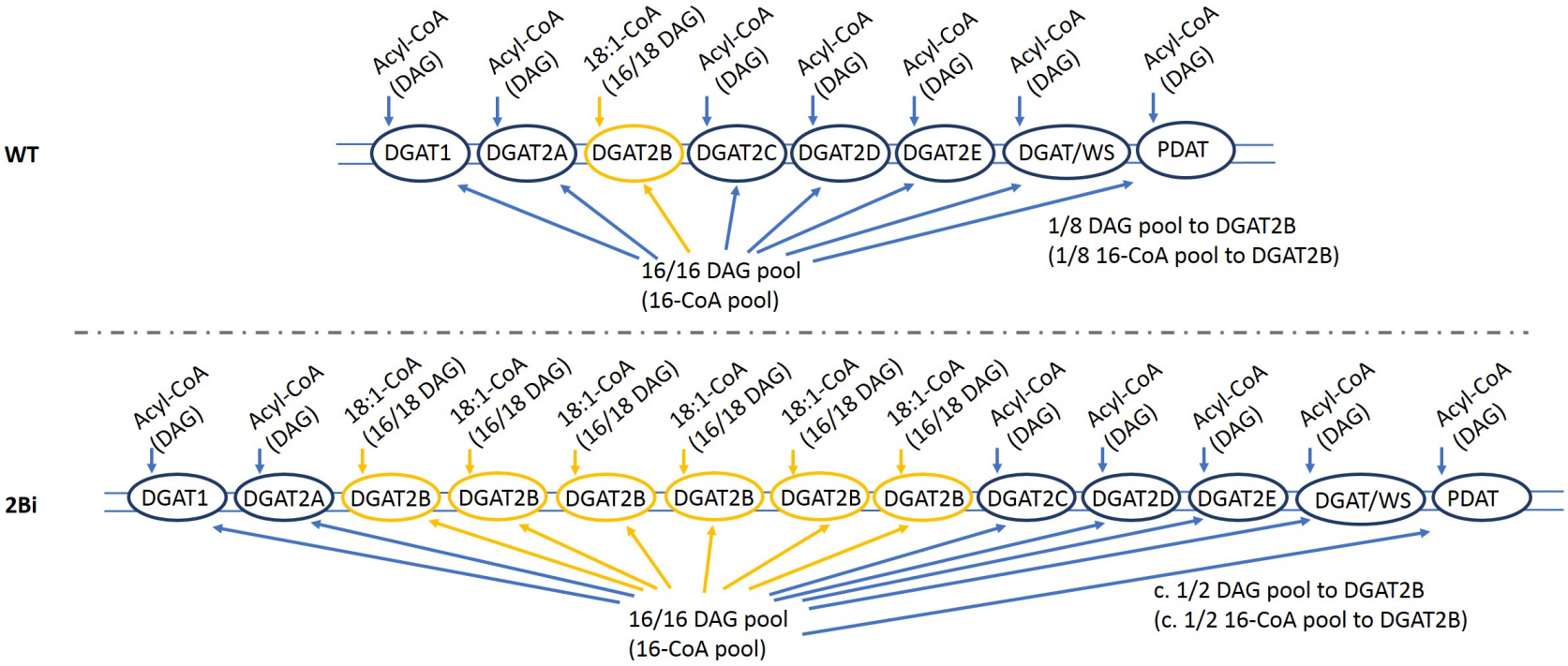


Figure 4.16 – Schematic of DGAT2B demonstrating a substrate preference for 18:1-CoA (or 16/18 DAG). The schematic shows two scenarios. The first scenario illustrates DGAT2B competing for 16:n/16:n DAG with the other DGATs and PDAT in the cell. In this scenario, DGAT2B prefers to incorporate 18:1-CoA onto 16:n/16:n DAG. Increased expression of DGAT2B in the DGAT2Bi lines is therefore expected to lead to incorporation of more 18:1 acyl chains onto 16:n/16:n DAG than in the wt. The second scenario, shown in parentheses, illustrates how DGAT2B competes for 16:n-CoA (n= 0 or 1) with the other DGATs and PDAT in the cell. In this scenario, DGAT2B prefers to use 16:n/18:1 DAG for the incorporation of 16:n-CoA. Increased expression of DGAT2B in the DGAT2Bi lines is therefore expected to lead to incorporation of more 16:n-CoA onto 16:n/18:1 DAG than in the wt. The amount of each DGAT or PDAT in the cell is symbolised by the number of labelled enzymes but do not represent the true proportion in the cell. DGAT and PDAT enzymes are shown situated in a single membrane but in reality may be localised in different compartments.

Regarding the altered proportions of FAs, an increase in 18:0, putatively observed in the current study, is supported by the results in Table 4.7, as it could contribute to T50:1 and T50:2 species. Furthermore, the decrease in 14 carbon acyl chains, observed by Dr Litvinenko for line DGAT2Bi-3 also finds support in the small decreases in T46:1 and T46:2. As detailed in Table 4.7, the TAG species that were decreased in extracts from DGAT2Bi lines could have given rise to 14:0/16:0, 14:0/16:1, 16:0/16:0, 16:0/16:1 and 16:1/16:1 DAG species. However, a 14/16 DAG that could arise from T46:1 or T46:2, could not produce a TAG species with 50 carbons and 3 or fewer double bonds using the FAs in *P. tricornutum* and so is unlikely to contribute to increased T50:1, T50:2 or T50:3. T46:1 and T46:2 DAGs are therefore not included in the table. The most likely FA compositions of T50:1, T50:2 or T50:3 contain 18 carbon FAs. Each of 16:0/16:0, 16:0/16:1 and 16:1/16:1 DAGs could be incorporated into one or more of these TAGs. However, the proportion of 16:0 in TAG decreases when DGAT2B is overexpressed and the TAG species with the largest proportional decrease, T48:2, could give rise to either 16:0/16:1 or 16:1/16:1. It is therefore likely that DGAT2B preferentially uses 16:1/16:1 and/or 16:1/16:0 over 16:0/16:0. The species that contribute the largest proportional increase to TAG are T50:2 and T50:3. Both T50:2 and T50:3 increase the proportion they constitute by 4.5%. T50:3 is likely only produced by 16:1/16:1 DAG. T50:2 may be produced by both 16:1/16:1 and 16:1/16:0 DAG. The significant increase in the proportion of 18:1 in TAG, that would combine with 16:1/16:0 to produce T50:2, indicates that this DAG species is used. Therefore, the information on the T50 TAG species also supports the use of 16:1/16:1 and 16:1/16:0 DAG species, but the contribution of each is not clear.

Scenario 2: A DAG preference

The alternative to a preference for a specific acyl-CoA is for DGAT2B to have a preference for specific DAG species. In this scenario, the increase in the proportion of 18:1 in TAG would derive from the use of DAG containing an 18 carbon acyl chain. As shown in Figure 4.1, the likely FA compositions of the proportionally increased T50:2 species are 16:1/16:1/18:0 and 16:1/16:0/18:1. The DAG species containing an 18:1 acyl chain would therefore be 16:1/18:1 or 16:0/18:1 and use either 16:0-CoA or 16:1-CoA. T50:3 was also proportionally increased and its likely composition is 16:1/16:1/18:1, hence it would require a 16:1/18:1 DAG and 16:1-CoA. Assuming

that the DGATs and PDATs demonstrate different preferences for DAG species, it is then competition for the acyl-CoA pool that, despite overexpression of DGAT2B, prevents increases in total TAG accumulation. The information in parentheses in Figure 4.16 illustrates how an increase in the expression of DGAT2B would increase the proportion of the 16:0-CoA and 16:1-CoA pool into 16/18 DAGs at the expense of DAG molecules preferred by other DGATs and PDAT.

Strikingly, in this scenario the DAG species that would not be used by the proportionally decreased TAG species are not the same as those DAG species used to synthesise the proportionally increased TAG species, as depicted in Table 4.7. As mentioned, there is support for proportional increases in 18:0 and of T50:1. Therefore, in addition to the DAG species 16:1/18:1 or 16:0/18:1 that may be used to produce T50:2 and T50:3, the DAG species 16:1/18:0 and 16:0/18:0 may be incorporated into T50:1 and 16:1/18:0 into T50:2. However, the major DAG species used under this scenario would be 16:0/18:1 and 16:1/18:1. The FAs that would be used in these proportionally increased TAG species are 16:0 and 16:1 and these could be supplied by the FAs that would not be used in the synthesis of the TAG species that are proportionally decreased.

4.2.2.4 Interpretations of substrate preference of PDAT

Scenario 1: An acyl-glycerolipid preference

Analogous reasoning can be applied to the PDAT overexpressing lines but here the preference is either for a specific acyl chain on a membrane glycerolipid or for a specific DAG species. Regarding the acyl-glycerolipid preference, increased expression of PDAT would be expected to lead to the incorporation of more 20:5 onto DAG and the proportion of DAG used by native DGATs would be decreased, as illustrated in Figure 4.17.

Table 4.8 contains the TAG species that were altered in the PDATfi lines and their likely FA composition as determined by Yu et al. (2009). The proportional decrease in T48:1 and T48:2 were statistically significant while the other included changes were not. However, there was support for a decrease in T48:0 and increases in T52:5 and T52:7 in strain PDATfi in the current study and the relative comparison of overexpressing lines shown in Figure 4.10. Furthermore, although increases in

Table 4.8 – Changes to TAG species and FA content in strain PDATfi lines and likely FA content and DAG species used or released under the two substrate preference scenarios. Changes were calculated relative to pEV. TAG FA composition composition from Yu et al. (2009).

			Change			Acyl chain preference				Preference for C16/C20 DAGs			
						DAG species usage		FA usage		DAG species usage		FA usage	
	FA species	TAG species	µmol g^{-1}	%	FA composition	Decreased	Increased	Decreased	Increased	Decreased	Increased	Decreased	Increased
Pfi	16:0			-4									
	20:5			+3									
		48:0[b]		-0.5	16:0/16:0/16:0	16:0/16:0		16:0		16:0/16:0		16:0	
		48:1		-5	16:0/16:0/16:1	16:0/16:1		16:0		16:0/16:1		16:0	
						16:0/16:0		16:1		16:0/16:0		16:1	
		48:2		-3.5	16:0/16:1/16:1	16:0/16:1		16:1		16:0/16:1		16:1	
						16:1/16:1		16:0		16:1/16:1		16:0	
		52:5[c]			16:0/16:0/20:5		16:0/16:0		20:5		16:0/20:5		16:0
		52:6[b]	+1.5	+1	16:1/16:0/20:5		16:1/16:0		20:5		16:1/20:5		16:0
											16:0/20:5		16:1
		52:7[c]			16:1/16:1/20:5		16:1/16:1		20:5		16:1/20:5		16:1

[b]Change in quantity and proportion is insignificant

[c]Significance of change not tested

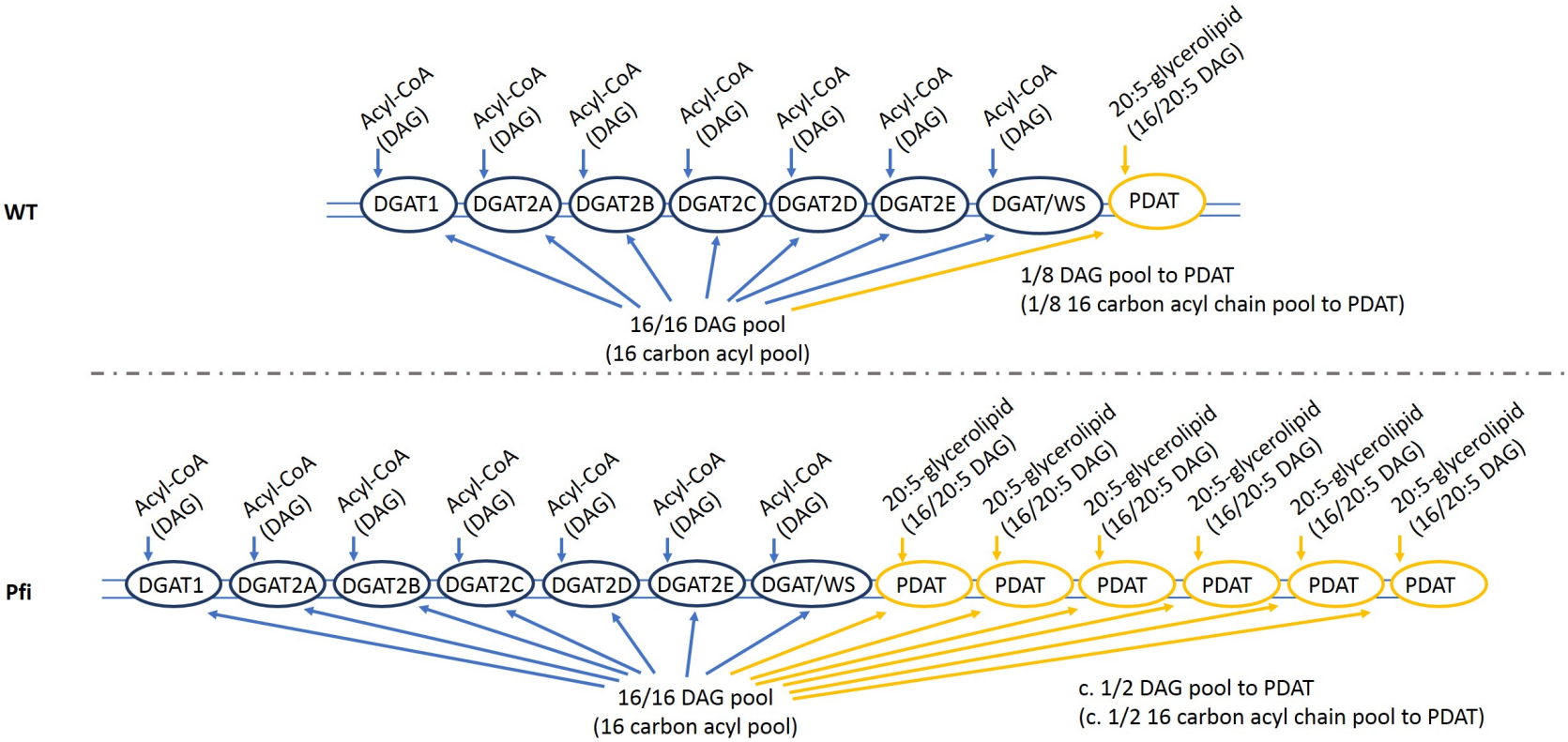


Figure 4.17 – Schematic of PDAT demonstrating a substrate preference for 20:5-glycerolipid (or 16/20 DAG). The schematic shows two scenarios. The first illustrates PDAT competing for 16:n/16:n (n= 0 or 1) DAG with the DGATs and PDAT in the cell. In this scenario, PDAT prefers to incorporate 20:5 onto 16:n/16:n DAG. Increased expression of PDAT in the PDATfi lines is therefore expected to lead to incorporation of more 20:5 acyl chains onto 16:n/16:n DAG than in the wt. The second scenario, in parentheses, illustrates how PDAT competes for 16:n acyl chain pool (n= 0 or 1) with the other DGATs and PDAT in the cell. PDAT uses acyl-glycerolipids as a substrate while DGAT uses acyl-CoA and so this competition is not direct. In this scenario, PDAT prefers to use 16:n/20:5 DAG for the incorporation of 16:n acyl chains. Increased expression of PDAT in the PDATfi lines is therefore expected to lead to incorporation of more 16:n onto 16:n/20:5 DAG than in the wt.

T52:5 and T52:7 are made less certain by the poor detection of low abundance species, the observed increase is substantial. These species were therefore included in Table 4.8. The DAG species and FAs that could constitute these species were derived and are also shown in the table. T48:1 and T48:2 are synthesised from 16:0/16:1, 16:0/16:0 or 16:1/16:1 DAG species and so their proportional decrease in the TAG of PDATfi strains means smaller amounts of these DAG species are used compared to pEV. These are also the DAG species that would be used in the production of the proportionally increased species T52:5, T52:6 and T52:7, in conjunction with 20:5-glycerolipid. Therefore, under this scenario, these are the three DAG species to which PDAT would add 20:5.

Scenario 2: A DAG preference

The alternative is for PDAT to have a preference for DAGs that contain a 20:5 acyl chain. Under this scenario, the DAG species used to produce T52:5, T52:6 and T52:7 would be 16:0/20:5 and 16:1/20:5. These would be combined with 16:1 or 16:0 from a glycerolipid. Overexpression of PDAT would therefore lead to incorporation of a higher proportion of the 16:0 and 16:1 from a glycerolipid into the DAG species preferred by PDAT, rather than into the DAGs preferred by the natively expressed DGATs. It should be noted that competition with DGATs for the 16 carbon acyl chains would be indirect, as DGATs make use of acyl-CoA as their acyl donor as illustrated in Figure 4.17. Additionally, as indicated in Table 4.8, the DAG species that would be used in synthesising the increased TAG species are different to those that are less used by the decreased species. The 16:0 and 16:1 acyl chains that are used under this scenario by the increased species are the same as those not used by the decreased species.

4.3 Discussion

4.3.0.1 Summary of functional knowledge of DGAT1, DGAT2A, DGAT2D and DGAT2E

In the work presented in this Chapter, the effects of individual overexpression of *P. tricornutum* DGAT1, DGAT2A, DGAT2B, DGAT2D, DGAT2E and PDAT on TAG were investigated. DGAT1 (Guihéneuf et al., 2011) and DGAT2D were previously

observed to complement TAG synthesis in TAG deficient yeast (Litvinenko, 2016) while DGAT2D and DGAT2A increased TAG content after overexpression under replete conditions in *P. tricornutum* (Niu et al., 2013; Dinamarca et al., 2017). Despite this, in the current study, *P. tricornutum* lines transformed with DGAT1, DGAT2A and DGAT2D in *P. tricornutum* did not have altered TAG fraction under nitrogen deplete conditions. Nor did lines containing DGAT2E but evidence of DGAT2E has not previously been noted. A similar lack of alteration to TAG in the same DGAT1, DGAT2A, DGAT2D and DGAT2E strains was noted by Dr Litvinenko. An exception to this was DGAT2Afi-6, which had a unique TAG profile among DGAT2Afi lines in the current study. Excepting DGAT2Afi-6, the lack of expression or activity of these enzymes in *P. tricornutum* could explain these observations. Another possibility is that one of these enzymes is already the main DGAT involved in TAG synthesis in nitrogen depletion. Evidence for this is not available from comparisons of TAG profiles produced in current overexpression studies (Dinamarca et al., 2017; Niu et al., 2013). A third option is that use of pEV as the control masked changes in some of these lines. Further analysis of the current strains, to include comparisons of TAG between wt and pEV and gene expression between transgene containing lines and controls would be useful. Additionally, work involving gene knockout studies by the application of CRISPR-Cas9 would provide insight into the contribution of DGATs and PDAT to TAG accumulation during nitrogen depletion. Furthermore, such strains would also enable the investigation of the contribution of these enzymes to TAG accumulation in other stress conditions.

4.3.0.2 Summary of DGAT2B function

DGAT2B was previously observed to act as a functional DGAT by Gong et al. (2013) and Dr Litvinenko previously noted that when DGAT2B was overexpressed in *P. tricornutum*, the proportion of T48 in TAG decreased and that of T50 increased. ESI-TQMS measurements presented in Section 4.2 indicate that the major TAG species responsible for these changes are T48:2, T50:2 and T50:3. The detected FA composition of TAG was altered as a result and a reduced proportion of 16:0 and increased 18:1 were identified as the two major alterations. These alterations support DGAT2B having a greater preference for incorporating 18:1 into TAG

and a lower preference for 16:0 compared to the natively expressed mixture of DGATs and PDATs in *P. tricornutum* that are expressed under nitrogen depletion. Notably, this preference for 18:1 contrasts with reports of PtDGAT1 demonstrating a preference for 16:0 and 18:0 over DGAT2D from *S. cerevisiae* (Guihéneuf et al., 2011). Previous expression of DGAT2B in yeast and measurements of line DGAT2Bi-3 TAG FAMEs (Gong et al., 2013; Litvinenko, 2016) support these DGAT2B substrate preferences. However, the current study provided the first robust evidence that, relative to the natively expressed TAG synthesis enzymes, DGAT2B has a preference for incorporating 18:1 into TAG to produce the species T50:2 and T50:3. These may not be the only TAG alterations to occur in the DGAT2Bi strain. Issues with detection of low abundance TAG species meant that alterations to a number of T50 TAG species could not be confidently identified. These T50 species contain 18 carbon acyl chains, as indicated in Table 4.1 and so alterations to these acyl chains may also have been missed. The current measurements indicated that proportional alterations to T50:1 and 18:0 may be as substantial as for T50:2, T50:3 and 18:1 and so it would be useful to carry out remeasurement of the lipid extracts using ESI-TQMS with a focus on detection of low abundance species. Quantification of 14 carbon acyl chains in TAG should also be carried out in the same experiments, to assess previous observation of a decrease in the contribution of this acyl chain.

The information on the relative preference of DGAT2B for 18:1 adds to the knowledge of *P. tricornutum* DGAT substrate specificities provided by the acyl chain preference of DGAT1 (Guihéneuf et al., 2011). However, substrate specificities of the remaining DGATs still need to be characterised, in addition to the further characterisation required for DGAT2B and DGAT1. Recently, a study of *N. oceanica* identified three NoDGATs with complementary substrate preferences (Xin et al., 2017). The ancestral origins of these DGAT isoforms were also identified. *N. oceanica* is a stramenopile and likely originates from the same secondary endosymbiosis as *P. tricornutum* (see Figure 1.1). This relationship indicates that DGAT isoforms in distantly related species such as these may share common ancestry. Comparison of current and future phylogenies of DGAT genes therefore hold potential for bringing together DGAT functional data that is often incomplete in an individual species. To what extent this is possible could be investigated by comparing functionally

characterised orthologues.

4.3.0.3 Summary of PDAT function

Characterisation of PtPDAT has not previously been published. The work presented in this Chapter therefore continues the first characterisation of this enzyme. Dr Litvinenko previously observed that overexpression of PDAT in the PDATfi strain led to decreased proportions of T48 and increased proportions of T52-T56 species. The ESI-TQMS results presented here provide robust support for these alterations resulting from changes to T48:1, T48:2 and T52:6. As a result, the proportion of 16:0 in TAG decreases and that of 20:5 (EPA) increases. As for strain DGAT2Bi, repeated ESI-TQMS measurements of PDATfi extracts with a focus on detection of low abundance species may enable alterations to other high mass TAG species, such as T52:5 and T52:7 to be confirmed. Additionally, investigation of the full length PDAT, containing the first 25 amino acids, will be required.

Identification of an enzyme that increases the proportion of the LC-PUFA EPA in TAG is of interest to biotechnology, although PDATfi increases incorporation of EPA by only a few percent (Figure 4.14c) in the conditions used. Engineering of the strain may enable this to be improved. Assuming that EPA is attached to a glycerolipid and not a DAG, the identification of a preference for incorporation of EPA also provides insight into the membrane lipids from which PDAT may source this acyl chain. PDATs use acyl chains from the sn-2 position of glycerolipids and in *P. tricornutum* the lipids that contain EPA at the sn-2 position are PC, DGTA and PE (Abida et al., 2015). This suggests that strain PDATfi may contain altered levels of PC, DGTA and PE and this could be investigated using ESI-TQMS. Alterations in these lipids in PDATfi relative to the other overexpressing lines, particularly DGTA, were noted in 4.2.1 and this suggests these measurements may be productive. PC, DGTA and PE are predominantly extraplastidial lipids. Furthermore, in *P. tricornutum*, major species of MGDG contain EPA at the sn-1 position and therefore similar PDAT activity on MGDG to that observed in *C. reinhardtii* (Yoon et al., 2012) is unlikely to be carried out by PtPDAT (Phatr3_J8860). However, in addition to the investigation of the lipids identified above and as mentioned in Section 5.4, it is important to experimentally characterise location of PtPDAT to better understand

the substrates it is exposed to.

4.3.0.4 Considering the substrate preferences of DGAT2B and PDAT: a preference for specific acyl chains on a donor molecule or a preference for specific DAG molecules

Two general substrate preference scenarios were outlined in Section 4.2.2.3 that could explain the alterations in TAG in both PDATfi and DGAT2Bi under nitrogen depletion. In the first scenario, the overexpressed enzymes had a preference for an acyl chain attached to either CoA or a glycerolipid and would compete with natively present enzymes for 16/16 DAG. In the second scenario, the enzymes would be specific for particular DAG molecules and compete for 16 carbon acyl chains to add to these. Based on likely DAG and FA constituents of the TAG species that were altered in each strain, either scenario is possible.

Acyl-CoA composition of wt cells at stationary phase indicate that acyl-CoAs are present in the following approximate amounts: 30% 20:5, 20% 16:0, 20% 16:1, 5% 18:0 and 7.5% 18:1 (Hamilton et al., 2014). 20:5 is also present in substantial proportions in some of the major lipids in the cell. It is present at the sn-2 position in around 24%, 11% and 20% of PE, DGTA and PC respectively (Abida et al., 2015). These are the lipids from which PDAT would likely source EPA under the first scenario. There are therefore substantial amounts of acyl-CoAs and a selection of lipids from which DGAT2B and PDAT could respectively draw 18:1 and 20:5 for incorporation into TAG. This supports the first scenario, in which DGAT2B and PDAT demonstrate substrate preference for C18-CoA and 20:5-glycerolipids respectively. Considering the second scenario, the DAG pool has been observed to have the following composition: 16:1 c.45%, 16:0 c.25%, 14:0 c.8%, 18:0 c. 4%, 18:1 c.3%, 20:5 c. 2% (Abida et al., 2015) under replete conditions. If DAG profiles are similar under nitrogen depletion then 16/20 and 16/18 DAGs are likely to be minor species. Despite this, given the small increases in TAG species that are seen, it is not clear that large quantities of DAG would be needed. Both scenarios would theoretically result in production of excess molecules that would not be used in the synthesis of the proportionally decreased TAG species. For PDAT, the second scenario does seem less likely as DGAT and PDAT do not directly compete for the

same substrate and so it is not clear how competition for 16 carbon acyl chains would prevent increases in overall TAG accumulation. However, overall it is not clear whether DGAT2B and PDAT demonstrate preferences for the acyl chain donor substrate or DAG. However, investigation of broader lipid metabolism may provide some insight. For example, measurement of total FAs in the cell would provide information on metabolic responses to overexpression of the enzymes. Other information that may provide insight are the position of FAs on the glycerol backbone of TAG and acyl-CoA pool in these overexpression lines.

4.3.0.5 A TAG synthesis deficient *P. tricornutum* strain

As mentioned above, knock out of individual DGAT and PDAT enzymes would enable identification of their roles under specific culture conditions. Expanding on this, a *P. tricornutum* strain in which all DGAT and and PDAT enzymes are knocked out would be useful for testing substrate specificity of TAG synthesis enzymes of LC-PUFA producing organisms. Furthermore, this would provide a useful biotechnological strain from which to tailor FA composition of TAG through selective introduction of DGATs and PDATs. The pursuit of a strain that is fully deficient in TAG synthesis may be difficult due to the physiological roles that TAG plays. However, strains with incrementally less ability to produce TAG would be commensurately more useful for production of tailored oils or assessment of DGAT and PDAT substrate specificity. The identification of a preference for incorporating EPA into TAG also opens up the possibility of using PtPDAT to engineer a strain to accumulate more EPA in TAG. This will depend upon whether PDAT has a substrate preference for 20:5-glycerolipid or 20:5 containing DAG but engineering a strain to produce more PDAT substrates would be a useful step in either case. Combining PDAT with a TAG synthesis deficient strain may also enable more cellular flux of carbon to be directed into TAG as EPA rather than shorter, more saturated acyl chains.

5

Investigating the contribution of chloroplast membrane lipid recycling to TAG accumulation in *P. tricornutum*

5.1 Introduction

Although TAG accumulation is majorly caused by *de novo* synthesis of FAs during nitrogen depletion, a number of strands of evidence support a contribution of remodeling of chloroplast membrane lipids. Within *P. tricornutum*, MGDG is decreased during nitrogen depletion and TAG accumulates (Abida et al., 2015; Levitan et al., 2015). Adding to this, an increase in EPA in TAG of a magnitude that corresponded to the decrease in MGDG has been observed in one study (Abida et al., 2015), supporting a transfer of acyl chains. Furthermore, in addition to diatoms, a decrease in MGDG and a concomitant increase in TAG has been observed in green algae, land plants and eustigmatophytes. Indeed, a movement of acyl chains from MGDG to TAG during nitrogen depletion has been observed for *C. reinhardtii* and is highly likely in *A. thaliana* (Li et al., 2012; Lippold et al., 2012). The role of MGDG in proper structuring of the light harvesting complexes and photosystems highlights the likely function of decreased MGDG levels in reducing light capture and electron excitation, preventing the production of damaging reactive oxygen species (ROS) (Nakamura and Li-Beisson, 2016; Du and Benning, 2016). Collectively, this supports the likely widespread occurrence of MGDG to TAG recycling during nitrogen depletion and hence the presence of the phenomena in *P. tricornutum*.

Presented in this chapter are experiments to investigate whether MGDG to TAG recycling occurs during nitrogen depletion in *P. tricornutum*. Also presented is work undertaken to identify and characterise enzymes that are candidates for catalysing this process. It was reasoned that identification of MGDG to TAG recycling and the involved enzymes would not only further understanding of the response to nitrogen depletion but also provide biotechnologically useful information. For example, the enzymes may be useful in engineering the accumulation of TAG or targeting EPA to TAG.

5.1.1 Investigating sources of triacylglycerol production under nitrogen depletion

This work focused on inhibiting fatty acid synthesis and measuring TAG accumulation during nitrogen depletion. Simply, if TAG is produced by just two processes during

nitrogen depletion, completely preventing one should leave TAG produced by the alternative mechanism. This rationale has been used in previous literature to identify the origin of TAG accumulation during nitrogen starvation in *C. reinhardtii*, specifically by inhibiting *de novo* FA synthesis (Fan et al., 2011), which plays the major role in TAG accumulation during nitrogen depletion in *P. tricornutum*.

FA synthesis can be prevented by FAS inhibitors such as cerulenin and triclosan. These chemicals are specific inhibitors of Ketoacyl-ACP synthase I (FabB), ketoacyl-ACP synthase II (FabF) and Enoyl-ACP reductase I (FabI) respectively. Both inhibitors block substrate access to the active site. Cerulenin does this through the formation of covalent bonds (Zhang et al., 2006) and triclosan carries this out by binding tightly through non-covalent interactions (Levy et al., 1999). FabB, FabF and FabI are members of the type II fatty acid synthesis pathway that consists of discrete enzymes. It is found in bacteria, plants and ochrophytes such as *P. tricornutum* (Ryall et al., 2003). When FabB and FabF are inhibited by cerulenin, fatty acid synthesis cannot occur past a chain length of four as illustrated in Figure 5.1. When FabI is inhibited by triclosan the final reduction to produce acyl-ACP cannot proceed. I therefore expect that no TAG accumulation based on *de novo* synthesised FAs would occur in the presence of these inhibitors.

Experimentally however, the effects of cerulenin and triclosan addition on *P. tricornutum* have not been well characterised. When *P. tricornutum* has been plated with cerulenin, reduced colony counts were observed with increasing concentrations of cerulenin and entirely cease to form colonies in the presence of between 60 and 80μM (Kaur, 2014). Little is published on the effects of triclosan on diatoms. Although, cerulenin and triclosan are thought to be specific inhibitors of FA synthesis, indirect chemical and physical interactions could result from the inhibitors' structures (Borrull et al., 2015; Fretland and Omiecinski, 2000). Furthermore, inhibiting FA synthesis may have downstream and systemic effects of its own. These possibilities therefore need to be taken into account during inhibition experiments. Recording cell viability and rate of division provides broad information on the the impact of inhibitors upon a cell while supplementing cells with FAs has been previously used to confirm that TAG synthesis still functions (Fan et al., 2011), supporting overall metabolic functioning despite the presence of inhibitors.

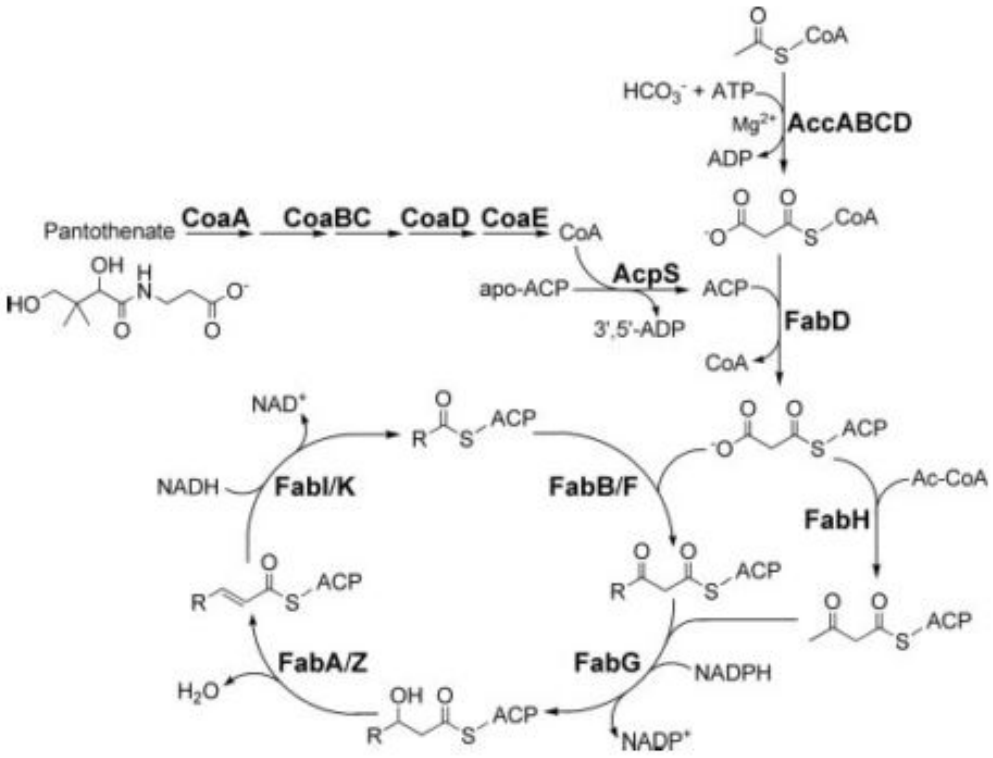


Figure 5.1 – Type II fatty acid synthesis pathway. In Type II FA synthesis, acetyl-CoA is converted to malonyl-CoA by ACC, followed by transacylation to ACP. Four monomeric enzymes constitute the Type II FAS system, catalysing a cyclical sequence of condensation, reduction, dehydration and a further reduction, in which two carbons from malonyl-ACP are transferred onto the growing acyl chain with each round. Figure reproduced from (Zhang et al., 2006)

To detect TAGs produced from MGDG components, it is essential that FA synthesis is completely inhibited. Cerulenin was chosen as the main inhibitor given its prior use in microalgae. To ensure complete inhibition, increasing concentrations of cerulenin were added with the expectation that the amount of FAs synthesised and accumulated in TAG, would decrease proportionally. If multiple mechanisms contribute to TAG synthesis then the amount of TAG accumulated should remain constant at a none-zero value, despite further addition of cerulenin.

The multiple inhibitor concentrations and controls required measurement of a large number of samples. Therefore, Nile red staining was used as it is a high throughput method of TAG measurement. Nile red is a dye that fluoresces much more strongly in hydrophobic enviroments compared to hydophilic ones and good correlation between Nile red fluorescence and neutral lipid content has been observed for a range of microalgae (Rumin et al., 2015). However, Nile red is capable of

staining other cellular constituents and therefore it is important to tailor the applied excitation and detected emission wavelengths to make use of its metachromatic properties, thereby increasing the stain's specificity to the desired lipid. Furthermore, a number of staining and measurement parameters need to be optimised to ensure measured fluorescence is proportional to TAG content.

5.1.2 Identifying and characterising enzyme candidates for production of TAG from MGDG under nitrogen depletion

Concurrent to work investigating the contribution that chloroplast membrane lipids make to TAG accumulation during nitrogen depletion, work on identifying enzyme candidates that may catalyse this role was begun. I reasoned that, regardless of the outcome of investigations into the origin of TAG, identifying and characterising novel lipid metabolism enzymes and thereby expanding our knowledge of this model diatom would be useful in its own right.

One common method of associating genes to a process is to utilise reverse genetic techniques to directly manipulate specific sequences and thereby investigate the roles of a list of candidate genes (Griffiths et al., 2000). The effort required to identify candidate genes has been much reduced in recent years due to the availability of bioinformatic resources. However, often an individual technique will not provide well supported candidates and so multiple must be used. One method for identifying candidate genes involves searching for orthologues to characterised genes via sequence similarity (Nichio et al., 2017). This enables function to be predicted for gene models identified through genome sequencing. The basic local alignment tool (BLAST) algorithm is one of the commonly used tools for identifying orthology (Nichio et al., 2017). However, a challenge to identifying orthologues in diatoms results from the distant phylogenetic relationship of diatoms to better studied photosynthetic model organisms, such as *C. reinhardtii* and *A. thaliana*.

A complementary method for predicting the function of a gene model is to use annotation associated with protein families, domains and functional sites to predict function. The main database that stores this information is InterPro (Finn

et al., 2017). InterPro member databases have identified signatures of these protein relationships and features that can be used to identify them in unannotated protein sequences. The functional information that can be derived from these annotations varies greatly, with some features highly specific to a certain function while others represent a broad range. Typically, manual review of these definitions is required to decide upon the specificity of the annotation and indeed, for many genes it is not possible to identify a specific function.

Efforts have been made to automate derivation of functional meaning from protein features and one of the main attempts is that of the Gene Ontology consortium (Drabkin et al., 2015). Gene Ontology (GO) terms provide a hierarchy of relationships and a controlled vocabulary for biological processes, molecular functions and cellular components. Along with other information, orthology data and shared domains are integrated to assign GO terms to genes. The major model organisms benefit from manual curation of GO term assignment while for others, including *P. tricornutum*, GO terms are added automatically. This results in the assignment of many high level GO annotations to *P. tricornutum* and hence the imparted functional information is often unspecific.

In addition to efforts that identify candidate genes based on sequence homology and functional annotation, parameters that are specific to a particular enzyme can also be used. With regards to enzymes that act on MGDG, the location of the enzyme is an important factor. MGDG is present at high concentrations within the thylakoid but is also found in the outer leaflet of the chloroplast (Nakamura and Li-Beisson, 2016; Moellering et al., 2010). The chloroplast of *P. tricornutum* is surrounded by four membranes and therefore candidates that act on MGDG will be targeted to the chloroplast or to the periplastidial space. Within the ochrophytes, proteins are targeted to the plastid via a bipartite targeting peptide at the N terminus of the protein (Gruber et al., 2015). It is unclear how proteins are transported to the periplastidial space but this is expected to require at least a signalling peptide that transports the protein into the endoplasmic reticulum (Keeling, 2013). Enzymes that act on MGDG would therefore be expected to contain a signalling peptide or bipartite targeting peptide at their N terminus. Further to targeting, I expect the expression of genes to be useful in identifying candidates. As MGDG is essential to

proper photosynthesis, enzymes that degrade it would be expected to be maintained at a low level during normal growth and upregulated during nitrogen depletion. As introduced in Chapter 3, multiple expression datasets exist for *P. tricornutum* cultured without nitrogen and can therefore supply the required information.

To begin identification of candidate genes, insight into the likely activities of enzymes is required. Possible pathways of TAG synthesis from MGDG can be identified using knowledge of previously characterised enzymes. Based on this knowledge, MGDG components could be transferred into TAG via four routes. Acyl tranferases, acyl hydrolases and galactosyl transferases illustrated respectively by the characterised CrPDAT, CrPGD1 and AtSFR2 enzymes that were introduced in Chapter 1 contribute to three of these mechanisms (Li et al., 2012; Yoon et al., 2012; Moellering et al., 2010). Galactosyl hydrolase is involved in the fourth. The reactions required to incorporate into TAG the DAG and FA that are liberated by these reactions are shown in Figure 5.2: an acyl transferase could directly transfer an acyl chain from MGDG to DAG, an acyl hydrolase could release a free fatty acid for activation and subsequent incorporation by a DGAT or finally, a galactosyl transferase or a galactosyl hydrolase could release DAG for subsequent incorporation by a DGAT. This DAG could also be incorporated by a PDAT, although this is not shown in the diagram.

Once candidate genes have been compiled, their functions can be investigated using reverse genetic techniques. Many of the genetic tools, resources and procedures required to carry out this work have been developed for *P. tricornutum*. The resulting modified lines can then be investigated using common screening methods such as lipid analysis or confocal microscopy. The cloning of genetic constructs is aided by the availability of plasmids and protocols for modular cloning of multi-gene constructs, for example, using the Golden Gate Modular Cloning toolbox (Engler et al., 2014a). Multiple characterised genetic parts are also available, including promoters, terminators, untranslated regions, selectable markers and fluorescent tags. High efficiency transformation methods based on electroporation, biolistics and conjugation have been developed. Transformation via electroporation and biolistics is dependent upon random integration and therefore to identify a consensus phenotype, multiple independent transformants must be screened.

Activity	Order	Reaction	Characterised examples
Acyl transferase	1st	MGDG + DAG -> lyso-MGDG + TAG	CrPDAT
Acyl hydrolase	1st 2nd 3rd	MGDG -> lyso- MGDG + FFA FFA + CoA -> acyl-CoA acyl-CoA + DAG -> TAG + CoA	CrPGD1, AtHIL1
Galactosyl hydrolase	1st 2nd	MGDG -> DAG + galactose DAG + acyl-CoA -> TAG + CoA	
Galactosyl transferase	1st 2nd	MGDG+MGDG -> DAG+ DGDG DAG + acyl-CoA -> TAG	AtSFR2

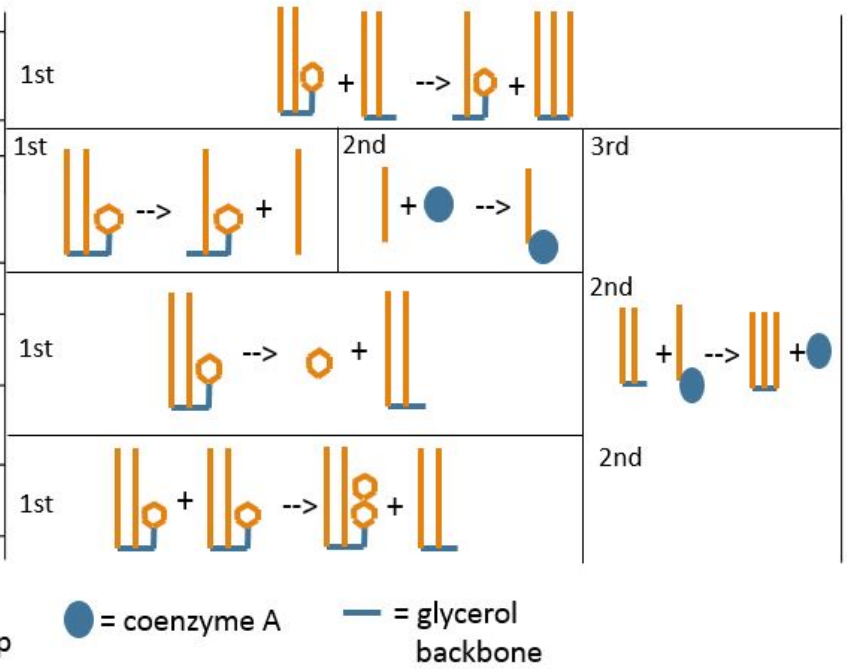


Figure 5.2 – Putative pathways of TAG accumulation from recycled MGDG. Four minimal pathways of TAG accumulation are outlined. An ideogram is provided based on the horizontally adjoined pathways. 1st, 2nd and 3rd in the ideogram correspond to the reactions detailed in the corresponding pathways

One well established screening method that can quickly provide information on the presence of a candidate gene is the detection of fluorescently tagged proteins. A common method to do this involves expression of a candidate gene in frame with a 3' fluorescent protein gene to produce fusion proteins. By screening such lines with confocal microscopy, subcellular targeting of the protein can be identified. In photosynthetic organisms, autofluorescence from chlorophyll can complicate detection of the fluorescent protein and so selection of proteins with substantially different excitation and emission spectra is required.

5.2 Investigating the contribution of membrane lipid recycling to TAG accumulation during acclimatisation to nitrogen depletion in *P. tricornutum*

5.2.1 Adapting Nile red staining for high-throughput neutral lipid measurements in *P. tricornutum*

Nile red staining was optimised for use in *C. reinhardtii* by Freddy Bunbury in the Plant Metabolism research group. Experiments were therefore undertaken to optimise the measurement for use in *P. tricornutum*, specifically by ensuring linear increases in Nile red fluorescence with increasing *P. tricornutum* cell density. It has been shown that linearity of Nile red staining in *C. reinhardtii* requires bleaching of pigments, which are otherwise thought to quench Nile red fluorescence and so the impact of bleaching was investigated at the same time. A culture of *P. tricornutum* was grown to stationary phase, at which time previous observations indicate TAG will have accumulated (Hamilton et al., 2014). The culture was diluted and treated with final concentrations of sodium hypochlorite (NaOCl) of 0, 0.04, 0.08 and 1.6% vol/vol. Samples were Nile red stained as detailed in Section 2.7.1 and fluorescence measured at time points after bleaching.

Treatment of cells with bleach reduced chlorophyll a levels in less than 10min as depicted in Figure 5.3A. Despite this, addition of bleach did not increase sample

fluorescence as seen in Figures 5.3 B & C. In fact, addition of 0.16% NaOCl reduced fluorescence. Furthermore Figures 5.3 B & C highlight that treatment with bleach did not improve linearity of fluorescence with increasing cell density. Therefore, chlorophyll a does not appear to quench Nile red fluorescence in *P. tricornutum* and addition of bleach is not required for a linear fluorescence response.

Good proportionality was seen particularly for cell densities above $1.5x10^6$ cells ml^{-1} in untreated cells. A final cell density of $3x10^6$ was therefore chosen for future Nile red measurement to ensure that fluorescent measurements would be in the linear part of the response curve.

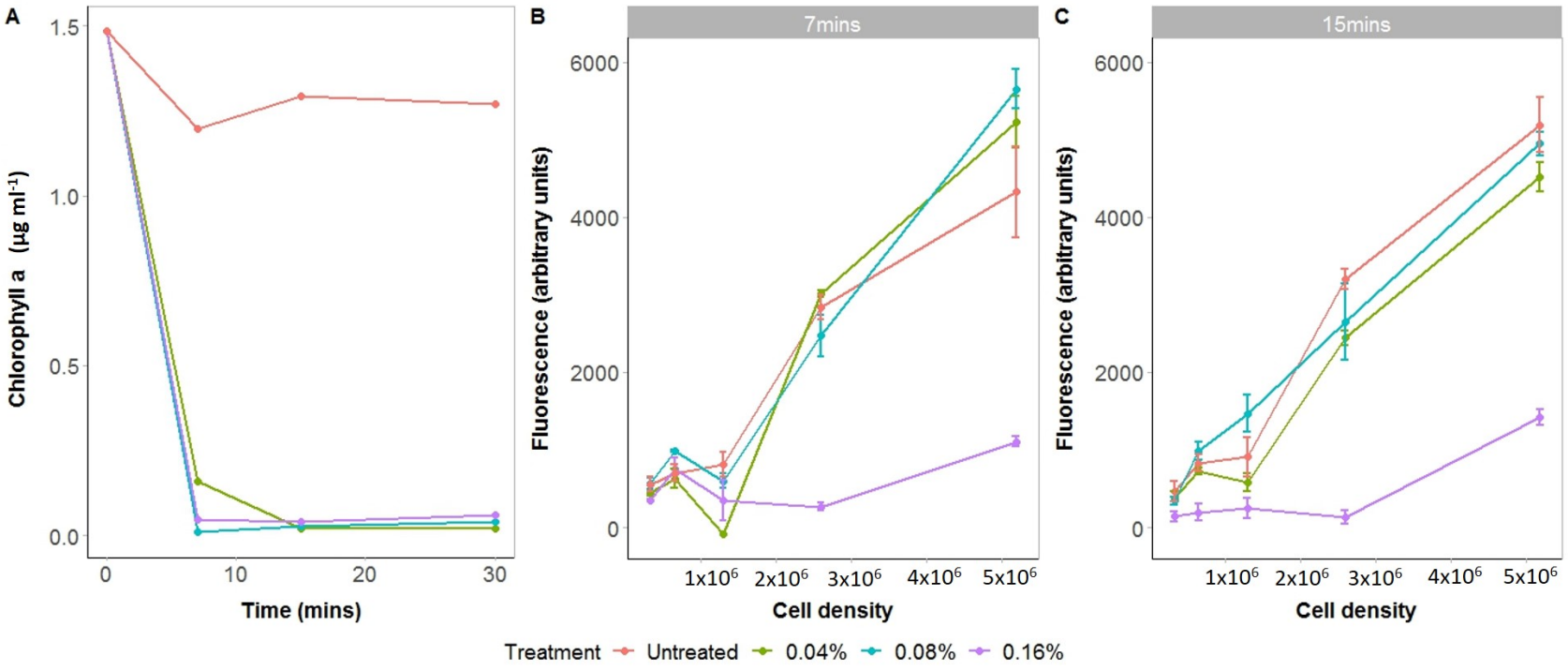


Figure 5.3 – The effect of pigment content on nile red fluorescence of *P. tricornutum* cells and linearity of fluorescence with increasing cell density. A) Chlorophyll a content of stationary phase cells. Measurements were taken for cells bleached for 7, 15 and 30 minutes and untreated cells at corresponding timepoints of 0, 7, 15 and 30 minutes. B & C) Nile red fluorescence of stationary phase cells treated as in A) after 7 (B) and 15 minutes (c) of treatment exposure. Treated cultures were diluted to produce a range of cell densities. Final concentrations of bleach were 0.04% (green), 0.08% (blue) and 0.16% (purple).

5.2.2 Preliminary investigation into the impact of cerulenin addition on TAG accumulation, pigment levels and cell density of *P. tricornutum*

5.2.2.1 Experimental outline

As explained in the introduction, cerulenin prevents TAG synthesis. To confirm that it alters TAG synthesis in *P. tricornutum* and to investigate its impact upon other cell properties, a preliminary time course was established. The growth conditions are as outlined in Section 2.3 and summarised along with the experimental procedure in Figure 5.4. During the experimental period *P. tricornutum* was cultured in media without nitrogen (f/2 0N10P) to promote TAG accumulation and with nitrogen (f/2 10N10P) for use as a control. When present, 10 times the usual amount of phosphate and nitrate were used to prevent depletion of these nutrients. At 24 hours of incubation, cerulenin was added at 10μM to some of the N-replete and N-deplete cultures so that there were four treatments in total: N-deplete with cerulenin, N-deplete control, N-replete with cerulenin and N-replete control. Three biological repeats were grown per treatment. In the samples without cerulenin Nile red fluorescence, cell density and chlorophyll a content were measured every 12 hours up to 60 hours. For cultures treated with cerulenin, only cell density and chlorophyll a content were measured up to 24h while after 36h, Nile red fluorescence was also measured. To investigate the effect of cerulenin on cells' ability to divide, aliquots of each of the cultures from the 36 and 60h time points were recultured and cell density measurements were taken at 0 and 48h after this separate reculture.

5.2.2.2 Effect of addition of cerulenin upon cell density and pigment content in cells grown with nitrogen and without nitrogen

Figure 5.5A presents the Nile red fluorescence of the cultures over the time course. N-deplete samples showed much more Nile Red fluorescence than N-replete samples from prior to the 14h time point, likely indicating that TAG began to accumulate sometime before 12h in N-deplete cells. The addition of cerulenin to cultures at 24h reduced subsequent Nile red fluorescence of both N-replete and N-deplete cells, indicating that it reduced synthesis of TAG. The renewed accumulation of Nile red

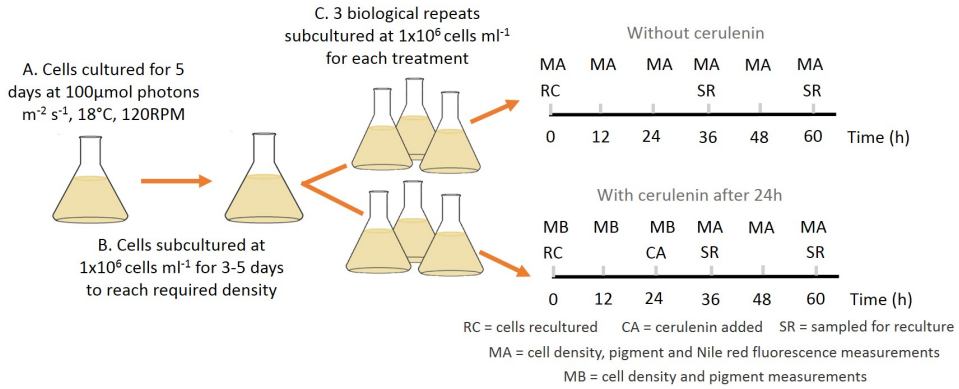


Figure 5.4 – Outline of the preliminary investigation into the physiological impacts of cerulenin addition to *P. tricornutum*. Cells were acclimatised to 100μmol photons m^{-2} s^{-1}, 18°C and 120RPM of shaking in f/2 10N10P media prior to reculture in f/2 0N10P of f/2 10N10P for the experiment. Cerulenin was added to three cultures each of N-replete and N-deplete media. Measurements were taken as described in the figure. Times at which samples were collected for further reculturing are indicated.

fluorescence in cerulenin treated, N-deplete cultures, between 48 and 60h is consistent either with FA synthesis inhibition that is insufficient to prevent TAG accumulation or membrane lipid recycling between these times.

The cell density of the cultures over time can be seen in Figure 5.5B. Addition of cerulenin led to a smaller increase in density of N-replete cultures between 48 and 60h compared to untreated cultures. As expected for N-deplete cells, cell density plateaued. This occurred from 36h onwards, obscuring any potential impact of cerulenin on cell density. Figures 5.5D and E depict the change in cell number over two days from samples recultured at 36 and 60h respectively. Of the samples that were recultured from aliquots taken at 36h, more cells accrued in N-replete cultures compared to N-replete samples with cerulenin. Conversely, for cultures sampled at 60h, the opposite was true.

Chlorophyll a content per cell is presented in Figure 5.5C. Chlorophyll a content was higher in N-replete cells from sometime prior to 24h. Treatment with 10μM cerulenin reduced the amount of chlorophyll in N-replete cells sometime between 12 and 24h after it was added.

As both chlorophyll a and MGDG are co-located, decreased chlorophyll a may indicate a decrease in MGDG. In this experiment, differences between N-deplete and N-replete cells in the level of chlorophyll a began prior to 24h. TAG began

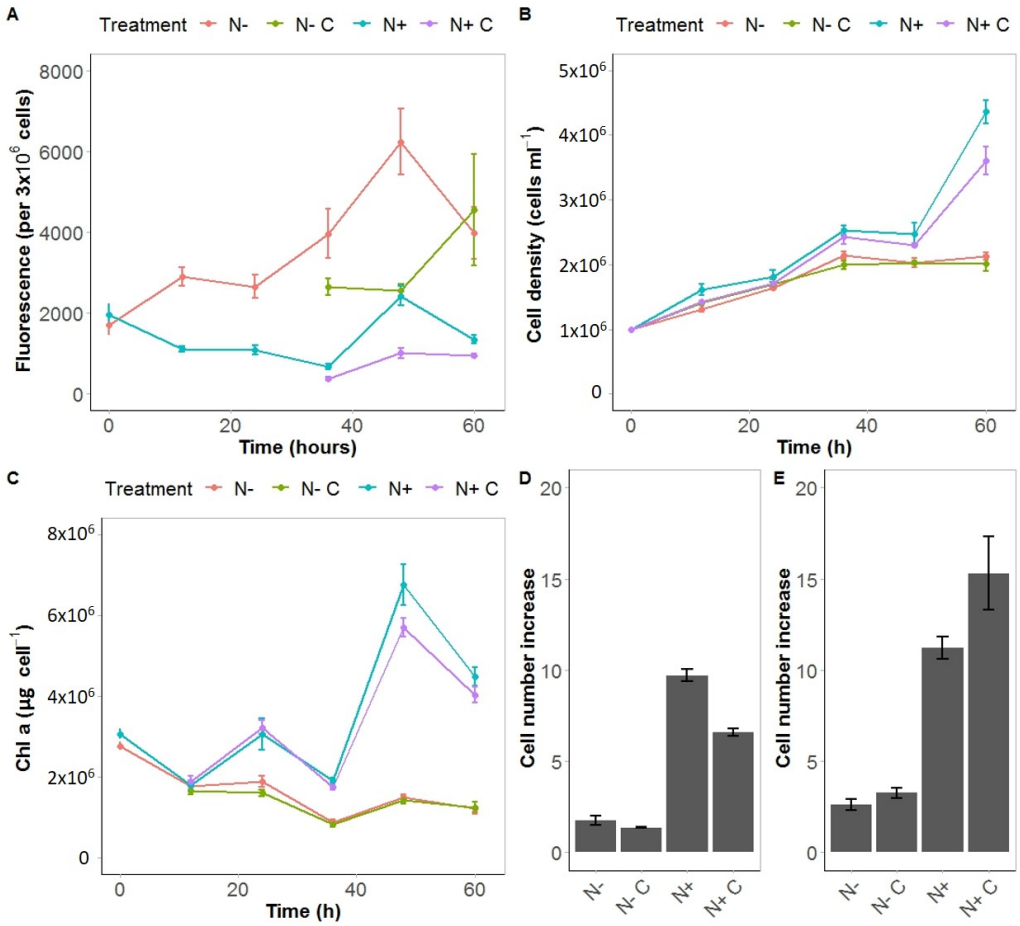


Figure 5.5 – The physiological impact of cerulenin on *P. tricornutum* cultured with and without nitrogen. *P. tricornutum* was cultured and sampled as described in Section 5.2.2.1. Measurements were: A) Nile red fluorescence, B) culture cell density, C) chlorophyll a content and the increase in cell number over 2 days normalised to original cell number, for samples recultured at 36 (D) and 60h (E). Cells were cultured in f/2 10N10P (N+) or f/2 0N10P (N-) media and cerulenin was added to some N-replete (N+ C) and N-deplete (N- C) cultures at 24h. Means and the one standard error shown are from three biological repeats.

to accumulate prior to 12h. These results are consistent with a movement of fatty acids from MGDG to TAG occurring in the first 24hr of acclimatisation to nitrogen depletion, under these experimental conditions. TAG also continued to accumulate after 24h. In following experiments, the impact of cerulenin on cells was therefore investigated over a 72h time course. The addition of cerulenin also has wider impacts on cell physiology, illustrated by the change in cell density and pigment content it causes. Future time course experiments were therefore divided into shorter exposure periods to reduce any potential impact cerulenin would have on acclimatisation to nitrogen depletion.

5.2.3 Investigating TAG accumulation during nitrogen depletion with inhibition of FA synthesis

5.2.3.1 Experimental design

To investigate the sources of TAG accumulation under nitrogen depletion, cells were grown with various levels of cerulenin over an 80 hour time course. Throughout the experiment *P. tricornutum* cells were grown and resuspended in either f/2 10N10P or f/2 0N10P media as described in Section 2.3 and this is summarised, along with the experimental outline in Figure 5.6. To limit unintended effects, treatment exposure time was limited to around 36 hours. The time course was therefore divided into three treatment exposure periods that, as a shorthand, I will refer to as the 2-39h, 28-59h and 49-80h periods. Respectively for the 2-39h, 28-59h and 49-80h time periods, treatments were made 2, 28 and 49h after reculture and cell density and Nile red measurements were taken 14 and 39h, 39 and 59h and 59 and 80h after experimental reculture. These measurement times are sometimes used to present the below results e.g 14-39h time period. Three cultures were established for each treatment and for each exposure period. For example, nine N-deplete cultures without cerulenin were cultivated: three biological repeats for each exposure period. Despite overlap of the treatment periods in time as illustrated in 5.6, it is important to note that treatments within the periods were carried out on different cultures.

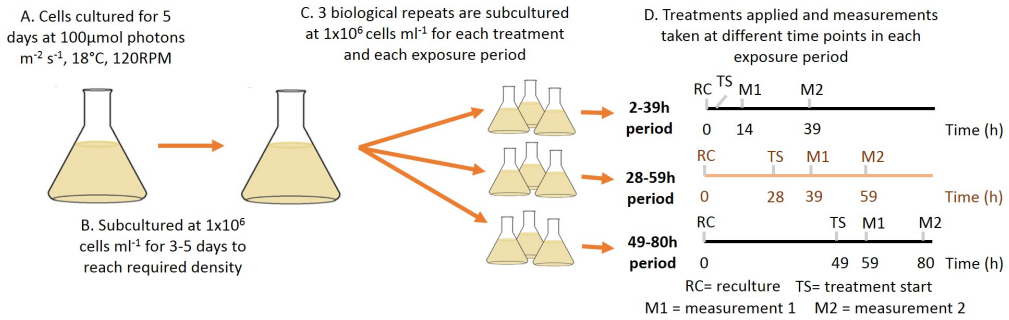


Figure 5.6 – Experimental outline used in investigating TAG accumulation during FA synthesis inhibition in Section 5.2.3. Cultures of *P. tricornutum* were acclimated in f/2 10N10P prior to reculture in f/2 10N10P or f/2 0N10P for the experiment. 3 biological repeats were grown in separate flasks for each treatment in each of the treatment exposure periods. Exposure periods were from 2-39h, 28-59h and 49-80h. Treatments at TS, cell density and Nile red fluorescence measurements were taken at M1 and M2. Treatments consisted of exposure of cells grown with or without nitrogen to various combinations of cerulenin, triclosan, and DMSO.

Numerous treatments were made and so it is necessary to explain their purpose. Cerulenin was added to N-depleted cultures at increasing concentrations of 10, 20, 40 and 60 µM to identify a concentration that completely inhibits FA synthesis. A control of N-depleted cells was included. N-replete cultures with 20 µM of cerulenin were added to investigate the effects of cerulenin on *P. tricornutum* cells and allow comparison with N-deplete cultures that contain cerulenin. A control of N-replete cells was grown. Triclosan was added at 20 µM to cultures that contained 40 µM of cerulenin, with the aim of ensuring FA synthesis was completely inhibited. Triclosan was also added to N-deplete cultures at 20 µM as a control for the impact of triclosan on cells. The various chemicals were dissolved in DMSO and so a control of 15 µL of DMSO per 15ml of N-deplete culture was included.

A short hand was devised to identify the numerous culture conditions and treatments. Cultures containing nitrogen were deemed N+ and those without N-. The identity of added FA synthesis inhibitors and their various concentrations in µM are given by Cx and T for cerulenin and triclosan respectively, where x is the added concentration. When present, triclosan was added at fixed concentrations. The components of the name are then separated by an underscore. For example, cells grown in N-replete media with 20 µM of both cerulenin and triclosan would be named N+_C20_T whereas cells grown in N-deplete media without inhibitors would be labelled N-.

5.2.3.2 Nile red fluorescence of N-deplete and N-replete cells treated with cerulenin

As depicted in Figure 5.6, cerulenin was added to cultures of *P. tricornutum* grown in N-replete or N-deplete media to final concentrations of 10, 20 , 40 or 60μM. 15μL of DMSO, which was used as the solvent for the cerulenin stocks, was added to N-deplete cells in the 49-80h treatment period as a control. Figures 5.7A, B and C display the Nile red fluorescence that was measured for these cultures grown between A) 2-39h, B) 28-59h and C) 49-80h treatment periods.

During the experiment, a number of general trends were identified with regard to fluorescence. Of those cells treated with cerulenin, nile red fluorescence increased much more strongly in cells lacking nitrogen. Indeed fluorescence of N-replete cells

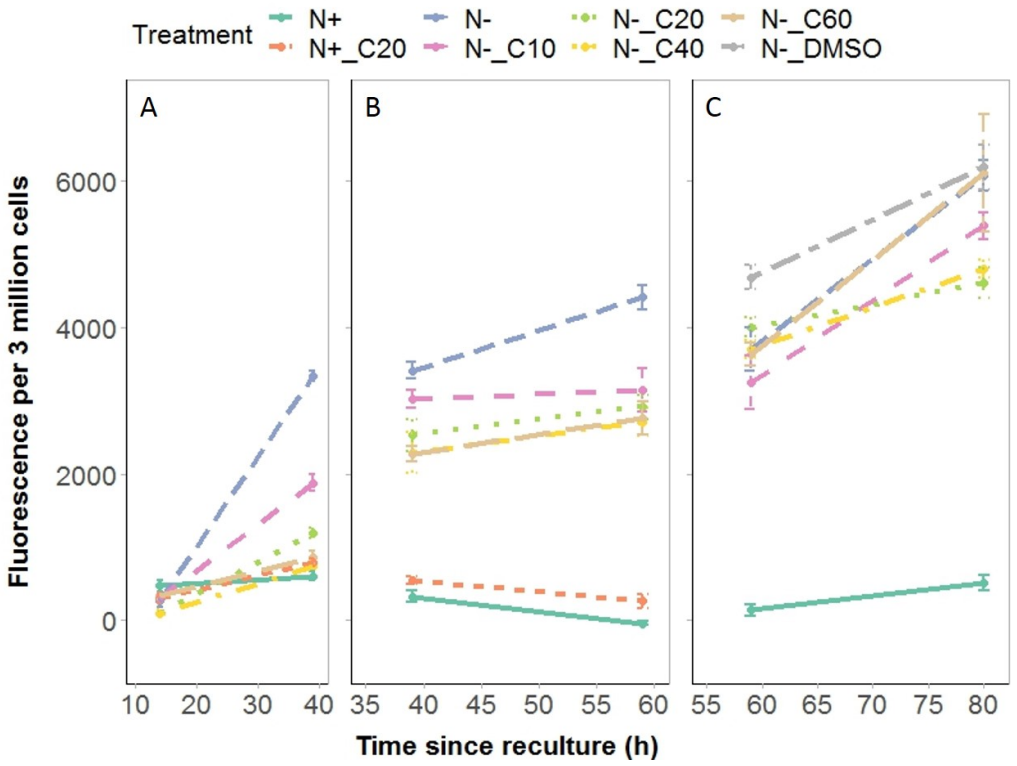


Figure 5.7 – Nile red fluorescence of *P. tricornutum* grown in N-deplete or N-replete cultures and treated with cerulenin. Nile red fluorescence were measured for cultures grown in the (A) 2-39h, (B) 28-59h and (C) 49-80h treatment periods. Cerulenin was added to cultures of *P. tricornutum* grown in N-replete (N+) or N-deplete (N-) media to final concentrations of 10 (C10), 20 (C20), 40 (C40) or 60µM (C60). 15µL of DMSO, which was used as the solvent for the cerulenin stocks, was added to N-deplete cells in the 49-80h treatment period as a control. Cultures of *P. tricornutum* were grown as described in Figure 5.6. Three biological repeats were used to calculate mean values and one standard error.

remained low throughout the time course, indicating that TAG did not accumulate. The increase in Nile red fluorescence in N-deplete samples began sometime between 14 and 39 hours after reculturing. Furthermore, there were alterations to how Nile red fluorescence changes between time periods. Fluorescence increased in the 2-39h (A) and 49-80h (C) time periods for all samples, except N-_C60 in the 49-80h period. In the 28-59h (B) period, increase in fluorescence slowed and may have stopped for N-deplete cultures while fluorescence in N-replete cultures decreased.

5.2.3.3 Change in Nile red fluorescence of cultures treated with cerulenin

To understand the impact of increasing cerulenin concentration on TAG accumulation, the changes in fluorescence in each time period were investigated. Figure 5.8 displays

the change in fluorescence that was calculated between measurements. The figure presents a grid with column and row labels indicating the contained data. Change in fluorescence was normalised based on the number of hours between measurements and is referred to as Δfluorescence (ΔF). N-replete cultures were untreated or grown with 20µM cerulenin while N-deplete cultures were either untreated or exposed to cerulenin at final concentrations of 10, 20, 40 or 60µM.

Between 14 and 39h in Figure 5.8, ΔF of cells grown without nitrogen decreased in an inversely proportional manner to the amount of cerulenin added. The decrease in ΔF between treatments was significant up to 20μM of cerulenin (Tukey HSD, $p< 0.05$). However, even at the highest level of inhibitor, a positive ΔF was still seen (one-sided t test, $p<0.05$). This inversely proportional relationship may also be the case between 59 and 80h with cerulenin treatments equal to or above 20μMol reducing ΔF, although the difference between cerulenin treatments is not significant (Games-Howell [non-parametric, equal variance not required], $p>0.05$). ΔF recorded at 60μMol goes against this trend but the variance of these measurements was very high. A significant increase in fluorescence was seen between 59 and 80h for all N-deplete treatments (one-sided t test, $p<0.05$) except N-_C60. Between 39 and 59h in Figure 5.8 however, the inversely proportional relationship is not seen. The presence of 10μM cerulenin visually reduces ΔF, but additional increases in cerulenin concentration had no further effect. These changes in ΔF between treatments were not significant however. Furthermore, statistically significant increases in ΔF were not seen for any N-deplete treatments over this time period (one-sided t test, $p>0.05$). Therefore, in at least one time period, the TAG content of N-depleted cells decreased with increasing cerulenin. However, an increase in fluorescence was seen even at the highest inhibitor concentrations.

Considering the ΔF for N-replete cells, between 14 and 39h and 59 and 80h, ΔF was positive, with the reverse true between 39 and 50h. This decrease in ΔF between 39 and 59h was significant for cells with 20μM of cerulenin (one-sided t test, $p<0.05$). N-replete cells grown with cerulenin showed greater ΔF compared to those not exposed to cerulenin, significantly so between 14 and 39h (t test, $p<0.05$), as seen in Figure 5.8. There were also differences in the ΔF between N-deplete and N-replete cells. Between 39 and 59h and 59 and 80h ΔF of N-replete cells was lower

than for N-deplete cells at all levels of cerulenin addition. However, between 14 and 39h, there was not a significant difference in fluorescence of N-replete cells and N-deplete cells treated with the highest concentrations of cerulenin.

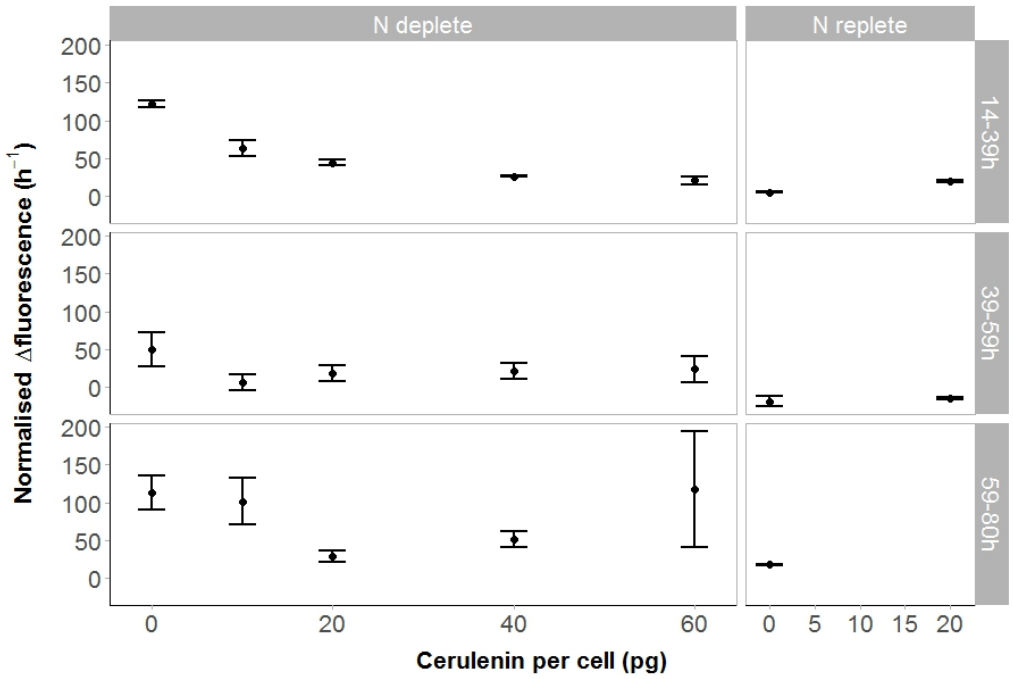


Figure 5.8 – The impact of cerulenin on changes in Nile red fluorescence of *P. tricornutum* cultured with or without nitrogen. Cultures of *P. tricornutum* were grown, treated and measured as described in Figure 5.6. Results are shown for cultures grown with (N-replete) and without nitrogen (N-deplete) and treated with various levels of cerulenin. Change in fluorescence was calculated between measurements taken at 14 and 39h (14-39h), 39 and 59h (39-59h) and 59 and 80h (59-80h) from cultures exposed to treatment for 2-39h, 28-59h and 49-80h after reculture in experimental media respectively. The figure is represented as a grid with column and row labels indicating the contained data, for example, the middle panel on the right side contains data on N-replete cultures calculated between 39 and 59h. Change in fluorescence was normalised based on the number of hours between measurements. Three biological repeats were used to calculate mean values and one standard error.

In summary, when applied to N-depleted cells, there is evidence that cerulenin reduced TAG accumulation throughout the time course in an inversely proportional manner. However, the low TAG accumulation and high variation of measurements in the 28-39h and 49-80h treatment periods respectively makes this less certain for these periods. Furthermore, N-_C60 cells may have produced a different fluorescence response between 49 and 80h, but the high variation in these measurements would require further evidence of this. In contrast, cerulenin increased Nile red fluorescence

of N-replete cells in early stages of the time course, between 14 and 39h but this was not seen at later stages.

5.2.3.4 Measurements of cell density of N-deplete and N-replete cells in which FA synthesis was inhibited with cerulenin

Cell density data was collected alongside the fluorescence data discussed above. Figure 5.9A depicts the 2-39h treatment period, B, the 28-59h period and C, that of 49-80h. In Figures 5.9A, B and C, cell density increased over time during the course of the experiment. Cultures grown with nitrogen divided more rapidly, markedly so from the 28-59h time period onwards with an increase in the rate of division but a decrease for N-deplete cells. Addition of cerulenin had a consistent effect on N-deplete and N-replete cultures with an inversely proportional relationship between the amount of cerulenin added and change in cell density. Despite the added cerulenin, maintenance or an increase of cell density is seen, except for in the 49-80h period for N-deplete cells. Cerulenin addition to N-deplete cells at levels equal to or above 40μM between 49 and 80h leads to cell lysis, given the decrease in cell number.

5.2.3.5 Nile red fluorescence and cell density measurements of cultures treated with the FA synthesis inhibitor triclosan

Triclosan inhibits FAS through an alternative method to cerulenin as outlined in Section 5.1.1. To ensure complete cessation of FA synthesis it was added to N-deplete cultures and also to N-deplete cultures with 40μM of cerulenin as detailed in Figure 5.6. The impact of triclosan upon Nile red fluorescence is presented in Figures 5.10A, B and C and cell density in Figures 5.10D, E and F. Triclosan sharply reduced cell density in all time periods. This was already notable by the first measurement point in each period but this decline continued to the second measurement. The addition of triclosan at this concentration therefore caused cell lysis and this increased over the culturing period, suggesting it is toxic to cells at this concentration.

In contrast to cell density, Nile red fluorescence increased upon addition of triclosan. In the 2-39h period this increase in fluorescence of samples with addition of triclosan was intermediate between that seen for N-_C0 and N-_C40 but was much higher in successive treatment periods, displayed in Figure 5.10A, B and C. Nile

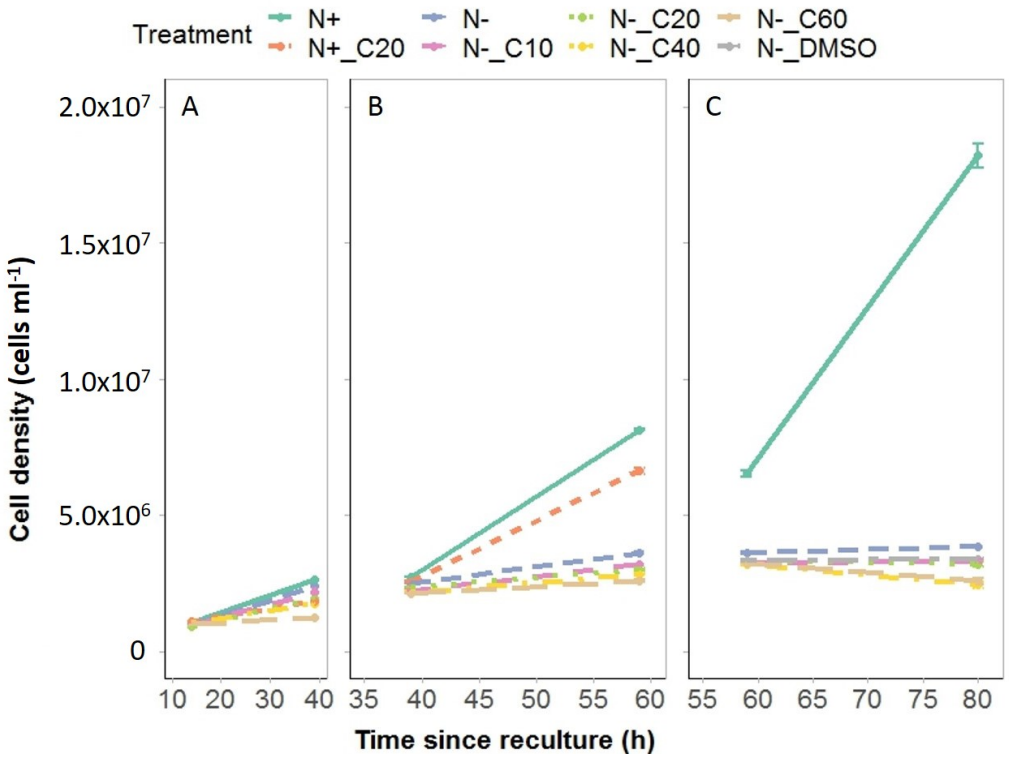


Figure 5.9 – Cell density of *P. tricornutum* grown in N-deplete or N-replete cultures and treated with cerulenin. Cell density was measured for cultures grown in the A) 2-39h, B) 28-59h and C) 49-80h treatment periods. Cerulenin was added to cultures of *P. tricornutum* grown in N-replete (N+) or N-deplete (N-) media to final concentrations of 10 (C10), 20 (C20), 40 (C40) or 60µM (C60). 15µL of DMSO, which was used as the solvent for the cerulenin stocks, was added to N-deplete cells in the 49-80h treatment period as a control. Cultures of *P. tricornutum* were grown as described in Figure 5.6. Three biological repeats were used to calculate mean values and one standard error.

red fluorescence is influenced by the solvent it is dissolved in. As I found that cells exposed to triclosan lysed, Nile red fluorescence may have been altered due to the changed environment and is therefore less likely to correspond to changes in TAG content in a manner comparable with other treatments. This is supported in Figure 5.10, where cell lysis was inversely proportional to increasing Nile red fluorescence.

5.2.3.6 Measurements of chlorophyll a per cell in N-deplete and N-replete cultures treated with cerulenin and triclosan

Cultures were grown and treated with different concentrations of cerulenin and triclosan as detailed in Figure 5.6. Chlorophyll a was extracted from cells 12h after treatment in the 2-39h treatment period and quantified spectrophotometrically, as described in Section 2.2. The amount of chlorophyll a per cell for the variously treated

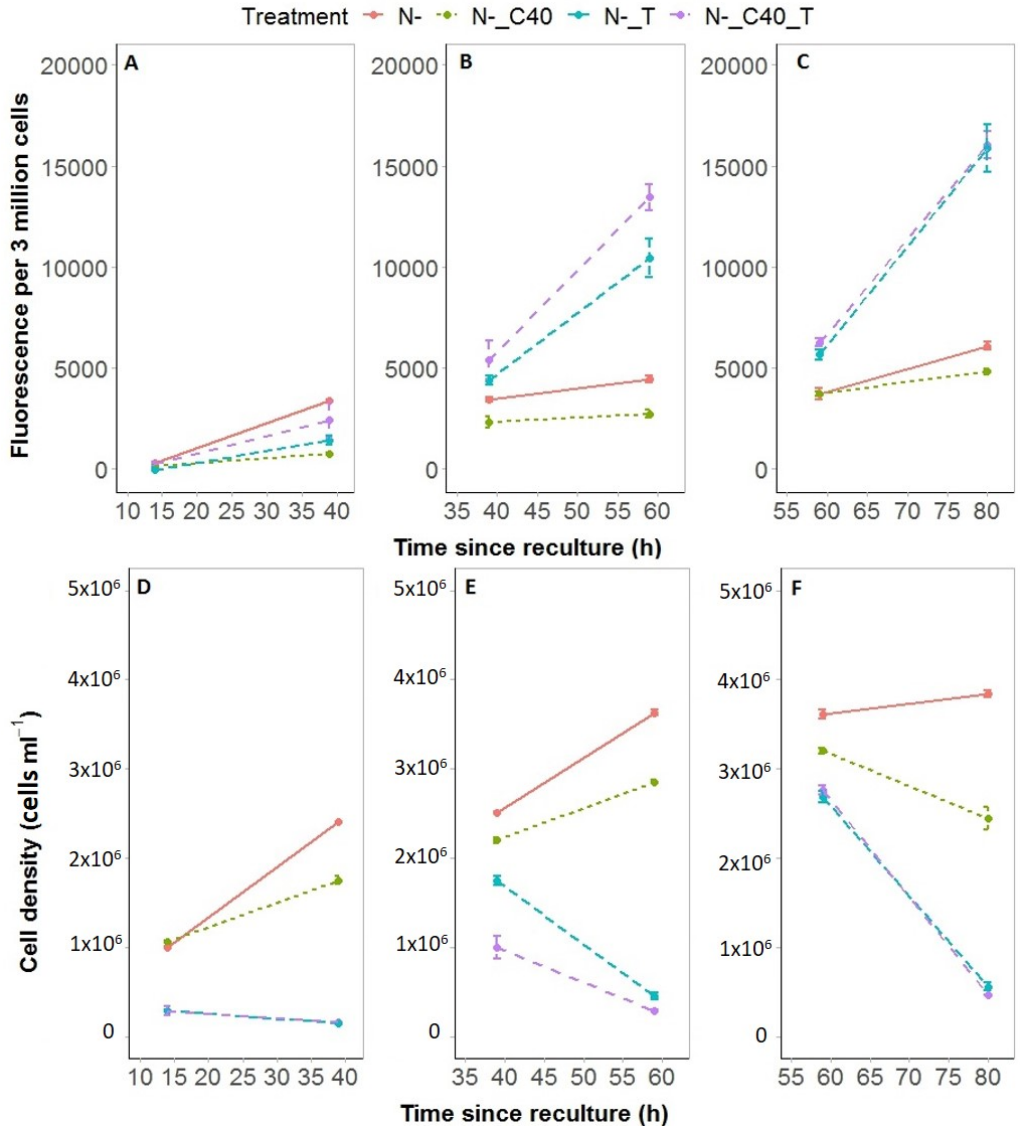


Figure 5.10 – Nile red fluorescence and cell density of *P. tricornutum* exposed to triclosan and cerulenin during nitrogen inhibition. Cultures of *P. tricornutum* were grown, treated and measured as depicted in Figure 5.6. Nile red fluorescence (A, B, C) and cell density (D, E, F) were measured for cultures grown in the 2-39h (A, D), 28-59h (B, E) and 49-80h (C, F) treatment periods. Treatments were grown in N-deplete media (N-) and some cultures were further treated with 40µM cerulenin (N-_C40), 20µM triclosan (N-_T) or a combination of 40µM cerulenin and 20µM triclosan (N-_C40_T). Common ordinate axes are used for Figures A, B and C and Figures D, E and F. Three biological repeats were used to calculate mean values and one standard error.

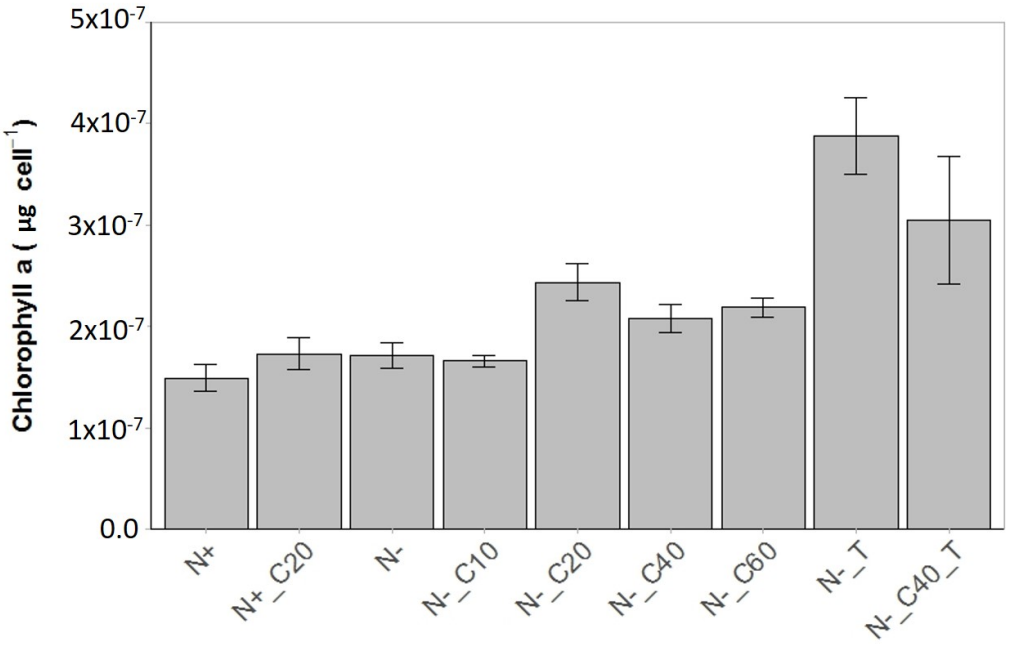


Figure 5.11 – Chlorophyll a content of *P. tricornutum* cultured in N-replete or N-deplete media and treated with cerulenin. Subculturing and treatments were carried out as depicted for the 2-39h treatment period in Figure 5.6 except chlorophyll a was only measured at 14h. Cells were grown with the following treatments: N-replete (N+), N-replete with 20µM cerulenin (N+_C20), N-deplete (N-) and N-deplete with 10 to 60µM of cerulenin (N-_C10, N-_C20, N-_C40, N-_C60). Significant differences were seen between the chlorophyll a levels of N- and N-_C20 ($p<0.05$), N-_C10 and N-_C20 ($p<0.05$), N+ and N-_C20 ($p<0.01$), N+_C20 and N-_C20 ($p<0.05$) and N+_C0 and N-_C60 ($p<0.05$). Statistical differences were calculated using Tukey's honest difference test and error bars are one standard error calculated from three biological repeats.

cultures is presented in Figure 5.11. Nitrogen content of media had no impact on pigment levels. The presence of cerulenin had an impact on chlorophyll a levels but only at or above 20 µM in cells grown without a nitrogen source. Statistically, only cells from the N-_C20 and N-_C60 cultures contained more chlorophyll a than other cultures, N-_C20 containing more than N-_C0, N-_C10 (Tukey's HSD, chl: $p<0.05$), N+_C0 (Tukey's HSD, chl: $p<0.01$) and N+_C20 (Tukey's HSD, chl: $p<0.05$) cultures while cells in N-_C60 culture contained more chlorophyll a than N+_C0 (Tukey's HSD, $p<0.05$).

5.3 Identifying and characterising enzymes putatively involved in chloroplast membrane degradation during nitrogen depletion

Concurrent to investigations into recycling of chloroplast membrane lipids to TAG, work was undertaken to identify enzymes that could catalyse this process. It was reasoned that improved knowledge of lipid metabolism would be useful, irrespective of whether TAG synthesis from membrane lipids was identified.

5.3.1 Identifying candidate genes

Transfer of MGDG constituents to TAG could occur via one of four mechanisms as outlined in Figure 5.2. One mechanism involves an acyltransferase transferring a fatty acid from MGDG to DAG to produce TAG. A second would employ an acylhydrolase to remove a fatty acid from MGDG. The released fatty acid could subsequently be activated by coenzyme A (CoA) prior to acting as a substrate in the acylation of DAG. The third and fourth mechanisms would release DAG either through hydrolysis of the galactosyl moiety from MGDG or transfer of the moiety to another MGDG molecule respectively. The resulting DAG from either scenario could then be acylated to form TAG.

Candidates were sought for the enzymes that putatively act on MGDG as these represent the first step in the outlined mechanisms of chloroplast lipid degradation and TAG accumulation. These enzymes are acyltransferases, acylhydrolases, galactosyl hydrolase and galactosyl transferase. To identify promising candidates, it was decided to use a combination of approaches to compile a list of as many of the potential candidates as possible and then to refine this based upon several sets of relevant information. Generally, two strategies were used to identify candidates. One focused on finding homologues to genes that demonstrated these functions when previously characterised. The other made use of functional annotation assigned to genes based on their sequences. Refining the list of candidates was carried out using protein targeting predictions, expression data and functional annotation.

Candidate identification was carried out in two rounds with a slightly varied

procedure each time. I will demark these different procedures and resulting candidates using the descriptions 'first identification round' and 'second identification round'. Chapter 2 provides detailed descriptions of the bioinformatic procedures. In the first identification round, protein sequences of CrPDAT (Yoon et al., 2012) and CrPGD1 (Du et al., 2018; Li et al., 2012) from *C. reinhardtii* were used as queries in TBLASTN searches against gene models in the Ensembl protist *P. tricornutum* database (Kinsella et al., 2011; Kersey et al., 2018) to identify genes with sequence similarity. The returned candidates were subsequently used as query gene models in another TBLASTN search. In the second identification round, the AtSFR2 protein (Moellering et al., 2010; Roston et al., 2014) was used as a TBLASTN query in addition to CrPDAT and CrPGD1. The amino acid sequences of these query proteins are shown in the appendix in Figure A.3.

Adding to the candidates identified through BLAST searches were those found using sequence annotation, specifically, domain and structure annotation associated with relevant functions. In the first identification round, *P. tricornutum* gene models in Ensembl Protist were searched by potential function using Interpro domain identifiers (Finn et al., 2017), Gene Ontology biological process terms (Ashburner et al., 2000; The Gene Ontology Consortium, 2017) and the term lipase. The terms were chosen based on the relevance of their descriptions to galactolipid acylhydrolase functionality. In the second identification round, domain and structure annotation from the query genes, CrPDAT, CrPGD1 and AtSFR2 was compiled and used to query the *P. tricornutum* database in Ensembl Protist. Identifiers from both rounds are shown in Table 5.1.

Several strands of information were used to narrow the candidates for experimental investigation. Gene models deemed irrelevant were screened out based on InterPro annotation. Of the returned genes, 26 from the first round of identification and 129 from the second were of potential interest. These gene models are shown alongside their annotation in the appendix Tables A.1 to A.5 in the appendix. These genes were annotated with transcriptomic data from four published studies (Alipanah et al., 2015; Levitan et al., 2015; Yang et al., 2013; Matthijs et al., 2016) and that of Huete-Ortega et al. (unpublished). Three of the authors carried out differential expression analysis, providing fold changes for transcripts and p values or p values

Table 5.1 – Domain and structure indentifiers used as query terms to identify candidate genes. Candidate identification was carried out on two occasions. In the first identification round, identifiers for acylhydrolases were used while those of acylhydrolases, acyltransferases, galactosyl hydrolases and galactosyl transferases were used in the second round. Terms originated from various databases: IPR - Interpro, GO - gene ontology, PF - Pfam, G3DSA - gene3D, PTHR - Panther.

ID - first round	Description
IPR001227	Acyltransferase domain superfamily
IPR002641	Patatin-like phospholipase domain
IPR002921	Fungal lipase-like domain
IPR003140	Phospholipase/carboxylesterase/thioesterase
IPR007751	Domain of unknown function DUF676, lipase-like
IPR008947	Phospholipase C/P1 nuclease domain superfamily
IPR014043	Acyltransferase
IPR016035	Acyltransferase/acylhydrolase/lysophospholipase
IPR020801	Polyketide synthase, acyltransferase domain
IPR024632	Phospholipase D, C-terminal
IPR032093	Phospholipase D, N-terminal
IPR006693	Partial AB-hydrolase lipase domain
IPR016090	Phospholipase A2 domain
IPR001531	Zinc-dependent phospholipase C
IPR014815	Phospholipase C-beta, C-terminal domain
IPR029002	Phospholipase C/D
IPR016035	Acyltransferase/acylhydrolase/lysophospholipase
IPR001087	GDSL lipase/esterase
IPR013818	Lipase/vitellogenin
IPR002642	Lysophospholipase, catalytic domain
IPR032341	MIT, C-terminal phospholipase D-like domain
IPR005592	Mono-/di-acylglycerol lipase, N-terminal
IPR025202	Phospholipase D-like domain
GO:0019374	galactolipid metabolic process
GO:0046461	neutral lipid catabolic process
GO:0046466	membrane lipid catabolic process
GO:0046503	glycerolipid catabolic process
GO:0009395	phospholipid catabolic process
GO:0008300	isoprenoid catabolic process
GO:0046460	neutral lipid biosynthetic process

ID - second round	Description
PF02450	Lecithin:cholesterol acyltransferase
PTHR11440	Lecithin-cholesterol acyltransferase-related
PF01764	Lipase (class 3)
G3DSA:3.40.50.1820	Alpha/Beta hydrolase fold
PTHR21493	CGI-141-related/lipase containing protein
G3DSA:3.20.20.80	Glycosidases
PF00232	Glycoside hydrolase family 1
PTHR10353	Glycosyl hydrolase

adjusted using the Benjamini and Hochberg method (Benjamini and Yosef, 1995). Fold change is stated for N-deplete cultures relative to N-replete ones. Therefore, fold changes larger than 1 indicate up-regulation under nitrogen depletion whereas values below 1 indicate down-regulation. Further to functional annotation and differential expresson, the subcellular localisation of candidates was predicted using ASAFind (Gruber et al., 2015) in the first round of identification and both ASAFind and HECTAR (Gschloessl et al., 2008) in the second.

Candidates to investigate experimentally were selected using the collated localisation, expression and functional annotation data. Candidates were deemed of interest if they were predicted to be localised in the chloroplast or endoplasmic reticulum, were upregulated under nitrogen depletion and were annotated with domains that could function as acylhydrolases in the first round or acylhydrolases, acyltransferases, galactosyl transferases and galactosyl hydrolases in the second round. Up regulation was considered across the range of time points in the first round of identification. In the second round, the focus was on early transcriptional changes (prior to 20h) as MGDG levels decrease early in nitrogen depletion as shown in the appendix in Figure A.2 and as supported by early degradation of chlorophyll a. As annotation indicating activity on phospholipids has also been assigned to ezymes observed to act on glycolipids (Yoon et al., 2012), functional annotation suggesting activity on phospholipids was not used to exclude candidates. Final selection was based on either the aggregate of these characteristics or one particularly interesting feature.

The candidates that were selected for experimental characterisation are displayed in Table 5.2. The attributes used in their selection are also displayed. The following genes were candidates of particular interest. Phatr3_J44028 is a promising candiate for a galactolipid acylhydrolase, given its predicted localisation to the chloroplast, its up-regulation thoughout nitrogen depletion and its appropriate functional annotation. Phatr3_J48297 and Phatr3_EG01099 are other interesting galactolipid acylhydrolase candidates. Phatr3_J48297 is not strongly upregulated but is predicted to be located in the chloroplast. Phatr3_EG01099 is not predicted to be localised to a relevant compartment. However, targeting predictions are fallible. Its high up-regulation under nitrogen depletion was therefore used to select it as an interesting candidate. Phatr3_J54569 was identified as a potential galactolipid galactosyl hydrolase on the

basis of its very high upregulation in one of the transcriptomic studies. Phatr3_J8860 was chosen as a putative galactolipid acyltransferase given its annotation as a PDAT and the impact of its overexpression on TAG species composition, as detailed in Chapter 5. The other candidates were chosen as the best options after those described above.

Table 5.2 – Putative galactolipid acylhydrolase, galactolipid acyltransferase, galactolipid galactosyl hydrolase and galactolipid galactosyl transferase genes that were chosen for experimental investigation. Subcellular location predictions were made using ASAFind and HECTAR. SignalP positive or signal peptide indicates a signal peptide was predicted to be present, targeting the protein to the ER. Signal anchor indicates the protein is expected to attach to a membrane, anchoring the enzyme in the lumen of the ER or outside the cell. Other localisation indicates no targeting peptide was identified. Fold change of transcripts from a number of studies and time points is presented. Fold change of transcripts in cells depleted of nitrogen was calculated relative to transcript quantity in cells grown with nitrogen.

	Gene ID	Phatr3_J31492	Phatr3_J41624	Phatr3_J44028	Phatr3_J8860	Phatr3_J49702	Phatr3_J45579	Phatr3_EG01099	Phatr3_J48297	Phatr3_J54569	Phatr3_J44873	Phatr3_EG02432
	Identification round	1	1	1 and 2	2	2	2	2	2	2	2	2
	ASAFind prediction	Not plastid, SignalP negative	Not plastid, SignalP positive	Plastid, low confidence	Not plastid, SignalP negative	Not plastid, SignalP negative	Plastid, high confidence	Not plastid, SignalP negative	Plastid, low confidence	Not plastid, SignalP negative	Not plastid, SignalP positive	Not plastid, SignalP negative
	HECTAR prediction	NA	NA	signal peptide	signal anchor	other localisation	signal peptide	other localisation	chloroplast	other localisation	signal peptide	other localisation
	Interpro_description	AB_hydrolase and Fungal_lipase-like	AB_hydrolase Fungal_lipase-like	Fungal_lipase-like and AB_hydrolase	LCAT/PDAT and Alpha/Beta hydrolase fold	LCAT/PDAT and Alpha/Beta hydrolase fold	PGAP1-like and Alpha/ Beta hydrolase fold	AB_hydrolase	DUF676_lipase-like and AB_hydrolase	Glycoside hydrolase superfamily	Glucanosyl-transferase and Glycoside hydrolase superfamily	Beta galactosidase small chain/ domain 5
Fold change (p value)	0h Huete-Ortega et al.	1.15	0.85	1.15	0.70	0.62	0.81	1.51	0.89	0.83	0.87	0.94
	4h Huete-Ortega et al.	1.64	1.27	1.38	1.04	0.70	0.70	2.86	0.98	2.03	0.70	0.78
	12h Huete-Ortega et al.	1.92	0.99	1.83	1.31	0.55	1.16	3.26	1.33	6.04	0.96	0.88
	18h Huete-Ortega et al.	1.46	1.31	3.72	1.53	0.33	1.51	7.27	0.96	21.35	0.99	0.84
	4h Matthijs et al. (2016)	1.91	2.65	2.35 (2x10-6)	1.52 (0.04)	1.11 (0.75)	1.52 (0.10)	NA	1.24 (0.41)	3.57 (1x10-3)	1.35 (0.17)	NA
	8h Matthijs et al. (2016)	1.96	2.55	2.07 (3x10-3)	1.53 (0.15)	1.37 (0.36)	1.18 (0.74)	NA	1.41 (0.33)	2.13 (0.02)	1.16 (0.76)	NA
	20h Matthijs et al. (2016)	2.02	NA	0.63 (0.02)	0.03 (1x10-12)	0.90 (0.73)	0.54 (0.03)	NA	1.49 (0.06)	0.11 (0)	1.31 (0.18)	NA
	48h Levitan et al. (2015)	4.00	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	48h Yang et al. (2013)	3.01	0.06	NA	NA	NA	NA	NA	NA	NA	NA	NA
	48h Alipanah et al. (2015)	2.25	3.05	NA	NA	NA	NA	NA	NA	NA	NA	NA
	72h Alipanah et al. (2015)	3.08	3.58	NA	NA	NA	NA	NA	NA	NA	NA	NA

5.3.2 Overexpression of candidate genes

5.3.2.1 Genetic construct design

Constructs were designed to enable the investigation of subcellular targeting of the candidate proteins and the impact of their overexpression. The general schematics for these designs can be seen in Figure 5.12. Constructs 1, 2 and 3 in this Figure were designed for expression of Phatr3_J41624, Phatr3_J31492 and Phatr3_J44028, the galactolipid acylhydrolase candidates identified in the first identification round. Design and cloning of these constructs occurred prior to those for the candidates from the the second identification round. Three variants were produced for these constructs. Construct 2 in Figure 5.12 shows a design where the candidate gene was fused to mVenus directly. In efforts to prevent interaction between the candidate and mVenus, construct 3 was designed to separate the two by a glycine linker. Construct 1 was designed to achieve the same goal, fusing the predicted targeting peptide or the N-terminal section of these candidates directly to mVenus.

Constructs 3, 4 and 5 in Figure 5.12 are the designs for the *P. tricornutum* candidates selected in the second round: Phatr3_J8860, Phatr3_J49702, Phatr3_J45579, Phatr3_EG01099, Phatr3_J44028, Phatr3_J48297, Phatr3_J44873 and Phatr3_EG02432. As described in Chapter 1, CrPDAT and CrPGD1 demonstrate the ability to catabolise MGDG in *C. reinhardtii*. They were therefore selected for heterologous expression in *P. tricornutum* for roles as potential positive controls, in addition to providing information on their function, if any, in *P. tricornutum*. Constructs 6, 7 and 8 in Figure 5.12 were designed for the expression of CrPDAT and CrPGD1 and their subsequent targeting to the chloroplast by a *P. tricornutum* chloroplast targeting peptide fused to the N terminus of the protein. Each of constructs 1 to 6 was designed to express the full candidate gene, except for CrPDAT which was designed to have the N terminus removed, prior to fusion with the *P. tricornutum* targeting peptide. Designs 1 and 4 were produced to investigate overexpression or heterologous expression of the candidates respectively. Designs 2, 3, 5 and 6 were created to investigate subcellular targeting of the proteins, either through fusion to YFP or a HA tag.

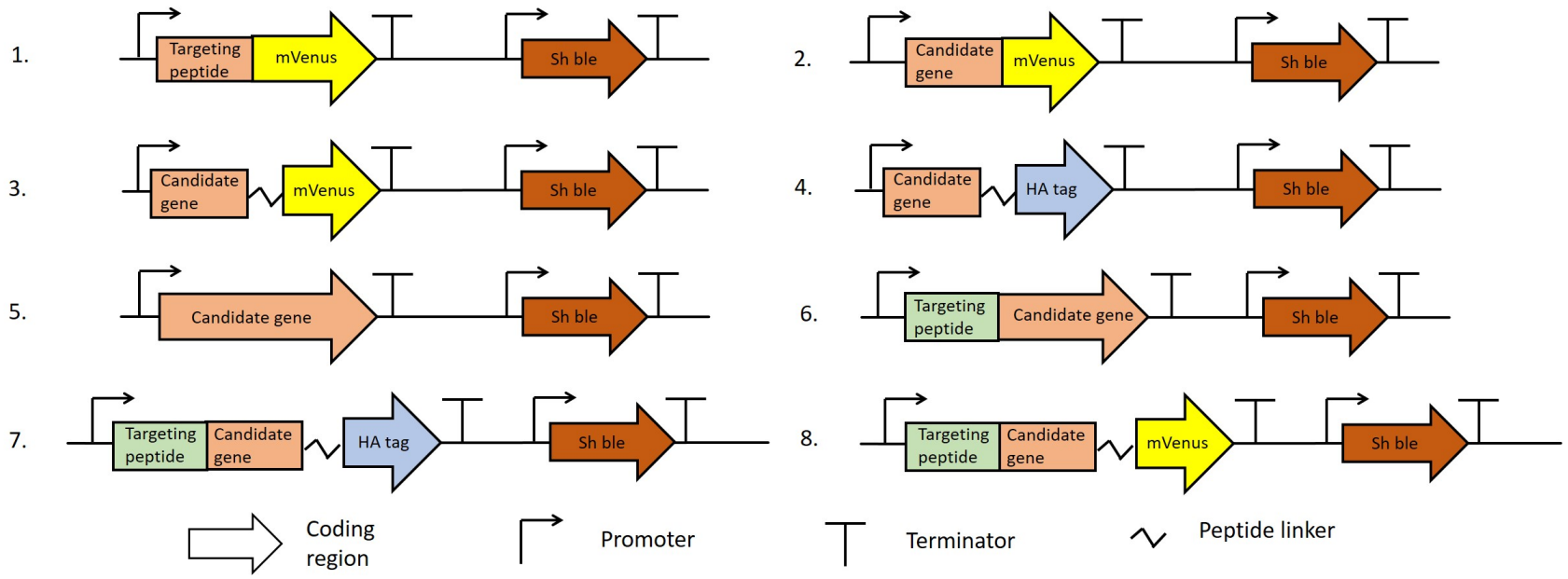


Figure 5.12 – Genetic construct schematics for candidate overexpression. The constructs were designed for the candidates presented in Table 5.2. Construct schematics for acylhydrolase candidates identified in round 1 are shown as 1, 2 and 3 and those identified for acylhydrolase, galactosyl hydrolase, acyltransferase and galactosyl transferase candidates in round 2 are schematics 3, 4, 5, 6, 7, 8. The Sh ble selectable marker was included in constructs, providing resistance to zeocin.

The designs in Table 5.3 depict the expression cassettes of the candidate genes and are level 1 constructs as defined by the Modular Cloning syntax that has previously been described by Engler et al. (2014a). The expression sequences within these constructs are designed to be cloned into the level 2 constructs shown in Figure 5.12. Design variants 1 to 3 in Table 5.3 shows the detailed design of the candidate expression regions of constructs 1 to 3 in Figure 5.12. Design variants 4 to 9 show those of constructs 3 to 8 in Figure 5.12. These level 1 constructs are designed to be formed through restriction of sequences from level 0 constructs that contain the desired genetic part. This is followed by ligation into a level 1 plasmid backbone. The naming of the cloning positions used in Table 5.3 refer to the overhangs that flank the level 0 sequences and dictate the acceptor plasmids that amplicons are cloned into. The appropriate overhangs and acceptor plasmids have been detailed elsewhere (Engler et al., 2014b).

The promoter and 5' untranslated region (UTR) from the fucoxanthin-chlorophyll a-c binding protein A (FcpA) gene of *P. tricornutum* were used in the design of constructs 1 to 3 from Table 5.3 as this promoter is commonly used to drive expression in *P. tricornutum*. Observations have demonstrated that expression driven by FcpA can vary during batch culture and is very low or ceases under nitrogen depletion (Adler-Agnon et al., 2017). Preliminary work carried out by Patrick Hickland in the Plant Metabolism group suggests that the B12 independent methionine synthase (METE) promoter from *P. tricornutum* drives high expression of transgenes in the absence of B_{12}. Nitrogen deprivation induced expression, driven by the ammonium transporter 1 (AMT1) promoter of *P. tricornutum* has been demonstrated by Dr Litvinenko, also while in the Plant Metabolism group. To enable high expression under both N-replete and N-deplete conditions, two versions of each of constructs 4 to 9 were designed, to use the promoter and 5' from either the METE or AMT1 genes. The commonly used 3'UTR and terminator from the *P. tricornutum* FcpC gene were used in all construct designs.

As displayed in Table 5.3, all construct designs for the expression of candidate gene:YFP or candidate gene:HA tag fusions incorporated a flexible glycine linker, except construct 2. The linker was either incorporated into the coding sequence (CDS) level 0 part through the primers used or as a separate level 0 part. Construct

Table 5.3 – Genetic constructs designed for overexpression or heterologous expression of candidate genes in *P. tricornutum*. Design is for a level 1 construct as defined in the Modular Cloning syntax (Engler et al., 2014a) and the cloning positions refer to those outlined in (Engler et al., 2014b). Designs utilised the promoter and 5'UTR of the FcpA, METE or AMT1 genes and the terminator and 3'UTR of the fucoxanthin-chlorophyll a-c binding protein C (FcpC) gene. Variants 1 to 3 correspond to the candidate expression units in constructs 1 to 3 in Figure 5.12. Variants 4 to 9 correspond to the candidate expression units in constructs 3 to 8 in Figure 5.12.

		Cloning position											
Design variant	Protein expressed	A1	A2	A3	B1	B2	B3	B4		B5	B6	C1	
1	targeting peptide: YFP fusion	FcpA Pro + 5' UTR					Targeting peptide	mVenus			FcpC 3'UTR + Ter		
2	full gene: YFP fusion	FcpA Pro + 5' UTR					CDS minus stop	mVenus			FcpC 3'UTR + Ter		
3	full gene: linker: YFP fusion	FcpA Pro + 5' UTR					CDS minus stop: glycine linker	mVenus			FcpC 3'UTR + Ter		
4	full gene: YFP fusion	METE or AMT1 Pro + 5' UTR					CDS minus stop	glycine linker		mVenus	FcpC 3'UTR + Ter		
5	full gene: HA tag fusion	METE or AMT1 Pro + 5' UTR					CDS minus stop	glycine linker		HA	FcpC 3'UTR + Ter		
6	full gene	METE or AMT1 Pro 5' UTR					CDS				FcpC 3'UTR + Ter		
7	Chloroplast target peptide: candidate gene fusion	METE or AMT1 Pro 5' UTR				DXR targeting peptide	CDS				FcpC 3'UTR + Ter		
8	Chloroplast target peptide: candidate gene fusion	METE or AMT1 Pro 5' UTR				DXR targeting peptide	CDS minus stop	glycine linker		HA	FcpC 3'UTR + Ter		
9	Chloroplast target peptide: candidate gene fusion	METE or AMT1 Pro 5' UTR				DXR targeting peptide	CDS minus stop	glycine linker		mVenus	FcpC 3'UTR + Ter		

3 in Table 5.3 made use of the former design, while constructs 4, 5, 8 and 9 make use of the latter. The YFP variant used was a *P. tricornutum* codon optimised version of mVenus (Lauersen et al., 2015) that was incorporated into the design of constructs 1, 2, 3, 4 and 9.

1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) is a member of the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway, which is predicted to be localised to the chloroplast of diatoms (Bertrand, 2010) and has been demonstrated to do so in the Plant Metabolism Group (unpublished research). The constructs for the expression of *C. reinhardtii* genes are outlined by 7-9 in Table 5.3 and were designed to use the targeting peptide of DXR to target the proteins to the chloroplast to enable them to access MGDG.

5.3.2.2 Identifying targeting peptide sequences in the selected candidate genes

The predicted targeting peptides sequences in Phatr3_J31492, Phatr3_J41624 and Phatr3_J44028, detailed in Figure 5.2, were required to clone the constructs based on design 1 in Figure 5.12. In addition, the removal of the targeting peptide sequence of CrPDAT was required to produce the constructs that used this gene, based on designs 6, 7 and 8 in Figure 5.12. Several methods were used to select the N terminal sequence of each gene that would be considered the targeting peptide. Firstly, targeting peptide prediction programs were used. Table 5.4 shows predicted subcellular locations and target or signal peptide lengths of these genes that were produced by the ASAFind (Gruber et al., 2015), TargetP (Emanuelsson et al., 2007) and PredAlgo (Tardif et al., 2012) programs. ASAFind was not used on *C. reinhardtii* as it is specific for algae with secondary plastids. The longest predicted targeting peptides from the three programs were 80, 38, 71 and 91 for Phatr3_J31492, Phatr3_J41624, Phatr3_J44028 and CrPDAT respectively.

Secondary structure and transmembrane domain predictions were used to evaluate structures surrounding the targeting peptide. Transmembrane domain predictions were made for CrPDAT, in an effort to prevent disruption of functionally important parts of the protein by the removal of the predicted targeting peptide. CrPDAT was predicted to contain a transmembrane domain, beginning from 62 residues

Table 5.4 – Predicted subcellular location and targeting peptide length of candidate genes. The predictions were made using three programs (Gruber et al., 2015; Emanuelsson et al., 2007; Tardif et al., 2012). Lengths refer to number of amino acid residues prior to the cleavage site.

Gene	ASAFind prediction	ASAFind length	TargetP plant prediction	TargetP plant length	PredAlgo prediction	PredAlgo length
Phatr3_J31492			mitochondrion	80		
Phatr3_J44028	chloroplast	38	secretory pathway	17	secretory pathway	11
Phatr3_J41624	secretory pathway	23	mitochondrion	71		
CrPDAT			chloroplast	91	chloroplast	65

into the protein. It has been previously hypothesised that this domain may play a functional role in the protein (Yoon et al., 2012) and so it was decided to maintain the transmembrane region. Removing the 65 residues of the shorter targeting peptide prediction from Table 5.4 left the transmembrane region intact and so was selected as the targeting peptide to remove. The secondary structure was evaluated for Phatr3_J31492, Phatr3_J41624, Phatr3_J44028 to reduce the chance that strong structures would form at the end of the targeting peptide and so disrupt the function of proteins to which they were fused. Secondary structures were predicted using the Chou-Fasman algorithm (Kumar, 2013) and transmembrane domain predictions were made using TMHMM2.0 (Krogh et al., 2001). Transmembrane domains were not predicted to be present for these proteins. The predicted target peptide lengths and knowledge that targeting peptides of these genes are predicted to be shorter than 100 residues (Emanuelsson et al., 2007) was used to search the secondary structure of these proteins. Unstructured regions that were more than 100 residues into the candidate proteins were identified as suitable 3' ends of the targeting peptides. Figure 5.13 shows the predicted secondary structures around this region for the three candidates. The predicted secondary structure of a particular locus is marked beneath the query amino acid sequence. Regions predicted to take on random coil structure are marked with C. The centre of regions of predicted random coil were selected as the end of the targeting peptides and resulted in peptide lengths of 111, 112 and 116 for Phatr3_J31492, Phatr3_J41624, Phatr3_J44028 respectively.

Figure 5.13 – Secondary structure predictions of candidate genes used in determining the end position of targeting peptides. Secondary structure predictions were produced for Phatr3_J31492 (a), Phatr3_J41624 (b) and Phatr3_J44028 (c) using the Chou-Fasman algorithm (Kumar, 2013). The query line presents the amino acid sequence of the proteins while the other lines display the predicted secondary stucture of each residue, which are either helices (H), beta sheets (E), beta turns (T) or random coil (C).

5.3.2.3 Cloning of candidate:YFP fusion plasmids and transformation of *P. tricornutum*

As with the candidate gene identification, cloning was undertaken in two rounds. Due to time constraints, cloning was only completed for candidates from the first round of identification. Details of the identification of second round candidates was provided for record. Cloning of Phatr3_J44028, Phatr3_J31492 and Phatr3_J41624 used in constructs 1, 2 and 3 from Table 5.3 and their subsequent incorporation into constructs 1, 2 and 3 in Figure 5.12 is described below.

To clone constructs produced in Figure 5.12 and Table 5.3, level 0 constructs were either sourced from members of the Plant Metabolism group or sequences were amplified and cloned for the production of these constructs. Level 0 plasmids for the FcpA, METE and AMT1 promoters and 5'UTRs, DXR targeting peptide, glycine linker, mVenus, HA tag and FcpC terminator and 3'UTR, *P. tricornutum* phospholipid:diacyglycerol acyl transferase (PtPDAT) and a level 1 plasmid for a Ble resistance cassette were kindly provided by Katrin Geisler, Patrick Hickland, Astrid Stubbusch, Maria Huete-Ortega and Aleix Gorchs-Rovira. The sequences of these constructs are shown in Figures A.5 and A.6 in the Appendix, except for the METE

promoter and 5'UTR and DXR targeting peptide, for which work is ongoing.

Primers to produce constructs from all identified candidates are stored in Table A.6. For designs 4, 5, 6, 7, 8 and 9 in Table 5.3, modifications to existing level 0 plasmids were required to introduce the necessary overhangs for cloning. Of those sourced from the lab, PtPDAT, METE promoter and 5'UTR and DXR targeting peptide were altered. Primers for alteration of the overhangs in the plasmid containing Phatr3_J44028 that was produced for use in design 2 in Table 5.3 are also included.

Amplification of the coding sequences and targeting peptides of Phatr3_J44028, Phatr3_J31492 and Phatr3_J41624 for constructs 1, 2 and 3 in appendix Table 5.3 was carried out with the primers detailed in Table A.6 and using gDNA as a substrate. For each gene, multiple amplicons were produced to remove BpiI and BsaI restriction sites. These amplicons were then restriction-ligated into the level 0 acceptor plasmid, pICH41258. Sanger sequencing was used to confirm the nucleotide sequence of the candidates in the level 0 parts. As many single nucleotide polymorphism (SNP)s have been detected within the candidate genes, a single reference sequence was not available and this led to the production of varying amplicons for some of the genes. The sequences that were selected for cloning are shown in Figure A.4.

Two versions of each candidate gene were successfully incorporated into acceptor plasmids, referred to as level 0 in the MoClo system. These were the predicted targeting peptide and full gene of Phatr3_J44028, Phatr3_J31492 and Phatr3_J41624. Despite the lack of predicted targeting peptide, an N-terminal sequence was cloned for Phatr3_J31492 to evaluate the prediction. As summarised in Table 5.3, level 1 plasmids were cloned through digestion and subsequent ligation of level 0 plasmids containing candidate regions, the FcpA promoter and 5' UTR, mVenus and the FcpC 3'UTR and terminator and the acceptor plasmid pICH47732. Presence of the correct products was checked by confirming polymerase chain reaction (PCR) fragments were the expected size and sequencing across regions where sequences were ligated together. These methods confirmed formation of level 1 constructs containing the correct sequences.

Level 2 plasmids were constructed by restriction and ligation of the level 1 constructs together with a Ble resistance gene expression cassette and the acceptor plasmid pAGM4723. To check the sequences, level 2 plasmids were digested with

BspHI. Several digestions were carried out but fragments of the expected size were not observed. The plasmids were therefore sequenced and PCR carried out. Sequencing across joins between the genetic parts indicated that all expected parts were present for the plasmids containing the selected targeting peptides of genes Phatr3_J41624, Phatr3_J31492, Phatr3_J44028. Sequencing also demonstrated that all parts were present in the plasmids containing the targeting peptides, except mVenus. The presence of mVenus was confirmed by colony PCR that produced amplicons of the expected size.

Control plasmids for the expression of mVenus were obtained, called pAKS001. These were previously cloned by Astrid Stubbusch in the Plant Metabolism group and were identical to the level 2 constructs described above, except the mVenus was expressed unfused to a section of candidate gene. All level 2 plasmids were transformed into *P. tricornutum* using electroporation as outlined in Chapter 2. Plasmids were first linearised to increase the efficiency of the process. pAKS001 and plasmids containing the signal peptides of the three genes and the full gene of Phatr3_J41624 were linearised by digestion with EcoRV. Plasmids containing the full genes of Phatr3_J44028 and Phatr3_J31492 were linearised with NdeI. Cells were then electroporated in the presence of the plasmids, incubated in the dark overnight and plated on f/2 gel media with zeocin. Individual colonies were picked after three weeks of incubation and then subcultured twice in f/2 media with zeocin to ensure stability of integration into the genome.

After subculturing, direct PCR was used to assay for the presence of transgene in *P. tricornutum* lines. Primers were designed to amplify a region beginning 5' of the FcpA promoter and ending towards the 3' on the mVenus sequence. With reference to Table 5.3, excepting the FcpA terminator, the majority of the candidate expression cassette was therefore subject to this assay. Figure 5.14 contains example gels confirming the presence of expected bands of 1212bp for the targeting peptide of Phatr3_J44028 in gel A and 2663bp of Phatr3_J41624 in gel B.

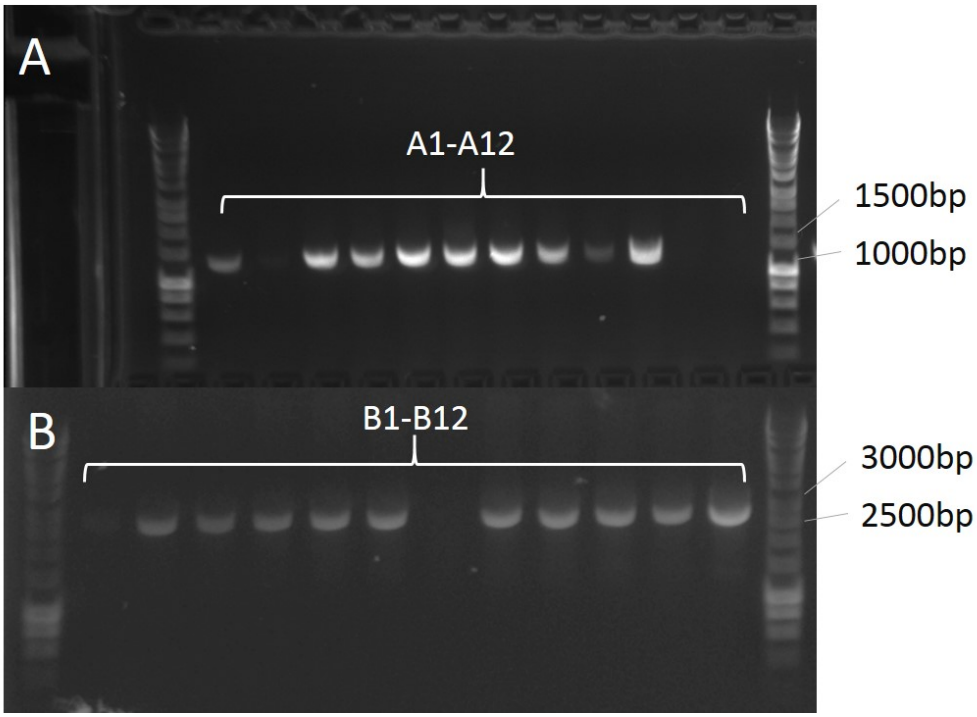


Figure 5.14 – Investigating presence of transgenes after transformation. Direct PCR using primers to amplify between 5' of the FcpA promoter and the 3' end of mVenus was carried out on *P. tricornutum* lines transformed with level 2 constructs. Lines were subcultured twice prior to PCR. PCR was carried out on Phatr3_J44028 (A) targeting peptide and Phatr3_J41624 (B) containing plasmids. 1kb ladders were used

5.3.3 Screening *P. tricornutum* lines that contain candidate: YFP fusion plasmids using fluorescent confocal microscopy

Confocal microscopy was carried out as described in Section 2.8. Samples were excited at 514nm. Emission from mVenus was measured between 520 and 550nm, subsequently referred to as the mVenus channel, while emission from chlorophyll autofluorescence was detected between 680 and 700nm, subsequently referred to as the chlorophyll channel. The information on how many independent transformant lines were obtained, how many contained the transgenes and the subsequent number that contained mVenus fluorescence is shown in Table 5.15. Despite the fact that large proportions of the transformants contained the transgenes, only a small proportion of these went on to fluoresce in constructs containing the Phatr3_J41624:mVenus and Phatr3_44028 targeting peptide:mVenus fusions. In contrast to the candidate

Gene part	Independent transformants	No. PCR screened	PCR positive transformants	No. viewed with confocal	No. Fluorescent Transformants
Phatr3_J41624 targeting peptide : mVenus	122	38	34	12	0
Phatr3_J41624 : mVenus	147	59	48	8	2
Phatr3_J31492 targeting peptide : mVenus	163	74	58	13	0
Phatr3_J31492 : mVenus	31	20	14	6	0
Phatr3_J44028 targeting peptide : mVenus	183	102	81	37	3
Phatr3_J44028 : mVenus	108	76	37	14	0
mVenus	209	88	63	11	9

Figure 5.15 – Summary of detection of the presence of transgenes and fluorescence in *P. tricornutum* transformed with candidate plasmids. The number of independent transformant lines (No. wells) is shown. A number of these were screened via direct PCR and a proportion of these cells produced the amplicon of expected size (PCR positive). PCR positive lines were then screened using fluorescent confocal microscopy and the number that were fluorescing is shown.

gene fusion constructs, the positive control was found to fluoresce in almost all of the screened transformants.

Images from the fluorescent Phatr3_J44028 targeting peptide:mVenus lines are shown in Figure 5.16. Images from two clonal lines and a WT are shown. The contrast and brightness of clonal line 1 images were altered to enable the mVenus signal that was observable on computer monitors to be visible when printed. The pattern of fluorescence was maintained. Images of the wildtype were altered in the same manner to provide a comparison. Line 2 fluoresced strongly and so the images were not altered. In the experimental lines, fluorescence was seen in the central region of the cell and in small regions towards the ends of the cell. In some cells, the mVenus fluorescence partly overlapped with the autofluorescence of the chloroplast. mVenus fluorescence was also observed to be spatially segregated from chlorophyll fluorescence in other cells. In one cell of line 2, fluorescence from mVenus runs along the outside of the chloroplast. As the images are a 2D representation of a 3D structure, it is not possible to identify whether the mVenus fluorescence is originating from outside the chloroplast in all cases. However, across the images, mVenus is not present throughout chloroplast and is particularly concentrated in strips or a network in the centre of the cell in images of line 2.

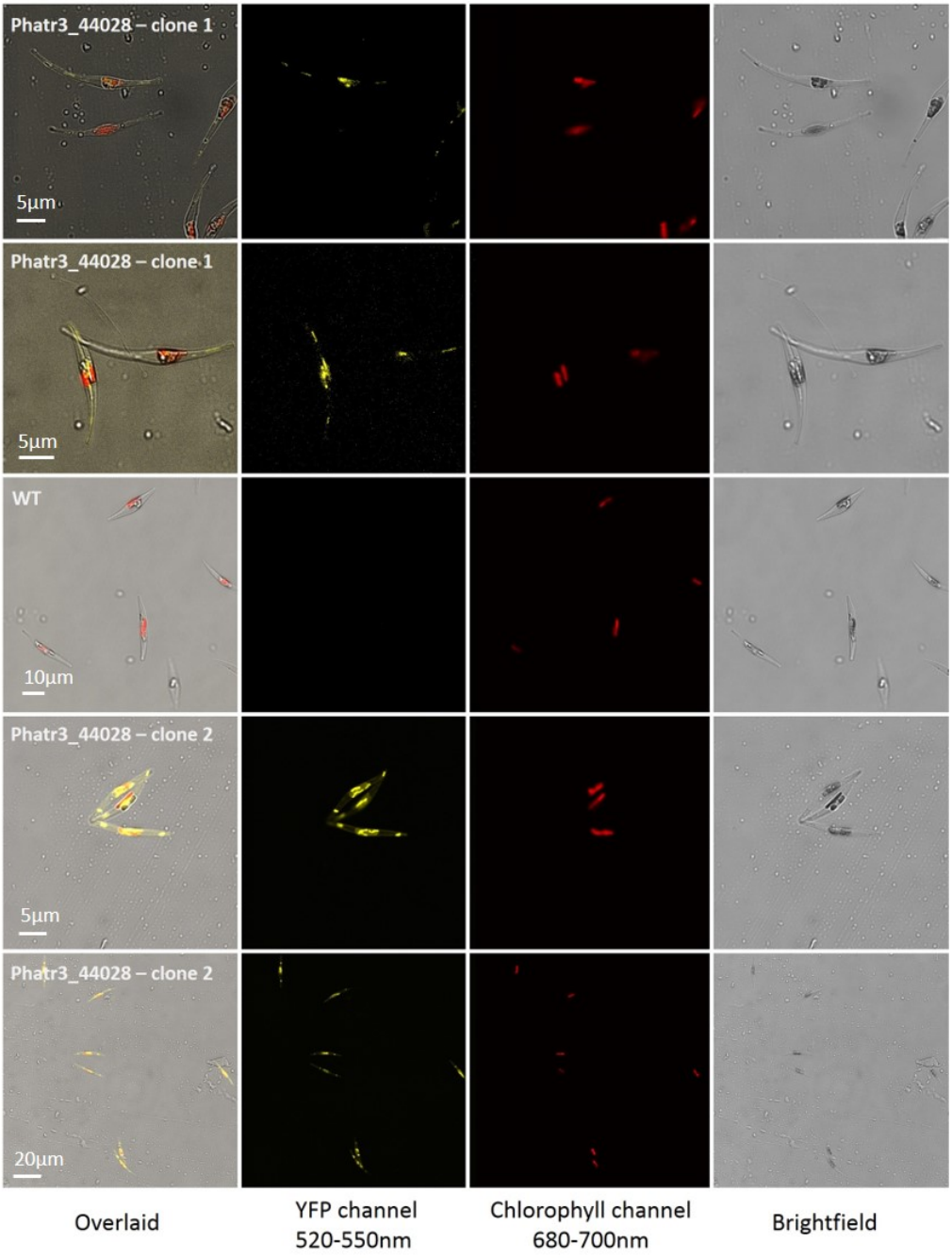


Figure 5.16 – Confocal microscope images of *P. tricornutum* transformed with plasmids for the expression of Phatr3_J44028 targeting peptide:mVenus fusions. Two positive fluorescence lines were identified with fluorescent confocal microscopy. The brightness and contrast of images of line 1 were altered to make the mVenus fluorescence visible when printed. WT images were altered in the same way while line 2 images were not altered. WT - wild type.

Images from the fluorescent Phatr3_J41624:mVenus lines are shown in Figure 5.17. Two clonal lines and a WT are shown. The contrast and brightness of the mVenus signals were altered in all images to render the mVenus emission visible when printed. This resulted in a visually stronger signal than was originally observed, although the pattern of fluorescence was maintained. The image of the wild type cells were adjusted in the same manner as the experimental lines to provide a comparison. In the experimental lines, faint mVenus fluorescence was observed to co-occur with chlorophyll fluorescence. The fluorescence was stronger in experimental lines compared to the wildtype. Additionally, stronger mVenus fluorescence was also observed in small regions at the side of the chloroplast, as observed in images of line 1, or the end of the chloroplast in images of line 2.

Figure 5.18 displays fluorescent microscopy images from two independent lines containing the mVenus positive control. Fluorescence from independent lines was observed in different locations compared to the experimental lines. Fluorescence occurred throughout the length of the cell and in a concentrated region next to or overlapping the autofluorescence of the chloroplast.

Liu et al. (2016a) identified the subcellular locations of a number of proteins in *P. tricornutum* that were targeted to cytoplasmic, endosomal, nuclear, mitochondrial and plastidial locations. Patterns of Phatr3_J44028:mVenus fluorescence were similar to those observed for endosomal membranes. The fluorescence at the ends of the cells and in the central region is similar to that observed for an endocytic marker, FM4-64. The fluorescence was also similar to that of the proteins, Sec61 and hDer1, said to be targeted to the ER, chloroplast ER (cER) and nuclear envelope and MIP1a and MIP1b, located at the cER and nuclear envelope. This was particularly the case for clone 1 that demonstrated more narrowly defined fluorescence. Patterns of Phatr3_J41624 fluorescence also find some similarity with that of proteins from (Liu et al., 2016a), demonstrating the centrally located parts of the fluorescence seen for ER targeted proteins. However, similarity is also seen with proteins located in other areas, particularly the cytoplasm, as can be seen by comparison with Figure 5.18. With the low level of fluorescence in Phatr3_J41624 lines, it is unclear whether fluorescence seen in the chloroplast represents fluorescence from the targeted protein or chlorophyll.

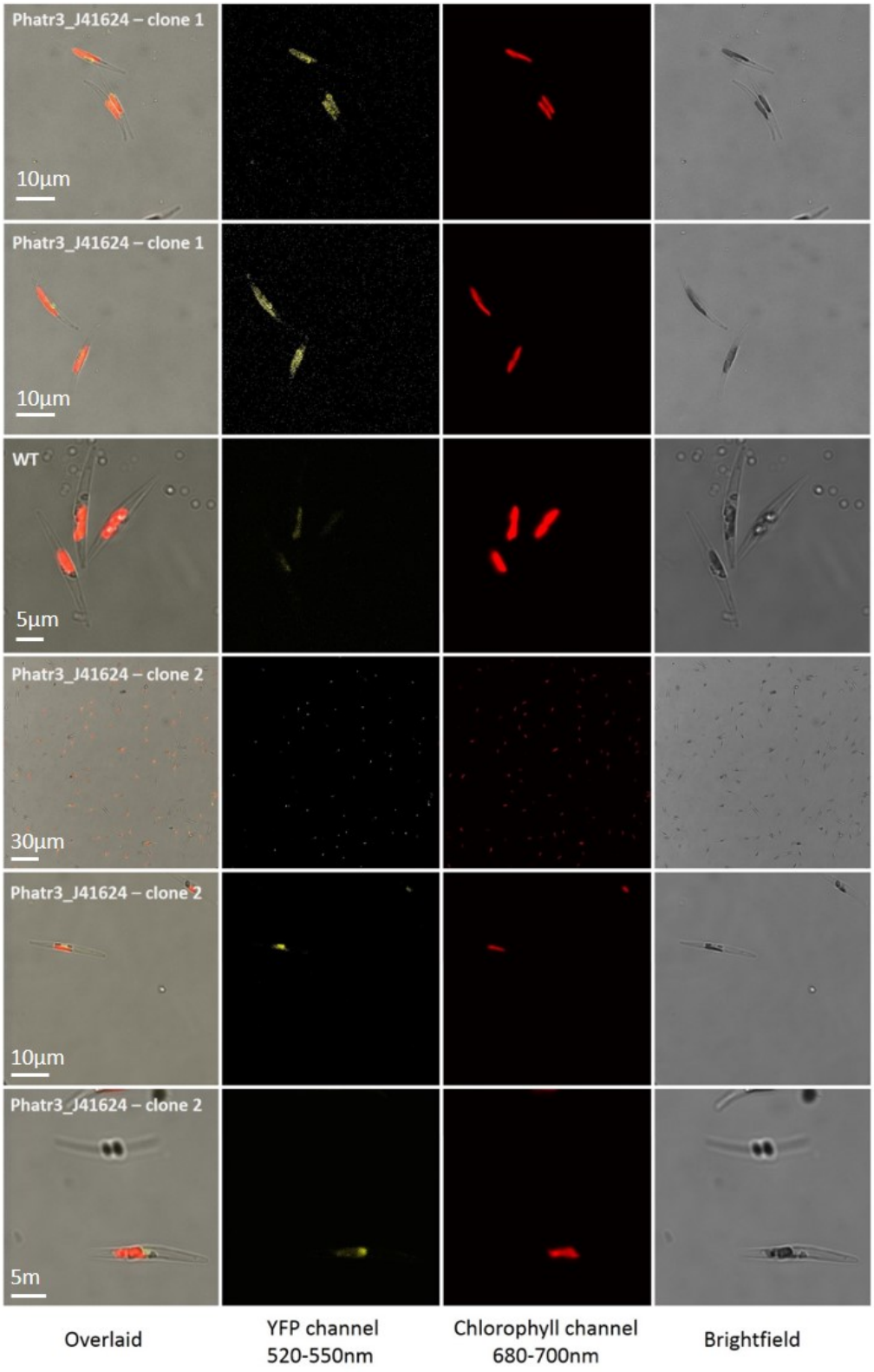


Figure 5.17 – Confocal microscope images of *P. tricornutum* transformed with plasmids for the expression of Phatr3_J41624:mVenus fusions. Two positive fluorescence lines were identified with fluorescent confocal microscopy. The brightness and contrast of the images were uniformly altered to make the mVenus fluorescence visible when printed. WT - wild type.

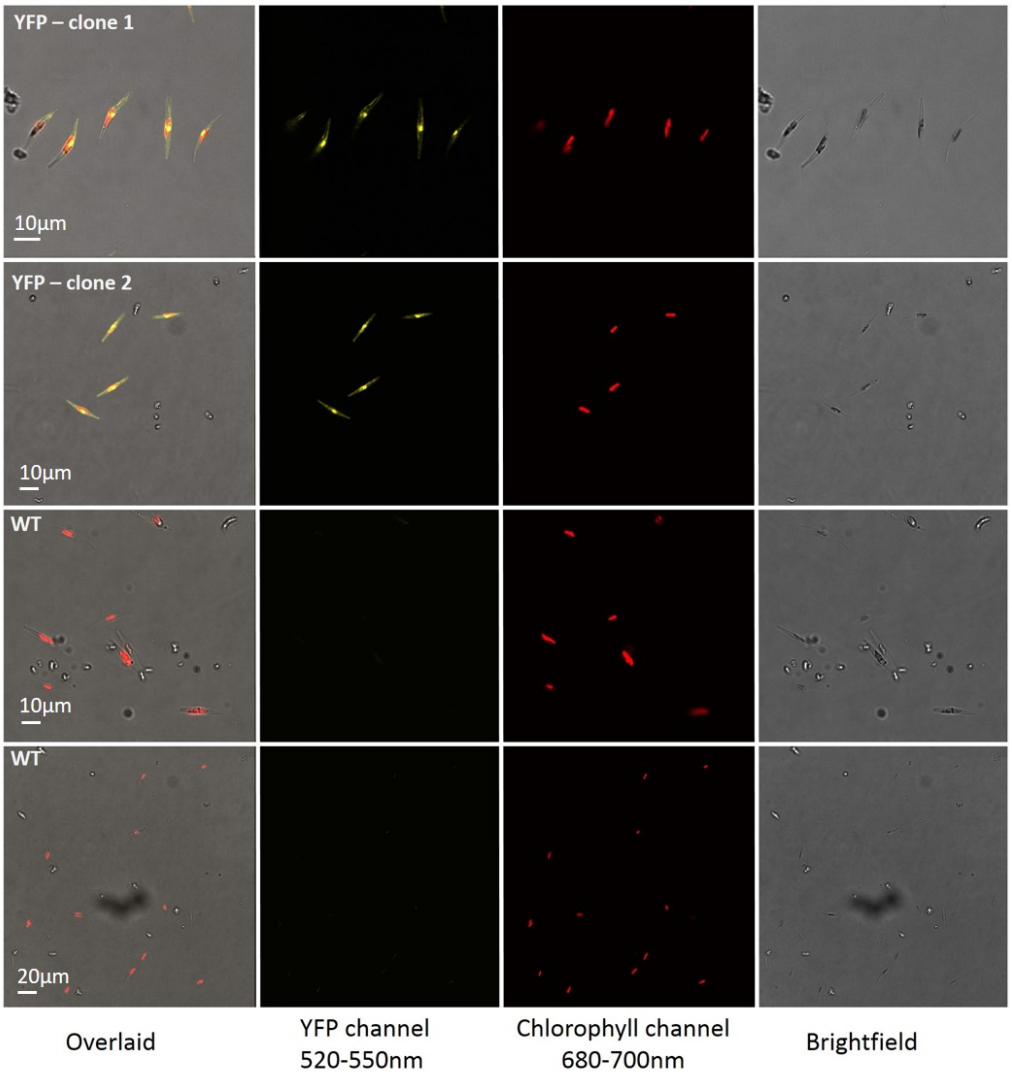


Figure 5.18 – Confocal microscope images of *P. tricornutum* transformed with plasmids for the expression of the mVenus positive control. Two independent mVenus expressing lines are shown. YFP - mVenus, WT - wild type.

5.4 Discussion

5.4.1 The purpose and potential mechanisms of decreases in MGDG during nitrogen depletion in *P. tricornutum*

5.4.1.1 The purpose of decreased MGDG and an outline of potentially involved processes

In this chapter I have presented work that is focused upon identifying TAG synthesis that uses constituents from recycled chloroplast membrane glycerolipids, in particular, MGDG. In this discussion, I frame my results in the broader questions of how and why a decrease in MGDG is brought about during nitrogen depletion in *P. tricornutum*. Ultimately, I propose that decreased MGDG during nitrogen depletion is important for preventing generation of ROS. During nitrogen depletion, it is thought that there is an oversupply of captured light energy compared to the needs of the cell (Wagner et al., 2017). Subsequently, this leads to generation of ROS mainly originating from the Mehler reactions at PSI but also from triplet state chlorophyll in PSII and over-reduction of the photosynthetic electron transport chain (Asada, 2006; Pospíšil, 2009, 2016). Production of more ROS than can be detoxified may cause photoinhibition or other cellular damage and hence the cell requires mechanisms to prevent this. In *A. thaliana* MGDG plays an essential role in the proper formation of the PSII and light harvesting complex II, in interactions between these complexes and in the efficiency of electron transport within PSII and between the two photosystems (Nakamura and Li-Beisson, 2016). These functions and the decreased MGDG content during nitrogen depletion suggest decreased amounts of the lipid play a role in reducing light capture and/or transfer of electrons through the electron transfer chain, thus preventing the generation of ROS.

There are a number of possible mechanisms that could decrease MGDG levels. The work in this chapter is based upon the assumption that MGDG is degraded. Previous studies had found that decreases in MGDG and increases in TAG both occurred during nitrogen depletion in *P. tricornutum* (Abida et al., 2015; Bower et al., 2015) and further to this, that the decrease in EPA in MGDG was matched

by the increase in TAG (Abida et al., 2015). Transfer of MGDG constituents to TAG during nitrogen depletion has also been seen in *C. reinhardtii* and *A. thaliana*. Furthermore, storage of EPA in TAG during nitrogen depletion would enable its rapid remobilisation upon resupply with nitrogen and might provide *P. tricornutum* with a comparative advantage in native habitats. It therefore seemed likely that MGDG would be degraded in *P. tricornutum* and that the released molecules would be recycled to produce TAG. Assuming MGDG is degraded, there are alternative fates that could befall its constituents other than incorporation into TAG. These fates include catabolism of the generated FAs by β-oxidation and incorporation into phytol esters. In this chapter, a number of enzymes were investigated that may specifically breakdown MGDG. However, part of the chloroplast and its lipids may also be degraded by the more general processes of autophagy. As an alternative to degradation, it is also possible that MGDG synthesis is decreased. With regard to the findings of the chapter, these various possibilities are discussed below, except for the potential roles of β-oxidation and autophagy that were discussed in Section 3.3.

5.4.1.2 Regarding altered synthesis of MGDG

Decreased amounts of MGDG during nitrogen depletion could originate from decreased synthesis of the lipid. In the systematic review presented in Chapter 3 two isoforms of MGDG synthase were downregulated under nitrogen depletion while a third was unregulated, as seen in Table 3.7. The impact of these changes on MGDG per cell are not clear however, given the decreasing rate of cell division. Separately, the ability of nitric oxide to inhibit MGDG synthase has recently been demonstrated (Dolch et al., 2017). The authors suggested that nitric oxide is formed by the activity of nitrate reductase (NR) on nitrate. In the study, nitric oxide associated (NOA) overexpression was found to upregulate NR expression and nitric oxide production. A role for NOA in nitrogen depletion was not observed, given its late upregulation under these conditions. However, nitric oxide levels were not measured under nitrogen depletion. Indeed, NR expression increases substantially and at an early time point under these conditions. This leaves open the possibility that nitric oxide is being produced in the early stages of nitrogen depletion in *P. tricornutum* and hence could inhibit MGDG synthase. A nitrite based production of nitric oxide during

nitrogen depletion seems counterintuitive but diatoms are known to store nitrogen sources and so this may be possible (Armbrust, 2009). As a result, a simple way to investigate potential inhibition of MGDG synthase during nitrogen depletion would be to measure nitric oxide levels.

5.4.1.3 Recycling of chloroplast membrane glycerolipids to TAG

During nitrogen depletion, TAG may be synthesised from potential sources of fatty acids and acylated glycerol backbones, namely MGDG and PG. N-depleted cells in which fatty acid synthesis was inhibited still fluoresced when stained with Nile Red stain in the early stages of nitrogen depletion. This fluorescence remained constant at the highest levels of cerulenin, indicating that fatty acid synthesis was maximally inhibited. The proportion of total TAG synthesis that could be represented by this fluorescence is commensurate to the proportion of TAG that MGDG degradation could provide, based on data in previous studies (Levitan et al., 2015; Abida et al., 2015). However, fluorescence of a similar level in the N-replete control treated with cerulenin challenges whether the fluorescence originates from TAG accumulation. Further investigation based on lipid extraction and subsequent TAG quantification using GC-FID would prevent this uncertainty. In the case that inhibition of FAS completely prevents TAG accumulation, it would still be useful to provide further evidence that a transfer of FAs from the chloroplast to TAG does not occur. The destination of degraded lipid components is important in understanding the response to nitrogen depletion. Furthermore, complete prevention of TAG accumulation by fatty acid synthesis inhibition does not preclude fatty acid transfer from the membrane to TAG. During nitrogen depletion, *de novo* synthesised fatty acids in *C. reinhardtii* are thought to first pass through MGDG prior to storage in TAGs (Li et al., 2012; Du et al., 2018).

Based on analogy of the system in *C. reinhardtii*, one of the candidate genes most likely to utilise fatty acids from MGDG for TAG synthesis is PDAT (Yoon et al., 2012). PtPDAT, Phatr3_J8860, was upregulated in one study prior to cessation of cell division, supporting a potential role in TAG synthesis during nitrogen depletion. PDATs transfer an acyl chain from the sn-2 position of lipids to the sn-3 position of DAGs (Xu et al., 2018; Yoon et al., 2012). As mentioned in Section 4.3, in *P.*

tricornutum, EPA within MGDG is primarily located at the sn-1 position (Abida et al., 2015). This makes it unlikely that PDAT directly transfers EPA from MGDG to DAG but does not preclude its role in transferring other FAs. The location of Phatr3_J8860 requires experimental characterisation. A mitochondrial location was predicted in the current study. However, this contrasts with predictions of a former member of the Plant Metabolism Group, Dr Litvinenko (Litvinenko, 2016). Dr Litvinenko did not identify a signal peptide in the N terminus that would be required for transport into the chloroplast endoplasmic reticulum but did identify the second part of a bipartite targeting signal, the chloroplast transit peptide. The location of Phatr3_J8860 in the cell is therefore ambiguous. In Chapter 4, characterisation was undertaken of the acyl chain that Phatr3_8860 prefers in its substrates. While the implications of this specificity for the identity of the acyl donor are presented, it should be noted that characterisation of the location of Phatr3_8860 is of import to this interpretation.

Interestingly, a second enzyme annotated as a PDAT/LCAT, Phatr3_J49702, was identified during the course of this work. During nitrogen depletion it was mostly unregulated but was transcriptionally down regulated in one study. Identification of a second putative PDAT is of interest as green algae and monocots have only a single isoform whereas dicots contain multiple (Pan et al., 2015). Multiple PDAT isoforms would add to the presumed flexibility of TAG synthesis that algae derive from multiple DGAT isoforms (Chen and Smith, 2012). Characterisation of Phatr3_J49702 as a functional PDAT would therefore increase our appreciation of the capabilities of diatom lipid metabolism.

An acylhydrolase is another enzyme class likely to degrade MGDG. In *C. reinhardtii*, CrPGD1 cleaves the acyl chain from the sn-1 position of MGDG (Li et al., 2012; Du et al., 2018). Out of the acylhydrolase candidates identified in *P. tricornutum*, a number were selected for investigation. Subcellular locations of two putative acylhydrolases indicated potential location in the endomembrane system for Phatr3_J44028 and in either the endomembrane system or cyotoplasm for Phatr3_J41624 while the ASAFind targeting software predicted their location in plastid and endomembrane system respectively. The localisation of Phatr3_J44028 fluorescence could include the cER (Liu et al., 2016b) and it is possible that MGDG

could be accessed as a substrate of the enzyme. However, further work to investigate the location of these two enzymes is still needed. In particular, strong, controllable expression from an inducible promoter would overcome potentially low levels of protein and/toxicity that may have contributed to the low number of clonal lines demonstrating fluorescence. Further to improving the genetic system, more informative screening methods could be used. Identification of location in specific subcellular compartments in *P. tricornutum* is complicated by the complex nature of the inner membranes. Therefore, colocalisation with dyes specific to certain compartments and immuno-electron microscopy would provide further clarity.

Out of the putative acylhydrolases, Phatr3_J44028, demonstrates the most promising transcript regulation and targeting and individual characterisation of this gene is worth pursuing as it is likely that it plays a role in acclimatisation to nitrogen depletion. Further to the activity of individual acylhydrolases, in order to completely breakdown MGDG the activity of an sn-1 specific acylhydrolase would have to be combined with an acylhydrolase or transferase that targeted the sn-2 position, e.g a PDAT and additionally, a galactosyl hydrolase. Galactosyl hydrolase candidates were identified bioinformatically and one of these, Phatr3_J54569 was upregulated 22 fold during nitrogen depletion in one study, see Table 5.2. This enzyme may actually be involved in metabolising chrysolaminarin but both possibilities warrant further investigation.

5.4.1.4 Phytol ester synthesis

A promising mechanism of MGDG degradation that was not pursued in this thesis is observed in *A. thaliana*. During nitrogen depletion, FA phytyl esters are synthesised from fatty acids and phytol. As in *P. tricornutum*, both chlorophyll a and MGDG levels decrease in *A. thaliana* during nitrogen depletion and MGDG likely supplies at least a portion of the FAs for synthesis of these fatty acid phytyl esters (FAPE) and may additionally supply DAG and FA for TAG synthesis (Gaude et al., 2007; Lippold et al., 2012). Phytol is released by the catabolism of chlorophyll during nutrient depletion and senescence in plants. Free phytol is toxic due to its detergent-like properties and it is therefore assumed that its release is carefully regulated by the cell. Chlorophyll a is also catabolised during nitrogen depletion in *P. tricornutum*

and it is therefore likely that diatoms possess a phytol detoxification mechanism. In general, detoxification of phytol is carried out by converting it to tocopherol or FAPE. The produced FAPE accumulates in lipid monolayer bound droplets that store neutral lipids on the thylakoid, called plastoglobuli. Generally, these stored lipids can include TAG, phytyl esters and tocopherol (Rottet et al., 2015; Spicher and Kessler, 2015) and *P. tricornutum* also accumulates plastoglobuli during nitrogen depletion (Balamurugan et al., 2017). Overall then, the production of phytol esters in *P. tricornutum* under nutrient depletion is possible based on current knowledge. Furthermore, although FAPEs have been detected in green algae, plants and dinoflagellates (Krauß and Vetter, 2018; Cranwell et al., 1990), detection in living diatoms would be a novel finding. Currently, the few studies on phytyl esters in diatoms have focused on the phytol chains of chlorophyll itself and have not detected the production of FAPEs in living cells (Louda et al., 2011; Mizoguchi et al., 2017). In addition to potential transfer of FAs from MGDG to TAG, the production of FAPEs in *P. tricornutum* during nitrogen depletion provides another potential commonality in the nitrogen depletion response with the distantly related archaeplastida lineage.

Focusing on FAPE synthesis as a means to investigate MGDG degradation under nitrogen depletion has several advantages over the approaches employed to investigate the fate of MGDG in this thesis. Firstly, detection of FAPEs offers a clear signal of FAPE synthesis during nitrogen depletion as the chemical is primarily thought to result from phytol detoxification in photosynthetic organism and therefore is not expected to be produced by other processes. Additionally, the release of phytol from chlorophyll a places enzymes involved in this function within the chloroplast. The PES 1 and 2 enzymes that carry out FAPE synthesis in *A. thaliana* have been characterised. These enzymes demonstrate MGDG acylhydrolase, fatty acid phytyl ester synthase (FAPES) and DGAT activities in vitro and FAPES and DGAT activities have been observed in vivo (Lippold et al., 2012). A PES homologue in *P. tricornutum* may therefore also be expected to act on MGDG and evidence from Chapter 3 supports MGDG and chlorophyll a degradation occurring contemporaneously in *P. tricornutum*. A BLAST search of the PES1 and 2 enzymes against the *P. tricornutum* proteome returns a limited number of strong hits: Phatr3_J43541,

Phatr3_J47663, Phatr3_EG00369, Phatr3_J34489, Phatr3_J50641, Phatr3_J37756 and therefore eases further investigation compared to the 50 hits returned for acylhydrolases. Phatr3_EG00369 is DGAT2E that has previously been identified (Chen, 2013). Expression of DGAT2E in *P. tricornutum* during nitrogen depletion did not have significant impacts upon TAG quantity or composition, as reported in Chapter 4, but more extensive characterisation work with a focus on FAPEs would be valuable. Additionally, Phatr3_J50641 and Phatr3_J37756 both contain PROSITE domain PS51257 at the N terminus. This domain is annotated as a localisation signal for prokaryotic membranes and would support a thylakoid location. Investigating the production of fatty acid phytyl esters during nitrogen depletion is therefore a promising course of action for furthering knowledge of MGDG degradation and its potential role in TAG accumulation in *P. tricornutum*.

6

Discussion

6.1 Introduction

Algae offer a rich field of study. The group embodies broad biotechnological potential and deep ecological significance while providing the means to answer important evolutionary and biological questions. Through this thesis, I have aimed to improve knowledge of a representative of one particular group of algae, the diatom *P. tricornutum*, particularly in areas related to TAG accumulation and chloroplast membrane recycling during nitrogen depletion. This discussion presents broader insights derived from my work. Algal research questions sit across scales and this is reflected in the structure of this discussion, beginning with aspects relevant to multiple lineages before focusing to diatom specific features. In the first part of this discussion, a potential commonality in the response to nitrogen depletion across photosynthetic lineages is discussed. The second part discusses challenges to investigating diatom lipid metabolism. Identifying how lipid metabolism operates in algal lineages is difficult due to their diverse origins. Work to provide more specific functional predictions of gene models would therefore prove useful. In addition, the benefits of using metabolic models to understand processes that lead to TAG accumulation are discussed. More specifically still, there are a number of open questions in diatom lipid metabolism that need addressing and these are detailed. I then turn to an industrial application and consider photosynthetic microalgal production of omega-3 LC-PUFAs. Finally, the key findings and recommendations for future work presented in this thesis are summarised in Tables 6.1 and 6.2 in Section 6.6.

6.2 Regarding the mechanism of chloroplast lipid decrease during nitrogen depletion in photosynthetic lineages

Thinking broadly, algae have in common a photosynthetic mode of life, ultimately derived from a common primary endosymbiosis. This provides a thread to tie these otherwise disparate lineages together and has underpinned my approach to studying plastid based responses to nitrogen depletion. In Section 5.4 the likely basis of the

decreased levels of MGDG in *P. tricornutum* during nitrogen depletion were outlined. More broadly, decreased levels of MGDG have been observed across numerous photosynthetic lineages during nitrogen depletion as outlined in Section 1. The same principle likely underlies this in both *P. tricornutum* and the other organisms: a decreased need for the products of photosynthesis during nitrogen depletion require mechanisms to prevent cellular damage caused by the production of ROS. Although, given the high quantities of MGDG in the thylakoid and its close association with photosystems (Rast et al., 2015), it is also plausible that decreased levels of MGDG are required to allow recycling of nitrogen from thylakoid proteins. As a result I hypothesise that decreased levels of MGDG form a conserved component of the response to nitrogen depletion across photosynthetic lineages. Multiple mechanisms by which this decrease can be brought about are outlined in Section 5.4 and include decreased synthesis of MGDG, complete catabolism of MGDG, transfer of MGDG constituents to TAG or the combination of fatty acids from MGDG with phytol from chlorophyll to form fatty acid phytol esters. I further suggest that these and potentially a limited number of other mechanisms will be responsible for decreased levels of MGDG and hence form a collection of processes that an organism can be screened for when investigating nitrogen depletion. There are several benefits that could arise from investigating this hypothesis. Firstly, it would highlight a process by which photosynthesis is reduced that is conserved in nitrogen depletion across photosynthetic lineages. Secondly, it would show the importance of MGDG to this process and so highlight questions surrounding the role of this lipid. Thirdly, in some microalgae such as *P. tricornutum*, MGDG is rich in EPA. Pursuing an understanding of its degradation in different lineages may therefore highlight enzymes of biotechnological interest. Furthermore, collation of these enzymes would provide data from which to investigate sequence determinants of the specificity of enzymes that act on galactolipids and so improve functional annotation of lipid metabolism genes. As the processes involved in decreased MGDG are hypothesised to be common to photosynthetic organisms as a whole and as these processes have already been identified in green algae-derived lineages, investigating whether the mechanisms occur in cyanobacteria and subsequently red algae would be an interesting place to start. Thinking more broadly than a decrease in MGDG during nitrogen depletion, this

approach of identifying processes that are likely common to plastids provides a useful way to frame investigations into different photosynthetic lineages.

6.3 Challenges in understanding lipid metabolism in *P. tricornutum*

6.3.1 Annotation of lipid catabolism genes

Genes have been annotated for many of the important lipid biosynthesis pathways in *P. tricornutum* (Fabris et al., 2012; Zulu et al., 2018; Rastogi et al., 2018). In contrast, many lipid catabolism genes remain imprecisely defined. This impairs knowledge of the presence of lipid catabolic features and interpretation of the importance of these processes under particular conditions, for example, acclimatisation to nitrogen depletion. In this study, I made use of functional annotation combined with expression and other data to identify putative acylhydrolases and lipid catabolic enzymes. Acylhydrolases are important genes in lipid catabolism but specificity of their annotation was generally limited to α/β-hydrolase. Many α/β-hydrolases were detected in *P. tricornutum* as illustrated in Tables A.1-A.3. However, α/β-hydrolases are a particularly diverse superfamily of enzymes that are defined by the presence of an α/β fold but otherwise differ substantially in sequence and function. They can act as acetylcholinesterases, lipases, thioesterases, and proteases among numerous other catalytic and non-catalytic functions.

The lack of a defined function engendered by α/β-hydrolase annotation offers an opportunity to improve the functional annotation of a broad swathe of lipid catabolism enzymes in *P. tricornutum*. Firstly, more refined annotation of α/β-hydrolases are available in the ESTHER database (Lenfant et al., 2013), in the CLADE and DAMA annotations made available by Rastogi et al. (2018) and as Interpro annotations, available in the Ensemble Protist *P. tricornutum* database. These annotations can separate lipase activity from other functions and provide more specific functional information. This information varies in informativeness but its compilation and assessment would provide a resource for further knowledge generation. Secondly, comparative studies of characterised α/β-hydrolases may enable better functional

prediction in *P. tricornutum*. One particularly useful sub-category of α/β-hydrolase is phospholipase A (PLA). PLAs hydrolyse the acyl chains from the sn-1 and sn-2 position of glycerolipids. Multiple PLAs have been characterised in plants and variously demonstrate activity on phospholipids and glycolipids (Burke and Dennis, 2009; Chen et al., 2013). Identification of PLAs in *P. tricornutum* and production of a phylogeny that includes experimentally characterised PLAs may therefore provide insight into the lipid classes that those in *P. tricornutum* act on. Furthermore, such a study may enable identification of sequence or structural elements that dictate specificity to substrate class and would further improve functional annotation of PLAs. Due to the presence of EPA in the galactolipids of some microalgae, development of annotation that predicts activity on MGDG may be useful for biotechnology. In this regard, improved functional annotation of PDATs should also be undertaken as CrPDAT has been demonstrated to act on MGDG. This activity is atypical for a PDAT and in combination with the substantially different sequence of CrPDAT compared to plant homologues (Pan et al., 2015) highlights the value of investigating sequence-lipid class specificity determinants for these enzymes. Indeed, PDATs are related to PLAs (Xu et al., 2018) and so both groups of enzymes may be studied collectively. Information from the characterised PLAs may provide insight into PDAT substrate specificity.

6.3.2 Understanding the mechanisms that cause TAG accumulation

As discussed in Section 3.3, TAG accumulation has now been attributed to alterations in expression of numerous enzymes under various conditions. Among these are TAG synthesis enzymes, branched chain amino acid degradation enzymes, genes involved in increased flux from malate to oxaloacetate and intermediate metabolism genes more widely (Niu et al., 2013; Cui et al., 2018; Dinamarca et al., 2017; Levitan et al., 2015; Ge et al., 2014). In addition, the systematic review of nitrogen depletion omic studies presented in Chapter 3 noted numerous alterations to intermediate metabolism and lipid metabolism alike. Currently, developing insight into the effect of the changes that nitrogen depletion has wrought has primarily been based on human expertise. However, given the numerous metabolic changes that may cause

TAG accumulation and the fact that metabolism is a network of interdependent reactions, the use of tools that can take these factors into account would enable improved understanding of the metabolic regulation that causes TAG accumulation. Whole genome metabolic models that simulate the network of reactions within a cell are such a tool and two have been developed for *P. tricornutum* (Levering et al., 2016; Kim et al., 2016). These metabolic networks are derived from genomic and experimental information and carrying out flux balance analysis (FBA) enables the impact of changes in gene expression on theoretical metabolite quantity to be investigated. Currently, FBA investigations of acclimatisation to nitrogen depletion have focused on intermediate metabolism (Levitan et al., 2015). Given the impact that regulation of lipid metabolism can have on TAG accumulation, it would be useful to expand these analyses to include this area of metabolism through use of the metabolic models. Furthermore, modelling may provide insight into how manipulation of the expression of lipid metabolism genes in previous studies, such as DGAT2D (Dinamarca et al., 2017) altered flux in wider metabolism and thus caused TAG to accumulate.

6.4 Open questions in diatom lipid metabolism

6.4.1 Location of TAG accumulation in diatoms

The location of TAG synthesis and storage has implications for efforts to engineer lipid accumulation in *P. tricornutum* and for the understanding of lipid transport. In plants, yeast and animals, TAG is mainly synthesised on the ER (Chapman and Ohlrogge, 2012; Xu and Shanklin, 2016; Moriyama et al., 2018) and accumulated in monolayer surrounded lipid droplets (LD) that bud from the ER. In the last decade, several studies found evidence to support plastidial accumulation of LDs in *C. reinhardtii* (Fan et al., 2011; Goodson et al., 2011; Goold et al., 2016) and so it was suggested that the location of LD formation in algae may differ to that in other organisms. In *P. tricornutum*, possible plastidial production of LDs is also supported by a 3D imaging study in which LDs appeared to be either in close association or within the chloroplast (Cavonius et al., 2015). In contrast, a recent analysis concluded that the observations of LDs within chloroplasts in *C. reinhardtii* were

the result of imaging cross sections of lipid droplets that are located in invaginations of the chloroplast but not in the organelle itself (Moriyama et al., 2018). This explanation would also account for the ambiguous observations in *P. tricornutum*. Indeed, LDs are clearly visible in the cytoplasm of *P. tricornutum* and the membrane lipid composition of isolated LDs suggests an origin at the outermost membrane of the chloroplast, the chloroplast endoplasmic reticulum (Balamurugan et al., 2017; Lupette et al., 2019). Supporting ER localisation of LD synthesis is the predicted presence of DGAT1, DGAT2B, DGAT2C and DGAT3 within the ER membrane, as outlined in Chapter 3.

However, there are also reasons to think TAG synthesis may occur in the plastid. Firstly, DGAT2D is predicted to be localised to the plastid, caused accumulation of TAG under replete conditions (Dinamarca et al., 2017) and is transcriptionally upregulated after cessation of cell division during nitrogen depletion (Levitan et al., 2015; Alipanah et al., 2015). Preliminary fluorescent label work carried out in the Plant Metabolism Group also supported plastidial targeting of DGAT2D. Collectively, this supports the ability of a plastid targeted DGAT to play a substantial role in TAG accumulation. Additional to DGAT2D, TAG synthesis in the plastid could be catalysed by PDAT or DGAT2E. A plastidial location is a possibility for PDAT, as outlined in Section 5.4. Additionally, DGAT2E is one of the genes with highest sequence similarity to the *A. thaliana* fatty acid phytyl ester synthesis genes PES1 and PES2, also discussed in Section 5.4, and so a plastidial location would be expected based on this function. Further, the overexpression of plastidial GPAT1 and LPAT1 in *P. tricornutum* causes accumulation of TAG (Wang et al., 2018b). This may indicate that TAG is synthesised within the plastid using DAG produced from the prokaryotic pathway or that DAG or TAG are exported from the plastid and respectively incorporated into TAG and lipid droplets on the ER.

Plastoglobules are an alternative neutral lipid storage organelle to LDs and form at the thylakoid. Plastoglobules have been observed in the plastids of *P. tricornutum* and indeed are found in a wide range of plastid containing organisms (Balamurugan et al., 2017; Moriyama et al., 2018; Lohscheider and Río Bártulos, 2016). Synthesis of TAG in the plastid of *P. tricornutum* may therefore lead to its accumulation in plastoglobules rather than lipid droplets. Plastoglobules also contain other neutral

lipids such as prenol lipids and those observed in *P. tricornutum* are smaller than the LDs accumulated during nitrogen depletion. These factors, combined with their thylakoid location indicate that the TAG stored in plastoglobules may originate in specialised responses such as remodelling of the thylakoid membrane rather than bulk storage of TAG. This leads to the hypothesis that at least one of the multiple DGAT isoforms in *P. tricornutum* is involved in the production of TAG that is stored in plastoglobules. Indeed, a role in synthesis of small amounts of TAG may explain why overexpression of DGAT2D and DGAT2E did not cause a detectable increase in TAG accumulation under nitrogen depletion, as noted in Chapter 4. Furthermore, this would suggest that attention should be paid to plastoglobuli when investigating DGAT2E, PDAT and particularly DGAT2D.

Still, the subcellular location of LDs in *P. tricornutum* is not resolved and more analysis making use of 3D imaging systems and/ or markers of subcellular location would be useful. With the possiblity of differentiated roles of the multiple DGATs, such a measurement system could be applied to several TAG accumulating conditions to investigate whether the location of LDs varies. Assuming LDs are extraplastidial, their close association with the chloroplast noted in Cavonius et al. (2015) would raise questions about the origin of DAG backbones and possibly TAG: a contribution from plastidial TAG or DAG to extraplastidial LDs is a possibility and would have implications for engineering efforts.

6.4.2 Transport of lipids over the membranes of the plastid

Movement of lipids over membranes is an important part of lipid metabolism in *P. tricornutum*. Furthermore, increasing knowledge of lipid transport in *P. tricornutum* may provide insight into their transport across the ochrophytes. The most salient example of lipid transport in *P. tricornutum* is the export of 16 and/or 18 carbon FAs from the site of *de novo* synthesis in the plastid for further elongation and desaturation to produce EPA. The enzymes involved in EPA synthesis are predicted to be located in the extraplastidial membranes. Therefore, the EPA that is found in chloroplast lipids must be re-imported after synthesis. Observations that require a movement of glycerolipids over membranes have not been made in *P. tricornutum* but nevertheless this would would fit with current understanding of lipid metabolism

and considerations of the location of TAG synthesis as outlined in the section above. Abida et al. (2015) also concluded that the presence of 16 carbon acyl chains at the sn-2 position in a portion of extraplastidial lipids indicated that either extraplastidial LPAT had a broad acyl chain substrate specificity or that glycerolipid backbones were exported from the plastid.

In *P. tricornutum*, the typically mentioned method of fatty acid export from the plastid is hydrolysis by a thioesterase followed by diffusion or transport over the membrane and subsequent attachment to CoA (Zulu et al., 2018; Abida et al., 2015). A thioesterase has been identified that is suggested to function in either import or export of FAs from the plastid (Hao et al., 2018; Gong et al., 2011). This enzyme was knocked out in experimental lines and although the rate of cell division was reduced and lipid profiles were altered, MGDG and EPA were both still present in the chloroplast, indicating that multiple mechanisms of glycerolipid and/or FA transport between the ER and plastid are employed by *P. tricornutum*. This is therefore an important and relatively untapped area for research. In recent years, knowledge of transport of FAs and lipids in plants has improved and a number of mechanisms have been characterised including the involvement of protein-mediated transmembrane transport, contact sites between membranes, diffusion within membranes and vesicle trafficking (Salminen et al., 2016; Du et al., 2016; Bates, 2016; Block and Jouhet, 2015; Hurlock et al., 2014; Li-Beisson et al., 2017). These known mechanisms offer a broader starting point from which to approach investigations in *P. tricornutum* than the current focus on thioesterases.

6.5 Regarding the commercial production of EPA and DHA in autotrophic microalgae

One unifying theme in this thesis has been the development of knowledge that will enable autotrophic microalgal production of the omega-3 LC-PUFAs EPA and DHA. Based on nutritional recommendations, the demand for these valuable fatty acids is already greater than current production and actual demand is expected to grow with increasing global income and population (Salem and Eggersdorfer, 2015). Aquaculture is the major user of omega-3 LC-PUFAs and the major commercial source used

in this industry is wild caught fish and fish oil (Sprague et al., 2017). Marine fish are a limited resource and this has led to decreased inclusion in aquaculture feed and as a result lowered omega-3 LC-PUFAs in aquaculture grown fish. Several other sources of EPA and DHA are available to meet nutritional and aquaculture requirements, of which those thought most likely to substantially fulfill demand include marine microbes, genetically modified oil crops and autotrophic microalgae. Of these three, heterotrophic production of oils is the primary commercially available source, with commercial GM plant oil production expected in 2019 (Sprague et al., 2017). Autotrophic algal strains that produce both EPA and DHA are not yet ready for industrial production.

The future availability of several omega-3 LC-PUFA sources prompts the question of which will be most desirable from an economic and environmental perspective. Although autotrophic production will not require an input of carbon source, algal and plant production could have negative characteristics such as higher costs and agricultural land requirements respectively. Several studies have carried out techno-economic analyses of autotrophic algal production and found that algal production of omega-3 PUFAs will be economically viable in the near future but that this may require co-production with other compounds in some cases (Chauton et al., 2015; Ruiz et al., 2016). However, studies directly comparing the heterotrophic microbe, plant and phototrophic algal production methods do not appear to have been carried out. Therefore, to evaluate the economic and environmental case for autotrophic microalgal production of omega-3 PUFAs comparative techno-economic and life cycle analyses of the different methods of EPA and DHA production would prove useful. Although commercial factors may ultimately guide the choice of production method, an understanding of economic costs related to each option would support companies in deciding whether to produce the fatty acids with a given system. Furthermore, the heightened awareness of environmental sensitivity that has resulted from the focus on climate change may see commercial decision makers place increasing importance upon other environmental considerations. Additionally, these types of analyses would provide EPA and DHA yield targets for scientists developing strains and cultivation processes.

When this thesis was begun, targeting of these omega-3 LC-PUFAs to TAG to

use developed extraction techniques was thought to be desirable in the development of commercial strains of autotrophic microalge sources of EPA and DHA (Khozin-Goldberg et al., 2011; Hamilton et al., 2014). Recently, this has been questioned by studies that have demonstrated the ability of commercially relevant fish species to derive omega-3 LC-PUFAs from whole algae, although in some cases this had a slight detrimental impact on growth (Sprague et al., 2017). One recent study found that when up to 15% of feed was replaced with *P. tricornutum*, a species of bass grew comparably to when fed a control diet that included LC-PUFA rich oil but the impact of higher inclusions of *P. tricornutum* were not tested (de Cruz et al., 2018). This highlights the need for further investigations on whether whole algal feed is sufficient for aqualculture needs. Despite this, direction of omega-3s to oil is still desirable for certain applications, for example, to provide oil to the baby formula industry (Liu et al., 2016c). Within this thesis, PtPDAT demonstrated a preference for incorporation of EPA into TAG and so is a candidate for biotechnological applications in which targeting to omega-3 LC-PUFAs to oil is required. Furthermore, during nitrogen depletion, down regulation of Lands' cycle genes, elongases and desaturases involved in EPA synthesis was observed, presumably resulting in the increased synthesis of 16:0 and 16:1. This highlights the potential for overexpression of these enzymes to increase the presence of EPA in the cell as has previously been reported by Hamilton et al. (2014). A combination of upregulation of Lands' cycle enzymes, elongases and desauturases in combination with PtPDAT may therefore be a useful strategy for increasing the direction of EPA to TAG.

6.6 Summary of thesis findings and recommendations for future work

The key findings from this thesis are summarised in Table 6.1 and recommendations for future work are presented in Table 6.2, with the most noteworthy recommendations highlighted with asterisks. Table entries are associated with the discussion section in which they were outlined.

Table 6.1 – Key knowledge takeaways.

Chapter 3 Discussion	Chapter 4 Discussion	Chapter 5 Discussion
Synthesis of 16:0 and 16:1 favoured due to downregulation of the Lands cycle enzymes, elongases, fatty acid desaturase 2, long chain acyl-CoA synthetases and a possibly by upregulation of Δ9 desaturase	DGAT2B demonstrates a preference for incorporation of 18:1 in TAG	The transfer of MGDG constituents to TAG has not been ruled out
	PDAT demonstrates a preference for incorporation of 20:5 into TAG	The acyl hydrolase Phatr3_J44028 is likely located in the endomembrane system
TAG accumulation can result from changes in both intermediate metabolism and lipid metabolism and these areas of metabolism may work cooperatively to accumulate TAG during nitrogen depletion		Putative lipid catabolism genes of *P. tricornutum* were compiled
		Domain annotation indicates the presence of two PDATs in *P. tricornutum*
Culture conditions likely impact adaptation to nitrogen depletion		
Autophagy is likely activated during nitrogen depletion in *P. tricornutum*		
A framework was outlined to aid understanding of TAG accumulation during nitrogen depletion		

Table 6.2 – Recommended future work. Work that forms the next steps in developing this thesis are marked with asterisks

Chapter 3 Discussion	Chapter 4 Discussion	Chapter 5 Discussion	Chapter 6
Investigate impacts of culture conditions on adaptation to nitrogen depletion. Initial acclimatisation state of particular interest.	*Knock out the different DGATs and PDATs to understand their function	Measure nitric oxide levels during nitrogen depletion to evaluate effects on MGDG synthase	Pursue the hypothesis that reduced MGDG is a component of nitrogen depletion across photosynthetic lineages
Identify the major source of acetyl-CoA for FA synthesis during nitrogen depletion	Use knowledge of characterised DGATs across the stramopiles to predict roles of *P. tricornutum* DGATs	*Using GC-FID, measure TAG accumulation during nitrogen depletion when FA synthesis is completely inhibited.	Compile annotation of α/β-hydrolase genes in *P. tricornutum*
Further investigate role of LPAT , DGAT2D, plastid Δ9 desaturase, malonyl CoA:ACP transacylase and PDAT in TAG accumulation during nitrogen depletion.	*Measure the glycerolipid levels in the PDAT overexpression line	*Experimentally characterise the location of PDATs and DGATs	*Produce a phylogeny of phopholipas A genes and PDATs containing characterised examples and those in *P. tricornutum*
*Upregulate fatty acid synthase enzymes to investigate its role in TAG accumulation	Measure the FA content of PDAT and DGAT2B ovexpressing lines	*Functionally characterise Phatr3_J44028	Carry out metabolic modelling techniques to investigate the causes of TAG accumulation during nitrogen depletion and in genetically modified strains
Delineate function of enzymes that may be involved in beta oxidation or BCAA degradation	Measure the FA position in TAG in the PDAT and DGAT2B ovexpressing lines	Express the PDAT of C. reinhardtii in the plastid of *P. tricornutum* and evaluate its impact on TAG	*Investigate the location of TAG synthesis with particular focus on transport of glycerolipids and plastoglobuli
Investigate how upregulation of DGAT2D causes alterations to the transcript levels of other genes.	*Develop a TAG synthesis deficient *P. tricornutum* strain or one that produces minimal TAG	Investigate the function of a putative galactosyl hydroalase, Phatr3_J54569, that is highly upregulated during nitrogen depletion.	*Outline the potential FA and glycerolipid transport processes present in *P. tricornutum* and investigate these
Investigate substrate preference of Δ9 desaturase Determine the plastid proteome in *P. tricornutum*		*Attempt to detect FA phytol esters in *P. tricornutum* during nitrogen depletion and subsequently investigate the identified candidate genes	Carry out techno-economic and environmental life cycle comparisons of different sources of EPA and DHA
*A strategy to increase omega-3 PUFAs content: increase activity of Lands cycle, FA elongation, FA desaturase 2 and long chain acyl-CoA synthetase activity *Investigate the role of autophagy in chloroplast membrane breakdown.			*Pursue an increase in EPA and DHA in *P. tricornutum* through alterations to expression of enzymes involved in the Lands cycle, FA elongation, FA desaturation and long chain acyl-CoA synthesis.
Investigate the role of beta oxidation in the decrease of MGDG			

Bibliography

Abida, H., L.-J. Dolch, C. Meï, V. Villanova, M. Conte, M. A. Block, G. Finazzi, O. Bastien, L. Tirichine, C. Bowler, F. Rébeillé, D. Petroutsos, J. Jouhet, and E. Maréchal

2015. Membrane glycerolipid remodeling triggered by nitrogen and phosphorus starvation in Phaeodactylum tricornutum. *Plant physiology*, 167(1):118–36.

Adl, S. M., A. G. Simpson, C. E. Lane, J. Lukeš, D. Bass, S. S. Bowser, M. W. Brown, F. Burki, M. Dunthorn, V. Hampl, A. Heiss, M. Hoppenrath, E. Lara, L. L. Gall, D. H. Lynn, H. McManus, E. A. Mitchell, S. E. Mozley-Stanridge, L. W. Parfrey, J. Pawlowski, S. Rueckert, L. Shadwick, C. L. Schoch, A. Smirnov, and F. W. Spiegel

2012. The revised classification of eukaryotes. *Journal of Eukaryotic Microbiology*, 59(5):429–493.

Adler-Agnon, Z., S. Leu, A. Zarka, S. Boussiba, and I. Khozin-Goldberg

2017. Novel promoters for constitutive and inducible expression of transgenes in the diatom Phaeodactylum tricornutum under varied nitrate availability. *Journal of Applied Phycology*, 30(5):2763–2772.

Aebersold, R. and M. Mann

2003. Mass spectrometry-based proteomics. *Pharmacogenomics*, 4(4):463–476.

Alboresi, A., G. Valle, J. Jouhet, G. Perin, A. Meneghesso, T. Morosinotto, M. A. Block, N. Vitulo, G. Giuliano, G. Diretto, and E. Maréchal

2016. Light Remodels Lipid Biosynthesis in Nannochloropsis gaditana by Modulating Carbon Partitioning Between Organelles. *Plant Physiology*, 171:2468–2482.

Alcolombri, U., S. Ben-Dor, E. Feldmesser, Y. Levin, D. S. Tawfik, and A. Vardi
2015. Identification of the algal dimethyl sulfide-releasing enzyme: A missing link in the marine sulfur cycle. *Science (New York, N.Y.)*, 348(6242):1466–9.

Alipanah, L., J. Rohloff, P. Winge, A. M. Bones, and T. Brembu
2015. Whole-cell response to nitrogen deprivation in the diatom Phaeodactylum tricornutum. *Journal of Experimental Botany*, 66(20):6281–6296.

Alipanah, L., P. Winge, J. Rohloff, J. Najafi, T. Brembu, and A. M. Bones
2018. Molecular adaptations to phosphorus deprivation and comparison with nitrogen deprivation responses in the diatom Phaeodactylum tricornutum. *PLOS ONE*, 13(2):e0193335.

Allen, A. E., C. L. Dupont, M. Oborník, A. Horák, A. Nunes-Nesi, J. P. McCrow, H. Zheng, D. A. Johnson, H. Hu, A. R. Fernie, and C. Bowler
2011. Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *Nature*, 473(7346):203–207.

Alonso, D. L., E. H. Belarbi, J. M. Fernández-Sevilla, J. Rodríguez-Ruiz, and E. Molina Grima
2000. Acyl lipid composition variation related to culture age and nitrogen concentration in continuous culture of the microalga Phaeodactylum tricornutum. *Phytochemistry*, 54(5):461–71.

Alonso, D. L., C. I. del Castillo, J. L. Sánchez, J. A. Sánchez Pérez, and F. G. Camacho
1994. Quantitative Genetics of Fatty Acid Variation in Isochrysis Galbana (Prymnesiophyceae) and Phaeodactylum Tricornutum (Bacillariophyceae). *Journal of Phycology*, 30(3):553–558.

Arabolaza, A., H. Gramajo, H. Alvarez, E. Rodriguez, and S. Altabe
2008. Multiple Pathways for Triacylglycerol Biosynthesis in Streptomyces coelicolor. *Applied and Environmental Microbiology*, 74(9):2573–2582.

Arao, T., A. Kawaguchi, and M. Yamada
1987. Positional distribution of fatty acids in lipids of the marine diatom Phaeodactylum tricornutum. *Phytochemistry*, 26(9):2573–2576.

Arao, T. and M. Yamada

1994. Biosynthesis of polyunsaturated fatty acids in the marine diatom, Phaeodactylum tricornutum. *Phytochemistry*, 35(5):1177–1181.

Armbrust, E. V.

2009. The life of diatoms in the world's oceans. *Nature*, 459(7244):185–192.

Asada, K.

2006. Production and Scavenging of Reactive Oxygen Species in Chloroplasts and Their Functions. *Plant Physiology*, 141(2):391–396.

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, and G. Sherlock

2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, 25(1):25–9.

Aznar-Moreno, J. A. and T. P. Durrett

2017. Simultaneous Targeting of Multiple Gene Homeologs to Alter Seed Oil Production in Camelina sativa. *Plant and Cell Physiology*, 58(7):1260–1267.

Badger, M. R., S. Von Caemmerer, S. Ruuska, H. Nakano, A. Laisk, J. F. Allen, K. Asada, H. C. Matthijs, and H. Griffiths

2000. Electron flow to oxygen in higher plants and algae: Rates and control of direct photoreduction (Mehler reaction) and rubisco oxygenase. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1402):1433–1446.

Bagnato, C., M. B. Prados, G. R. Franchini, N. Scaglia, S. E. Miranda, and M. V. Beligni

2017. Analysis of triglyceride synthesis unveils a green algal soluble diacylglycerol acyltransferase and provides clues to potential enzymatic components of the chloroplast pathway. *BMC Genomics*, 18(223).

Balamurugan, S., X. Wang, H. L. Wang, C. J. An, H. Li, D. W. Li, W. D. Yang, J. S. Liu, and H. Y. Li

2017. Occurrence of plastidial triacylglycerol synthesis and the potential regulatory

role of AGPAT in the model diatom Phaeodactylum tricornutum. *Biotechnology for Biofuels*, 10(1):1–14.

Ball, M. and M. Weeda

2015. The hydrogen economy - Vision or reality? *International Journal of Hydrogen Energy*, 40(25):7903–7919.

Barka, F., M. Angstenberger, T. Ahrendt, W. Lorenzen, H. B. Bode, and C. Büchel

2016. Identification of a triacylglycerol lipase in the diatom Phaeodactylum tricornutum. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1861(3):239–248.

Barney, B. M., B. D. Wahlen, E. Garner, J. Wei, and L. C. Seefeldt

2012. Differences in substrate specificities of five bacterial wax ester synthases. *Applied and Environmental Microbiology*, 78(16):5734–5745.

Bates, P. D.

2016. Understanding the control of acyl flux through the lipid metabolic network of plant oil biosynthesis. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1861(9):1214–1225.

Baurain, D., H. Brinkmann, J. Petersen, N. Rodríguez-Ezpeleta, A. Stechmann, V. Demoulin, A. J. Roger, G. Burger, B. F. Lang, and H. Philippe

2010. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Molecular Biology and Evolution*, 27(7):1698–1709.

Benjamini, Y. and H. Yosef

1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300.

Bertrand, M.

2010. Carotenoid biosynthesis in diatoms. *Photosynthesis Research*, 106:89–102.

Block, M. A. and J. Jouhet

2015. Lipid trafficking at endoplasmic reticulum-chloroplast membrane contact sites. *Current Opinion in Cell Biology*, 35:21–29.

Bojko, M., K. Brzostowska, P. Kuczyska, D. Latowski, M. Olchawa-Pajor, W. Krzeszowiec, A. Waloszek, and K. Strzaka

2013. Temperature effect on growth, and selected parameters of Phaeodactylum tricornutum in batch cultures. *Acta Biochimica Polonica*, 60(4):861–864.

Bolte, K., N. Gruenheit, G. Felsner, M. S. Sommer, U. G. Maier, and F. Hempel

2011. Making new out of old: Recycling and modification of an ancient protein translocation system during eukaryotic evolution. *BioEssays*, 33(5):368–376.

Borowitzka, M. A.

2013. High-value products from microalgae—their development and commercialisation. *Journal of Applied Phycology*, 25(3):743–756.

Borrull, A., G. López-Martínez, M. Poblet, R. Cordero-Otero, and N. Rozès

2015. New insights into the toxicity mechanism of octanoic and decanoic acids on Saccharomyces cerevisiae. *Yeast*, 32(5):451–460.

Bower, I., M. L. Hamilton, R. P. Haslam, and O. Sayanova

2015. MRes report: Changes in the lipidome of Phaeodactylum tricornutum grown under nitrogen deplete conditions. Technical report.

Bowler, C., A. E. Allen, J. H. Badger, J. Grimwood, K. Jabbari, A. Kuo, U. Maheswari, C. Martens, F. Maumus, R. P. Otillar, E. Rayko, A. Salamov, K. Vandepoele, B. Beszteri, A. Gruber, M. Heijde, M. Katinka, T. Mock, K. Valentin, F. Verret, J. A. Berges, C. Brownlee, J.-P. Cadoret, A. Chiovitti, C. J. Choi, S. Coesel, A. De Martino, J. C. Detter, C. Durkin, A. Falciatore, J. Fournet, M. Haruta, M. J. J. Huysman, B. D. Jenkins, K. Jiroutova, R. E. Jorgensen, Y. Joubert, A. Kaplan, N. Kröger, P. G. Kroth, J. La Roche, E. Lindquist, M. Lommer, V. Martin–Jézéquel, P. J. Lopez, S. Lucas, M. Mangogna, K. McGinnis, L. K. Medlin, A. Montsant, M.-P. O. Secq, C. Napoli, M. Obornik, M. S. Parker, J.-L. Petit, B. M. Porcel, N. Poulsen, M. Robison, L. Rychlewski, T. A. Rynearson, J. Schmutz, H. Shapiro, M. Siaut, M. Stanley, M. R. Sussman, A. R. Taylor, A. Vardi, P. von Dassow, W. Vyverman, A. Willis, L. S. Wyrwicz, D. S. Rokhsar, J. Weissenbach, E. V. Armbrust, B. R. Green, Y. Van de Peer, and I. V.

Grigoriev

2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature*, 456(7219):239–244.

BP

2018. 2018 BP Energy Outlook. *BP Energy Outlook*, Pp. 1–125.

Brayner, R., A. Couté, J. Livage, C. Perrette, and C. Sicard

2011. Micro-algal biosensors. *Analytical and Bioanalytical Chemistry*, 401(2):581–597.

Buhman, K. K., S. J. Smith, S. J. Stone, J. J. Repa, J. S. Wong, F. F. Knapp, B. J. Burri, R. L. Hamilton, N. A. Abumrad, and R. V. Farese

2002. DGAT1 is not essential for intestinal triacylglycerol absorption or chylomicron synthesis. *Journal of Biological Chemistry*, 277(28):25474–25479.

Burke, J. E. and E. A. Dennis

2009. Phospholipase A2 biochemistry. *Cardiovascular Drugs and Therapy*, 23(1):49–59.

Burki, F., N. Okamoto, J. F. Pombert, and P. J. Keeling

2012. The evolutionary history of haptophytes and cryptophytes: Phylogenomic evidence for separate origins. *Proceedings of the Royal Society B: Biological Sciences*, 279(1736):2246–2254.

Burki, F., M. Sakaguchi, S. Kumar, T. Hashimoto, A. Horak, K. S. Jakobsen, T. Cavalier-Smith, K. Shalchian-Tabrizi, P. J. Keeling, J. Bråte, J. M. Archibald, J. Pawlowski, D. Klaveness, and Y. Inagaki

2009. Large-Scale Phylogenomic Analyses Reveal That Two Enigmatic Protist Lineages, Telonemia and Centroheliozoa, Are Related to Photosynthetic Chromalveolates. *Genome Biology and Evolution*, 1:231–238.

Burrows, E. H., N. B. Bennette, D. Carrieri, J. L. Dixon, A. Brinker, M. Frada, S. N. Baldassano, P. G. Falkowski, and G. C. Dismukes

2012. Dynamics of Lipid Biosynthesis and Redistribution in the Marine Diatom Phaeodactylum tricornutum Under Nitrate Deprivation. *BioEnergy Research*, 5(4):876–885.

Byrdwell, W. C. and W. E. Neff

2002. Dual parallel electrospray ionization and atmospheric pressure chemical ionization mass spectrometry (MS), MS/MS and MS/MS/MS for the analysis of triacylglycerols and triacylglycerol oxidation products. *Rapid Communications in Mass Spectrometry*, 16(4):300–319.

Caballero, M. A., D. Jallet, L. Shi, C. Rithner, Y. Zhang, and G. Peers

2016. Quantification of chrysolaminarin from the model diatom Phaeodactylum tricornutum. *Algal Research*, 20:180–188.

Cagliari, A., R. Margis, F. Dos Santos Maraschin, A. C. Turchetto-Zolet, G. Loss, and M. Margis-Pinheiro

2011. Biosynthesis of triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology*, 2(1):40–52.

Campbell, P. K., T. Beer, and D. Batten

2011. Life cycle assessment of biodiesel production from microalgae in ponds. *Bioresource Technology*, 102(1):50–56.

Cao, H., J. M. Shockey, K. T. Klasson, D. C. Chapital, C. B. Mason, and B. E. Scheffler

2013. Developmental Regulation of Diacylglycerol Acyltransferase Family Gene Expression in Tung Tree Tissues. *PLoS ONE*, 8(10).

Cao, J., S. Hahm, M. Shi, S. Stahler, J. McKew, V. Suri, H. Peng, T. Gareski, M. Perreault, R. E. Gimeno, A. Qadri, J. F. Tobin, J. Jones, X. Xu, X. Huang, and Y. Zhou

2011. Targeting Acyl-CoA:Diacylglycerol Acyltransferase 1 (DGAT1) with Small Molecule Inhibitors for the Treatment of Metabolic Diseases. *Journal of Biological Chemistry*, 286(48):41838–41851.

Carneiro, M. L. N., F. Pradelle, S. L. Braga, M. S. P. Gomes, A. R. F. Martins, F. Turkovics, and R. N. Pradelle

2017. Potential of biofuels from algae: Comparison with fossil fuels, ethanol and biodiesel in Europe and Brazil through life cycle assessment (LCA). *Renewable and Sustainable Energy Reviews*, 73:632–653.

Cases, S., S. J. Stone, P. Zhou, E. Yen, B. Tow, K. D. Lardizabal, T. Voelker, and R. V. Farese

2001. Cloning of DGAT2, a Second Mammalian Diacylglycerol Acyltransferase, and Related Family Members. *Journal of Biological Chemistry*, 276(42):38870–38876.

Cavonius, L., H. Fink, J. Kiskis, E. Albers, I. Undeland, and A. Enejder

2015. Imaging of Lipids in Microalgae with Coherent Anti-Stokes Raman Scattering Microscopy 1[OPEN]. *Plant physiology*, 167:603–616.

Chapman, K. D. and J. B. Ohlrogge

2012. Compartmentation of triacylglycerol accumulation in plants. *Journal of Biological Chemistry*, 287(4):2288–2294.

Chauton, M. S., K. I. Reitan, N. H. Norsker, R. Tveterås, and H. T. Kleivdal

2015. A techno-economic analysis of industrial production of marine microalgae as a source of EPA and DHA-rich raw material for aquafeed: Research challenges and possibilities. *Aquaculture*, 436:95–103.

Chen, G., M. S. Greer, and R. J. Weselake

2013. Plant phospholipase A: Advances in molecular biology, biochemistry, and cellular function. *Biomolecular Concepts*, 4(5):527–532.

Chen, J. E.

2013. *The identification and characterisation of diacylglycerol acyltransferases (DGATs) in microalgae*. PhD thesis, University of Cambridge.

Chen, J. E. and A. G. Smith

2012. A look at diacylglycerol acyltransferases (DGATs) in algae. *Journal of Biotechnology*, 162(1):28–39.

Chi, X., R. Hu, X. Zhang, M. Chen, N. Chen, L. Pan, T. Wang, M. Wang, Z. Yang, Q. Wang, and S. Yu

2014. Cloning and functional analysis of three diacylglycerol acyltransferase genes from peanut (Arachis hypogaea L.). *PLoS ONE*, 9(9):1–15.

Chrismadha, T. and M. A. Borowitzka

1994. Effect of cell density and irradiance on growth, proximate composition and

eicosapentaenoic acid production ofPhaeodactylum tricornutum grown in a tubular photobioreactor. *Journal of Applied Phycology*, 6(1):67–74.

Clarens, A. F., E. P. Resurreccion, M. A. White, and L. M. Colosi

2010. Environmental Life Cycle Comparison of Algae to Other Bioenergy Feedstocks. *Environmental Science & Technology*, 44(5):1813–1819.

Cortassa, S., B. O'Rourke, R. L. Winslow, and M. A. Aon

2009. Control and regulation of integrated mitochondrial function in metabolic and transport networks. *International journal of molecular sciences*, 10(4):1500–13.

Courchesne, N. M. D., A. Parisien, B. Wang, and C. Q. Lan

2009. Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of Biotechnology*, 141(1-2):31–41.

Cranwell, A., G. H. Jaworski, and H. M. Bickley

1990. Hydrocarbons, sterols, esters and fatty acids in six freshwater chlorophytes. *Phytochemistry*, 29(1):145–151.

Cruz de Carvalho, M. H., H.-X. Sun, C. Bowler, and N.-H. Chua

2016. Noncoding and coding transcriptome responses of a marine diatom to phosphate fluctuations. *New Phytologist*, 210(2):497–510.

Cui, Y., J. Zhao, Y. Wang, S. Qin, and Y. Lu

2018. Characterization and engineering of a dual-function diacylglycerol acyltransferase in the oleaginous marine diatom Phaeodactylum tricornutum. *Biotechnology for Biofuels*, 11(1):32.

Cui, Y., G. Zheng, X. Li, H. Lin, P. Jiang, and S. Qin

2013. Cloning and characterization of a novel diacylglycerol acyltransferase from the diatom Phaeodactylum tricornutum. *Journal of Applied Phycology*, 25(5):1509–1512.

Curien, G., C. Giustini, M. Matringe, M. Kuntz, L. Magneschi, D. Petroutsos, V. Villanova, G. Finazzi, G. Forti, and S. Flori

2016. The Water to Water Cycles in Microalgae. *Plant and Cell Physiology*, 57(7):1354–1363.

Daboussi, F., S. Leduc, A. Maréchal, G. Dubois, V. Guyot, C. Perez-Michaut, A. Amato, A. Falciatore, A. Juillerat, M. Beurdeley, D. F. Voytas, L. Cavarec, and P. Duchateau

2014. Genome engineering empowers the diatom Phaeodactylum tricornutum for biotechnology. *Nature communications*, 5:3831.

Dahlqvist, A., H. Ronne, U. Stahl, A. Banas, M. Lee, S. Stymne, M. Lenman, and L. Sandager

2000. Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proceedings of the National Academy of Sciences*, 97(12):6487–6492.

Dautermann, O. and M. Lohr

2017. A functional zeaxanthin epoxidase from red algae shedding light on the evolution of light-harvesting carotenoids and the xanthophyll cycle in photosynthetic eukaryotes. *Plant Journal*, 92(5):879–891.

de Cruz, C. R., A. Lubrano, and D. M. Gatlin

2018. Evaluation of microalgae concentrates as partial fishmeal replacements for hybrid striped bass Morone sp. *Aquaculture*, 493(May):130–136.

De Martino, A., A. Bartual, A. Willis, A. Meichenin, B. Villazán, U. Maheswari, and C. Bowler

2011. Physiological and Molecular Evidence that Environmental Changes Elicit Morphological Interconversion in the Model Diatom Phaeodactylum tricornutum. *Protist*, 162(3):462–481.

De Martino, A. D., A. Meichenin, J. Shi, K. Pan, and C. Bowler

2007. Genetic and phenotypic characterization of Phaeodactylum tricornutum (Bacillariophyceae) accessions. *Journal of Phycology*, 43(5):992–1009.

De Riso, V., R. Raniello, F. Maumus, A. Rogato, C. Bowler, and A. Falciatore

2009. Gene silencing in the marine diatom Phaeodactylum tricornutum. *Nucleic acids research*, 37(14).

Delorme-Axford, E. and D. J. Klionsky

2018. Transcriptional and post-transcriptional regulation of autophagy in the yeast Saccharomyces cerevisiae. *Journal of Biological Chemistry*, 293(15):5396–5403.

Dinamarca, J., O. Levitan, G. K. Kumaraswamy, D. S. Lun, and P. G. Falkowski
2017. Overexpression of a diacylglycerol acyltransferase gene in Phaeodactylum tricornutum directs carbon towards lipid biosynthesis. *Journal of Phycology*, 53(2):405–414.

Diner, R. E., V. A. Bielinski, C. M. Noddings, A. E. Allen, J. Jablanovic, J. Jansson, M. A. Anzelmatti, N. C. Lian, B. J. Karas, N. A. Nguyen, C. L. Dupont, P. D. Weyman, J. L. Espinoza, A. K. Kang, and J. B. McQuaid
2017. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proceedings of the National Academy of Sciences*, 114(29):E6015–E6024.

Diner, R. E., C. L. Dupont, V. A. Bielinski, P. D. Weyman, A. E. Allen, and R. E. Diner
2016. Refinement of the Diatom Episome Maintenance Sequence and Improvement of Conjugation-Based DNA Delivery Methods. *Frontiers in Bioengineering and Biotechnology*, 4(65).

Dolch, L.-J. and E. Maréchal
2015. Inventory of Fatty Acid Desaturases in the Pennate Diatom Phaeodactylum tricornutum. *Marine Drugs*, 13(3):1317–1339.

Dolch, L.-J., C. Rak, G. Perin, G. Tourcier, R. Broughton, M. Leterrier, T. Morosinotto, F. Tellier, J.-D. Faure, D. Falconet, J. Jouhet, O. Sayanova, F. Beaudoin, and E. Maréchal
2017. A Palmitic Acid Elongase Affects Eicosapentaenoic Acid and Plastidial Monogalactosyldiacylglycerol Levels in Nannochloropsis. *Plant physiology*, 173(1):742–759.

Domergue, F., J. Lerchl, U. Zähringer, and E. Heinz
2002. Cloning and functional characterization of Phaeodactylum tricornutum front-end desaturases involved in eicosapentaenoic acid biosynthesis. *European Journal of Biochemistry*, 269(16):4105–4113.

Domergue, F., P. Spiekermann, J. Lerchl, C. Beckmann, O. Kilian, P. G. Kroth, W. Boland, U. Zaehringer, and E. Heinz

2003. New Insight into Phaeodactylum tricornutum Fatty Acid Metabolism. Cloning and Functional Characterization of Plastidial and Microsomal delta-12Fatty Acid Desaturases. *Plant physiology*, 131:1648–1660.

Dorrell, R. G., G. Gile, G. McCallum, R. Méheust, E. P. Bapteste, C. M. Klinger, L. Brillet-Guéguen, K. D. Freeman, D. J. Richter, and C. Bowler

2017. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. *eLife*, 6(e23717).

Draaisma, R. B., R. H. Wijffels, P. (Ellen) Slegers, L. B. Brentner, A. Roy, and M. J. Barbosa

2013. Food commodities from microalgae. *Current Opinion in Biotechnology*, 24(2):169–177.

Drabkin, H., D. Hill, S. Carbon, C. Mungall, M. Munoz-Torres, J. Lomax, D. Osumi-Sutherland, and P. Roncaglia

2015. Gene Ontology Consortium : going forward. 43(November 2014):1049–1056.

Du, Z. Y., T. Arias, W. Meng, and M. L. Chye

2016. Plant acyl-CoA-binding proteins: An emerging family involved in plant development and stress responses. *Progress in Lipid Research*, 63(78):165–181.

Du, Z.-Y. and C. Benning

2016. Triacylglycerol Accumulation in Photosynthetic Cells in Plants and Algae. In *Lipids in Plant and Algae Development*, Y. Nakamura and Y. Li-Beisson, eds., chapter 8, Pp. 179–205. Springer, Cham.

Du, Z.-Y., B. F. Lucker, K. Zienkiewicz, T. E. Miller, A. Zienkiewicz, B. B. Sears, D. M. Kramer, and C. Benning

2018. Galactoglycerolipid Lipase PGD1 Is Involved in Thylakoid Membrane Remodeling in Response to Adverse Environmental Conditions in Chlamydomonas.

Dyrløv Bendtsen, J., H. Nielsen, G. Von Heijne, and S. Brunak

2004. Improved prediction of signal peptides — SignalP 3.0. *J. Mol. Biol.*

El-Gebali, S., J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. Tosatto, and R. D. Finn

2019. The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1):D427–D432.

Emanuelsson, O., S. Brunak, G. Von Heijne, and H. Nielsen

2007. Locating proteins in the cell using TargetP, SignalP and related tools.

EMBL-EBI

2019. EBI: Next generation sequencing practical course. www.ebi.ac.uk/training/online/course/ebi-next-generation-sequencing-practical-course.

Engler, C., M. Youles, R. Gruetzner, T.-M. Ehnert, S. Werner, J. D. G. Jones, N. J. Patron, and S. Marillonnet

2014a. A Golden Gate Modular Cloning Toolbox for Plants. *ACS Syn*, 3:839–843.

Engler, C., M. Youles, R. Gruetzner, T.-M. Ehnert, S. Werner, J. D. G. Jones, N. J. Patron, and S. Marillonnet

2014b. Quick Guide to Type IIS Cloning, TSL SynBio.

Ensembl Protist

. Ensembl Protist: Phaeodactylum tricornutum.

Eppink, M. H. M., G. Olivieri, J. H. Reith, C. van den Berg, M. J. Barbosa, and R. H. Wijffels

2017. From Current Algae Products to Future Biorefinery Practices: A Review. In *Advanced Biochemical Engineering Biotechnology*, volume 1, Pp. 99–123.

EU

2001. White Paper: European Transport Policy for 2010: time to decide.

EU

2011. Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system.

Fabris, M., M. Matthijs, S. Rombauts, W. Vyverman, A. Goossens, and G. J. Baart
2012. The metabolic blueprint of Phaeodactylum tricornutum reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *The Plant Journal*, 70(6):1004–1014.

Fahy, E., F. Spener, E. A. Dennis, C. R. H. Raetz, M. Nishijima, S. Subramaniam, G. van Meer, T. Shimizu, M. J. O. Wakelam, and R. C. Murphy
2008. Update of the LIPID MAPS comprehensive classification system for lipids. *Journal of Lipid Research*, 50(Supplement):S9–S14.

Falciatore, A., R. Casotti, C. Leblanc, C. Abrescia, and C. Bowler
1999. Transformation of nonselectable reporter genes in marine diatoms. *Marine Biotechnology*, 1(3):239–251.

Falciatore, A., K. Lease, E. Ingham, J. C. Walker, and C. O. Plant
2012. Perception of Environmental Signals by a Marine Diatom Perception of Environmental Signals by a Marine Diatom. 2363(2000):2363–2367.

Falkowski, P. G., R. T. Barber, and V. Smetacek
1998. Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, 281(5374):200–206.

Fan, J., C. Andre, and C. Xu
2011. A chloroplast pathway for the de novo biosynthesis of triacylglycerol in Chlamydomonas reinhardtii. *FEBS Letters*, 585(12):1985–1991.

Fawley, M. W.
1984. Effects of light intensity and temperature interactions on growth characteristics of Phaeodactylum tricornutum (Bacillariophyceae). *Journal of Phycology*, 20(1):67–72.

Fazal, M., A. Haseeb, and H. Masjuki
2011. Biodiesel feasibility study: An evaluation of material compatibility; performance; emission and engine durability. *Renewable and Sustainable Energy Reviews*, 15(2):1314–1324.

Field, C. B., M. J. Behrenfeld, J. T. Randerson, and P. Falkowski

1998. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science (New York, N.Y.)*, 281(5374):237–40.

Finco, A. M. d. O., L. D. G. Mamani, J. C. de Carvalho, G. V. de Melo Pereira, V. Thomaz-Soccol, and C. R. Soccol

2017. Technological trends and market perspectives for production of microbial oils rich in omega-3. *Critical Reviews in Biotechnology*, 37(5):656–671.

Finn, R. D., T. K. Attwood, P. C. Babbitt, A. Bateman, P. Bork, A. J. Bridge, H.-Y. Chang, Z. Dosztányi, S. El-Gebali, M. Fraser, J. Gough, D. Haft, G. L. Holliday, H. Huang, X. Huang, I. Letunic, R. Lopez, S. Lu, A. Marchler-Bauer, H. Mi, J. Mistry, D. A. Natale, M. Necci, G. Nuka, C. A. Orengo, Y. Park, S. Pesseat, D. Piovesan, S. C. Potter, N. D. Rawlings, N. Redaschi, L. Richardson, C. Rivoire, A. Sangrador-Vegas, C. Sigrist, I. Sillitoe, B. Smithers, S. Squizzato, G. Sutton, N. Thanki, P. D. Thomas, S. C. E. Tosatto, C. H. Wu, I. Xenarios, L.-S. Yeh, S.-Y. Young, and A. L. Mitchell

2017. InterPro in 2017-beyond protein family and domain annotations. *Nucleic acids research*, 45(D1):D190–D199.

Flori, S., P.-H. Jouneau, G. Finazzi, E. Maréchal, and D. Falconet

2016. Ultrastructure of the Periplastidial Compartment of the Diatom Phaeodactylum tricornutum. *Protist*, 167(3):254–267.

Francius, G., B. Tesson, E. Dague, V. Martin-Jézéquel, and Y. F. Dufrêne

2008. Nanostructure and nanomechanics of live Phaeodactylum tricornutum morphotypes. *Environmental Microbiology*, 10(5):1344–1356.

Fretland, A. J. and C. J. Omiecinski

2000. Epoxide hydrolases: Biochemistry and molecular biology. *Chemico-Biological Interactions*, 129(1-2):41–59.

Fu, W., K. Wichuk, and S. Brynjólfsson

2015. Developing diatoms for value-added products: challenges and opportunities. *New Biotechnology*, 32(6):547–551.

Gatenby, C. M., D. M. Orcutt, D. A. Kreeger, B. C. Parker, V. A. Jones, and R. J.

Neves

2003. Biochemical composition of three algal species proposed as food for captive freshwater mussels. *Journal of Applied Phycology*, 15(1):1–11.

Gaude, N., C. Bréhélin, G. Tischendorf, F. Kessler, and P. Dörmann

2007. Nitrogen deficiency in Arabidopsis affects galactolipid composition and gene expression and results in accumulation of fatty acid phytyl esters. *Plant Journal*, 49(4):729–739.

Ge, F., W. Huang, Z. Chen, C. Zhang, Q. Xiong, C. Bowler, J. Yang, J. Xu, and H. Hu

2014. Methylcrotonyl-CoA Carboxylase Regulates Triacylglycerol Accumulation in the Model Diatom Phaeodactylum tricornutum. *The Plant Cell*, 26:1681–1697.

Geider, R., B. Osborne, and J. Raven

1985. Light dependence of growth and photosynthesis in Phaeodactylum tricornutum (Bacillariophyceae). 21:609–619.

Gibbs, S. P.

1981. The Chloroplast Endoplasmic Reticulum: Structure, Function, and Evolutionary Significance. *International Review of Cytology*, 72:49–99.

Gong, Y., X. Guo, X. Wan, Z. Liang, and M. Jiang

2011. Characterization of a novel thioesterase (PtTE) from Phaeodactylum tricornutum. *Journal of Basic Microbiology*, 51(6):666–672.

Gong, Y., J. Zhang, X. Guo, X. Wan, Z. Liang, C. J. Hu, and M. Jiang

2013. Identification and characterization of PtDGAT2B, an acyltransferase of the DGAT2 acyl-coenzyme A: diacylglycerol acyltransferase family in the diatom Phaeodactylum tricornutum. *FEBS letters*, 587(5):481–7.

Goodson, C., R. Roth, Z. T. Wang, and U. Goodenough

2011. Structural Correlates of Cytoplasmic and Chloroplast Lipid Body Synthesis in Chlamydomonas reinhardtii and Stimulation of Lipid Body Production with Acetate Boost. *Eukaryotic Cell*, 10(12):1592–1606.

Goold, H. D., S. Cuiné, B. Legeret, Y. Liang, S. Brugière, P. Auroy, H. Javot, M. Tardif, B. J. Jones, F. Beisson, G. Peltier, and Y. Li-Beisson

2016. Saturating Light Induces Sustained Accumulation of Oil in Plastidal Lipid Droplets in Chlamydomonas reinhardtii. *Plant Physiology*, 171(August):pp.00718.2016.

Gould, S. B., R. F. Waller, and G. I. McFadden

2008. Plastid Evolution. *Annual Review of Plant Biology*, 59(1):491–517.

Grattepanche, J.-D., C. E. Lane, C. F. Delwiche, D. L. Paim Pinto, L. M. Walker, B. M. Ott, and L. A. Katz

2018. Microbial Diversity in the Eukaryotic SAR Clade: Illuminating the Darkness Between Morphology and Molecular Data. *BioEssays*, 40(4):1700198.

Griffiths, A., J. Miller, and D. Suzuki

2000. *An Introduction to Genetic Analysis*, 7th edition. New York: W.H Freeman.

Gruber, A. and P. G. Kroth

2017. Intracellular metabolic pathway distribution in diatoms and tools for genome-enabled experimental diatom research. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372(1728).

Gruber, A., G. Rocap, P. G. Kroth, E. V. Armbrust, and T. Mock

2015. Plastid proteome prediction for diatoms and other algae with secondary plastids of the red lineage. *The Plant journal : for cell and molecular biology*, 81(3):519–28.

Gruber, A., S. Vugrinec, F. Hempel, S. B. Gould, U.-G. Maier, and P. G. Kroth

2007. Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif. *Plant molecular biology*, 64(5):519–30.

Gschloessl, B., Y. Guermeur, and J. M. Cock

2008. HECTAR: A method to predict subcellular targeting in heterokonts. *BMC Bioinformatics*, 9(1):393.

Guerra, L. T., O. Levitan, M. J. Frada, J. S. Sun, P. G. Falkowski, and G. C.

Dismukes

2013. Regulatory branch points affecting protein and lipid biosynthesis in the diatom Phaeodactylum tricornutum. *Biomass and Bioenergy*, 59:306–315.

Guihéneuf, F., S. Leu, A. Zarka, I. Khozin-Goldberg, I. Khalilov, and S. Boussiba

2011. Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase1-like gene (PtDGAT1) from the diatom Phaeodactylumtricornutum. *FEBS Journal*, 278(19):3651–3666.

Guillard, R. R. L. and J. H. Ryther

1962. Studies of marine planktonic diatoms: I. Cyclotella nanahusdedt, and Detonula confervea (cleve) gran. *Canadian Journal of Microbiology*, 8(2):229–239.

Guo, Y., Z. Fu, and J. E. Van Eyk

2007. A Proteomic Primer for the Clinician. *Proceedings of the American Thoracic Society*, 4(1):9–17.

Haider, S. and R. Pal

2013. Integrated Analysis of Transcriptomic and Proteomic Data. *Current Genomics*, 14:91–110.

Hamilton, M. L., R. P. Haslam, J. A. Napier, and O. Sayanova

2014. Metabolic engineering of Phaeodactylum tricornutum for the enhanced accumulation of omega-3 long chain polyunsaturated fatty acids. *Metabolic Engineering*, 22:3–9.

Hao, X., L. Luo, J. Jouhet, F. Rébeillé, E. Maréchal, H. Hu, Y. Pan, X. Tan, Z. Chen, L. You, H. Chen, F. Wei, and Y. Gong

2018. Enhanced triacylglycerol production in the diatom Phaeodactylum tricornutum by inactivation of a Hotdog-fold thioesterase gene using TALEN-based targeted mutagenesis. *Biotechnology for Biofuels*, 11(1):312.

Haseeb, A., M. Fazal, M. Jahirul, and H. Masjuki

2011. Compatibility of automotive materials in biodiesel: A review. *Fuel*, 90(3):922–931.

Hosp, F. and M. Mann

2017. A Primer on Concepts and Applications of Proteomics in Neuroscience. *Neuron*, 96(3):558–571.

Houten, S. M., S. Violante, F. V. Ventura, and R. J. Wanders

2016. The Biochemistry and Physiology of Mitochondrial Fatty Acid β-Oxidation and Its Genetic Disorders. *Annual Review of Physiology*, 78(1):23–44.

Hrdlickova, R., M. Toloue, B. Tian, M. Genetics, and R. New

2017. RNA-Seq methods for transcriptome analysis. 8(1):1–24.

Huang, A., L. He, and G. Wang

2011. Identification and characterization of microRNAs from Phaeodactylum tricornutum by high-throughput sequencing and bioinformatics analysis. *BMC Genomics*, 12.

Huang, R., L. Tirichine, J. Ding, M. H. Cruz de Carvalho, K. Gao, C. Bowler, and X. Lin

2018. A Potential Role for Epigenetic Processes in the Acclimation Response to Elevated pCO2 in the Model Diatom Phaeodactylum tricornutum. *Frontiers in Microbiology*, 9(January):1–12.

Huete-Ortega, M., K. Okurowska, R. V. Kapoore, M. P. Johnson, D. J. Gilmour, and S. Vaidyanathan

2018. Effect of ammonium and high light intensity on the accumulation of lipids in Nannochloropsis oceanica (CCAP 849/10) and Phaeodactylum tricornutum (CCAP 1055/1). *Biotechnology for Biofuels*, 11(1):60.

Hurlock, A. K., R. L. Roston, K. Wang, and C. Benning

2014. Lipid trafficking in plant cells. *Traffic*, 15(9):915–932.

Illumina

2016. The Next Generation of Gene Expression Profiling with RNA-Seq. Technical report.

Ishida, H., M. Izumi, S. Wada, and A. Makino

2014. Roles of autophagy in chloroplast recycling. *Biochimica et Biophysica Acta - Bioenergetics*, 1837(4):512–521.

ITF

2018. Transport carbon dioxide and the Paris Climate Agreement: Reviewing the impacts of nationally determined contributions. Technical report.

Izumi, M., H. Ishida, S. Nakamura, and J. Hidema

2017. Entire Photodamaged Chloroplasts Are Transported to the Central Vacuole by Autophagy. *The Plant Cell*, 29(2):377–394.

Jarvis, P. and J. Soll

2001. Toc, Tic, and chloroplast protein import. *Biochimica et Biophysica Acta - Molecular Cell Research*, 1541(1-2):64–79.

Jeennor, S., M. Veerana, J. Anantayanon, S. Panchanawaporn, C. Chutrakul, and K. Laoteng

2017. Diacylglycerol acyltransferase 2 of Mortierella alpina with specificity on long-chain polyunsaturated fatty acids: A potential tool for reconstituting lipids with nutritional value. *Journal of Biotechnology*, 263(September):45–51.

JGI

2008. Joint Genome Institute: Phaeodactylum tricornutum server.

Jia, J., D. Han, H. G. Gerken, Y. Li, M. Sommerfeld, Q. Hu, and J. Xu

2015. Molecular mechanisms for photosynthetic carbon partitioning into storage neutral lipids in Nannochloropsis oceanica under nitrogen-depletion conditions. *Algal Research*, 7:66–77.

Jiang, H. and K. Gao

2004. Effects of lowering temperature during culture on production of polyunsaturated fatty acids in the marine diatom Phaeodactylum tricornutum. *Journal of Phycology*, 40(4):651–654.

Jiang, M., B. Guo, X. Wan, Y. Gong, Y. Zhang, and C. Hu

2014. Isolation and Characterization of the Diatom Phaeodactylum Δ5-Elongase

Gene for Transgenic LC-PUFA Production in Pichia pastoris. *Marine Drugs*, 12(3):1317–1334.

Johnston, M., J. S. Flick, and T. Pexton

1994. Multiple mechanisms provide rapid and stringent glucose repression of GAL gene expression in Saccharomyces cerevisiae. *Molecular and Cellular Biology*, 14(6):3834–3841.

Jones, P., D. Binns, H.-Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S.-Y. Yong, R. Lopez, and S. Hunter

2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9):1236–1240.

Kadono, T., A. Miyagawa-Yamaguchi, N. Kira, Y. Tomaru, T. Okami, T. Yoshimatsu, L. Hou, T. Ohama, K. Fukunaga, M. Okauchi, H. Yamaguchi, K. Ohnishi, A. Falciatore, M. Adachi, D. G. Mann, S. J. M. Droop, D. M. Nelson, P. G. Falkowski, V. Martin-Jézéquel, M. Hildebrand, M. A. Brzezinski, R. H. Wijffels, M. J. Barbosa, E. V. Armbrust, C. Bowler, M. Fabris, U. Maheswari, T. Mock, E. V. Armbrust, C. Bowler, K. E. Apt, P. G. Kroth-Pancic, A. R. Grossman, L. A. Zaslavskaia, J. C. Lippmeier, P. G. Kroth, A. R. Grossman, K. E. Apt, A. Miyagawa, M. Miyahara, M. Aoi, N. Inoue-Kashino, Y. Kashino, K. Ifuku, P. D. Weyman, F. Daboussi, S. Seo, H. Jeon, S. Hwang, E. Jin, K. S. Chang, H. Fischer, I. Robl, M. Sumper, N. Kröger, N. Poulsen, N. Kröger, T. G. Dunahay, E. E. Jarvis, P. G. Roessler, V. Sabatino, N. Poulsen, P. M. Chesley, N. Kröger, A. Falciatore, R. Casotti, C. Leblanc, C. Abrescia, C. Bowler, A. Miyagawa-Yamaguchi, K. Ifuku, M. Muto, B. J. Karas, Y. Lehahn, K. Kimura, Y. Tomaru, C. A. Suttle, C. de Vargas, Y. Tomaru, Y. Shirai, H. Suzuki, T. Nagasaki, T. Nagumo, Y. Tomaru, Y. Tomaru, H. Harada, D. Nakatsuma, M. Ishida, Y. Matsuda, R. Radakovits, P. M. Eduafo, M. C. Posewitz, Y. F. Niu, E. M. Trentacoste, M. L. Hamilton, R. P. Haslam, J. A. Napier, O. Sayanova, T. Kadono, A. Oeltjen, J. Marquardt, E. Rhiel, M. T. Russo, R. Annunziata, R. Sanges, M. I. Ferrante, A. Falciatore, E. V. Schmidt, G. Christoph, R. Zeller, P. Leder, P. N. Benfey, N. H. Chua, K. Sakaue, H. Harada, Y. Matsuda, D. R. Stevens, J. D. Rochaix, S. Purton, K. Shimogawara, S. Fujiwara,

A. Grossman, H. Usuda, K. Higo, Y. Ugawa, M. Iwamoto, T. Korenaga, M. Lescot, S. Ambawat, P. Sharma, N. R. Yadav, R. C. Yadav, M. Jakoby, T. Laloum, S. D. Mita, P. Gamas, M. Baudin, A. Niebel, R. Rezsohazy, A. J. Saurin, C. Maurel-Zaffran, Y. Graba, N. Zheng, E. Fraenkel, C. O. Pabo, N. P. Pavletich, E. Rayko, F. Maumus, U. Maheswari, K. Jabbari, C. Bowler, A. Mathelier, A. T. Kwon, D. J. Arenillas, R. W. Hunt, W. W. Wasserman, T. Okumura, H. Makiguchi, Y. Makita, R. Yamashita, K. Nakai, S. L. Hobbs, T. D. Warkentin, C. M. DeLong, B. T. Wakimoto, A. Mitra, D. W. Higgins, G. R. Smerdon, S. J. Aves, E. F. Walton, Y. Tomaru, K. Kimura, H. Yamaguchi, N. Venayak, N. Anesiadis, W. R. Cluett, R. Mahadevan, O. Pulz, W. Gross, E. C. Theriot, M. Ashworth, E. Ruck, T. Nakov, R. K. Jansen, M. Dutt, S. A. Dhekney, L. Soriano, R. Kandel, J. W. Grosser, T. Juven-Gershon, J. T. Kadonaga, A. McLeod, C. D. Smart, W. E. Fry, D. R. Liston, P. J. Johnson, D. Bhaya, A. R. Grossman, N. Ohno, R. Yoshinaga, M. Niwa-Kubota, H. Matsui, Y. Matsuda, C. Leblanc, A. Falciatore, M. Watanabe, C. Bowler, M. Siaut, T. Brakemann, B. Frank, K. Peter, R. Erhard, R. L. Guillard, W. Smith, M. Chanley, L. C.-M. Chen, T. Edelstein, J. McLachian, D. S. Gorman, R. P. Levine, P. J. Kersey, D. Smedley, P. R. Sanders, J. A. Winter, A. R. Barnason, S. G. Rogers, R. T. Fraley, and M. Boshart

2015. Characterization of marine diatom-infecting virus promoters in the model diatom Phaeodactylum tricornutum. *Scientific Reports*, 5:1–13.

Kalisch, B., P. Dörmann, and G. Hölzl

2016. DGDG and Glycolipids in Plants and Algae. In *Lipids in Plant and Algae Development*, Y. Nakamura and Y. Li-Beisson, eds., chapter 3, Pp. 51–83. Springer, Cham.

Kanehisa, M. and S. Goto

2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic acids research*, 28(1):27–30.

Karas, B. J., R. E. Diner, S. C. Lefebvre, J. Mcquaid, A. P. R. Phillips, C. M. Noddings, J. K. Brunson, R. E. Valas, T. J. Deerinck, J. Jablanovic, J. T. F. Gillard, K. Beeri, M. H. Ellisman, J. I. Glass, C. A. Hutchison Iii, H. O. Smith,

J. C. Venter, A. E. Allen, C. L. Dupont, and P. D. Weyman

2015. Designer diatom episomes delivered by bacterial conjugation. *Nature Communications*, 6:1–10.

Kaur, S.

2014. *Genetic and biotechnological development of the pennate marine diatom Phaeodactylum tricornutum for high-value bioproducts and carbon bio-mitigation*. PhD thesis, NUI Galway.

Keeling, P. J.

2010. The endosymbiotic origin, diversification and fate of plastids. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1541):729–48.

Keeling, P. J.

2013. The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. *Annual Review of Plant Biology*, 64(1):583–607.

Kersey, P. J., J. E. Allen, A. Allot, M. Barba, S. Boddu, B. J. Bolt, D. Carvalho-Silva, M. Christensen, P. Davis, C. Grabmueller, N. Kumar, Z. Liu, T. Maurel, B. Moore, M. D. McDowall, U. Maheswari, G. Naamati, V. Newman, C. K. Ong, M. Paulini, H. Pedro, E. Perry, M. Russell, H. Sparrow, E. Tapanari, K. Taylor, A. Vullo, G. Williams, A. Zadissia, A. Olson, J. Stein, S. Wei, M. Tello-Ruiz, D. Ware, A. Luciani, S. Potter, R. D. Finn, M. Urban, K. E. Hammond-Kosack, D. M. Bolser, N. De Silva, K. L. Howe, N. Langridge, G. Maslen, D. M. Staines, and A. Yates

2018. Ensembl Genomes 2018: an integrated omics infrastructure for non-vertebrate species. *Nucleic acids research*, 46(D1):D802–D808.

Khozin-Goldberg, I., U. Iskandarov, and Z. Cohen

2011. LC-PUFA from photosynthetic microalgae: occurrence, biosynthesis, and prospects in biotechnology. *Applied Microbiology and Biotechnology*, 91(4):905–915.

Kim, H. U., J. H. Jung, M.-C. Suh, K.-R. Lee, J. B. Kim, and Y. S. Go

2011. Endoplasmic Reticulum-Located PDAT1-2 from Castor Bean Enhances

Hydroxy Fatty Acid Accumulation in Transgenic Plants. *Plant and Cell Physiology*, 52(6):983–993.

Kim, J., M. Fabris, G. Baart, M. K. Kim, A. Goossens, W. Vyverman, P. G. Falkowski, and D. S. Lun

2016. Flux balance analysis of primary metabolism in the diatom Phaeodactylum tricornutum. *Plant Journal*, 85(1):161–176.

Kinsella, R. J., A. Kahari, S. Haider, J. Zamora, G. Proctor, G. Spudich, J. Almeida-King, D. Staines, P. Derwent, A. Kerhornou, P. Kersey, and P. Flicek

2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database*, 2011(0):bar030–bar030.

Kobayashi, K. and H. Wada

2016. Role of Lipids in Chloroplast Biogenesis. In *Sub-cellular biochemistry*, volume 86, Pp. 103–125.

Kong, F., I. T. Romero, J. Warakanont, and Y. Li-Beisson

2018. Lipid catabolism in microalgae. *New Phytologist*, 218(4):1340–1348.

Krank, J., R. C. Murphy, R. M. Barkley, E. Duchoslav, and A. McAnoy

2007. Qualitative Analysis and Quantitative Assessment of Changes in Neutral Glycerol Lipid Molecular Species Within Cells. *Methods in Enzymology*, 432(07):1–20.

Krauß, S. and W. Vetter

2018. Phytol and Phytyl Fatty Acid Esters: Occurrence, Concentrations, and Relevance. *European Journal of Lipid Science and Technology*, 120(7).

Kröger, N. and N. Poulsen

2008. Diatoms—From Cell Wall Biogenesis to Nanotechnology. *Annual Review of Genetics*, 42(1):83–107.

Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer

2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580.

Kroon, J. T. M., W. Wei, W. J. Simon, and A. R. Slabas

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(23):2541–2549.

Kroth, P. G., A. M. Bones, F. Daboussi, M. I. Ferrante, M. Jaubert, M. Kolot, M. Nymark, C. Río Bártulos, A. Ritter, M. T. Russo, M. Serif, P. Winge, and A. Falciatore

2018. Genome editing in diatoms: achievements and goals. *Plant Cell Reports*, 37(10):1401–1408.

Kumar, T. A.

2013. CFSSP: Chou and Fasman Secondary Structure Prediction server. *Wide Spectrum*, 1(9).

Lauersen, K. J., O. Kruse, and J. H. Mussgnug

2015. Targeted expression of nuclear transgenes in Chlamydomonas reinhardtii with a versatile, modular vector toolkit. *Applied Microbiology and Biotechnology*, 99(8):3491–3503.

Lehner, R.

1993. Transacylase from Rat Intestinal Microsomes *. *Journal of Biological Chemistry*, 268(12):8781–8786.

Lenfant, N., T. Hotelier, E. Velluet, Y. Bourne, P. Marchot, and A. Chatonnet

2013. ESTHER, the database of the α/β-hydrolase fold superfamily of proteins: Tools to explore diversity of functions. *Nucleic Acids Research*, 41(D1):423–429.

Levering, J., J. Broddrick, C. L. Dupont, G. Peers, K. Beeri, J. Mayers, A. A. Gallina, A. E. Allen, B. O. Palsson, and K. Zengler

2016. Genome-scale model reveals metabolic basis of biomass partitioning in a model diatom. *PLoS ONE*, 11(5):1–22.

Levitan, O., J. Dinamarca, E. Zelzion, D. S. Lun, L. T. Guerra, M. K. Kim, J. Kim, B. A. S. Van Mooy, D. Bhattacharya, and P. G. Falkowski

2015. Remodeling of intermediate metabolism in the diatom Phaeodactylum

tricornutum under nitrogen stress. *Proceedings of the National Academy of Sciences*, 112(2):412–417.

Levy, C. W., A. Roujeinikova, S. Sedelnikova, P. J. Baker, A. R. Stuitje, A. R. Slabas, D. W. Rice, and J. B. Rafferty
1999. Molecular basis of triclosan activity. *Nature*, 1;398(6726.

Li, C., L. Li, J. Lian, R. Watts, R. Nelson, B. Goodwin, and R. Lehner
2015. Roles of Acyl-CoA: Diacylglycerol acyltransferases 1 and 2 in triacylglycerol synthesis and secretion in primary hepatocytes. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 35(5):1080–1091.

Li, R., K. Yu, and D. F. Hildebrand
2010. DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. *Lipids*, 45(2):145–157.

Li, X., E. R. Moellering, B. Liu, C. Johnny, M. Fedewa, B. B. Sears, M.-H. Kuo, and C. Benning
2012. A Galactoglycerolipid Lipase Is Required for Triacylglycerol Accumulation and Survival Following Nitrogen Deprivation in Chlamydomonas reinhardtii. *The Plant Cell*, 24(11):4670–4686.

Li, X., Y. Pan, and H. Hu
2018. Identification of the triacylglycerol lipase in the chloroplast envelope of the diatom Phaeodactylum tricornutum. *Algal Research*, 33(May):440–447.

Li-Beisson, Y., J. Neunzig, Y. Lee, and K. Philippar
2017. Plant membrane-protein mediated intracellular traffic of fatty acids and acyl lipids. *Current Opinion in Plant Biology*, 40:138–146.

Lippold, F., P. Dormann, K. vom Dorp, M. Abraham, F. Kessler, S. Stymne, C. Montandon, G. Holzl, J. L. Yilmaz, C. Besagni, V. Wewer, and I. Lager
2012. Fatty Acid Phytyl Ester Synthesis in Chloroplasts of Arabidopsis. *The Plant Cell*, 24(5):2001–2014.

Litvinenko, A.

2016. *Building better biodiesel : characterisation of neutral lipid synthesis enzymes from marine algae*. PhD thesis, University of Cambridge.

Liu, Q., R. M. Siloto, R. Lehner, S. J. Stone, and R. J. Weselake

2012. Acyl-CoA:diacylglycerol acyltransferase: Molecular biology, biochemistry and biotechnology. *Progress in Lipid Research*, 51(4):350–377.

Liu, X., F. Hempel, S. Stork, K. Bolte, D. Moog, T. Heimerl, U. G. Maier, and S. Zauner

2016a. Addressing various compartments of the diatom model organism Phaeodactylum tricornutum via sub-cellular marker proteins. *Algal Research*, 20:249–257.

Liu, X.-Y., L.-L. Ouyang, and Z.-G. Zhou

2016b. Phospholipid: diacylglycerol acyltransferase contributes to the conversion of membrane lipids into triacylglycerol in Myrmecia incisa during the nitrogen starvation stress. *Scientific Reports*, 6(1):26610.

Liu, Y., A. Beyer, and R. Aebersold

2016c. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, 165(3):535–550.

Lohscheider, J. N. and C. Río Bártulos

2016. Plastoglobules in algae: A comprehensive comparative study of the presence of major structural and functional components in complex plastids. *Marine Genomics*, 28:127–136.

Longworth, J., D. Wu, M. Huete-Ortega, P. C. Wright, and S. Vaidyanathan

2016. Proteome response of Phaeodactylum tricornutum, during lipid accumulation induced by nitrogen depletion. *Algal Research*, 18:213–224.

Louda, J. W., P. Mongkhonsri, and E. W. Baker

2011. Chlorophyll degradation during senescence and death-III: 3-10yr experiments, implications for ETIO series generation. *Organic Geochemistry*, 42(6):688–699.

Lupette, J., A. Jaussaud, K. Seddiki, C. Morabito, S. Brugière, H. Schaller, M. Kuntz, J. L. Putaux, P. H. Jouneau, F. Rébeillé, D. Falconet, Y. Couté, J. Jouhet,

M. Tardif, J. Salvaing, and E. Maréchal

2019. The architecture of lipid droplets in the diatom Phaeodactylum tricornutum. *Algal Research*, 38(January):101415.

Ma, Y.-H., X. Wang, Y.-F. Niu, Z.-K. Yang, M.-H. Zhang, Z.-M. Wang, W.-D. Yang, J.-S. Liu, and H.-Y. Li

2014. Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom Phaeodactylum tricornutum. *Microbial Cell Factories*, 13(1):100.

Manandhar-Shrestha, K. and M. Hildebrand

2015. Characterization and manipulation of a DGAT2 from the diatom Thalassiosira pseudonana: Improved TAG accumulation without detriment to growth, and implications for chloroplast TAG accumulation. *Algal Research*, 12:239–248.

Mann, D.

2008. Tree of life: Pennate diatoms.

Marañón, E., M. P. Lorenzo, P. Cermeño, and B. Mouriño-Carballido

2018. Nutrient limitation suppresses the temperature dependence of phytoplankton metabolic rates. *ISME Journal*, 12(7):1836–1845.

Marioni, J. C., C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad

2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 18(9):1509–1517.

Marmon, S., Z.-Y. Du, I. Feussner, K. Vollheyde, K. Zienkiewicz, C. Herrfurth, E. Poliner, E. M. Farré, C. Benning, and A. Zienkiewicz

2017. Nannochloropsis, a rich source of diacylglycerol acyltransferases for engineering of triacylglycerol content in different hosts. *Biotechnology for Biofuels*, 10(1):1–20.

Martin, G. J. O., D. R. A. Hill, I. L. D. Olmstead, A. Bergamin, M. J. Shears, D. A. Dias, S. E. Kentish, P. J. Scales, C. Y. Botté, and D. L. Callahan

2014. Lipid Profile Remodeling in Response to Nitrogen Deprivation in the Microalgae Chlorella sp. (Trebouxiophyceae) and Nannochloropsis sp. (Eustigmatophyceae). *PLoS ONE*, 9(8):e103389.

Matthijs, M., M. Fabris, S. Broos, W. Vyverman, and A. Goossens

2016. Profiling of the Early Nitrogen Stress Response in the Diatom Phaeodactylum tricornutum Reveals a Novel Family of RING-Domain Transcription Factors. *Plant physiology*, 170(1):489–98.

Matthijs, M., M. Fabris, T. Obata, I. Foubert, J. M. Franco-Zorrilla, R. Solano, A. R. Fernie, W. Vyverman, and A. Goossens

2017. The transcription factor bZIP14 regulates the TCA cycle in the diatom Phaeodactylum tricornutum. *The EMBO Journal*, 1(36):1559–1576.

Melis, A.

2009. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science*, 177(4):272–280.

Minhas, A. K., P. Hodgson, C. J. Barrow, and A. Adholeya

2016. A review on the assessment of stress conditions for simultaneous production of microalgal lipids and carotenoids. *Frontiers in Microbiology*, 7(MAY):1–19.

Miyahara, M., Y. Kashino, K. Ifuku, M. Aoi, and N. Inoue-Kashino

2013. Highly Efficient Transformation of the Diatom Phaeodactylum tricornutum by Multi-Pulse Electroporation. *Bioscience, Biotechnology, and Biochemistry*, 77(4):874–876.

Mizoguchi, T., M. Isaji, N. Yamano, J. Harada, R. Fujii, and H. Tamiaki

2017. Molecular Structures and Functions of Chlorophylls-a Esterified with Geranylgeranyl, Dihydrogeranylgeranyl, and Tetrahydrogeranylgeranyl Groups at the 17-Propionate Residue in a Diatom, Chaetoceros calcitrans. *Biochemistry*, 56(28):3682–3688.

Moellering, E. R., B. Muthan, and C. Benning

2010. Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science (New York, N.Y.)*, 330(6001):226–8.

Moriyama, T., M. Toyoshima, M. Saito, H. Wada, and N. Sato

2018. Revisiting the Algal "Chloroplast Lipid Droplet": The Absence of an Entity That Is Unlikely to Exist. *Plant physiology*, 176(2):1519–1530.

Msanne, J., T. Awada, H. Cerutti, A. R. Konda, D. Xu, E. B. Cahoon, and J. A. Casas-Mollano

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.

Mueller, S. P., M. J. Mueller, A. Fekete, L. Guender, and M. Unger

2017. Phospholipid:Diacylglycerol Acyltransferase-Mediated Triacylglyerol Synthesis Augments Basal Thermotolerance. *Plant Physiology*, 175(1):486–497.

Mus, F., J.-P. Toussaint, K. E. Cooksey, M. W. Fields, R. Gerlach, B. M. Peyton, and R. P. Carlson

2013. Physiological and molecular analysis of carbon source supplementation and pH stress-induced lipid accumulation in the marine diatom Phaeodactylum tricornutum. *Applied Microbiology and Biotechnology*, 97:3625–3642.

Nakamura, Y. and Y. Li-Beisson

2016. *Subcellular Biochemistry 86: Lipids in Plant and Algae Development*. Springer, Cham.

Napier, J. A., S. Usher, R. P. Haslam, N. Ruiz-Lopez, and O. Sayanova

2015. Transgenic plants as a sustainable, terrestrial source of fish oils. *European journal of lipid science and technology : EJLST*, 117(9):1317–1324.

Nichio, B. T. L., J. N. Marchaukoski, and R. T. Raittz

2017. New Tools in Orthology Analysis : A Brief Review of Promising Perspectives. 8(October):1–12.

Niu, Y.-F., X. Wang, D.-X. Hu, S. Balamurugan, D.-W. Li, W.-D. Yang, J.-S. Liu, and H.-Y. Li

2016. Molecular characterization of a glycerol-3-phosphate acyltransferase reveals key features essential for triacylglycerol production in Phaeodactylum tricornutum. *Biotechnology for Biofuels*, 9(1):60.

Niu, Y. F., M. H. Zhang, D. W. Li, W. D. Yang, J. S. Liu, W. B. Bai, and H. Y. Li

2013. Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by

overexpressing a type 2 diacylglycerol acyltransferase in marine diatom Phaeodactylum tricornutum. *Marine Drugs*, 11(11):4558–4569.

Niu, Y.-F., C.-C. Zhu, W.-D. Yang, M.-H. Zhang, Z.-K. Yang, H.-Y. Li, and J.-S. Liu

2012. Transformation of diatom Phaeodactylum tricornutum by electroporation and establishment of inducible selection marker. *BioTechniques*, 1(52).

Nookaew, I., M. Papini, N. Pornputtapong, G. Scalcinati, L. Fagerberg, M. Uhlén, and J. Nielsen

2012. A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: A case study in Saccharomyces cerevisiae. *Nucleic Acids Research*, 40(20):10084–10097.

Nymark, M., A. K. Sharma, T. Sparstad, A. M. Bones, and P. Winge

2016. A CRISPR/Cas9 system adapted for gene editing in marine algae. *Scientific Reports*, 6(1):24951.

Oelkers, P., D. Cromley, M. Padamsee, J. T. Billheimer, and S. L. Sturley

2002. The DGA1 gene determines a second triglyceride synthetic pathway in yeast. *Journal of Biological Chemistry*, 277(11):8877–8881.

Okamoto, N., C. Chantangsi, A. Horák, B. S. Leander, and P. J. Keeling

2009. Molecular phylogeny and description of the novel Katablepharid Roombia truncata gen. et sp. nov., and establishment of the Hacrobia taxon nov. *PLoS ONE*, 4(9):1–11.

Pan, X., F. Y. Peng, and R. J. Weselake

2015. Genome-Wide Analysis of Phospholipid:Diacylglycerol Acyltransferase (PDAT) Genes in Plants Reveals the Eudicot-Wide PDAT Gene Expansion and Altered Selective Pressures Acting on the Core Eudicot PDAT Paralogs . *Plant Physiology*, 167(3):887–904.

Pan, X., R. M. P. Siloto, A. D. Wickramarathna, E. Mietkiewska, and R. J. Weselake

2013. Identification of a pair of phospholipid: Diacylglycerol acyltransferases from developing flax (Linum usitatissimum L.) seed catalyzing the selective production of trilinolenin. *Journal of Biological Chemistry*, 288(33):24173–24188.

Pan, Y., J. Yang, Y. Gong, X. Li, and H. Hu
2017. 3-Hydroxyisobutyryl-coa hydrolase involved in isoleucine catabolism regulates triacylglycerol accumulation in phaeodactylum tricornutum. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1728).

Panaud, O., A. Zambounis, C. Boyen, A. Moustafa, G. Michel, J. Poulain, C. Leblanc, F. Cabello-Hurtado, S. A. Rensing, B. Porcel, T. Gabaldon, L. Cladiere, J. M. Cock, P. Nyvall Collen, L. Brillet, D. H. McLachlan, B. Noel, C. M. M. Gachon, F. Artiguenave, B. Kloareg, F. Partensky, T. Tonon, J. F. Barbosa-Neto, F.-Y. Bouget, F. Denoeud, S. Capella-Gutierrez, K. Jabbari, P. J. Lopez, M. Czjzek, S. Rousvoal, K. Valentin, K. Labadie, P. Wincker, L. Meslet-Cladiere, C. Colleoni, A. Arun, C. Chaparro, P. Deschamps, A. Groisillier, M. Elias, N. Kowalczyk, J.-M. Aury, J. Collen, B. Charrier, S. M. Dittami, G. Samson, C. Herve, M. Katinka, J. Weissenbach, C. Da Silva, Z. Nehr, A. Symeonidi, S. G. Ball, T. Barbeyron, L. Delage, S. M. Coelho, W. Carre, and J. H. Bothwell
2013. Genome structure and metabolic features in the red seaweed Chondrus crispus shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences*, 110(13):5247–5252.

Pérez-Pérez, M. E., I. Couso, L. G. Heredia-Martínez, and J. L. Crespo
2017. Monitoring Autophagy in the Model Green Microalga Chlamydomonas reinhardtii. *Cells*, 6(4).

Petroutsos, D., S. Amiar, H. Abida, L. J. Dolch, O. Bastien, F. Rébeillé, J. Jouhet, D. Falconet, M. A. Block, G. I. McFadden, C. Bowler, C. Botté, and E. Maréchal
2014. Evolution of galactoglycerolipid biosynthetic pathways - From cyanobacteria to primary plastids and from primary to secondary plastids. *Progress in Lipid Research*, 54(1):68–85.

Popko, J., C. Herrfurth, K. Feussner, T. Ischebeck, T. Iven, R. Haslam, M. Hamilton, O. Sayanova, J. Napier, I. Khozin-Goldberg, and I. Feussner
2016. Metabolome Analysis Reveals Betaine Lipids as Major Source for Triglyceride Formation, and the Accumulation of Sedoheptulose during Nitrogen-Starvation of Phaeodactylum tricornutum. *PloS one*, 11(10):e0164673.

Pospíšil, P.

2009. Production of reactive oxygen species by photosystem II. *Biochimica et Biophysica Acta - Bioenergetics*, 1787(10):1151–1160.

Pospíšil, P.

2016. Production of Reactive Oxygen Species by Photosystem II as a Response to Light and Temperature Stress. *Frontiers in Plant Science*, 7(December):1–12.

Radakovits, R., P. M. Eduafo, and M. C. Posewitz

2011. Genetic engineering of fatty acid chain length in Phaeodactylum tricornutum. *Metabolic Engineering*, 13(1):89–95.

Ragni, M. and M. R. D'Alcalà

2007. Circadian variability in the photobiology of Phaeodactylum tricornutum: pigment content. *Journal of Plankton Research*, 29(2):141–156.

Rast, A., S. Heinz, and J. Nickelsen

2015. Biogenesis of thylakoid membranes. *Biochimica et Biophysica Acta - Bioenergetics*, 1847(9):821–830.

Rastogi, A., U. Maheswari, R. G. Dorrell, F. R. J. Vieira, F. Maumus, A. Kustka, J. McCarthy, A. E. Allen, P. Kersey, C. Bowler, and L. Tirichine

2018. Integrative analysis of large scale transcriptome data draws a comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of diatoms. *Scientific Reports*, 8(1):4834.

Remmers, I. M., D. E. Martens, R. H. Wijffels, P. P. Lamers, Y. Li, and P. Schenk

2017. Dynamics of triacylglycerol and EPA production in Phaeodactylum tricornutum under nitrogen starvation at different light intensities. *PLOS ONE*, 12(4):e0175630.

Rice, D. W. and J. D. Palmer

2006. An exceptional horizontal gene transfer in plastids: Gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biology*, 4:1–15.

Robinson, D.

2018. Electric vehicles and electricity. Technical Report 36.

Rogato, A., H. Richard, A. Sarazin, B. Voss, S. Cheminant Navarro, R. Champeimont, L. Navarro, A. Carbone, W. R. Hess, and A. Falciatore

2014. The diversity of small non-coding RNAs in the diatom Phaeodactylum tricornutum. *BMC Genomics*, 15(1).

Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson, and D. J. Pappin

2004. Multiplexed Protein Quantitation in Saccharomyces cerevisiae Using Amine-reactive Isobaric Tagging Reagents . *Molecular & Cellular Proteomics*, 3(12):1154–1169.

Roston, R. L., K. Wang, L. A. Kuhn, and C. Benning

2014. Structural determinants allowing transferase activity in SENSITIVE TO FREEZING 2, classified as a family I glycosyl hydrolase. *The Journal of biological chemistry*, 289(38):26089–106.

Rottet, S., C. Besagni, and F. Kessler

2015. The role of plastoglobules in thylakoid lipid remodeling during plant development. *Biochimica et Biophysica Acta - Bioenergetics*, 1847(9):889–899.

Rottig, A. and A. Steinbuchel

2013. Acyltransferases in Bacteria. *Microbiology and Molecular Biology Reviews*, 77(2):277–321.

Ruiz, J., G. Olivieri, J. de Vree, R. Bosma, P. Willems, J. H. Reith, M. H. M. Eppink, D. M. M. Kleinegris, R. H. Wijffels, and M. J. Barbosa

2016. Towards industrial products from microalgae. *Energy & Environmental Science*, 9(10):3036–3043.

Rumin, J., H. Bonnefond, B. Saint-Jean, C. Rouxel, A. Sciandra, O. Bernard, J.-P. Cadoret, G. Bougaran, Y. Chisti, R. Wijffels, M. Barbosa, T. Chacón-Lee, G. González-Mariño, M. Daroch, S. Geng, G. Wang, Q. Hu, M. Sommerfeld,

E. Jarvis, M. Ghirardi, M. Posewitz, M. Seibert, O. Konur, N. Pragya, K. Pandey, P. Sahoo, J. Singh, S. Gu, J.-P. Cadoret, O. Bernard, Y. Chisti, X. Montagne, P. Porot, C. Aymard, C. Querleu, A. Bouter, D. Lorne, S. Ríos, C. Torres, C. Torras, J. Salvadó, J. Mateo-Sanz, L. Jiménez, Y. Han, Q. Wen, Z. Chen, P. Li, E. Bligh, W. Dyer, J. Folch, M. Lee, G. S. Stanley, M. Eltgroth, R. Watwood, G. Wolfe, A. Jara, H. Mendoza, A. Martel, C. Molina, L. Nordstron, V. Rosa, S. Lee, B.-D. Yoon, H.-M. Oh, K. Cooksey, J. Guckert, S. Williams, P. Callis, M. Cooper, W. Hardin, T. Petersen, R. Cattolico, T. Mutanda, D. Ramesh, S. Karthikeyan, S. Kumari, A. Anandraj, F. Bux, Y. Chisti, D. Elsey, D. Jameson, B. Raleigh, M. Cooney, G.-H. Huang, G. Chen, F. Chen, T. Govender, L. Ramanna, I. Rawat, F. Bux, J. Smith, L. Lison, Z. Menschik, J. Vergara, J. Cleine, K. Dixon, L. Dale, J. Smith, J. Slack, S. Fowler, P. Greenspan, P. Greenspan, E. Mayer, S. Fowler, P. Greenspan, S. Fowler, K. Ira, D. Sackett, J. Wolff, S. Ruvinov, X. Yang, K. Parris, U. Banik, S. Ahmed, E. Miles, K. Kimura, M. Yamaoka, Y. Kamisaka, W. Chen, C. Zhang, L. Song, M. Sommerfeld, Q. Hu, H. Guzmán, A. Valido, J. La, L. Duarte, K. Presmanes, D. Simionato, E. Sforza, E. C. Carpinelli, A. Bertucco, G. Giacometti, T. Morosinotto, G. Bougaran, C. Rouxel, N. Dubois, R. Kaas, S. Grouas, E. Lukomska, H. Pereira, L. Barreira, A. Mozes, C. Florindo, C. Polo, C. Duarte, D. Wong, A. Franz, N. Velmurugan, M. Sung, S. Yim, M. Park, J. Yang, K. Jeong, S. Mou, D. Xu, N. Ye, X. Zhang, C. Liang, Q. Liang, H. Guzmán, A. Valido, J. La, K. Presmanes, L. Duarte, L. Brennan, A. B. Fernández, A. Mostaert, P. Owende, J. Cirulis, B. Strasser, J. Scott, G. Ross, H. Vigeolas, F. Duby, E. Kaymak, G. Niessen, P. Motte, F. Franck, Z. Chen, Y. Gong, X. Fang, H. Hu, R. Gardner, K. Cooksey, F. Mus, R. Macur, K. Moll, E. Eustance, H.-Y. Ren, B.-F. Liu, C. Ma, L. Zhao, N.-Q. Ren, G. Hu, Y. Fan, L. Zhang, C. Yuan, J. Wang, W. Li, M. Roleda, S. Slocombe, R. Leakey, J. Day, E. Bell, M. Stanley, T.-T. Doan, J. Obbard, G.-D. Feng, F. Zhang, L.-H. Cheng, X.-H. Xu, L. Zhang, H.-L. Chen, T.-T. Doan, J. Obbard, T. Doan, J. Obbard, M. Montero, M. Aristizabal, R. Garcia, D. Xu, Z. Gao, F. Li, X. Fan, X. Zhang, N. Ye, H. Guzmán, A. Valido, J. La, L. Duarte, K. Presmanes, S. Hoz, W. Ayidzoe, A. Ben-Zvi, R. Burrell, W. McCaffrey, M. Isleten-Hosoglu, I. Gultepe, M. Elibol, L.-C. Su, Y.-H. Hsu, H.-Y. Wang, A. Loudet, K. Burgess, I. Elle, L. Olsen, D. Pultz, S. Rødkær, N. Færgeman, P. Gocze,

D. Freeman, M. Cooper, L. D'Amico, C. Henry, A. Gottardi, M. Vinciguerra, A. Sgroi, M. Moukil, F. Ravier-Dall'Antonia, V. Pazienza, P. Greenspan, E. Mayer, S. Fowler, N. Ghoneim, F. Alonzo, P. Mayzaud, A. Dutta, K. Kamada, K. Ohta, W. Brown, T. Sullivan, P. Greenspan, E. O'Rourke, A. Soukas, C. Carr, G. Ruvkun, E. Wolfrum, L. Laurens, G. Diaz, M. Melis, B. Batetta, F. Angius, A. Falchi, L. Castell, R. Mann, U. Pick, T. Rachutin-Zalogin, K. Suman, T. Kiran, U. Devi, N. Sarma, J. Priscu, L. Priscu, A. Palmisano, C. Sullivan, A. Vines, G. McBean, A. Blanco-Fernández, R. Davis, J. Volponi, H. Jones, B. Carvalho, H. Wu, S. Singh, E. Bertozzini, L. Galluzzi, A. Penna, M. Magnani, W. Chen, M. Sommerfeld, Q. Hu, I. Ahmad, Z. Fatma, S. Yazdani, S. Kumar, R. Notman, M. Noro, B. O'Malley, J. Anwar, I. Sitepu, L. Ignatia, A. Franz, D. Wong, S. Faulina, M. Tsui, T. Doan, B. Sivaloganathan, J. Obbard, A. Massart, E. Aubry, A.-L. Hantson, H. Azencott, G. Peter, M. Prausnitz, U. Pick, H. Benbouzid, S. Floch, L. Stephan, R. Olier, M. Privat, G. Castro, R. Kamdar, B. Panilaitis, D. Kaplan, J. Murphy, H. Furusho, R. Paton, K. Nomura, P. Hyka, S. Lickova, P. Přibyl, K. Melzoch, K. Kovar, K. Manandhar-Shrestha, M. Hildebrand, J. Traller, M. Hildebrand, M. Bader, R. Gardner, P. Peters, B. Peyton, K. Cooksey, G. Carrier, M. Garnier, L. Cunff, G. Bougaran, I. Probert, and C. Vargas

2015. The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae. *Biotechnology for Biofuels*, 8(1):42.

Ryall, K., J. T. Harper, and P. J. Keeling

2003. Plastid-derived Type II fatty acid biosynthetic enzymes in chromists. *Gene*, 313:139–148.

Sakaue, K., H. Harada, and Y. Matsuda

2008. Development of gene expression system in a marine diatom using viral promoters of a wide variety of origin. *Physiologia Plantarum*, 133(1):59–67.

Salem, N. and M. Eggersdorfer

2015. Is the world supply of omega-3 fatty acids adequate for optimal human nutrition? *Current opinion in clinical nutrition and metabolic care*, 18(2):147–54.

Salminen, T. A., K. Blomqvist, and J. Edqvist

2016. Lipid transfer proteins: classification, nomenclature, structure, and function. *Planta*, 244(5):971–997.

Sandager, L., M. H. Gustavsson, U. Ståhl, A. Dahlqvist, E. Wiberg, A. Banas, M. Lenman, H. Ronne, and S. Stymne

2002. Storage lipid synthesis is non-essential in yeast. *Journal of Biological Chemistry*, 277(8):6478–6482.

Sarthou, G., K. R. Timmermans, S. Blain, and P. Tréguer

2005. Growth physiology and fate of diatoms in the ocean: a review. *Journal of Sea Research*, 53(1-2):25–42.

Sayanova, O., V. Mimouni, L. Ulmann, A. Morant-Manceau, V. Pasquet, B. Schoefs, and J. A. Napier

2017. Modulation of lipid biosynthesis by stress in diatoms. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 372(1728):20160407.

Seidel, C.

2008. Introduction to DNA Microarrays. In *Analysis of Microarray Data: A Network-Based Approach*, F. Emmert-Streib and M. Dehmer, eds., chapter 1, Pp. 1–26. Wiley-VCH.

Serif, M., G. Dubois, A.-L. Finoux, M.-A. Teste, D. Jallet, and F. Daboussi

2018. One-step generation of multiple gene knock-outs in the diatom Phaeodactylum tricornutum by DNA-free genome editing. *Nature Communications*, 9(1):3924.

Sharma, A. K., M. Nymark, T. Sparstad, A. M. Bones, and P. Winge

2018. Transgene-free genome editing in marine algae by bacterial conjugation – comparison with biolistic CRISPR/Cas9 transformation. *Scientific Reports*, 8(1):1–11.

Sharma, K. K., H. Schuhmann, P. M. Schenk, K. K. Sharma, H. Schuhmann, and P. M. Schenk

2012. High Lipid Induction in Microalgae for Biodiesel Production. *Energies*, 5(5):1532–1553.

Sheehan, J., T. Dunahay, J. Benemann, and P. Roessler

1998. A Look Back at the U.S. Department of Energy's Aquatic Species Program: Biodiesel from Algae. Technical report.

Shiva, S., H. S. Vu, M. R. Roth, Z. Zhou, S. R. Marepally, G. H. Nune, Daya Sagar , Lushington, M. Visvanathan, and R. Welti

2013. *Plant Lipid Signaling Protocols*, volume 77.

Shockey, J. M.

2006. Tung Tree DGAT1 and DGAT2 Have Nonredundant Functions in Triacylglycerol Biosynthesis and Are Localized to Different Subdomains of the Endoplasmic Reticulum. *the Plant Cell Online*, 18(9):2294–2313.

Siaut, M., M. Heijde, M. Mangogna, A. Montsant, S. Coesel, A. Allen, A. Manfredonia, A. Falciatore, and C. Bowler

2007. Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum. *Gene*, 406(1):23–35.

Simionato, D., M. A. Block, N. L. Rocca, J. Jouhet, E. Maréchal, G. Finazzi, and T. Morosinotto

2013. The Response of Nannochloropsis gaditana to Nitrogen Starvation Includes De Novo Biosynthesis of Triacylglycerols, a Decrease of Chloroplast Galactolipids, and Reorganization of the Photosynthetic Apparatus. *Eukaryotic cell*, 12(5):665–676.

Smith, S. J., B. Tow, R. V. Farese, R. H. Eckel, D. R. Jensen, J. Raber, D. A. Sanan, H. C. Chen, E. Sande, and S. Cases

2002. Obesity resistance and multiple mechanisms of triglyceride synthesis in mice lacking Dgat. *Nature Genetics*, 25(1):87–90.

Smol, J. and E. Stoermer

2010. *The Diatoms: Applications for the Environmental and Earth Sciences*, 1st edition. Cambridge: Cambridge University Press.

Song, M., H. Pei, W. Hu, and G. Ma

2013. Evaluation of the potential of 10 microalgal strains for biodiesel production. *Bioresource Technology*, 141:245–251.

Spicher, L. and F. Kessler

2015. Unexpected roles of plastoglobules (plastid lipid droplets) in vitamin K1 and E metabolism. *Current Opinion in Plant Biology*, 25:123–129.

Spilling, K., P. Ylöstalo, S. Simis, and J. Seppälä

2015. Interaction Effects of Light, Temperature and Nutrient Limitations (N, P and Si) on Growth, Stoichiometry and Photosynthetic Parameters of the Cold-Water Diatom Chaetoceros wighamii. *PloS one*, 10(5):e0126308.

Sprague, M., M. B. Betancor, and D. R. Tocher

2017. Microbial and genetically engineered oils as replacements for fish oil in aquaculture feeds. *Biotechnology Letters*, 39(11):1599–1609.

Stobart, K., M. Mancha, M. Lenman, A. Dahlqvist, and S. Stymne

1997. Triacylglycerols are synthesised and utilized by transacylation reactions in microsomal preparations of developing safflower (Carthamus tinctorius L.) seeds. *Planta*, 203(1):58–66.

Stukenberg, D., S. Zauner, G. Dell'Aquila, and U. G. Maier

2018. Optimizing CRISPR/Cas9 for the Diatom Phaeodactylum tricornutum. *Frontiers in plant science*, 9:740.

Suresh Kumar, K., H.-U. Dahms, E.-J. Won, J.-S. Lee, and K.-H. Shin

2015. Microalgae – A promising tool for heavy metal remediation. *Ecotoxicology and Environmental Safety*, 113:329–352.

Suzuki, H., T. Kaizuka, N. Mizushima, and N. N. Noda

2015. Structure of the Atg101–Atg13 complex reveals essential roles of Atg101 in autophagy initiation. *Nature Structural & Molecular Biology*, 22(7):572–580.

Tardif, M., A. Atteia, M. Specht, G. Cogne, N. Rolland, S. Brugière, M. Hippler, M. Ferro, C. Bruley, G. Peltier, O. Vallon, and L. Cournac

2012. PredAlgo: A New Subcellular Localization Prediction Tool Dedicated to Green Algae. *Molecular Biology and Evolution*, 29(12):3625–3639.

Terezono, E.

2017. Fisheries Challenged by rising cost of fish oil in feed.

Terry, K. L., J. Hirata, and E. A. Laws

1985. Light-, nitrogen-, and phosphorus-limited growth of Phaeodactylum tricornutum Bohlin strain TFX-1: Chemical composition, carbon partitioning, and the diel periodicity of physiological processes. *Journal of Experimental Marine Biology and Ecology*, 86(1):85–100.

The Gene Ontology Consortium

2017. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic acids research*, 45(D1):D331–D338.

Thomas, D. M., J. Mechery, and S. V. Paulose

2016. Carbon dioxide capture strategies from flue gas using microalgae: a review. *Environmental Science and Pollution Research*, 23(17):16926–16940.

Tréguer, P. J. and C. L. De La Rocha

2013. The World Ocean Silica Cycle. *Annual Review of Marine Science*, 5(1):477–501.

Trentacoste, E. M., R. P. Shrestha, S. R. Smith, C. Glé, A. C. Hartmann, M. Hildebrand, and W. H. Gerwick

2013. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49):19748–53.

Troncoso-Ponce, M. A., A. Kilaru, X. Cao, T. P. Durrett, J. Fan, J. K. Jensen, N. A. Thrower, M. Pauly, C. Wilkerson, and J. B. Ohlrogge

2011. Comparative deep transcriptional profiling of four developing oilseeds. *Plant Journal*, 68(6):1014–1027.

Turchetto-Zolet, A. C., F. S. Maraschin, G. L. De Morais, A. Cagliari, C. M. Andrade, M. Margis-Pinheiro, and R. Margis

2011. Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis. *BMC Evolutionary Biology*, 11(1):1–14.

Turchetto-Zolet, A. C., R. Margis, G. Loss-Morais, M. Margis-Pinheiro, F. R. Kulcheski, and A. P. Christoff

2016. Diversity and evolution of plant diacylglycerol acyltransferase (DGATs)

unveiled by phylogenetic, gene structure and expression analyses. *Genetics and Molecular Biology*, 39(4):524–538.

Tynan, C. T.

1998. Ecological importance of the Southern Boundary of the Antarctic Circumpolar Current. *Nature*, 392(6677):708–710.

UN

2015. Convention on Climate Change: Climate Agreement of Paris.

Valenzuela, J., A. Mazurie, R. P. Carlson, R. Gerlach, K. E. Cooksey, B. M. Peyton, and M. W. Fields

2012. Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. *Biotechnology for Biofuels*, 5(1):40.

van Erp, H., J. Shockey, J. Burgal, P. D. Bates, and J. Browse

2011. Castor Phospholipid:Diacylglycerol Acyltransferase Facilitates Efficient Metabolism of Hydroxy Fatty Acids in Transgenic Arabidopsis. *Plant Physiology*, 155(2):683–693.

Veluchamy, A., B. Lombard, A. Rastogi, J. McCarthy, S. Ott, M. Rivarola, X. Liu, A. E. Allen, P. D. Rabinowicz, D. Loew, X. Lin, C. Bowler, Y. Thomas, F. Dingli, O. Murik, Y. Sun, and L. Tirichine

2015. An integrative analysis of post-translational histone modifications in the marine diatom Phaeodactylum tricornutum. *Genome Biology*, 16(1):1–18.

von Dassow, P. and M. Montresor

2011. Unveiling the mysteries of phytoplankton life cycles: patterns and opportunities behind complexity. *Journal of Plankton Research*, 33(1):3–12.

Wagner, H., T. Jakob, A. Fanesi, and C. Wilhelm

2017. Towards an understanding of the molecular regulation of carbon allocation in diatoms: The interaction of energy and carbon allocation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1728).

Wagner, M., T. Czabany, G. Daum, K. Hoppe, I. Feussner, M. Heilmann, and

M. Fulda

2010. Identification and characterization of an acyl-CoA:diacylglycerol acyltransferase 2 (DGAT2) gene from the microalga O. tauri. *Plant Physiology and Biochemistry*, 48(6):407–416.

Walsh, J. J.

1981. A carbon budget for overfishing off Peru. *Nature*, 290(5804):300–304.

Wang, C., Y. Li, J. Lu, X. Deng, H. Li, and Z. Hu

2018a. Effect of overexpression of LPAAT and GPD1 on lipid synthesis and composition in green microalga Chlamydomonas reinhardtii. *Journal of Applied Phycology*, 30(3):1711–1719.

Wang, D., K. Ning, J. Li, J. Hu, D. Han, H. Wang, X. Zeng, X. Jing, Q. Zhou, X. Su, X. Chang, A. Wang, W. Wang, J. Jia, L. Wei, Y. Xin, Y. Qiao, R. Huang, J. Chen, B. Han, K. Yoon, R. T. Hill, Y. Zohar, F. Chen, Q. Hu, and J. Xu

2014. Nannochloropsis Genomes Reveal Evolution of Microalgal Oleaginous Traits. *PLoS Genetics*, 10(1).

Wang, X., H.-P. Dong, W. Wei, S. Balamurugan, W.-D. Yang, J.-S. Liu, and H.-Y. Li

2018b. Dual expression of plastidial GPAT1 and LPAT1 regulates triacylglycerol production and the fatty acid profile in Phaeodactylum tricornutum. *Biotechnology for Biofuels*, 11(1):318.

Weber, E., C. Engler, R. Gruetzner, S. Werner, and S. Marillonnet

2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*, 6(2):1–11.

Wellburn, A. R.

1994. The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, 144(3):307–313.

Wijffels, R. H., M. J. Barbosa, and M. H. M. Eppink

2010. Microalgae for the production of bulk chemicals and biofuels. *Biofuels, Bioproducts and Biorefining*, 4(3):287–295.

Winkler, U., W. Säftel, and H. Stabenau

1988. β-Oxidation of fatty acids in algae: Localization of thiolase and acyl-CoA oxidizing enzymes in three different organisms. *Planta*, 175(1):91–98.

Wolf, J. B.

2013. Principles of transcriptome analysis and gene expression quantification: An RNA-seq tutorial. *Molecular Ecology Resources*, 13(4):559–572.

Woodfield, H. K., A. Cazenave-Gassiot, R. P. Haslam, I. A. Guschina, M. R. Wenk, and J. L. Harwood

2018. Using lipidomics to reveal details of lipid accumulation in developing seeds from oilseed rape (Brassica napus L.). *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1863(3):339–348.

Xin, Y., Y. Lu, Y. Y. Lee, L. Wei, J. Jia, Q. Wang, D. Wang, F. Bai, H. Hu, Q. Hu, J. Liu, Y. Li, and J. Xu

2017. Producing Designer Oils in Industrial Microalgae by Rational Modulation of Co-evolving Type-2 Diacylglycerol Acyltransferases. *Molecular Plant*, 10(12):1523–1539.

Xu, C. and J. Shanklin

2016. Triacylglycerol Metabolism, Function, and Accumulation in Plant Vegetative Tissues. *Annual Review of Plant Biology*, 67(1):179–206.

Xu, Y., J. Ozga, R. J. Weselake, K. M. P. Caldo, M. J. Lemieux, D. Pal-Nath, and G. Chen

2018. Properties and Biotechnological Applications of Acyl-CoA:diacylglycerol Acyltransferase and Phospholipid:diacylglycerol Acyltransferase from Terrestrial Plants and Microalgae. *Lipids*, 53(7):663–688.

Xue, J., Y.-F. Niu, T. Huang, W.-D. Yang, J.-S. Liu, and H.-Y. Li

2015. Genetic improvement of the microalga Phaeodactylum tricornutum for boosting neutral lipid accumulation. *Metabolic Engineering*, 27:1–9.

Xue, J., W.-D. Yang, H.-Y. Li, K.-T. Peng, X.-Y. Chen, C.-N. Zheng, W. Bai, and J.-S. Liu

2014. Delta 5 Fatty Acid Desaturase Upregulates the Synthesis of Polyunsaturated

Fatty Acids in the Marine Diatom Phaeodactylum tricornutum . *Journal of Agricultural and Food Chemistry*, 62(35):8773–8776.

Yang, Z.-K., Y.-F. Niu, Y.-H. Ma, J. Xue, M.-H. Zhang, W.-D. Yang, J.-S. Liu, S.-H. Lu, Y. Guan, and H.-Y. Li

2013. Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation. *Biotechnology for biofuels*, 6(1):67.

Yen, C.-L. E., S. J. Stone, S. Koliwad, R. V. Farese, and C. Harris

2008. Thematic Review Series: Glycerolipids. DGAT enzymes and triacylglycerol biosynthesis. *Journal of Lipid Research*, 49(11):2283–2301.

Yen, W.-L., J. E. Legakis, U. Nair, and D. J. Klionsky

2007. Atg27 is required for autophagy-dependent cycling of Atg9. *Molecular biology of the cell*, 18(2):581–93.

Yodsuwan, N., S. Sawayama, and S. Sirisansaneeyakul

2017. Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom Phaeodactylum tricornutum. *Agriculture and Natural Resources*, 51(3):190–197.

Yongmanitchai, W. and O. P. Ward

1991. Growth of and omega-3 fatty acid production by Phaeodactylum tricornutum under different culture conditions. *Applied and environmental microbiology*, 57(2):419–25.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto, and D. Bhattacharya

2004. A Molecular Timeline for the Origin of Photosynthetic Eukaryotes. *Molecular Biology and Evolution*, 21(5):809–818.

Yoon, K., D. Han, Y. Li, M. Sommerfeld, and Q. Hu

2012. Phospholipid:Diacylglycerol Acyltransferase Is a Multifunctional Enzyme Involved in Membrane Lipid Turnover and Degradation While Synthesizing Triacylglycerol in the Unicellular Green Microalga Chlamydomonas reinhardtii. *The Plant Cell*, 24(9):3708–3724.

Yu, E. T., F. J. Zendejas, P. D. Lane, S. Gaucher, B. A. Simmons, and T. W. Lane
2009. Triacylglycerol accumulation and profiling in the model diatoms Thalassiosira pseudonana and Phaeodactylum tricornutum (Baccilariophyceae) during starvation. *Journal of Applied Phycology*, 21(6):669–681.

Zaslavskaia, L. A., J. C. Lippmeier, P. G. Kroth, A. R. Grossman, and K. E. Apt
2000. Transformation of the diatom Phaeodactylum tricornutum (Bacillariophyceae) with a varietty of selectable marker and reporter genes 1. *J. Phycol*, 36:379–386.

Zhang, M., J. Fan, D. C. Taylor, and J. B. Ohlrogge
2009. DGAT1 and PDAT1 Acyltransferases Have Overlapping Functions in Arabidopsis Triacylglycerol Biosynthesis and Are Essential for Normal Pollen and Seed Development. *The Plant Cell*, 21(12):3885–3901.

Zhang, Y.-M., S. W. White, and C. O. Rock
2006. Inhibiting Bacterial Fatty Acid Synthesis *. *The Journal of Biological Chemistry*, 281(26):17541–17544.

Zhang, Z., D. Sun, K. W. Cheng, and F. Chen
2018. Inhibition of autophagy modulates astaxanthin and total fatty acid biosynthesis in Chlorella zofingiensis under nitrogen starvation. *Bioresource Technology*, 247(July 2017):610–615.

Zhao, S., W. P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu
2014. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*, 9(1).

Zhou, X. and P. J. Wangersky
1985. Copper complexing capacity in cultures of Phaeodactylum tricornutum. 17(4):301–312.

Zhu, B.-H., C.-C. Tu, H.-P. Shi, G.-P. Yang, and K.-H. Pan
2017. Overexpression of endogenous delta-6 fatty acid desaturase gene enhances eicosapentaenoic acid accumulation in Phaeodactylum tricornutum. *Process Biochemistry*, 57:43–49.

Zienkiewicz, K., Z.-Y. Du, W. Ma, K. Vollheyde, and C. Benning

2016. Stress-induced neutral lipid biosynthesis in microalgae — Molecular, cellular and physiological insights. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1861(9):1269–1281.

Zou, J., Y. Wei, C. Jako, A. Kumar, G. Selvaraj, and D. Taylor

1999. The Arabidopsis TAG1 mutant has a mutation in a diacylglycerol acyltransferase gene. *The Plant Journal*, 19(6):645–653.

Zulu, N. N., K. Zienkiewicz, K. Vollheyde, and I. Feussner

2018. Current trends to comprehend lipid metabolism in diatoms. *Progress in Lipid Research*, 70:1–16.

A

Appendix

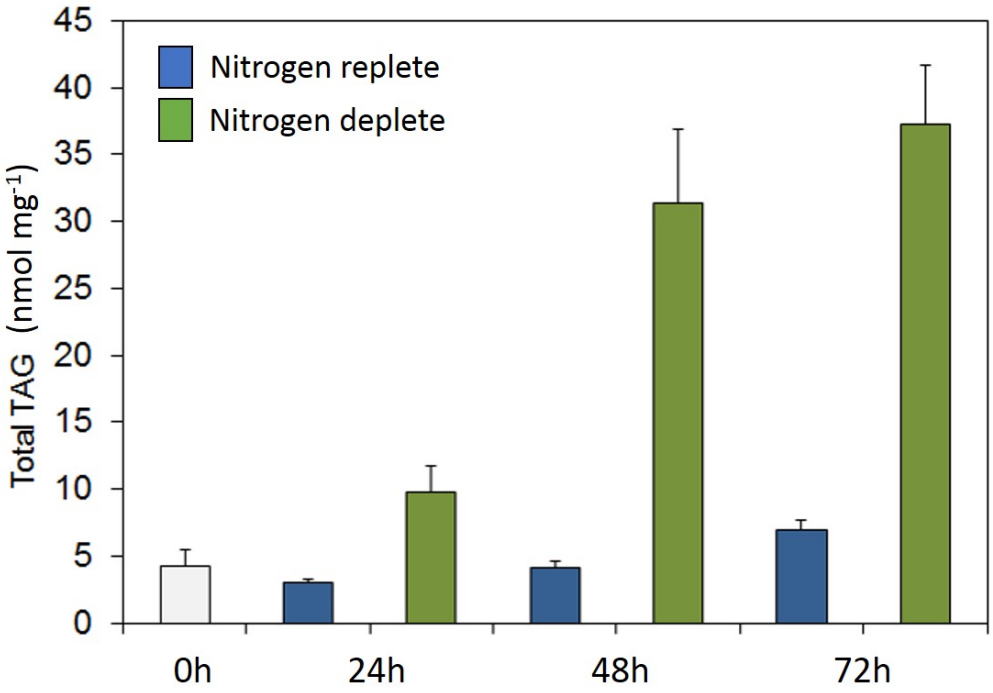


Figure A.1 – MRes thesis data showing TAG accumulation in *P. tricornutum* during nitrogen depletion. Total lipid extractions were carried out on strain Pt4 (described in Chapter 2) that was grown in f/2 with or without nitrogen for up to 72h. TAG was quantified using ESI-TQMS (Chapter 2) and is normalised to dry weight. The 0h time point was a culture of exponentially growing cells in f/2 with nitrogen.

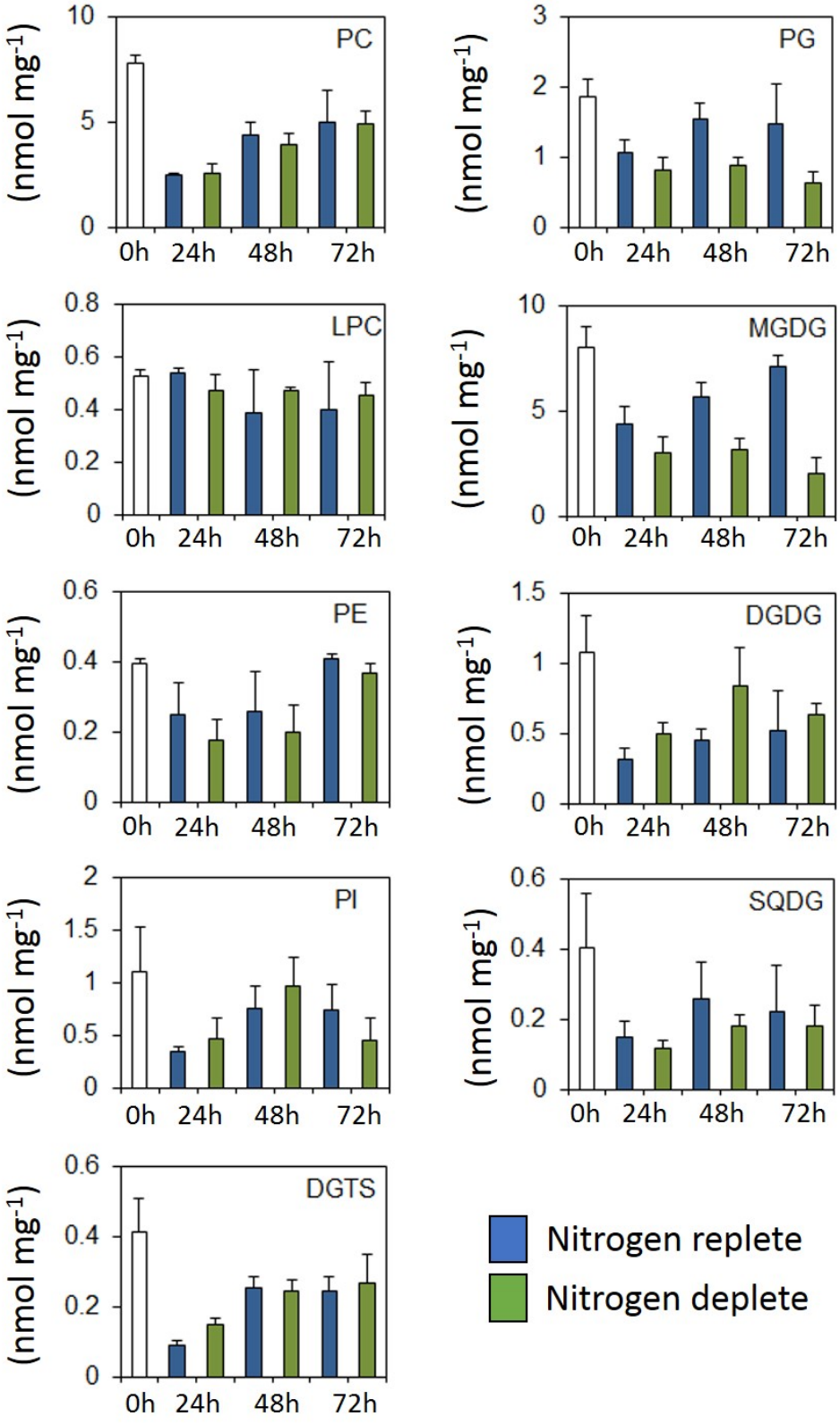


Figure A.2 – MRes thesis data showing glycerolipids in *P. tricornutum* Pt4 during nitrogen depletion. Cells cultured and lipids grown as in Figure A.1. PC- phosphatidylcholine, PG- phosphatidylglycerol, LPC- lysophosphatidylcholine, MGDG- monogalactosyl diacylglycerol, PE- phosphatidylethanolamine, DGDG- digalactosyldiacylglycerol, PI- phosphatidylinositol, SQDG- sulphoquinovosyldiacylglycerol, DGTS- diacylglycerol trimethylhomoserine.

>*C. reinhardtii* PDAT - Cre02.g106400.t1.1

MTTPTKGNSNARQRKKGGTSAEASAATPSKAKPGRDHNVHATPSHSHSHSHSQSQQHRQGPHAAQPKSERRLVLWLAAAG
VVLLPLVLLPPALMVEEGPVAELLHDSLGLPRLNAYLPPLPSIVALDSLLASAGFAVPSSALNGTSAAAAAANSSAYNGLLALFGV
TGDMASAGMDGLTSRTVTAAQQLRAFLAQLPALSIPGGPAATGGGPAGAGAGAGAGAGAEVNATAAADEAAGLRPGQLM
ARRGYRAKHPVVIVPGFVTSGLELWRGLPCGQRYFRQRMWGTLAMVQAFLTDAACWFRHMELDTVSGLDPEGIKLRAALGL
EAVDYFIQGYWVWGKLVEALADVGYDSNSLVSMPYDWRLAVPLLEERDGYYTRLRRTIEQLVELTGERVVVTSHSYGENVFRA
FMHWVEAAAAEEEEGGKQPRSGGGSGHSGGGWVDRHIASTINIAGTSLGVPKSVSALLSGETRDTAQLGALAGFLTSNMVP
RAARTRVWRSWGASYAMLPVGGPGVWGNASWAPDDTPEMRANRRTFGSMVSLWPHNWQALMAAAPQQQREQLSQL
VAEVSQDVDAGSGSGSGSGSGSGSGSGSGSGSGSGSGSGEDAVNGTGGAGSAGGAALAPAEAAAVGTAGGGDNATAATA
VNSTTASNSTGSAGTGSSGSGGSSSSKAIAAIADTIKKLVGGGVDAAEFLNSDHVTRLDVSGFIALLREVGGPLVSANIAQWGAL
QLPAEEAAEAVRKRAAAASDRAAPAAAGSPAAQAVGGADRVKDSTAGGAGAAGIGRGGSSAGSGRVALFPDATRTPLPSAP
KTTMVCLYGVGLPTERGYHYLRPPPPTGAAATATAAAANAPAGNAAAAPAADAQPPPQAAGGGDGDGMATTDGTAEADA
ERGVPPRGATASGEEANPDAGGDASSGWMIMKDVTQAQSALDVGVHISDGDGTVPLLSLGLMCRGGWREAGHLNPGAM
RVVTREYKHKAVSMLQDARGGPAAAAHIDILGNDAVLRDVITVVAGRADELADIVVSDIDRIAAAVDWAALAA*

> *C. reinhardtii* PGD1 - Cre03.g193500

MSQLLSHFVRVPTFASPDQVLREARDKERELQNARAPTDVSGFLAPVGVWELKHLRKLSSLTSLTYYMHLVTPRRLQLMHGLD
LVVTSRACDVRPYEHNRTAEECGADGDGMAVSFAEARQVYAELKRGTGGASGSGSNGAAAAPVAVAAGVSNIVALPRELPFV
PLPGGATEGGEAGAEAAAGAAAAAAAAAGEGTGPGAQGQGAGGLQLPLASTEAIGRMLRSPAEVVSAKLAEAALAASAAAA
ASPLGAAAESFYAGLASLPIPLAGGLVGANNKAANTLLAPPNGAAASSSGGGGGSGGAAAASEVVGSSRGAQGADPADPGAP
NTSGKATAASIAAMATAELRSRRLGGTGPAKTGGSGAASSGSSSGGIGAGGMAPVTAGGMRLSPAATAFSAPPAATVSSLAST
DAGTALVPVASSSAASLTFSSSSAYSCPSEWFVVDDPASATRIFVIQGSDTLDHWKLNLTFDPVVFEEPALGVKVHRGVYEAALV
LYERFLPLVYEHLEASPFSKVTFTGHSIGGSMATLLMLMYRNRGVLPPHSIATVYTFGAPAVFCQQQQPAVADASSACCNGSSS
NGSSTPASGSSSPRSGSPGSASASSLPAGSGSGAGGMSLWALGLSGFGMGGASLAGGGSTSAPSSTAGLASVDGGAVAAGG
GGSGFNSSGLGFMSVEDPQAVSMPPGAAQAPSPASAPAPTAGPGHNSHSSHSKAGPAAAKSCACGVDGLLTRLGLAPHVVR
NVVMARDVVPRAFACDYSLVADILKGWGPAFREHCCLNRHGRKHLYYFVGRMCILQPDAWHSFVGGDPEHPMLPPGPELYA
LAEPEDAAAARAHYPALSDLPILNAVTSNGHTRGSGGNGANAAVNAAVNASGPSAAASGGGGGSQQPTAAAAVPSTANFGT
ALVASAAQRERDARGGGSRLQPRSVVEAVWEIMDNPHPLETLADPGAYLASGSISRYHNPEHYTKALGRLTHLKRLAERRQHP
HGQAQQKQAQPQAGEGGIRSMFAGRNIRSFGGGVRSGSGSGSAGRRGLLHQQAASNGTAADAVLASGAAGAAAAAAWGS
APQLADLVSGNGGRASAGYEGGVWDSSDGLDLHLSDFMGASAVGAADPHACR*

>*A .thaliana* - AT3G06510.2

MELFALLIKVAGLLATVTVGANVVSYSRFRRQNLAKFRSPIDESKEVLADFNSIEHEEGKFFFGLATAPAHAEDDLDDAWLQFAK
ETPCSAEEAEAADKKARRKKVKLAVGAITKGLAKNTHGKEDKNAADKPPSKNVAAWHNAPHAEDRLKFWSDPDKEVKLAKDT
GVTVFRMGVDWSRIMPVEPTKGIKEAVNYEAVEHYKWILKKVRSNGMKVMLTLFHHSLPPWAADYGGWKMEKTVDYFMD
FTRIVVDSMYDLVDSWVTFNEPHIFTMLTYMCGSWPGNNPDFLEIATSTLPMGVFHRALHWMAVAHSKAYDYIHGKISLKKP
LVGVAHHVSFMRPYGLFDIGAVTISNSLTIFPYIDSICEKLDFIGINYYGQVRELQVKIAIRSQILINNIAFSRISMLESDSRNQEAVC
GAGLKLVETDEYSESGRGVYPDGLYRVLLMFHERYKHLKVPFIVTENGVSDETDVIRRPYLIEHLLALYAAMLKGVPVLGYIFWTI
SDNWEWADGYGPKFGLVAVDRSHDLARTLRQSYHLFSKIVKSGKVTRKDRSLAWNELQKAAKAGKLRPFYRGVDNHNLMYA
DGLDKPQWRPFVDRDWRFGHYQMDGLQDPLSRVARTLLIWPLIMKKRIRKVKIKHTDDAGLVLHPALASPFD*

Figure A.3 – Protein sequences used to identify candidate genes for a role in degrading MGDG. Candidate genes were identified from *P. tricornutum* gene models in Ensemble Protist using a BLAST search with the following query sequences: CrPDAT, CrPGD1 and AtSFR2. AtSFR2 was not included in the first identification round. In the second identification round, *P. tricornutum* gene models were also queried using the domains and structures found within these genes.

Table A.1 – *P. tricornutum* gene model candidates for enzymes that degrade the chloroplast during nitrogen depletion. Candidates were identified as described in Section 5.3.1.

Gene ID - first round	Interpro domains
Phatr3_EG01544	AB_hydrolase; Fungal_lipase-like
Phatr3_EG01635	AB_hydrolase, Fungal_lipase-like
Phatr3_EG02610	AB_hydrolase, Fungal_lipase-like
Phatr3_J12431	C2_dom; PLipase_D/transphosphatidylase; PLipase_D_C; PLipase_D_fam
Phatr3_J24223	
Phatr3_J31492	AB_hydrolase; Fungal_lipase-like
Phatr3_J35589	AB_hydrolase; DUF676_lipase-like
Phatr3_J37652	Ac_transferase_dom; Acyl_transferase; Acyl_Trfase/lysoPLipase; Malonyl_CoA-ACP_transAc
Phatr3_J37711	AB_hydrolase; Fungal_lipase-like
Phatr3_J38196	Lipase_GDSL; SGNH_hydro
Phatr3_J39432	Alk_phosphatase_PhoD-like; Metallo-depent_PP-like; PhoD_N
Phatr3_J41624	AB_hydrolase; Fungal_lipase-like
Phatr3_J42683	C2_dom; EF-hand-dom_pair; PH_like_dom; PI-PLC_fam; PLC-like_Pdiesterase_TIM-brl; PLipase_C_PInositol-sp_X_dom; PLipase_C_Pinositol-sp_Y
Phatr3_J43678	Acyl_Trfase/lysoPLipase; Patatin/PLipase_A2-rel
Phatr3_J43856	AB_hydrolase; Fungal_lipase-like
Phatr3_J44028	AB_hydrolase; Fungal_lipase-like
Phatr3_J44231	AB_hydrolase; Fungal_lipase-like
Phatr3_J45518	Acyl_Trfase/lysoPLipase; EF_Hand_1_Ca_BS; EF_hand_dom; Patatin/PLipase_A2-rel; Triacylglycerol_lipase
Phatr3_J45895	AB_hydrolase; Fungal_lipase-like
Phatr3_J46753	
Phatr3_J49142	Acyl_Trfase/lysoPLipase
Phatr3_J49771	C2_dom; PH_like_domPH_like_dom; PI-PLC_fam; PLC-like_Pdiesterase_TIM-brl; PLipase_C_EF-hand-like; PLipase_C_Pinositol-sp_Y PLipase_C_PInositol-sp_X_dom
Phatr3_J9277	PLipase_C/P1_nuclease; S1/P1nuclease
Phatr3_J9794	DiacylglycerolacylTrfase1; Oat_ACAT_DAG_ARE; PH_domain; PH_like_dom
Phatr3_Jdraft1000	
Phatr3_Jdraft1611	C2_dom; EF_Hand_1_Ca_BS; EF_hand_dom; EF-hand-dom_pair; PI-PLC_fam; PLC-like_Pdiesterase_TIM-brl; PLipase_C_Pinositol-sp_Y PLipase_C_PInositol-sp_X_dom

Table A.2 – *P. tricornutum* gene model candidates for enzymes that degrade the chloroplast during nitrogen depletion. Candidates were identified as described in Section 5.3.1.

Gene ID - second round	InterPro description
Phatr3_EG02462	
Phatr3_EG02383	
Phatr3_J49702	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase and Alpha/Beta hydrolase fold
Phatr3_EG02223	
Phatr3_EG02212	
Phatr3_J11735	
Phatr3_J45122	
Phatr3_J45579	GPI inositol-deacylase PGAP1-like and Alpha/Beta hydrolase fold
Phatr3_J36597	
Phatr3_J46568	
Phatr3_J14397	
Phatr3_EG02395	
Phatr3_J28219	
Phatr3_J54642	
Phatr3_J769	
Phatr3_J8860	Lecithin:cholesterol/phospholipid:diacylglycerol acyltransferase and Alpha/Betahydrolase fold
Phatr3_J15323	AB_hydrolase_1 and AB_hydrolase
Phatr3_J44153	AB_hydrolase
Phatr3_J37756	AB_hydrolase_1 AB_hydrolase
Phatr3_J50641	AB_hydrolase
Phatr3_J48297	DUF676_lipase-like and AB_hydrolase
Phatr3_EG01919	AB_hydrolase
Phatr3_J49386	AB_hydrolase
Phatr3_J42896	AB_hydrolase
Phatr3_J44991	AB_hydrolase
Phatr3_EG00439	AB_hydrolase
Phatr3_J43541	AB_hydrolase
Phatr3_J49598	AB_hydrolase
Phatr3_J45895	Fungal_lipase-like and AB_hydrolase
Phatr3_J47448	AB_hydrolase
Phatr3_J25619	AB_hydrolase
Phatr3_J44931	AB_hydrolase
Phatr3_J35589	DUF676_lipase-like and AB_hydrolase
Phatr3_J44028	Fungal_lipase-like and AB_hydrolase
Phatr3_J11336	AB_hydrolase
Phatr3_J34949	DUF676_lipase-like and AB_hydrolase
Phatr3_J45940	AB_hydrolase
Phatr3_J10362	AB_hydrolase

Table A.3 – *P. tricornutum* gene model candidates for enzymes that degrade the chloroplast during nitrogen depletion. Candidates were identified as described in Section 5.3.1.

Gene ID - second round	InterPro description
Phatr3_Jdraft453	AB_hydrolase_3 and AB_hydrolase
Phatr3_J34884	AB_hydrolase
Phatr3_J41624	Fungal_lipase-like and AB_hydrolase
Phatr3_J33738	AB_hydrolase
Phatr3_J43666	DUF829_TMEM53 and AB_hydrolase
Phatr3_J49810	AB_hydrolase
Phatr3_J43025	AB_hydrolase_1
Phatr3_J45728	AB_hydrolase
Phatr3_J47445	Fungal_lipase-like and AB_hydrolase
Phatr3_EG01520	Fungal_lipase-like and AB_hydrolase
Phatr3_J49119	AB_hydrolase
Phatr3_J54974	AB_hydrolase_1
Phatr3_J49144	AB_hydrolase
Phatr3_J42481	AB_hydrolase_1 and AB_hydrolase
Phatr3_J8407	Znf_FYVE and Fungal_lipase-like and Znf_FYVE_PHD and Znf_RING/FYVE/PHD and Znf_C2H2_type and Znf_FYVE-rel and AB_hydrolase
Phatr3_EG02201	AB_hydrolase
Phatr3_J42422	AB_hydrolase
Phatr3_J31492	Fungal_lipase-like and AB_hydrolase
Phatr3_J43118	AB_hydrolase_3
Phatr3_J43295	LDAH and AB_hydrolase
Phatr3_J43463	Fungal_lipase-like and AB_hydrolase
Phatr3_EG02610	Fungal_lipase-like and AB_hydrolase
Phatr3_J10718	AB_hydrolase
Phatr3_J33720	PLipase/COase/thioEstase and AB_hydrolase
Phatr3_J44260	AB_hydrolase
Phatr3_J49795	AB_hydrolase
Phatr3_J40695	AB_hydrolase
Phatr3_J50327	AB_hydrolase
Phatr3_J41220	DUF1350 and AB_hydrolase
Phatr3_J45633	AB_hydrolase
Phatr3_J35081	AB_hydrolase
Phatr3_J45583	AB_hydrolase
Phatr3_J35564	PLipase/COase/thioEstase and AB_hydrolase
Phatr3_J45778	AB_hydrolase
Phatr3_J45964	AB_hydrolase
Phatr3_J46006	AB_hydrolase
Phatr3_J46456	AB_hydrolase and AB_hydrolase_5
Phatr3_J47043	AB_hydrolase
Phatr3_J37441	AB_hydrolase

Table A.4 – *P. tricornutum* gene model candidates for enzymes that degrade the chloroplast during nitrogen depletion. Candidates were identified as described in Section 5.3.1.

Gene ID - second round	Interpro domains
Phatr3_J21912	AB_hydrolase_1 and AB_hydrolase
Phatr3_J47524	AB_hydrolase_3 and AB_hydrolase
Phatr3_J52268	PLipase/COase/thioEstase and AB_hydrolase
Phatr3_J47663	AB_hydrolase_1 and AB_hydrolase
Phatr3_J48221	AB_hydrolase
Phatr3_J48227	AB_hydrolase
Phatr3_J48133	AB_hydrolase_1 and AB_hydrolase
Phatr3_J48226	AB_hydrolase
Phatr3_J43856	Fungal_lipase-like and AB_hydrolase
Phatr3_J44066	DDHD_dom and AB_hydrolase
Phatr3_J44231	Fungal_lipase-like and AB_hydrolase
Phatr3_J33163	DUF829_TMEM53 and AB_hydrolase
Phatr3_J46753	Fungal_lipase-like and AB_hydrolase
Phatr3_J46935	PGAP1-like and AB_hydrolase
Phatr3_EG01099	AB_hydrolase
Phatr3_EG01544	Fungal_lipase-like and AB_hydrolase and phospholipase A1-II
Phatr3_EG01635	Fungal_lipase-like and AB_hydrolase
Phatr3_EG01987	Fungal lipase-like domain and Ribonuclease H-like superfamily and Reverse transcriptase, RNA-dependent DNA polymerase and Alpha/Beta hydrolase fold
Phatr3_J3527	DUF1350 and AB_hydrolase
Phatr3_J42756	AB_hydrolase
Phatr3_J43961	DUF829_TMEM53 and AB_hydrolase
Phatr3_J44005	DDHD_dom and AB_hydrolase
Phatr3_Jdraft1423	DUF1350 and AB_hydrolase
Phatr3_J50397	Fungal_lipase-like and AB_hydrolase and C2 domain superfamily
Phatr3_J37711	Fungal_lipase-like and AB_hydrolase
Phatr3_J47614	Fungal_lipase-like and AB_hydrolase
Phatr3_J36381	AB_hydrolase
Phatr3_J48040	AB_hydrolase
Phatr3_J44873	Glucanosyltransferase and Glycoside hydrolase superfamily
Phatr3_EG02351	Glycoside hydrolase family 31 and Galactose mutarotase-like domain superfamily and Glycoside hydrolase superfamily and Glycoside hydrolase family 31, N-terminal domain
Phatr3_EG00719	Glycoside hydrolase superfamily and Uncharacterised protein family, glycosyl hydrolase catalytic domain
Phatr3_EG01004	Glycoside hydrolase, family 2, immunoglobulin-like beta-sandwich and Galactose-binding-like domain superfamily and Immunoglobulin-like fold and Glycoside hydrolase superfamily and Beta-Galactosidase/glucuronidase domain superfamily
Phatr3_J43302	Glycoside hydrolase superfamily
Phatr3_J49610	Glycoside hydrolase superfamily

Table A.5 – *P. tricornutum* gene model candidates for enzymes that degrade the chloroplast during nitrogen depletion. Candidates were identified as described in Section 5.3.1.

Gene ID - second round	InterPro description
Phatr3_EG02210	Glycoside hydrolase family 1 and Integrase, catalytic core and Ribonuclease H-like superfamily and Reverse transcriptase, RNA-dependent DNA polymerase and Glycoside hydrolase, family 1, beta-glucosidase and Glycoside hydrolase superfamily and Glycosyl hydrolases family 1, N-terminal conserved site and Ribonuclease H superfamily
Phatr3_J34543	Glycoside hydrolase superfamily
Phatr3_J45418	Glycoside hydrolase superfamily
Phatr3_J54569	Glycoside hydrolase superfamily
Phatr3_EG02430	Glycoside hydrolase, family 2, immunoglobulin-like beta-sandwich and Glycoside hydrolase family 2, catalytic domain and Glycosyl hydrolases family 2, sugar binding domain, Galactose-binding-like domain superfamily and Immunoglobulin-like fold and Glycoside hydrolase superfamily and Beta-Galactosidase/glucuronidase domain superfamily
Phatr3_EG02432	Beta galactosidase small chain/ domain 5 and Glycoside hydrolase family 2, catalytic domain and Galactose mutarotase-like domain superfamily and Immunoglobulin-like fold and Glycoside hydrolase-type carbohydrate-binding and Glycoside hydrolase superfamily and Beta-Galactosidase/glucuronidase domain superfamily

Phatr3_44028 targeting peptide

ATGAGGTTCTGCTTGCTCCTCTTTCTTACTCTGTTGGATAATATCGCCGGGAGCTTTATTCAGCGCTCGCACGCAGGTTCGTTTCGCACTTTGACGCTACGGACAC
ATAAAAAAACTGATACGGAAGAACATGGCGGACCGCAAAAGCTGAATATACAGCAGGAGTCATCGTTGTTTGATATAAGAGCGCAATGGGAGAAAACTACCTC
AGAGTTTGATCCAGTCGATTTCTGGAAGCAAATTGTACGAAATGACGAGACGATAAATGACTCAGTAGTCAACCGAGAAACAAACGAAGAGGGAAGAATAACA
GAATTTTGGGATGCGCTGAAACACAATACGGATAGT

Phatr3_44028 full coding sequence

ATGAGGTTCTGCTTGCTCCTCTTTCTTACTCTGTTGGATAATATCGCCGGGAGCTTTATTCAGCGCTCGCACGCAGGTTCGTTTCGCACTTTGACGCTACGGACAC
ATAAAAAAACTGATACGGAAGAACATGGCGGACCGCAAAAGCTGAATATACAGCAGGAGTCATCGTTGTTTGATATAAGAGCGCAATGGGAGAAAACTACCTC
AGAGTTTGATCCAGTCGATTTCTGGAAGCAAATTGTACGAAATGACGAGACGATAAATGACTCAGTAGTCAACCGAGAAACAAACGAAGAGGGAAGAATAACA
GAATTTTGGGATGCGCTGAAACACAATACGGATAGTTCTCATCTTGAGGACGCGTTCGGCAACGTGGACTTTCATGGCATCGGGCAGAATCTCATGGGACTGAT
TTCCGGCGGTTTTCACAGCGATGTCACGGTGAACGACATCATCGAACGCGTTCGTGACAGTACAAACCAAGACGATAAAGAAGATTCTAAAAGATTCGATATTA
TCATTGAAATGATTCAAAACTACAAAGGTCTCATCACAGAAGTCATGGCAAAATATGTGCTGGATACAGTTGATTTCAGCCGTTTCCGGCCAGCCTCACTTTTTTA
CTACATAGAAAAAGAGGACTCCATCAAGACACCGAGCTGGAAAAGGAGGCAGCACCGCTTTTGCCCTGAGGTCGACTTAAATATGGTAAAGAATCTCAACAAC
GACTTAGCGCTGGCGCGTCTTAGCTATGCAGATTCCTGCGATGAAATTCGTGAAGGACTACGTCTTTTCAAGATGCCTTACGAGCTGGTCTATGCAGAAGTTGAA
AGCGAACCAGGAAAACCAGGCCACTTCATTGCTCTAAAACGAAACCAGCCCTTCCGAGGCGGAGAGCTTGAAGCTATTTTGGTTGTCAGAGGGACCAAGACAG
TAGCCGACGTCGTTACGGATCTTTGCTGTGATATCGTTGAGTACAACGAGGGAAAGGCGCATTCCTTCATGCTTGCTAGCGGCCAGTATATTGCCGAAAAACAT
ACACAGCTATTCGCAGATCTGTTAGCCAAATCAGGTAAGAGTAAACTGAAGCTGACTCTTGTTGGGCACTCCCTGGGAGCTGGTGCTGCAGCAATAGCTGGCAT
GGAGCTTAACGCGCACCCTGATATCAATGTAGAAGGAGTTGGGTTTGGCTGTCCTTCTCTTGTGACCCAAGAGCTGGCAGAGTATACTTCTTGGTATGTTACCGT
CGTCAACGACTCCGACGTGGTCCCTCGTGCAAATCCAGTTACTGTTGCCAATGTTCTGCTCAACATCACAGAATACGATTGGGTCCCCAGCGCCAAACGTGATGC
CCTTCAAGTCCTGAATGAGCTTCACCGGAGACAACCCTTTATCTTTCAAGAAGGAGTTATCGAAAAGCTGCAGGAAATGCTTCAGCCTTTGATTGAATCAGCTGA
TAGCGGTCCGAGATCTAAATTGTGTAGTGACGGAGAGCGGATTCAACCCATTCTTGTCCCGCCAGGAAAATGCATTCATTTCTACCACGATGGCTACCGAATATC
AGGCCGTTACGTTCCGAACTCTTTTTTTTGTGAAATCGATGTAACTCGGCGCATGTTTAAAGATCACTTTTTGATAGCGGATACGAGCAGGTCTTTCTGGACCTC
ATGCGGCAGGAGAAAGGTGACCCCCACTTTCGATTCAAAGTGGAACAAACCGAATAT

Phatr3_41624 targeting peptide

ATGACGGTTCAGCGGTGGCGGAACGGGTGGGGGAGGGTAGGATGGTGCATTCTGTGGTACGGCGCGCTGCCGATCTACGGGACGGATATTCCCTCGGAGGA
GGGTTCCCCCGGGTCGGATCCCACGACGACCTCAACAAGAACGACGCCCACCCCCAGTACTGCCATTCCCAACGACTATGCCTGGCTCAATCTCGGGTCGTCCGT
TCCGTTCTTTTCGCAACTATTCGACAATCCCGATCCTTCCGATGATCCCTTTGATCGCCTTGCCGGTTCCCACGAATCGGAAGATTCCGATCTCTTTCGGAAGCTCG
CTTCCCTCTTTCGATCGACACCATCTCTGAGTGTGTTTCCGAACGACAACAGCAACAACAATTACGACAAGGTACACGACGAGAGTCAGAAATTCGTCAACCAAA
TTGTGAACACCTTTAGTTCTCCCGGGCAAGCACAAACTCGCCAACAAATCTTCAAATTGTTCCAGCAAAGCATCATCGCCATGCAGGAACAAATGCAGAAAACCT
TTGGAGACATGCAGAACGAACTTTGGGAGCGTTTCAATCTACTGCAACTCACCTATTTTCTACAGCACGAAGAATCCGTCAAGAATCCTGTTTGGAAACGACGCC
AGCATCGATTTATGACGCCACTGCCCGTCAGCGAGGCCGTACAACTGGGCGATGGACTCTTTCTCTCCCATCTCGCTTACGTGGACGACTGTCACCTCATTCAGG
AATATCTGCGGAGCTTCCATAACGGCGCCTTTGTCCTACGCAACTGCACCACCTCCAGTCAACCCAACCAACCGGCCCACTTTTTGGCCGTCCGTAAGATCGCCA
GCCCCTCCCAGACAACCACAAAATCCTTCCTGACGCCTGAGTGGAAATGGCTCATTCATCAGTACTTGCCCTTCTTCTTTCGTCCCGAACCGCTCGAAGTCGTCCT
CACGATCCGCGGCACCAAAGAAATAGGTGACTTCTTGTCCGACGCCATGTTAGCCGCCGCGAAGCATCGGAATGGTAAGGCACACGACGGTATACTCAAATCA
ACCCAATGGATGCTGAAAACATACACGGACGATTTGCAACAATTGTTAAAAGATTCTCAACGGGATAGAATGAATCTATGGTTGGTCGGACATTCTCTAGGAGG
CGGAACGGCGGCCTTAATGGCCATTGAGTTGTTCGAAACGCAAGACGGATGGGTTCAGCCCCACGCATTGGGCTTTGGTACGCCCTCCTTGGTTTCAGCGGAAC
TATCCCGCAAGTACAAACCAATTGTTAAGACCGTCATTAACGATGCCGATGCCGTGCCCCGCATGTCGGGTGCGTCCATTGCCAACGCTTGGCTACGAGTCGTCC
GTTTTAACTGGACCGATGCCTTCTTGCAAGACTTTGATCAAACCGTTGGTTTTTTGAGCGATTACTTCTTAAAATTCAAGATGAATACCGATTGGATCAATGAGTG
CTCGCGTGACATTCGACAGATACTCATCACTTGGCTTGATCAAAAGGTTTCTCCGGGCTTGAAGAACTTGCCTAAAGCGACCGAAAAAGAATCCGAACTCATTCC
TCCCGGAAACTGTGTCCACTTATACCGTGATGGAACTTCCTGGCAAGGAGCATACACCAGCTGTTACTTCTTTCAAGAGTTGGAAGTGGTTCGGCACTTGGTCGA
TGATCATTTGATTCCGGGCGGGTACTACCGTGGCCTTTTGAATTATGTTCGCGAGCAGAAACAGAATGTTTCATGGAAATTCGACCATGACCTGCTCACACTCCC
GGTGGGC

Phatr3_J31492 targeting peptide

ATGTCGTTAGTGGCGCAACAGCTGGAGACGCACAAGCGACTGCAAGCCGCGCGACAGGAACTATGGAGACTATATCTGCCAAACTTTCTGTGTCCCTACCGCG
AGTTCGTATTAGACCCTAGCCTTCAGTTCTGGACTATCGTTGTTACTCTCGCTTTACCCAAACGCTGGGATTTTTCTATCTCTCGAATTTCACAGACGTTGTTCATA
CGTCCAACTCTTGCTACTTTGCGATTCTTCTCCAAAGACCCTGATGGCGAAGATCTTCTGGTCACCGCTGGCAAAGAGCTGCTGCACCCAGTTGCGCCCCGTCCAC
TGCATTTTCAGCCTCCC

Phatr3_J31492 full CDS

ATGTCGTTAGTGGCGCAACAGCTGGAGACGCACAAGCGACTGCAAGCCGCGCGACAGGAACTATGGAGACTATATCTGCCAAACTTTCTGTGTCCCTACCGCG
AGTTCGTATTAGACCCTAGCCTTCAGTTCTGGACTATCGTTGTTACTCTCGCTTTACCCAAACGCTGGGATTTTTCTATCTCTCGAATTTCACAGACGTTGTTCATA
CGTCCAACTCTTGCTACTTTGCGATTCTTCTCCAAAGACCCTGATGGCGAAGATCTTCTGGTCACCGCTGGCAAAGAGCTGCTGCACCCAGTTGCGCCCCGTCCAC
TGCATTTTCAGCCTCCCTCCGACATACAGTCGGCTCTGTCGGGTGTTGCCGCGTTTGTCCTGAGTGTCGTCGCGGACGTGATTGGTGCATTTATGGATCCCGTTAA
AATGAAAAAATGGTTTTCCGCAATGTCGGCCTTCCGGGCGTACTTACAAGCATCCGGTGTAGGTGCAGAATTAGAGGAATCCCTCATCAAACCTCTTTGGAGAG
GGCGGCTTCTAGACAACCTAAAGATTCTCAATGACTGCCAAGAAATCTTAGACGAGGACCGTACGAAACTTGCGAATGACCTTGATTCAGAGGTAAGCTCGGA
GGACCTCGTCGTCGGTTGCAGCATGATGCGTTTCGCCACCGCCGCGTATGGTGTTGAAATGGTTCGCTCGGCAATTGATCGCGAAGCGAATTACGAACATGTCA
ACAGTGAGCGAAAGGCCATTGCATTTCACTGCAATATCCCAACCGAGGACGTCAAGTATATTTATATCCAGCCAGGAGACGAAATGCACACGATGCGTCATTTC
ATTGCGGTCGATGAAAAGACCAAATCCGTCGTTCTCGCCATCCGGGGAACGTTATCAATTTCCGGTGCCTTGGCGGACATGCAAGCTATGGATTTTGATTTTTGT
GGCGGCAAGGCACACATGGGTATAGCGGAACAAGCCAATTTACTTTGGCAGAAAACAGGACAACGCCTCCGCAGGATCGCTTCCGCATACTCGGAAGAATACC
GAATCATTTTTACGGGACATTCGCTTGGAGGAGGTGCCGCGTGCCTATTGCACGTGAAAGTGCACACAGAGAATCTGTTGCCGACGAGACAGGTCTACTGCTAC
GGCTTTGCACCCCCACCAACATATTGCAAGGGTAGCACTCCTTCGCCAGGTCTGGAAATGGCCGTCAAGAACTGTGTATGCTTTGTGCACGATAACGACTGTGTT
CCACTTCTGAGTGTGGCATCCATCCGTCGTCTGGCTTGCCTTATGGATGCGGTTGACAATTGCACGGAAAATCTCTGGTTCACGACACGTTTCCGAATCTTTTGGG
AGTTTGTCAAGGTCCCTGGCGATATCGTCAAAACTGTCTGCAGCGTTAAGCATGACTCGAAGGCAGTCGTTGGTGAGTCAGCCATGGTCATCCCAGCCCGTTGC
ATTGTTTGGATGAAGAAAACCTTAAGTGGACGTTTTGAGGCCCTTGCGTGTAGTTCAAAAGCCATGGCCTCCATGAATATCTTTGTCTGCCAAGATATGATTGCT
GATCACATGCCAGAGCAATACGAGGATGCCCTAGACAGTCTTGTAGCTAGAAGGTTTCAAGAGCAACTG

Figure A.4 – Candidate gene sequences cloned in Section 5.3.

FcpA promoter and 5' UTR

TTAATCACTCTGTGGTCTCAGGAGGGGCTGCAGGACGCAATGGAGGATTATCACCGCAAAAATGAACTTCGAAAAAAACTTTCGAGCGACCATGGAAAAGGAG
GATCAGATTCAGATTACAACAGTGGATTGCTCTGGTAGCAAATATCTTCTGCTAGATTGGCTCATGGTCGGTTTTGGACGTTCGAAGCTCACCGTCAAAAGAAAC
AAAAGAGAAGAATGACGTCCTCGTGACGTAGAATCTACGACTGTACTCGGATCTGGGAAATGAATTGACTCACGGTATTCTTCGAGTCCTGTTACAGGCCCTTG
GTCCGAACCCCCACACGATTTTTGCACCAAAGATTTGCTTCAATTTGCTGGATGTTTTGACTGCAAGATCAGCTGGCCTAGCAAGAGTGCTCGTGTTGCTTCGTCG
GGAATCCCTACGAATTTCAGTTCTGCACAAATTTGTCTGCCGTTTCGAAATGTGAGACCACGAA

FcpC terminator and 3' UTR

AATCACTCTGTGGTCTCAGCTTTTTTGTTACATTTACTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGATCCGAGGTTTCTATAGACTCTCTA
TAGACTCTGTTAACCTAATAGAATCAGACATACCTCTCCTGCTATTTTGTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTGAATGTTATTTTCCGCCAGGT
GTGTTAGTCGCTTGAGACCACGAAG

Ble resistance cassette

AGACAAGCAAGAATTCAAGCTTGGAGACATACCTTCAGCGTCGTGTTCACTGTCACAGTCAACTGACAGTAATCGTTGATCCGGAGAGATTCAAAATTCAATCT
GTTTGGACCTGGATAAGACACAAGAGCGACATCCTGACATGAACGCCGTAAACAGCAAATCCTGGTTGAACACGTATCCTTTTGGGGGCCTCCCGCTACGACGC
TCGCTCCAGCTGGGGCTTCCTTACTATACACAGCGCGCATATTTCACGGTTGCCAGATGTCAAGAATGGCCAAGTTGACCTCCGCCGTCCCCGTCCTCACCGCCC
GTGACGTCGCCGGAGCCGTCGAATTCTGGACCGACCGTCTCGGATTCTCCCGTGACTTCGTTGAGGACGACTTCGCCGGAGTCGTCCGTGACGACGTCACCCTC
TTCATTTCCGCCGTCCAGGACCAGGTCGTCCCCGACAACACCCTCGCCTGGGTCTGGGTCCGTGGACTCGACGAACTCTACGCCGAATGGTCCGAAGTCGTCTCC
ACCAACTTCCGTGACGCCTCCGGACCCGCCATGACCGAAATTGGAGAACAGCCCTGGGGACGTGAATTCGCCCTCCGTGACCCCGCCGGAAACTGCGTCCACTT
CGTCGCCGAAGAACAGGACTGAGCTTTTTGTTACATTTACTGACTTCAAGGAGTCGAGGAATCGATACTGCCGTCGTTTCCAGGATCCGAGGTTTCTATAGACT
CTCTATAGACTCTGTTAACCTAATAGAATCAGACATACCTCTCCTGCTATTTTGTTTTTATGAATTTGGCTTTTGCCTCTCTAGTCAGATTTGAATGTTATTTTCCGC
CAGGTGTGTTAGTCGCTACTATTGTCTTCTGCAC

Glycine linker

TTAATCACTCTGTGGTCTCAAGGTGGTGGTGGCGGTGGCGGAGGAGGTTCGTGAGACCACGAA

mVenus for MoClo positions B4-B5

GATTCATTAATCACTCTGTGGTCTCAAGGTCTGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGG
CCACAAGTTCAGCGTGAGCGGCGAGGGCGAGGGCGACGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCC
ACCCTGGTGACCACCCTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGT
GCAGGAGCGCACCATCTTCTTCAAGGACGACGGTAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGG
CATCGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCACCGCCGACAAGCAGAAGAACGGCATC
AAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGC
CCGACAACCACTACCTGAGCTACCAGAGCAAGCTGAGCAAGGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCAC
CCTGGGCATGGACGAGCTGTACAAGTAAGCTTTGAGACCACGAA

mVenus for MoClo position B5

CTGTGGTCTCATTCGGTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCATCCTGGTGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTG
AGCGGCGAGGGCGAGGGCGACGCCACCTACGGCAAGCTGACCCTGAAGCTGATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGACCACCC
TGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGTGCAGGAGCGCACCAT
CTTCTTCAAGGACGACGGTAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGA
GGACGGCAACATCCTGGGCCACAAGCTGGAGTACAACTACAACAGCCACAACGTGTACATCACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAG
ATCCGCCACAACATCGAGGACGGCGGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACC
TGAGCTACCAGAGCAAGCTGAGCAAGGACCCCAACGAGAAGCGCGACCACATGGTGCTGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGGGCATGGACG
AGCTGTACAAGTAAGCTTTGAGACCACGAA

HA tag

ATCACTGTGGTCTCATTCGTACCCCTACGACGTCCCCGACTACGCCTAGCTTTGAGACCACG

Figure A.5 – Sequences of DNA parts used in cloning in Section 5.3.

PDAT without introns

AGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATCACTCTGTGGTCTCAAATGTCGAA
ACACAAAACTCGCCACGCCAAAAGAAACACGACTAGCAATCAACAGCAAGCGTCACGAGCGTCGAAAATGGGATCCAAATCGTCGAGTCTCAAACAGGAGGAC
TCATCTTCGGAGGAAGGATCTTCTTCGGATGACGATGAGATGGTGTCGACCCGTGGATTCCTCTGTCGCCGCCGTCGTCGTCGCCGCAGTAGGAAACAATCTAC
CAGAGCAACCGTTGGATACTTGATTGGCTGCGCCTTTGCACTGTGGTTGCTCGAATTTCTTGTGGTACTCGGACTGAAACGTTTCGGTGTCCTGGATCCTCGGTTT
CGACTAGACGAGGCTGTGGCGCAGAAAGTACTTCCCCAACTGGAACAAGGTTGGAATCGACTGAACGAATCAATTCCTCCATCCTATCTCGCACAATTCGAACA
GGGTTGGAGTCGCTTGAATCAGTCTATTCCTCTGTCGTATCTGACACAAGAAACGAAGCGACCAGGTTTCCAGCTGGCTCAAAAAGGCGCAAAAGTAAACTATC
CTGTGGTCATGATGCCAGGCTTTGTCACATCGGGTCTTGAGGTTTGGGGCGGGAAAGAGTGTGCTCGCAGTCATTTTCGACAGCGCCTTTGGGCTGCTATTGGT
GGAGCAAGGAGCTTTCTGACTGACCGGGAATGCTGGAAGGAGCATATGATGTTGAATCTAGAAACCGGTGTCGATCCAGCTGACATTCGGCTACGAGCTGCCC
AAGGCTTTGAAGCGGCGGACTACTTCATGGCAAACTATTGGGTTTTCGGAAAGATCATTGAAAATTTGGCCGATTTGGGATATTCTCCATCCGAAATGACGATG
GAACCGTATGACTGGCGCCTGGCGTTTCCCTTACTCGAAAAGCGCGACGGATACCTTACTAAACTACGGCACACAATTGAAGCCATGCACAAAACTACCGGGAA
GAAGATTGTTTTAACCTCGCATTCCATGGGGGGAATGTTAGTCCACTATTTCTTTAAGTGGGTGACGACGAGTGAATCGAAGGGCGGAGGTGGGGGAGGCAAA
CACTGGGTTGACGAGCATATTCACGCTTACGTCAATATTGCCGGTTCCCATCTGGGCGTTGTGAAAGCGGCAACAGCATTGTTGTCCGGAGAAATGAGTGACAC
CATCCTCATGGGAACAATGGGCAGTATGCTGGAACAGTTCTTCGGACGACGACAGCGCCGAGATTTATGGACGACATGGGGATCGCTATGGACGATGCTTCCA
TTAGGAGGAAATTCAATCTGGGGTAGTGGTGCCGACATGTGCCTGAAACGTTCTGCGGAGGACCCAATGTGTCCAGAGGAAGGGTTGTCTCCATTAATGGTTA
TGACAGATGCAAAAGAGAACGCCAAATCCGAAATGAATGCAACGATCGACATAGATTCATCTCTCAAAGACTTCGCAGAGGACTTTGTATCATATAAATCGCAC
ACGGCCGAAGATGTCGCCGACTTTCTCATTGGATTCGGTGCCGCACGCGGGCACGAAGTTGCCAACCCAAAAATGGTATCTGTGTACGGTGACGACGAAAAAC
CGTCGTCCAGAACGTGGCACGACCCAACGCGAACTCCACTGCCTTACGCCCCAAACATGAAGATATATTGCATGTACGGTGTGGGGGTTGCAACGGAACGCGC
GTACTACTACCAGGGAAATCGCGAGGAGGCGAAAGATGAAGCAGGCCCAGGACAAGATCTTCAGGAGCCACCGGTCGTTCTGGATCCGACTGTAAACGATGC
AGAGCGCAATGTAACCCACGGGATTCGTTACAGTGACGGGGATGGATCTGTACCACTTATCTCGCTTGGGTATGTTTGCGTAGACTTGTGGAAGCGAAGAGAG
ACTGGCCTGAACCCATCGCAGACAGCCGTGCATACTAGAGAATACCATCACAGTCCGGGTTTCTGTGTAGATGATCCTATGCGTGGTGGCCCTTCGAGCTCCGA
TCATGTTGACATCCTGGGCAACATGAACATGATGGAAGATTTTTTGAGGGTGGTCAGTGACTTTGAAATCTCCGAAGTAAATAACGACAAGATTTCTAGCGACA
TAAAGCATCTTTCGGAAGAAATCACCGCCCACAAGCGGAAAACAACGAAAAAGAAAGCCTGGTTCACTTTTTGAGCTTTGAGACCACGAAGTGGCTCTTCAGTG
GACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT

AMT1 promoter and 5'UTR

ATTAATCACTCTGTGGTCTCAGGAGCCTCACCTCCAAGTCGTCAAGCCAAGATGAGTAACAGTGAGGATTTGTACTTTCCAAAGGCTTGATGACTGGCTCATCCA
AAACGATTTTTTGTAGTGGAATCCGACGTCTCTTACTATTGAAAGCAAAAGCTATTTATAGGCTATGGTTGATCGACGTAAGAACGGTCCTGTGCGAATTCTGTA
AAACGACTAGGACACAAGTTCGAGTGGTACTCTCTAATTAAGCTGGTATAGGCCTTGAGATGACTGACGCGCCACTGTTCTGCGACGAGCCAATCAAAACCTTT
CAATCTTACCTGGCCGTTGCCTTCTCTCGAACAATTAATCTGGAGTTCTGAATTCGATTTAATATTTTCTGTCTATGTCGACGCAATCTTATTGGATGACTAAGCTT
CGATAGGATTCTAGTGGCTGGAAATGTCGTTTGATTGTGCAAGCGTGTCAATTGCGATAACAGAGTTGATGATTGGCCCAGGGCAATTATATAGGATCAACCGT
CCTGTGAAGTTGGTAATGATTCTACATACCGGAAACTGCTACGCTGTGCTTCTACTAAACCTTCCCTCACGCCTCTTTGATCAGATCGGAATTTCCGTTTGCACCC
ACGCGGATCCGACCTCCGTAACTTTCATAAAACTCTGTCGACCCGTAGAGTTGTTCTCTCCTCTGTGACACACTCATCCTCTCAGCATTCACAATGTGAGACCACG
AA

Figure A.6 – Sequences of DNA parts used in cloning in Section 5.3.

Table A.6 – Primers used to amplify parts for golden gate cloning into level 0 constructs. The level 0 constructs correspond to the CDS, METE promoter and 5' UTR and DXR targeting peptide parts in Table 5.3. The column 'design' refers to the 'design variant' in Table 5.3. Aligning primers of the same design number against the amplification substrate enables identification of primer pairs. The glycine linker for Phatr3_J31492 and Phatr3_J44028 was first added by amplification from gDNA, before a second round of amplification, that used the resulting amplification product as a substrate, added a BpiI restriction site for cloning. The METE and DXR regions, Phatr3_J8860 and the Phatr3_J44028 used for designs 7, 8 and 9, were amplified from existing constructs.

Gene	Design	Direction	PCR substrate	Sequence
Phatr3_J41624	1, 2	Forward	gDNA	TTGAAGACATAATGACGGTTCAGCGGTGG
	2	Reverse	gDNA	TTGAAGACATAACCAAGGAGGGCGTACCA
	2	Forward	gDNA	TTGAAGACATACCTGCCCACCGGGAGTGTGAG
	2	Reverse	gDNA	TTGAAGACATGGTTTCAGCGGAACTATCCCGCAAG
	1	Reverse	gDNA	TTGAAGACATACCTTGGTGTCGATCGAAAGAGGGAAG
Phatr3_J31492	1, 2, 3	Forward	gDNA	TTGAAGACATAATGTCGTTAGTGGCGCAA
	1	Reverse	gDNA	TTGAAGACATACCTGGGAGGCTGAAAATGCAGTGG
	2, 3	Reverse	gDNA	TTGAAGACATCCTCCGAGCTTACCTCTGAATC
	2, 3	Forward	gDNA	TTGAAGACATGAGGACCTCGTCGTCGG
	2, 3	Reverse	gDNA	TTGAAGACATGTTTTCTTCATCCAAACAATGCAACG
	3	Reverse	gDNA	TTGAAGACATAAACCTTAAGTGGACGTTTTGAGGCCCTTGCGTGTAGT
	2	Reverse	gDNA	TTGAAGACATACCTCAGTTGCTCTTGAAACCTTCTAGC
	3	Forward	gDNA	TCCTCCTCCTCCTCCCCTAGGTCCCAGTTGCTCTTGAAACCTTCTAG
	2, 3	Forward	gDNA or Phatr3_J31492 with linker	TTGAAGACATAAACCTTAAGTGGACGTTTTGAGGC
	3	Reverse	Phatr3_J31492 with linker	TTGAAGACATACCTTCCTCCTCCTCCTCCCCTAG
Phatr3_J44028	1	Forward	gDNA	TTGAAGACATAATGAGGTTCTGCTTGCTCCTC
	1, 2, 3	Reverse	gDNA	TTGAAGACATGTTTTCTCCCATTGCGCTCTT
	1, 2, 3	Forward	gDNA	TTGAAGACATAAACTACCTCAGAGTTTGATCCA
	1	Reverse	gDNA	TTGAAGACATACCTACTATCCGTATTGTGTTTCAGC
	2, 3	Forward	gDNA	TTGAAGACATAATGAGGTTCTGCTTGCTC
	2, 3	Reverse	gDNA	TTGAAGACATCCTCAAGATGAGAACTATCCGTATTGTGTTTCAG
	2, 3	Forward	gDNA	TTGAAGACATGAGGACGCGTTCGGCAACG
	2, 3	Reverse	gDNA	TTGAAGACATCCTCTTTTTCTATGTAGTAAAAAAGTGAGGC
	2	Reverse	gDNA	TTGAAGACATACCTATATTCGGTTTGTTCCACTTTG
	2	Forward	gDNA	TTGAAGACATGAGGACTCCATCAAGACACCGA
	3	Reverse	gDNA	TCCTCCTCCTCCTCCCCTAGGTCCATATTCGGTTTGTTCCACTTTG
	3	Forward	gDNA	TTGAAGACATGAGGACTCCATCAAGACACCGAGCTGG
	3	Forward	Phatr3_J44028 with linker	TTGAAGACATGAGGACTCCATCAAGACACCGAGCT
	3	Reverse		TTGAAGACATACCTTCCTCCTCCTCCTCCCCTAG
Phatr3_J44028	4, 5, 6	Forward	Phatr3_J44028 without BsaI or BpiI restriction sites.	TTGAAGACATAATGAGGTTCTGCTTGCTC
	4, 5	Reverse		TTGAAGACATACCTCCATATTCGGTTTGTTCCACTTTGAATC
	6	Reverse		TTGAAGACATAAGCTTAATATTCGGTTTGTTCCACTTTG
Phatr3_J8860	4, 5, 6	Forward	Phatr3_J8860 without BsaI or BpiI restriction sites or introns	TTGAAGACATAATGTCGAAACACAAAACTCG
	6	Reverse		TTGAAGACATTTTCTAGATTCAACATCATATGCTCC
	6	Forward		TTGAAGACATGAAACCGGTGTCGATCCAGC
	6	Reverse		TTGAAGACATAAGCTCAAAAAGTGAACCAGGCTTTC
	4, 5	Reverse		TTGAAGACATACCTCCAAAAGTGAACCAGGCTTTC
Phatr3_J49702	4, 5, 6	Forward	cDNA	TTGAAGACATAATGATGAAGGCGTTTCG
	4, 5, 6	Reverse	cDNA	TTGAAGACATGAATCTTCCACGTTATCGTCC
	4, 5, 6	Forward	cDNA	TTGAAGACATATTCTATGTACGATCCTCTCCAAATC
	4, 5, 6	Reverse	cDNA	TTGAAGACATGGCATTCCTTCATGCAAGAAACG
	4, 5, 6	Forward	cDNA	TTGAAGACATTGCCTTCAGCTAGGATGGAATC
	4, 5, 6	Reverse	cDNA	TTGAAGACATCATCTTCAAGGTTTCTAGCTGC
	4, 5, 6	Forward	cDNA	TTGAAGACATGATGAAGCGAGCACTGATACTTCC
	4, 5, 6	Reverse	cDNA	TTGAAGACATCCAATCTTCTCGGTCAAGTAACAACG
	4, 5, 6	Forward	cDNA	TTGAAGACATTTGGGATTCAGATCAACCAATGG
	4, 5, 6	Reverse	cDNA	TTGAAGACATCTCCCACAAAGCACGCTG
	4, 5, 6	Forward	cDNA	TTGAAGACATGGAGGACGCCGGAGGAAC
	4, 5	Reverse	cDNA	TTGAAGACATACCTCCTACAATGTTCAAGTCGTCGTG
	6	Reverse	cDNA	TTGAAGACATAAGCTTATACAATGTTCAAGTCGTCGT
Phatr3_J54569	4, 5, 6	Forward	gDNA	TTGAAGACATAATGACGGACTTGTCAACAAATG
	6	Reverse	gDNA	TTGAAGACATAAGCCTATTCCTGCACATATTGCGG
	4, 5	Reverse	gDNA	TTGAAGACATACCTCCTTCCTGCACATATTGCGGATTG
Phatr3_EG01099	4, 5, 6	Forward	gDNA	TTGAAGACATAATGCTCGACAAGAATCGC
	4, 5, 6	Reverse	gDNA	TTGAAGACATCCTCAGCGGAGACTATCGAG
	4, 5, 6	Forward	gDNA	TTGAAGACATGAGGACTGGTTTGCTAGTGGC
	4, 5	Reverse	gDNA	TTGAAGACATACCTCCTTGTCGGACTGTAGACCAG
	6	Reverse	gDNA	TTGAAGACATAAGCTTATTGTCGGACTGTAGACCAG
Cre02.106400	4, 5, 6	Forward	cDNA	TTGAAGACATAATGCCAAAATCGGAGCGGCGTCTGGTGCTATGGC
	4, 5	Reverse	cDNA	TTGAAGACATACCTCCGGCGGCCAGTGCGGCCCAG
	6	Reverse	cDNA	TTGAAGACATAAGCTCAGGCGGCCAGTGCGGCCCAG
Cre03.g.193500	4, 5, 6	Forward	cDNA	TTGAAGACATAATGAGCCAGCTATTGTCGCAC
	4, 5, 6	Reverse	cDNA	TTGAAGACATGACGGCCTGGGGGTCCTCG
	4, 5, 6	Forward	cDNA	TTGAAGACATCGTCTCGATGCCGCCGGGC
	4, 5	Reverse	cDNA	TTGAAGACATACCTCCCCGGCAGGCGTGTGGGTC
	6	Reverse	cDNA	TTGAAGACATAAGCTCACCGGCAGGCGTGTGG
METE promoter with 5' UTR	7, 8, 9	Forward	METE promoter with 5' UTR construct	TTGAAGACATGGAGCCGAATAACTGAGTG
	7, 8, 9	Reverse		TTGAAGACATATGGTGTTGATCGAGATAGCTTTTTGGAAG
DXR targeting peptide	7, 8, 9	Forward	DXR targeting peptide construct	TTGAAGACATCCATGCGGTTAACGGAACTGTC
	7, 8, 9	Reverse		TTGAAGACATCATTCCCTGGAAATTGCTGCCTTCC