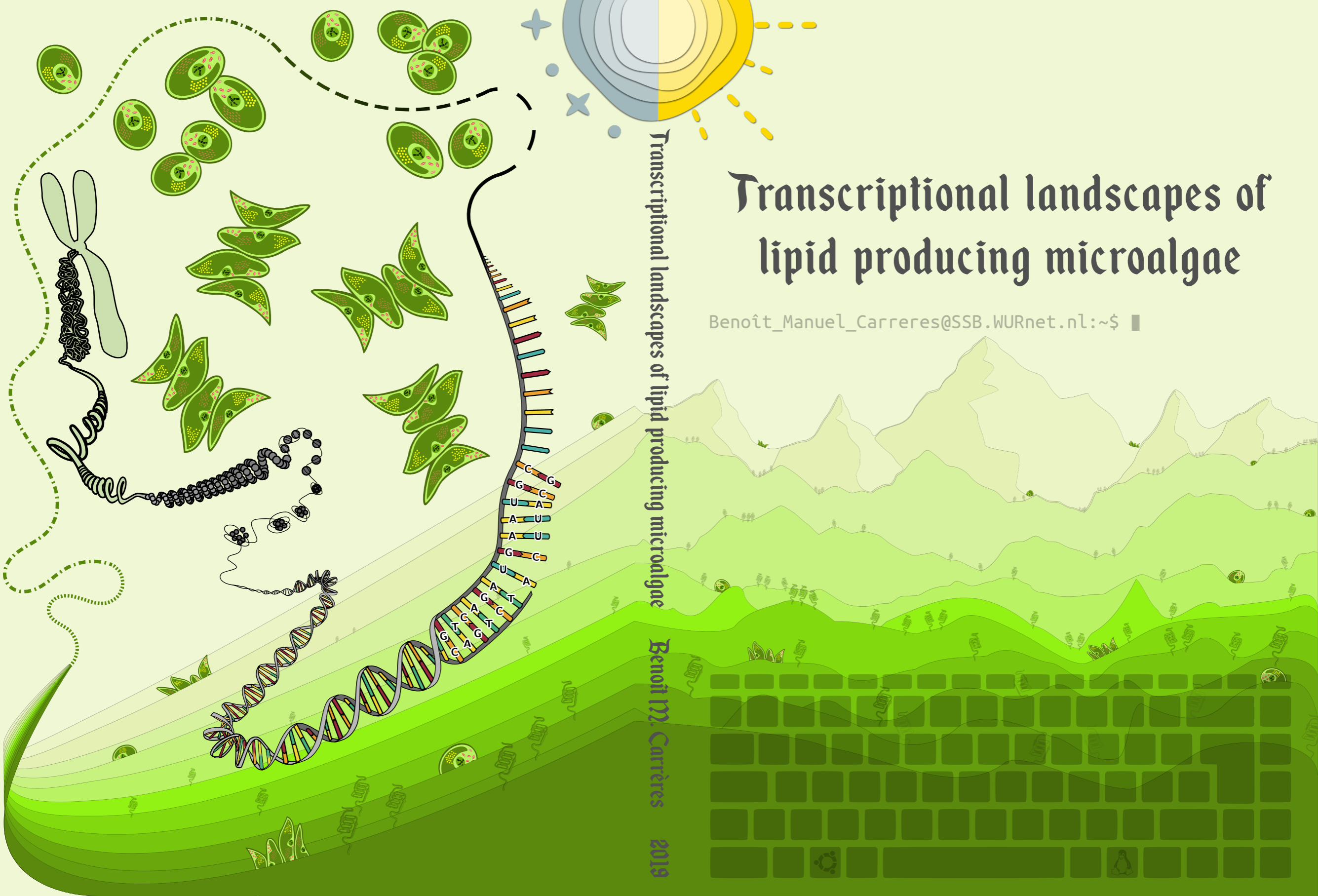


Transcriptional landscapes of lipid producing microalgae

Benoît_Manuel_Carreres@SSB.WURnet.nl:~\$ █

Transcriptional landscapes of lipid producing microalgae

Benoît M. Carrères 2019



Transcriptional landscapes of lipid producing microalgae

Benoît Manuel Carrères

Thesis committee

Promotors

Prof. Dr Vitor A. P. Martins dos Santos
Professor of Systems and Synthetic Biology
Wageningen University & Research

Prof. Dr René H. Wijffels
Professor of Bioprocess Engineering
Wageningen University & Research

Co-promotors

Dr Peter J. Schaap
Associate professor, Systems and Synthetic Biology
Wageningen University & Research

Dr Dirk E. Martens
Associate professor, Bioprocess Engineering
Wageningen University & Research

Other members

Prof. Dr Alison Smith, University of Cambridge
Prof. Dr. Dick de Ridder, Wageningen University & Research
Dr Aalt D.J. van Dijk, Wageningen University & Research
Dr Gabino Sanchez-Perez, Genetwister, Wageningen

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Transcriptional landscapes of lipid producing microalgae

Benoît Manuel Carrères

Thesis

submitted in fulfilment of the requirements for the degree of doctor

at Wageningen University

by the authority of the Rector Magnificus,

Prof. Dr A.P.J. Mol,

in the presence of the

Thesis Committee appointed by the Academic Board

to be defended in public

on Wednesday 20 November 2019

at 1.30 p.m in the Aula.

Benoît Manuel Carrères

Transcriptional landscapes of lipid producing microalgae,

256 pages

PhD thesis, Wageningen University, Wageningen, The Netherlands (2019)

With references, with summary in English

ISBN: 978-94-6395-183-8

DOI: 10.18174/504726

Contents

Chapter 1	Introduction and thesis outline	7
Chapter 2	Algal omics: the functional annotation challenge	21
Chapter 3	A systems approach to explore triacylglycerol production in <i>Neochloris oleoabundans</i>	39
Chapter 4	<i>Neochloris oleoabundans</i> is worth its salt: Transcriptomic analysis under salt and nitrogen stress	77
Chapter 5	Draft Genome Sequence of the Oleaginous Green Alga <i>Tetradismus obliquus</i> UTEX 393	103
Chapter 6	The diurnal transcriptional landscape of the microalga <i>Tetradismus obliquus</i>	107
Chapter 7	Evaluation of diurnal responses of <i>Tetradismus obliquus</i> under nitrogen limitation	145
Chapter 8	Discussion	181
	Summary	211
	Citations	219
	Acknowledgments	241
	Curriculum vitae	251
	Publication list	253
	Overview of completed training activities	255

Chapter 1

Introduction and thesis outline

1.1 Microalgae to produce biofuels

Microalgae are unicellular microscopic and photosynthetic organisms. Due to their long evolutionary history, their biodiversity is very broad, both phenotypically and genetically [1]. They can live individually as single cells, in clusters or communities, or in symbiosis with bacteria. They are found in all sorts of environments ranging from freshwater to hyper-saline conditions, nutrient rich and nutrient limited conditions, and in environments that differ largely in temperature and pH [2]. Just like other micro-organisms, their importance within the planet's ecosystem was overlooked for a long time. Among the many important roles algae play and have played on the planet are that they are in the origin of the planet's oxygen rich atmosphere that paved the road for the evolution of other eukaryotes, and they are currently still accounted for producing half of the atmospheric oxygen [3]. Microalgae have been the main contributors of the fossil fuels we use in everyday life, and they are now seen as a hope to reduce our dependence on these same fossil fuels [3–7]. Indeed, while they have been studied for their outstanding capacity to clean water from excessive content in nitrogen, phosphorus, and metals, the interest has now shifted towards biotechnological production of valuable products [8, 9]. Microalgae, and notably green algae, are valuable candidates for diverse biotechnological applications because they are fast growing photoautotrophic organisms that require minimal nutrition, do not require arable land to grow and thus do not compete with feed production, have superior areal productivity, and are more efficient at converting photons into biomass than plants [10–12]. Many microalgae are categorized as oleaginous because they can accumulate a large proportion of their dry-weight in fatty-acids (30-60%) [13–15]. The two species studied in this thesis are the oleaginous microalgae named *Ettlia oleoabundans*, formally known as *Neochloris oleoabundans*, and *Tetradesmus obliquus*, formally known as *Scenedesmus obliquus* and *Acutodesmus obliquus*. Like other oleaginous microalgae, they can produce triacylglycerides (TAG), which can be extracted and easily converted into biodiesel [4, 7, 11]. Besides, TAGs are free of unwanted chemical constituents unlike the phospholipids and the glycolipids, leading to purer biofuel [14]. The main interests here are the renewable and sustainable aspects of growing microalgae, leading to a carbon-neutral biofuel production. While promising, the current state of biofuel production using microalgae is not economically feasible [16]. Besides biofuels, other products with higher

added value can be produced, such as pigments, unsaturated fatty acids, bioplastics and proteins [6, 17–20].

Despite of its potential, the costs of lipid production in algae are currently still too high to make biofuel production with microalgae economically feasible considering the current oil prices. It was recently estimated that the cost of microalgal biomass could be as low as 0.50 euros per Kg, and at this price level, biofuel production would be profitable [16]. For the moment, commercial bioproduction from microalgae is still limited to high-value products. Applications for commodities such as bulk chemicals and biofuels require substantial improvements, such as the increase of the total solar-to-product conversion efficiency. One direct way to achieve this is to optimize the external conditions and to maintain these during scale-up. The production scale-up difficulties are studied at the Algae Production And Research Center (AlgaePARC) [21, 22]. Another way to increase efficiency is to modify the metabolism of the algae to reach a higher solar efficiency under production conditions. For this, a better understanding of algal metabolism and underlying mechanisms is needed. This is precisely the topic of this thesis, in which I analyze the internal phenotypes of microalgae in response to a number of growth conditions relevant to industrial applications. Specifically, I studied the effect of light-dark cycles, salt water, intense light, and nitrogen limited or depleted conditions.

1.2 Limiting factors for profitable bioproduction

Microalgae can be grown in very diverse type of bioreactors, each coming with their own advantages and disadvantages. The most used reactors are race-way ponds, diverse configurations of tubular systems, rectangular panels, and plastic foil based bags. Relying on freely available sunlight is an important factor to minimize costs. However, outdoor cultivation is subject to variations in temperature, light intensity, and a light-dark cycle. During the dark period, cellular respiration occurs and can lead up to 25% loss in daily produced biomass [23]. Commonly, this loss is due to the usage of a storage compound like starch that accumulates during the light period. Thus, starch is a transient energy and carbon storage compound that is used in the night for maintenance purposes and for preparation of the cells for the next day. As a result, mutants that cannot produce starch show a decrease in the efficiency in diurnal growth condition [24]. Lipids, another possible storage compound, are hardly used during the night, even by starchless mutants.

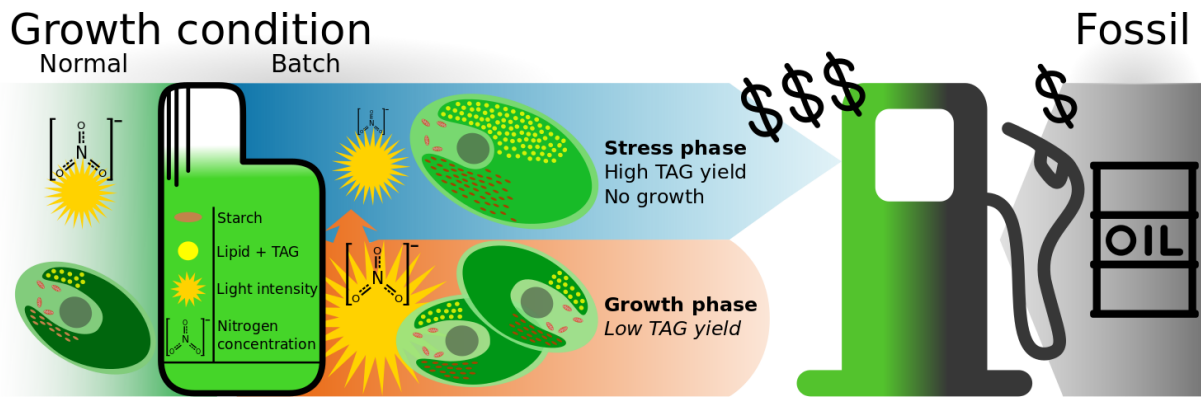


Figure 1.1: Batch cultivation to increase TAG yield for biodiesel production. Batch cultivation is done in two-step process with a “growth phase” and a “stress phase”. The stress phase is induced by nitrogen depletion, resulting in lipids and TAGs accumulation. Current solutions are not optimal for biodiesel production at a competitive price.

In order to achieve high concentrations of lipids and TAG (30-60%), microalgae need to be exposed to adverse growth conditions. Commonly this is done by applying nitrogen limitation or complete depletion of nitrogen from the medium, which results in overflow of the photon energy into carbon-based molecules like starch and TAG [25]. Starch and TAG are the main forms of carbon and energy storage and serve as an electron sink when the synthesis of reducing equivalents in the photosystems exceeds the requirement for cell growth [14]. This mechanism has led to a simple two-step batch strategy for TAG production, starting with a growth phase under replete conditions, followed by a stress phase during which usually nitrogen is depleted as displayed in Figure 1.1 [26]. However, upon complete depletion of nitrogen the efficiency of photo-energy conversion decreases during the accumulation of TAG, leading eventually to a decrease in the yield of TAG on photons. Therefore, instead of complete starvation, a less stressful condition of continuous nitrogen limitation was proposed [27]. In such a system there is simultaneous growth and accumulation of TAG for longer periods of time. Furthermore, constant photosynthetic efficiency can be maintained and the amount of down time for cleaning and restarting the system is much shorter. Additionally, such a condition allows more flexibility to adjust the system with the changing outdoor conditions. A recent comparison against continuous production, batch production displayed higher photon-to-TAG yield for both wild-type and the starchless mutant (*slm1*). Even though the starchless mutant had 20% lower photo-to-TAG yield, it accumulated 65% more TAG as part of its dry weight than the wild-type. Both strains were studied in this thesis in Chapters Error: Reference source not found and Error: Reference source not found [28].

For large-scale production of biofuels using microalgae, salt water is the preferred medium basis over fresh water, since fresh water is or may become a limiting resource. Waste water is a possible alternative to salt water, since it would be detoxified all the while providing nutrients for growth [29–31]. However, not all TAG-producing microalgae species can survive high salinity and oxidative stress caused by salt or waste water and if they can survive they may suffer from a strong decrease in TAG yield on light [32].

One way to reduce the production costs of biofuels is to increase the TAG yield on light. This can be done through the development of processes that are more efficient like described in the previous paragraph, by changing algal metabolism or a combination of these. With respect to algal metabolism there are two aspects that can be improved, being maintaining a high photosynthetic efficiency during the stress phase and preventing the formation of competing metabolites. The loss of photosynthetic efficiency is among others related to the breakdown of the photosystems and chlorophyll. However, many things are still unknown on the mechanisms that cause a decrease in photosynthetic efficiency during the stress phase and more research is needed before the photosynthetic efficiency can be improved in a rational way.

The most important competing metabolite in many green algae is starch. During the stress phase (Figure 1.1) starch accumulation may account for nearly half of the energy stored with the other half being directed towards TAG synthesis [33–35]. Removing the ability to synthesize such a competing metabolite could lead to a significant increase in the yield of TAG on light.

Last but not least, once the batch is ready for harvest, the microalgae cells should be collected and broken open in order to extract its content [4]. However, it would be most efficient to not destroy the batch and keep the cells in a production state. For this, the optimal solution would be to have TAG directly excreted by the cells, allowing continuous production without having to wait for growth and accumulation. The lipids excretion functionality could be genetically engineered based on observation from microalgae that already do it in nature, but on a more suitable organism regarding the aforementioned limiting factors.

1.3 Strain improvement

For strain improvement, a number of techniques exist with their own advantages and disadvantages, ranging from selective breeding, forced evolution, to metabolic engineering. Selective breeding is the most common method used to improve for example plants. The sexual reproduction of plants is an important factor that allows crossing breeds and combine wanted phenotypes. Forced evolution can be achieved by physically altering the DNA of an organism using UV light and selecting wanted phenotypes. Both methods are a mimic of natural processes, known factors of evolution and biological diversity, and as a consequence, are easier accepted by general public than metabolic engineering methods. As an example, forced evolution via random mutation (UV metagenesis) was used to generate the starchless mutant *slm1* of *Tetrademus obliquus* studied in this thesis. The metabolic engineering methods offer the best potential for strain improvement but can be very complex to design and require a detailed knowledge of the metabolism and regulation of the organism. Furthermore, the metabolic engineering approach requires the availability of a genetic modification toolbox for the specific organism. For example, CRIPR-CAS9 is currently the state-of-the-art tool for genome editing, and it could be used to replace, add or remove genes, effectively allowing to modify the internal circuitry of the metabolism. While these synthetic biology methods can be easily applied in the better characterized organisms, they need to be developed for the specific microalgae. As stated above, detailed knowledge of algal metabolism is required to find the gene targets to be engineered. This is a crucial step because trial and error in synthetic biology is laborious and time consuming in comparison to other methods. For this constraint based metabolic models can be used to predict the possible outcomes of metabolic engineering strategies thus increasing the chances of success.

1.4 Metabolic modeling

To help improving the knowledge from an organism's internal phenotype, constraint-based metabolic models (CBMM) can be of great help. Metabolic flux balance analysis performed using CBMM can provide valuable information on the physiological state of the cells [36]. With sufficient knowledge on the metabolic reactions occurring in the organism, it is possible to model the system as a network of these reactions and estimate their rates, based on measured rates of nutrient uptake (including light) and product formation (including biomass

and storage compounds). In addition, the models can be used to make predictions on, for example, the maximum yield of TAG on light and the associated flux distributions. As demonstrated in the Error: Reference source not found, predictions can be made on the potential gain in yield of TAG on light upon the removal of a metabolic pathway responsible for the synthesis of a competing metabolite. The development of CBMMs for algae faces a number of obstacles. These include the difficulty of identifying correct protein functions predicting protein localization in the cell, and of ascertaining transporters for all metabolites . We have shown that it is currently difficult to identify gene and protein functions in algae due to a lack of experimentally validated sequences of microalgae (Error: Reference source not found). Nevertheless, the number of these validated proteins is steadily growing. In parallel, better methods have been developed during the time of this work to identify protein functions, and so the quality of computational annotation has also increased for microalgae. The second problem is the localization of the different proteins in the different algal compartments like the cytosol, chloroplast, mitochondrion and peroxisome. This is an important piece of information that is currently hard to obtain for microalgae. This is due to a lack of experimental data and consequently a lack of reliable training sets specifically for microalgae. For example, the quality of the predictions made by PredAlgo, a multi-subcellular localization prediction tool dedicated to microalgae, is directly limited by the quality and the quantity of experimental data used to train its neural network [37].

1.5 Using RNA-seq to fill the knowledge gap

To do rational process design or apply rational metabolic engineering approaches, more knowledge is required on the algal physiological in responses to changing environmental conditions such as the influence of day-night cycles [38]. To learn more about this, there is a choice between different high-throughput methods, ranging from genomics to proteomics and fluxomics (Figure 8.1). Using transcriptomics to study the dynamics of the transcriptome landscape is a convenient middle-ground, as it can be easily obtained by RNA sequencing that is relatively cheap using second generation high throughput sequencing technology. Gene transcripts can be reconstructed *de novo* from RNA-seq reads using read assembly methods or through alignment of RNA-seq reads with high quality genome sequences. The large quantities of RNA-seq reads favors accurate expression estimation, and finally, the information obtained can be extrapolated to proteomics. However, to provide a functional context, the transcripts

should be functionally annotated, which is done by sequence alignment with previously annotated organisms and inference of homology. The main issue with microalgae is the large evolutionary distance between the currently studied microalgal species and even larger distance with its closest and well-enough studied organisms, being land plants. While RNA sequencing allows to quantify how features of the genome are expressed, it also allows to identify what protein sequence could be translated from it. While the quantity of RNA is not directly correlated to the absolute quantity of a protein in a cell, the relative difference in quantity between samples (variation) is very valuable information. The identified proteins sequences can be coding for (part of) an enzyme, and the absolute and relative change in expression can also be used to extrapolate whether there is more or less of the specific protein present. Although the variation in protein abundance can be a good indicator of relative enzymatic activity, some other factors can have a strong influence on the enzymatic activity, which is discussed in more detail in Error: Reference source not found. In this thesis, it was assumed that the relative changes in gene expression reflect relative changes in enzyme activity, unless there was complementary information.

1.6 Thesis aim and outline

The aim of this thesis is to obtain a better understanding of the physiology of lipid producing microalgae under conditions relevant for large-scale production by analyzing the transcriptional landscape.

To analyze the transcriptome, it is necessary to know the functions of the proteins translated from the expressed genes. In contrast to model organisms like the yeast *Saccharomyces cerevisiae* or the plant *Arabidopsis thaliana* whose genomes have been deeply annotated and studied for a long time, microalgae are far from that state, and even nowadays, microalgal species with a sequenced genome are not very common. For both organisms studied in this thesis, *Ettlia oleoabundans* and *Tetrademus obliquus*, it was necessary to annotate the genes and transcripts. To functionally annotate a gene, most methods rely on sequence similarities to identify the closest gene in known organisms. Green algae are a difficult case due to the large genetic distance between the currently studied species and the closest reference organisms being the land plants. The Error: Reference source not found discusses about the particular problem of protein annotation in microalgae. Here, we also discuss the prediction of protein

localization within the different subcellular compartment in microalgae. Finally, we stress the importance of a large-scale wet-lab efforts for a few selected microalgae in order to provide a solid foundation for the application of computational methods.

To obtain more insights into the metabolism of *Ettlia oleoabundans* in Error: Reference source not found, a constraint-based metabolic model of *Ettlia oleoabundans* was built around the central carbon metabolism. This model was built based on the knowledge of central metabolism of algae at that moment and was cross-checked with a *de novo* assembled annotated transcriptome. Experiments in controlled turbidostat were conducted in different combinations of light intensity and nitrogen supply. The measurements from the experimental conditions were used as constraints on the inputs and outputs of the model, effectively allowing us to estimate the metabolic flux distributions. In addition, RNA samples from the different experimental conditions were sequenced and analyzed. These data were used to validate the model structure, to correlate expression levels with flux distributions, and to get a better understanding of the effect of light and nutrient conditions on algal physiology.

Water is a precious resource and fresh water may not be a sustainable resource for microalgae bioproduction. In contrast, salt-water is abundantly available and would be cheaper to supply. Therefore, it is important to understand how biotechnologically interesting microalgae deal with a salt water environment under production conditions. In Error: Reference source not found, we studied how the microalgae *Ettlia oleoabundans* deals with high salinity conditions under nitrogen replete (growth) and nitrogen deplete (TAG accumulation) conditions using a transcriptomics approach. This oleaginous microalgae was chosen as a model alga, since it can accumulate large amounts of TAG and can grow in both fresh and salt-water. Experiments were done in fresh water and salt water, both in combination with nitrogen replete and nitrogen deplete conditions. In addition to the transcriptome, we analyzed the biomass composition including TAG and starch content and used the data obtained to study the different salt water resistance mechanisms.

Although the TAG content that can be reached in *Ettlia oleoabundans* is high, the volumetric TAG productivity in *Tetradismus obliquus* was evaluated to be clearly higher, while reaching the same TAG content. This was mainly due to the ability of *T. obliquus* to maintain photosynthetic efficiency for a longer time longer during the nitrogen depletion phase [13]. Therefore, it was decided to switch to *T. obliquus* as a model organism. To obtain an idea of the capabilities of *T. obliquus* and to make transcriptome experiments easier to analyze, the genome of *T. obliquus* was sequenced. In Error: Reference source not found, the sequencing of the genome of *T. obliquus* is presented. The assembly approach was unconventional as two

different methods were used, combining the higher coverage from one method with the precision of the other method. In this way, a high quality of draft assembly was obtained.

Production using microalgae will in many cases occur outdoors using sunlight. Consequently the algae will be exposed to the naturally occurring day night cycle. To better understand the effect of these day night cycles, in Error: Reference source not found and Error: Reference source not found, the transcriptional response of algae to diurnal cycles was studied under nitrogen replete conditions and nitrogen limiting conditions for the wild type and a mutant that can not synthesize starch. In Error: Reference source not found, hourly samples of RNA of *Tetradismus obliquus* UTEX 393 were taken from a turbidostat culture operated under nitrogen replete conditions over a diurnal cycle of 16 hour light and 8 hours dark, to obtain more insight in in the transcriptional response towards diurnal cycles. In addition, to understand the effects of a lack of starch, the major transient energy storage, we sequenced samples of the starchless mutant *slm1* that were collected every three hours under the same conditions. At the same time, samples were collected and measurements of the biochemical composition of biomass and the specific light absorption rate were performed [24]. The genome features were annotated using more than 30 RNA-seq time samples obtained in this study, the SAPP tool, a newly developed semantic framework for FAIR genome annotation [39], was extended for transcriptome analysis is described in the discussion Error: Reference source not found.

In Error: Reference source not found a similar experimental approach was taken, where samples were taken for biochemical and transcriptome analysis every 3 hours from a turbidostat culture operated at the same diurnal cycle of 16 hours light and 8 hours dark, but this time in nitrogen limiting conditions, resulting in TAG accumulation. Again this was done for the wild type and the starchless mutant. The transcriptional landscape and biochemical data are compared to the nitrogen replete condition in Error: Reference source not found, to evaluate the effect of nitrogen limitation in general and study how the lack of starch is affecting TAG accumulation under nitrogen limiting conditions.

In Error: Reference source not found, the results of this thesis are discussed. The importance of analyzing the transcriptome of microalgae is explained. Pros and cons of current microalgal annotation methods are discussed. Basic numbers of the *T. obliquus* annotation results are compared to the current status in UniProtKB. Particular points from the transcriptomics data are discussed, notably the expression of the *slm1*'s mutant gene, and the interests from using single-cell technologies. Suggestions are then made to improve the experimental conditions and the photobioreactors setups. The efforts made in the thesis to generate and store the data

according to the FAIR principles are explained. Finally, using the knowledge acquired during this thesis, suggestions are made to improve the growth conditions and to improve TAG production with diverse metabolic engineering strategies.

Chapter 2

Algal omics: the functional annotation challenge

This chapter is published as:

Maarten J.M.F. Reijnders[§], Benoit M. Carreres[§], Peter J. Schaap

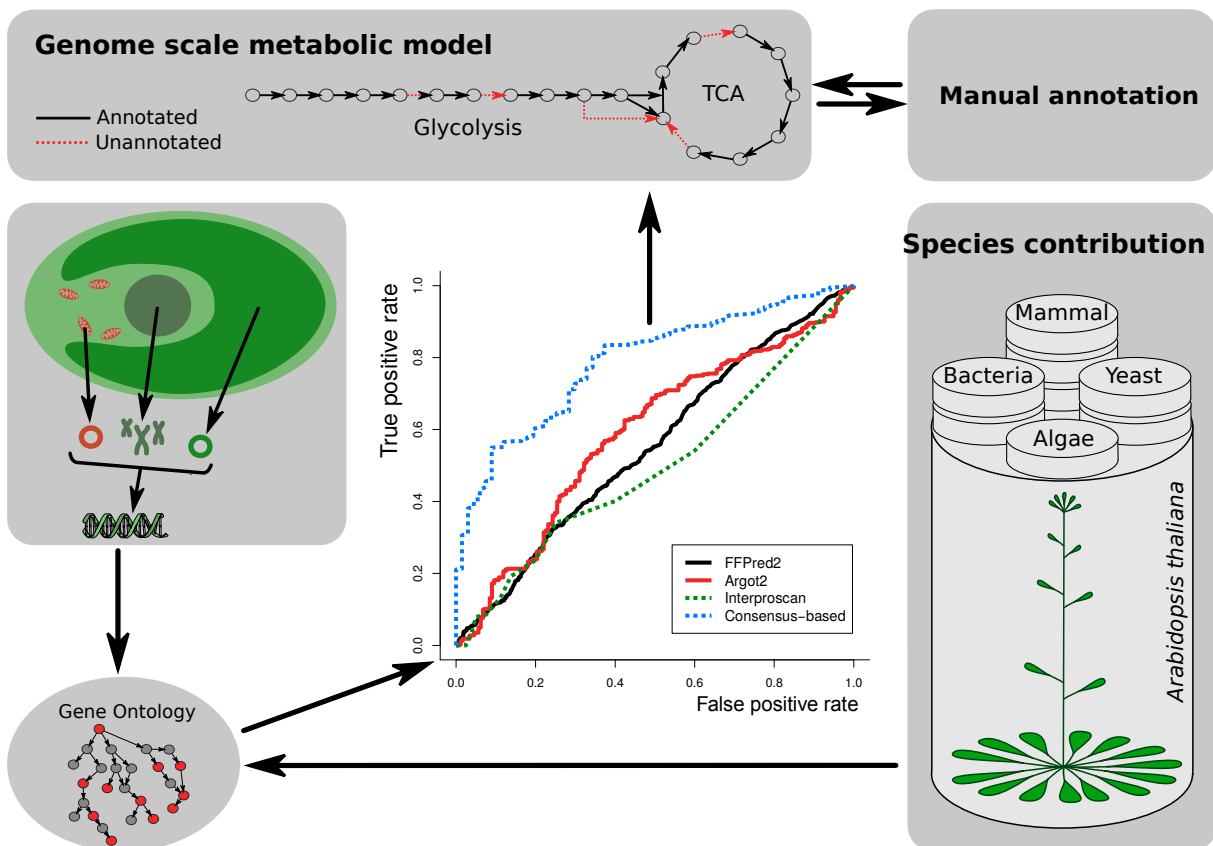
Algal Omics: The Functional Annotation Challenge.

Current Biotechnology (2015) 4(4) 457-463

§ Authors contributed equally

ABSTRACT

Background: To fully exploit the potential of microalgae as commercial green hosts, the scientific community has to improve their understanding of these organisms from a systems biology perspective. Compared to other model organisms, our genomic knowledge of the microalgae model species *Chlamydomonas reinhardtii* is very limited. Currently, almost 90% of the functional annotated proteins of *C. reinhardtii* and of other microalgal proteins are homologs of *Arabidopsis thaliana* proteins, which suggests that for the most part only the metabolic core conserved between these species is properly annotated. **Objective:** This review highlights how proteins outside of this core can be annotated by applying publically available tools and methods. These include the use of novel state-of-the-art prediction tools, combinations of these tools, and the use of metabolic modeling-assisted functional annotation. Furthermore, we discuss the need for data on the subcellular location of microalgal proteins. Finally, some remaining bottlenecks regarding functional annotation of microalgal proteins are discussed. **Conclusion:** We conclude that both large dry-lab and wet-lab efforts are required to generate reliable functional annotations of microalgae.



2.1 Introduction

Microalgae are considered as promising organisms for a bio-based economy and unlocking their power potentially holds solutions for achieving global sustainability. In order to cope with some of the most demanding nutritional and energetic challenges of the future, research has focused on the renewable oil that can be extracted in significant amounts from these microalgae to create sustainable consumer products. However, compared to the more traditional sources, economically interesting molecules, such as triacylglycerides and polysaccharides, are currently not produced at a cost competitive rate [40]. To increase the yield it is important to understand the genomic makeup of microalgae. More specifically, it is important to understand microalgae as biological systems at such a level of detail that mathematical models can be developed for these cell factories. These models can then predict the most optimal conditions for growth and production of interesting compounds, and can guide genetic precision engineering of these cell factories [41]. Such models, often in the form of genome-scale metabolic models, require a thorough functional annotation of the proteins encoded by the genomes.

In today's age of biology, computational annotation of protein functions is of vital importance. Sample throughput of the classical biochemical and genetic methods is simply too low to be considered as an alternative. However, there is large phylogenetic distance between microalgae and well characterized (model) species [42], and this distance hampers standard computational methods for genome annotation. Many of the popular computational methods for function prediction try to infer homology by calculating sequence-based statistical similarity scores with proteins of known function [43, 44]. This works fairly well for a comparison between a well-studied model organism with a large set of proteins validated by biochemical and genetic methods, such as *Arabidopsis thaliana* and *Escherichia coli*, and close by plants and bacteria, but the efficiency of a sequence similarity based annotation method decreases drastically when it is used between a group of species with little experimentally validated proteins, or when it is used for species that have a large phylogenetic distance to a well-studied homologous species. The most studied microalgae species *Chlamydomonas reinhardtii* became known early on as an excellent model species for microalgae because of its genetic amenability [45], but two decades later our genetic knowledge of this species still trails far behind that of other model species. Currently only some 150 proteins are characterized by direct biochemical methods.

Furthermore, due to the large phylogenetic distance to the closest well studied model organism [3], *Arabidopsis thaliana*, only the most conserved genes are properly computationally annotated. Subsequently, only this limited set can be used as a reference set in sequence similarity-based methods to annotate other algal species of interest. As a result, only the conserved core metabolism of various microalgae is functionally annotated with a high level of confidence. Most of these microalgae were, however, selected for their ability to produce interesting and novel compounds [46]. To truly exploit microalgae for a bio-based economy, it is therefore important to know the function of the proteins that are not part of this metabolic core. By gaining more detailed genomic knowledge we will be able to produce more accurate algae specific genome-scale metabolic models. This allows for the prediction of biomass composition and conditions for optimal growth rates of microalgae, as well as for diversifying between the unique characteristics and capabilities of different microalgae and strains. Recently, alternative methods to functionally annotate microalgae have been described [47]. In this review we assess the current state of microalgal functional annotation, standardly used methods and discuss some alternative methods and workflow based on novel annotation tools that are currently available to the scientific community. Finally, we address some bottlenecks that currently cannot be solved by computational methods.

2.2 Available data

From early on, *Chlamydomonas reinhardtii* was the only microalgae species that was extensively studied on a molecular scale. This species was first proposed a model organism for algal genetics in 2001 [45], and a draft genome sequence was available in 2003 [48]. However, due to “unusual challenges” in generating a high-quality genome [48], the genome was only published as late as 2007 [49]. For *C. reinhardtii* to serve as a model species to which other algae can be compared, it is important that many algal-specific protein functions and other key functions are based on experimental evidence, and not only inferred from electronic annotations. In the UniProt database (<http://www.uniprot.org>) [50] there are currently 148 proteins from *C. reinhardtii* with an experimentally validated function, compared to 5,766 for *Arabidopsis thaliana*, and 3,255 for *Escherichia coli*.

The electronic annotation of *C. reinhardtii* is an ongoing process, and so far out of a total of 15,000 proteins there are roughly 7,000 proteins available in the UniProt database that are functionally annotated with at least one GO term. However, when we take the reviewed

proteins of *Chlamydomonas* into account, there are only 299 proteins with a high-quality annotation available.

2.2.1 The state of microalgal annotations

For microalgae, the inability to obtain a high quality functional annotation for the majority of the proteins seems to be a returning trend. In Table 2.1 we bring some recently annotated microalgae, and show how deep they are annotated. All of these microalgae were annotated using standard homology-based methods [51–59]. For each of these microalgae roughly half of the proteins lack any form of functional annotation documented in their UniProt database entry [49]. This is likely the direct result of a lack of phylogenetically close well-annotated model species. That does not necessarily mean that the annotations obtained are unspecific or inaccurate, but it does imply that accurate electronic annotations are retrieved only for highly conserved proteins common amongst many microalgae.

Table 2.1: Number of computationally annotated and experimentally verified proteins in microalgae and reference species.

Species	Total Proteins	Annotated Proteins	Experimentally Validated Proteins
<i>Arabidopsis thaliana</i>	52178	32354	5766
<i>Escherichia coli</i> (strain K12)	6061	5399	2619
<i>Saccharomyces cerevisiae</i> (strain ATCC)	6729	5919	4162
<i>Chlamydomonas reinhardtii</i>	15.152	7.385	148
<i>Auxenochlorella protothecoides</i>	7.193	4.072	0
<i>Bathycoccus</i> sp.	7.889	4.203	0
<i>Chlorella variabilis</i>	9879	5409	0
<i>Coccomyxa subellipsodae</i>	9.802	5.27	0
<i>Micromonas pusilla</i>	10279	4782	0
<i>Nannochloropsis gaditana</i>	15361	7419	0
<i>Ostreococcus tauri</i>	7912	4555	2
<i>Ostreococcus lucimarinus</i>	7407	4326	0
<i>Volvox carteri</i>	14.833	6.292	18

*Model species are in bold. Data were taken from UniProt (<http://www.uniprot.org>) on August 1 2015. Total proteins were taken by querying the species taxonomy, total annotated proteins were retrieved by including the query for proteins to have a GO term using any assertion method, and experimentally validated proteins were obtained by only including proteins that have a GO term using any experimental assertion.

2.2.2 Diversity of Microalgal Annotations

The diversity of microalgae makes them unique biological reservoirs for bioprospecting, and it would be interesting to see how a good quality functional annotation can contribute to this process. By taking the Gene Ontology (GO) annotations from the microalgae species presented in Table 2.1 into account, and by checking the occurrence of these terms in the nearest well-studied model species *A. thaliana*, we can get hints about the diversity and origin of microalgal proteins annotations (Figure 2.1). The figure shows that 88% of the GO terms assigned to microalgal proteins also occur in *A. thaliana*. Overall, 85% of the microalgal GO terms are used in the annotation of *C. reinhardtii* protein, but only 7% of the specific *Chlamydomonas* GO terms do not occur in *A. thaliana*.

For all other species the amount of mapped GO terms is far less than in *C. reinhardtii*, showing an even less diverse annotation. With such little amount of microalgal GO terms, that are not also mapped to *Arabidopsis*, it becomes clear that the current annotation of microalgae largely describes the conserved core-metabolism shared between eukaryotic photosynthetic organisms, and as such will only provide a small contribution to the process of bioprospecting. To summarize, microalgal experimental protein data is very limited, and due to the large phylogenetic distance to the better-characterized model species large amounts of proteins remain unannotated (Table 2.1). To circumvent these bottlenecks it is necessary to use more advanced annotation methods.

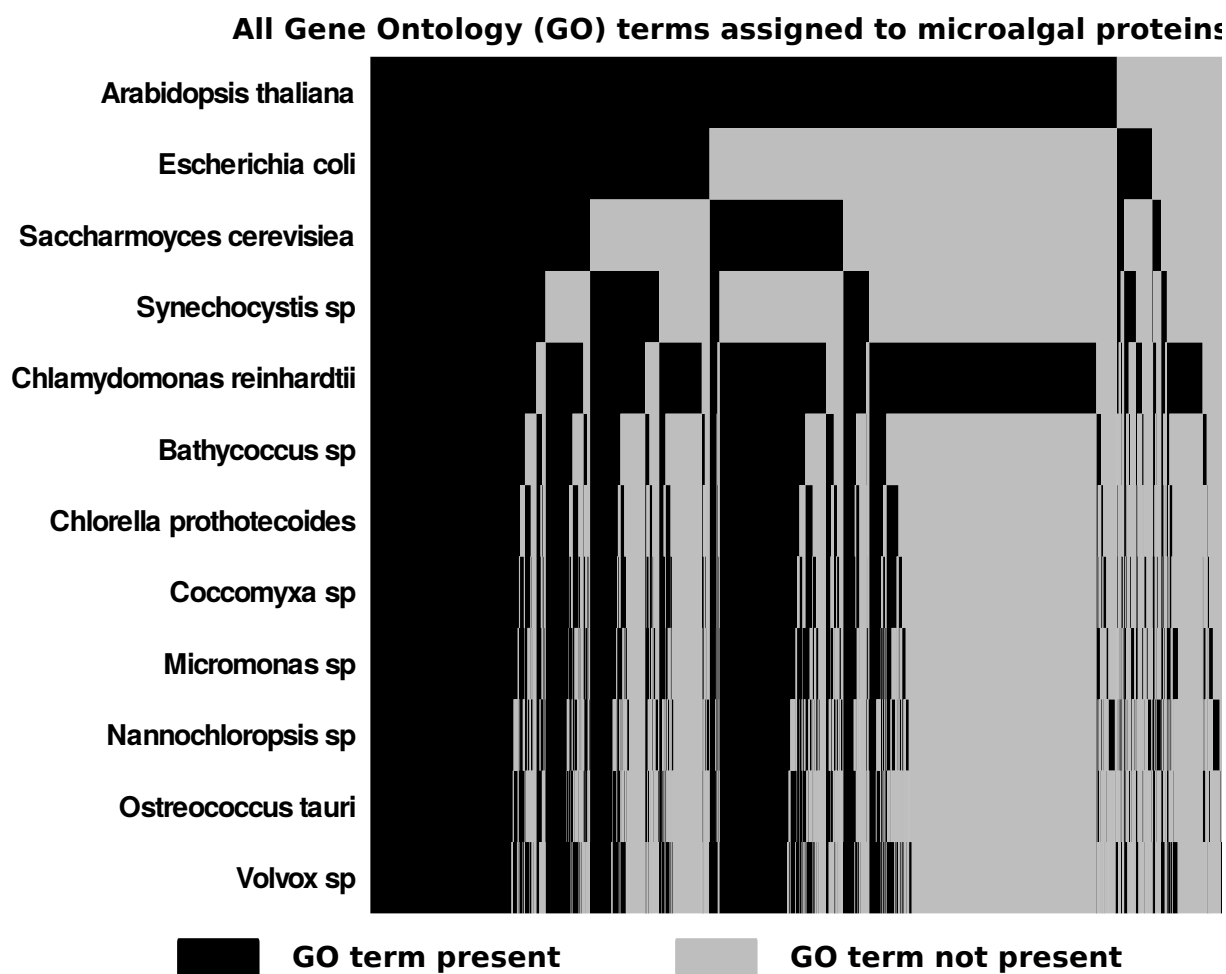


Figure 2.1: Heatmap showing the presence of microalgal GO term annotations in microalgae and in model species. Y-axis, species presented in Table 2.1; X-axis, GO terms annotated to microalgae, sorted by their assignment to at least one protein in descending order of the species list.

2.3 Improved functional annotations

2.3.1 Keeping up to date with annotation methods – the CAFA experiment

One way to improve the functional annotation of microalgae, is by applying the latest state-of-the-art tools. The classical sequence similarity-based annotation methods often remain the first method of choice because of their success in the past. However, functional annotation of

proteins is a hot topic in the bioinformatics community and new tools are published every year. To keep track of these tools and how well they perform, the Critical Assessment of Protein Function Annotation (CAFA) experiment attempts to rank them according to their performance [60]. The first edition showed that as many as 33 new methods outperformed the standard BLAST-based method. This can be explained by the fact that many of these tools apply sequence similarity-based predictions in different ways, for example by using one-to-many homology-based annotations, or by using context-aware principles, as described by [47]. An example of a new method that uses a one-to-many approach is Argot2 [61], which combines BLAST results with sequence analysis using profile hidden Markov Models (HMM) and subsequently compares predictions using a semantic similarity approach. An example of a context-aware approach is FFPred2 [62], which attempts to transfer functional annotations from known (human) proteins to unknown proteins with similar biophysical attributes. The CAFA experiment provides a good ranked overview of state-of-the-art annotation tools. However, most of these tools in part still rely on primary sequence similarity, and the structure and context based alternatives require extensive training sets. Thus, although these tools will most likely produce more reliable results than the classical mainstream functional annotation tools, they will still be unable to annotate many microalgal proteins.

2.3.2 Consensus-based annotation

An alternative way to improve the results of protein functional annotation is by using a set of complementary tools and combining individual predictions in a statistical solid manner. For example, by combining FFPred2 with Argot2 we combine a one-to-many homology-based annotation method with a context-aware annotation method. This can be further complemented with a protein domain homology-based transfer of annotation approach using InterProScan [44]. If we would then take the GO term predictions of each of these methods and compare predicted GO terms using a semantic similarity approach as applied in Argot2, we obtain a comparison between predictions of each of these methods, and the specificity for each predicted GO term. By then applying a machine-learning algorithm such as Random Forest we are able to reassess the validity of each of these predictions.

As a test-case we applied this method to a test-set of all new microalgal SwissProt (<http://www.uniprot.org>) proteins entries between the 1 st of July 2014 and 1 st of July 2015, using the UniProtKB [50], Uniref90 [63], and Pfam [64] databases from before the 1 st of July

2014 as reference. The experiment was set up with double 10-fold cross validation. Ten data sets were generated with 90% of the predicted GO terms assigned to the training set, and 10% as test-set. The training set was used to train a Random Forest model on the input using 10-fold cross-validation. In this way each final predicted GO term has no influence on the model used to predict these GO term, eliminating overfitting. The accuracy of this method largely improves over that of FFPred2, Argot2, and InterProScan (Figure 2.2A). A test-set of non-algal proteins (Figure 2.2B) was used to compare results with the algal data set. There is a noticeable difference in the performance of FFPred2, Argot2 and InterProScan. For microalgae, the latter two showed lower prediction accuracy.

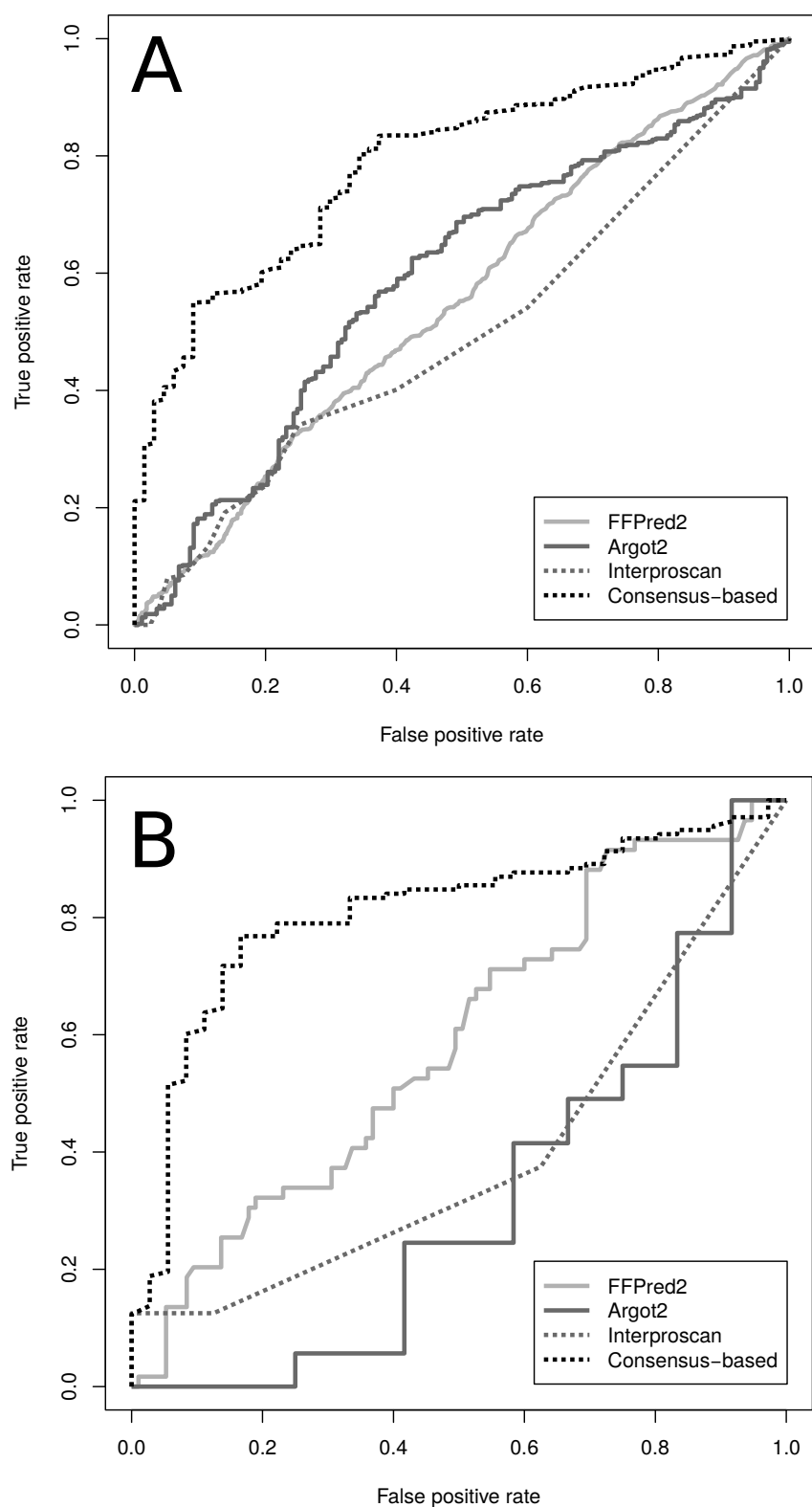


Figure 2.2: Receiver Operator Curve of three protein function prediction methods, as well as the consensus-based method that combines these. The test-sets used consisted of protein entries not present in UniProt (<http://www.uniprot.org>) before the 1st of July 2014. The test-set of (A) 246 reviewed microalgal proteins comprising 540 assigned GO terms, and (B) consisted of 2429 reviewed proteins comprising 11701 assigned GO terms. Predictions were done using database versions from before the 1st of July 2014.

2.3.3 Functional annotation with Hidden Markov Models

Profile Hidden Markov Models (HMM's) provide a statistical description of a sequence family consensus [65]. Effectively a profile HMM turns a multiple sequence alignment of a specific protein family into scoring system that takes into account position-dependent amino acid distributions and position-dependent insertion and deletion gap penalties, which makes this technique suitable for searching remote homologs. To obtain the best model while keeping a high specificity, it is important to build it from experimentally validated proteins only. Because there is no database that contains HMM's built from experimentally validated proteins only, for each specific function a new HMM has to be built, which makes this method not easily applicable for high throughput annotation. Moreover, when selecting only experimentally validated proteins of a specific function, the amount of experimentally verified sequences available can often be too limited.

2.3.4 Validation of matching proteins via 3D structure

An extension to sequence similarity techniques is 3D structure prediction. Once the best matching proteins are found, they can be used in homology-based modeling approaches. Homology-based modeling uses a reference protein as a base to predict the 3D structure of the protein of interest. The two 3D structures can then be scored for overall quality, and be compared. Several tools are capable to estimate the tertiary and quaternary structure of proteins in such a way. One example is SWISS-MODEL (<http://www.swissmodel.expasy.org>), a web-based tool aimed to provide easy access to predict protein 3D structure from its amino acid sequence, assisted by homology modeling techniques as explained above [66]. Regardless of the automated 3D modeling limitations, scores such as QMEAN, coverage, and identity, can provide an interesting addition to estimate the similarity between the protein of interest and the reference protein.

2.3.5 Model-assisted annotation

A genome scale metabolic reconstruction aims to integrate biochemical metabolic pathways in a single network and provides a structured platform to correspond metabolic genes with

metabolic pathways [67]. As an alternative to laborious manual metabolic model construction, tools such as SEED [68] and Pathway Tools [69] are capable of automatically generating metabolic maps from pathway databases and enzyme annotations. While these tools often provide only a basic overview of an organism's metabolic capabilities, as the topology and breadth of the network is largely dependent on available data, even for microalgae these models can provide valuable insights. For instance, an orphan reaction in metabolic pathways can simply be due to a missed or a too broad annotation. With this information in mind it becomes feasible to use more elaborate, case-by-case, manual annotation methods to close these gaps. A simple first step could be to take the GO term specific for the particular protein and link this to similar but more generic parental GO term(s). Proteins annotated with these less specific GO term(s) are then considered to be promising candidates for the missing GO term, and should be re-evaluated. One way to do this is by building a HMM based on UniProt proteins that are experimentally validated to have the specific GO term. This statistical model can then be applied to the candidate protein selection, and in this way we might be able to identify the correct protein.

2.3.6 Subcellular localization

Protein localization prediction is an important part of a proteins functional annotation. If two proteins involved in the same reaction are functionally assigned to a different subcellular compartment, the reaction cannot occur. On the other hand microalgae are known to possess multiple isofunctional proteins that essentially perform the same reaction but in different subcellular compartments. This information is crucial for the more elaborate compartmentalized genome scale models of microalgae. One robust way of figuring out the subcellular location of (isofunctional) proteins is by performing subcellular proteomics [70], but this is often technically difficult, expensive, and time consuming. The UniProt database currently contains 529 reviewed microalgal proteins with a subcellular location annotated, of which 54 are experimentally validated. Therefore it is necessary to computationally predict the subcellular location of proteins. For this purpose several tools are available, such as: Argot2 [61] TargetP [71], SignalP [72], PSORTb [73], and PredAlgo [37]. However, with the exception of PredAlgo, most of these tools are trained with different types of species in mind, resulting in predictions that do not take into account the specific cellular arrangements and compartments in microalgal species.

PredAlgo is a predictor specifically trained for microalgae, using a *C. reinhardtii* based training set of 79 chloroplast, 39 mitochondrial, 39 secretory pathway, and 89 cytosol proteins. It shows good prediction results for Chlamydomonas proteins and closely related microalgal proteins. However, for other more distantly related microalgal species predicting subcellular localization is difficult due to their polyphyletic nature. It is believed that different endosymbiotic events happened in parallel, forming the first microalgae [74]. This caused a difference in the arrangements of cellular compartments, or even different types of cellular compartments. Therefore, PredAlgo may not be accurate in predicting protein locations for microalgae not related to *C. reinhardtii*. To circumvent this, the PredAlgo algorithm will have to be trained with proteins from additional microalgal clades. Alternatively, results from multiple predictors possibly can be combined as described above. Finally it should be noted that PredAlgo only predicts to which compartment a protein is targeted. If the compartment where translation of the protein takes place is unknown it is still not possible to know the final location, or to which membrane it is targeted. Therefore, compartmentalized omics data is needed to accurately predict the final subcellular location of a protein. The UniProt database contains 8 Chlorophyta proteome sets based on genome sequencing data that are fully annotated. Additionally, GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) harbors many chloroplast and mitochondrial datasets that can be used for this purpose [75]. These datasets could be compared to assess the potential impact of the parallel endosymbiotic events, which in turn can be taken into account in cellular location predictions.

In conclusion by using consensus-based protein function prediction, and/or model-assisted annotations, many improvements can be made in functional annotation of microalgal proteins. Nevertheless, there will always be a set of species-specific proteins that will remain unannotated using computational methods.

2.4 Remaining bottlenecks

2.4.1 Unsupervised computational annotation can lead to error propagation

The GO project has become the standard way of annotating proteins [76]. All major databases use these terms for the documentation of protein functions and cellular locations. GO terms are accompanied with an evidence code, stating how a gene was assigned to a GO term [77]. In most cases the evidence code is “Inferred from Electronic Annotation”, meaning that an unsupervised computational method was used. Such annotations are error-prone. Furthermore, with an exponential-increasing amount of sequencing data being generated every day, the amount of unsupervised electronically assigned GO terms also increases exponentially. To illustrate the problem, Schaid et al. showed that already in 2010, 50% of the 200,000 human protein GO term assignments were done electronically. Consequently, these gene annotations were likely to contain a number of errors [78]. If such an electronic annotation is done using, for example, a standard BLAST based transfer of annotation method and proteins that also have their GO terms electronically assigned were used as a reference, this can easily lead to error propagation of GO term assignments. Recording the provenance of unsupervised annotations to GO terms is therefore essential. Several annotation tools are scoring GO term annotations based on the GO structure and evidence codes [79]. The evidence ontology (ECO) [77] provides more descriptive evidence-based annotation to proteins in UniProt database by describing, for example, evidence types, methods and data curation. A detailed provenance can help to obtain more precise evidence scores than is possible with the standard evidence codes.

2.4.2 Orphan reactions

One of the direct results of a small amount of experimentally validated proteins is a large amount of orphan reactions. These are reactions catalyzed by enzymes of which it is assumed that they must occur, for instance through phenotypic analysis or because they bridge a gap in a metabolic pathway, but which do not have an encoding gene assigned. The Orphan Enzymes

Project (<http://www.orphanenzymes.org>) is attempting to link sequences to such Orphan Enzymes, and a similar effort should be made regarding microalgal enzymes.

2.4.3 The lack of identified microalgal specific protein domains

An effective way of assigning GO terms to proteins is by linking GO terms to protein domains and searching proteins for the presence of these domains. However, as can be seen in Figure 2.2B (InterProScan results), microalgae proteins show a low level of sequence similarity to domains available in the PFAM database (<http://www.pfam.xfam.org>), which suggests that microalgae have accumulated many novel domains that are not yet identified by the scientific community. To start to identify these novel domains it might be useful to develop an algae specific domainome by routinely performing large-scale comparative genomics between all available microalgal genomics data, as was done with bacteria [80]. Recurring patterns can then be assigned to specific domains with presently unknown function. If specific domains keep recurring in proteins associated with specific traits, these domains can be linked to a function.

2.5 Concluding remarks

Systems biology approaches to unlock the potential power of microalgae are seriously hampered by lack of genomic knowledge. Genome annotations of recently sequenced species still heavily depend on sequence similarity based functional annotation methods, which are less suitable for species that have no close by well-studied and annotated homologous species. As a result almost 90% of microalgal functional protein annotations is still for the most part describing the metabolic core shared between algae and plant species. The application of novel state-of-the-art annotation methods, as well as approaches that combine multiple methods, may result in a more accurate and more diverse functional annotation. Genome scale modeling approaches could additionally help in identifying metabolic gaps, which can then be looked at more thoroughly. However, for microalgae to fulfill their promise as a biosynthetic host it is important to overcome at least some of the annotation bottlenecks that are not solvable by computational methods. We therefore suggest that a large-scale wet lab effort focused on a

number of selected microalgal reference species is essential. This would provide the computational methods with larger, more diverse set of reference genes, and would allow computational annotations methods to quickly tap into the promising biological reservoirs of industrially interesting algal species [81].

Chapter 3

A systems approach to explore triacylglycerol
production in *Neochloris oleoabundans*

Benoit M. Carreres[§], Anne J. Klok[§], Maria Suarez-Diez, L. de Jaeger, Mark H.J. Sturme, Packo
P. Lamers, René H. Wijffels, Vitor A.P. Martins dos Santos, Peter J. Schaap, Dirk E. Martens

§ Authors contributed equally

ABSTRACT

The microalga *Neochloris oleoabundans* accumulates triacylglycerol (TAG) upon nitrogen limitation. TAG can be used in food products, feedstocks or easily converted into an effective and clean biofuel. Commercialization of algae-derived TAG currently requires a further reduction in production costs, which can be reached by optimizing TAG yield on light.

A constraint based metabolic model of central (carbon) metabolism was constructed, parameterized, and supported using turbidostat experiments and transcriptome data. The model was used to study changes in metabolic flux distribution and carbon and energy partitioning as a result of nitrogen limitation and changes in light intensity. Under nitrogen limitation, *N. oleoabundans* redistributes photosynthetic energy supply by reducing linear electron transport and by redirecting carbon and energy towards starch and TAG synthesis. The metabolic model calculates a maximum TAG yield of *N. oleoabundans* on light of $1.06 \text{ g (mol photons)}^{-1}$, more than 3 times the current experimental yield under optimal conditions. Nitrogen limitation has direct effect on expression of nitrogen dependent pathways, while high light impacted gene expression of a wider number of pathways. The combination of high light and nitrogen limitation showed more changes in gene expression than the sum of the inductions alone. We found that under nitrogen limitation, starch accumulation correlates with down-regulation of starch degrading enzymes and that TAG accumulation correlates with up-regulation of few reactions in lipids and TAG biosynthesis pathways. Geranylgeranyl diphosphate reductase was found to have an opposite regulation than the other reactions from chlorophyll, terpenoid and carotenoid biosynthesis pathways.

Integration of predicted metabolic fluxes and expression data increased our understanding of the effect of the studied environmental stresses on microalgal metabolism. Our metabolic model shows that TAG yield on light can be more efficiently improved by optimizing photosynthetic conversion than by blocking competing pathways. Geranylgeranyl diphosphate reductase was identified as a potential regulator for photosynthetic capability. It complements the fine-tuning of chlorophyll levels from synthesis and degradation. Finally, we were able to identify some key reactions that could be targeted to improve TAG yield, by not only specifically increasing the flux within the lipids and TAG pathways, but also potentially redirect carbons from other pathways.

3.1 Background

The oleaginous microalga *Neochloris oleoabundans* is able to accumulate relative high amounts of triacylglycerol (TAG). Adverse growth conditions as for example nitrogen limitation are often applied to microalgae to induce production of industrially interesting compounds such as β -carotene or TAG [14, 82]. When nitrogen is limiting or absent, protein synthesis and growth are reduced or stopped, resulting in reduced demand for energy and electrons. Since light supply is maintained, this results in an excess of absorbed light energy, which in turn results in a surplus of electrons and reductive potential in the photosystems [83]. Accumulation of carbon-rich compounds, such as starch and TAG, provides a sink for these electrons [84] and can be seen as a strategy to safely channel away redundant electrons, while maintaining considerable carbon fixation rates.

There are three possible fates for the excitation energy resulting from light absorption by chlorophyll being: (i) photochemical transformation into energy carriers (NADPH and ATP), (ii) dissipation via fluorescence or (iii) dissipation as heat (non-photochemical quenching, NPQ). Nitrogen limitation causes a considerable decline in photosynthetic efficiency [85]. Thus, the excess of energy is not only balanced by channeling redundant electrons towards starch and highly reduced TAG but also by reducing electron generation itself by energy dissipation through heat production and fluorescence [86, 87].

High energy dissipation rates result in low areal TAG productivity and increased production costs [7]. Moreover, the synthesis of competing metabolites, such as a starch, direct additional energy away from TAG synthesis [88] again lowering the TAG productivity. This in particular poses a problem in industrial production of compounds for bulk applications, such as TAG intended for biofuels [89]. Knowledge on the interplay between photosynthesis, nitrogen assimilation, carbon fixation, and energy dissipation, is limited and would provide useful insights into how microalgae distribute energy and carbon fluxes that could be used to improve areal productivity of TAG.

The aim of this study is to obtain more insight into the molecular mechanism behind the loss of photosynthetic efficiency and the partitioning of carbon and energy towards TAG under nitrogen limiting conditions, using transcriptome analysis in combination with metabolic modeling.

Transcriptome and metabolic data were collected in steady-state conditions in controlled photobioreactors operated in turbidostat mode. Two different light absorption rates and several nitrogen supply rates were used resulting in 9 different steady states as described in [27] and depicted in Figure 3.1. Comparison of gene expression was done for four conditions representing a low and high light absorption rate in combination with a nitrogen replete and nitrogen limiting conditions.

A constraint based metabolic model (CBM) was build based on general biochemical considerations and on a previously published CBM of *Chlamydomonas reinhardtii* [90]. *De novo* transcript assembly and annotation provided evidence for the transcript-protein-reaction associations underlying the CBM and to some extent for their compartmentalization. For this, better transcriptome coverage was obtained by combining 16 sequenced samples obtained from five ([A][C][E][F] and [I], Figure 3.1) experiments described in [27] and in Table 3.1 and eight RNA-seq datasets generated in our lab [91]. Functional annotation was performed using tools able to bridge the great genetic distance with available databases [44, 91, 92].

In this work, it is shown how integration of expression data and metabolic flux analysis can be used to evaluate several scenarios, predict optimal TAG production on available light, find yield bottlenecks, and important reactions or mechanisms. These insights can be used to re-direct carbon and energy fluxes towards TAG, leading to a higher areal productivity and thus a lower cost price of microalgal TAG.

Table 3.1: Biomass composition given as weight fraction (A) and balanced rates (B) obtained for each experimental condition. Culture A and E in the gray shaded columns are the nitrogen replete cultures that are light limited at respectively low and high light. Figure 3.1 presents a graphical overview of all these experimental conditions. The condition balance were defined based on published model [88].

A^a	Biomass composition (%w/w)								
Experiment	A*	B	C*	D	E*	F	G	H	I*
Ash	6.2	5.4	4.9	3.7	6.7	5.9	5.7	4.3	3.9
TAG	1.5	3.7	7.5	12.4	1.7	2.7	4.3	7.0	8.7
Starch	14.2	23.4	29.1	34.7	25.2	28.2	30.9	39.5	42.4
Functional Biomass	78.1	67.5	58.6	49.2	66.3	63.3	59.1	49.3	45.0
Protein	38.9	33.5	29.2	23.0	35.5	30.7	28.1	23.0	20.3
Carbohydrate	16.8	16.3	15.0	16.9	11.6	13.9	15.5	15.5	15.6
Membrane lipid	10.5	10.2	9.3	6.8	8.8	8.7	7.7	6.7	6.8
Chlorophyll	1.9	1.2	1.0	0.7	1.5	0.9	0.7	0.5	0.4
Carotenoid	0.3	0.2	0.2	0.1	0.3	0.2	0.2	0.2	0.1
RNA	8.2	4.7	2.7	1.1	7.6	7.6	5.7	2.5	1.3
DNA	1.5	1.3	1.1	0.6	1.0	1.3	1.2	0.8	0.5
B	Balanced rates in mmol (g protein)⁻¹ h⁻¹								
Experiment	A*	B	C*	D	E*	F	G	H	I*
r_r	-124	-112	-107	-129	-271	-259	-259	-272	-304
$r_{NO_3^-}$	-0.83	-0.36	-0.21	-0.16	-1.34	-	-0.60	-0.41	-0.28
r_X^b	0.71	0.33	0.20	0.17	1.08	0.64	0.54	0.41	0.30
$r_{TAG} (\cdot 10^3)$	2.45	4.15	5.25	7.76	4.82	5.62	7.76	10.3	9.55
r_{STARCH}	0.13	0.14	0.11	0.12	0.38	0.32	0.30	0.31	0.25
r_{CO_2}	-5.48	-3.22	-2.24	-2.24	-9.48	-	-5.69	-5.06	-3.97
r_{O_2}	7.47	4.19	2.88	2.81	12.64	6.41	8.27	7.24	6.28

^a $N \geq 3$ for each measurement, with the exception of ash, which consisted of a single measurement.

^b specific functional biomass production rate

*: Experimental conditions used for comparison in transcriptome and flux analysis.

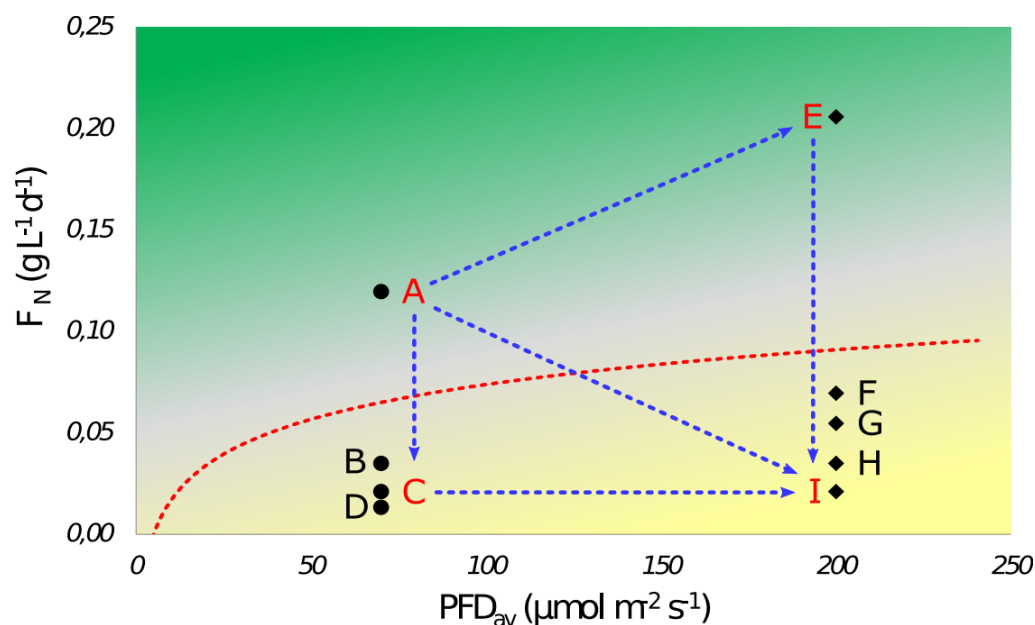


Figure 3.1: Experimental conditions. Average photon flux densities (PFD_{av}) of 70 (LL) and 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (HL) were combined with several nitrogen supply rates (F_N), resulting in nine different growth conditions ([A] – [I]). The red dotted line represents the settings where light supply rate and nitrogen supply rate are in balance. With F_N above this line (green area) the culture will be light limited, below the line it will be nitrogen limited (yellow area), where F_N is equal to r_N . The blue dashed arrows represent the five pair-wise comparisons performed when calculating differential flux and differential expression. The arrow-head side represents the nominator of the division. For example, $FC=2$ in $[E]/[A]$ comparison indicates a two-fold higher expression at [E] (HL) than at [A] (LL). The condition balances were defined as described in a previous article [88].

3.2 Methods

3.2.1 Reactor set-up and experiments

To test the effect of nitrogen limitation on the energy and carbon fluxes in oleaginous microalgae, *Neochloris oleoabundans* (UTEX 1185, The culture collection of Algae, university of Texas, Austin) was grown in a turbidostat operated flat panel photobioreactor under constant red-light conditions, with a nitrogen feed separate from the diluting medium. This type of system allows independent control of light absorption and nitrogen supply rates to the culture [27]. Setting the nitrogen supply rate below the nitrogen demand of the culture, which is dictated by the fixed light absorption rate, will result in nitrogen limited growth conditions. Light absorption in the turbidostat was controlled by setting a fixed photon flux density entering the system (PFD_{in} in $\mu\text{mol m}^{-2} \text{s}^{-1}$). The photon flux density leaving the system (PFD_{out} in $\mu\text{mol m}^{-2} \text{s}^{-1}$) was measured by a light sensor connected to the rear glass panel of the reactor and compared to a set point. When a value below this set point was detected, a diluting flow of nitrogen free medium was supplied to the system. This control loop ensured a constant PFD_{out} and combined with the fixed PFD_{in} , a constant light absorption (r_{γ} in $\text{mmol}^{-1} \text{L}^{-1} \text{h}^{-1}$) over the entire culture depth (δ_{pbr} in m):

$$r_{\gamma} = \frac{(PFD_{in} - PFD_{out}) \cdot 3600 \cdot 10^{-6}}{\delta_{pbr}} \quad (\text{Equation 2})$$

A nitrate containing feed was constantly supplied independently of the nitrogen free diluting medium. By setting the flow rate of this feed, the nitrate supply to the system ($F_{NO_3^-,in}$ in $\text{mmol L}^{-1} \text{h}^{-1}$) was controlled and could be set above (N replete) or below (N limited) the required nitrogen consumption rate of the culture. The nitrogen consumption rate $r_{NO_3^-}$ ($\text{mmol L}^{-1} \text{h}^{-1}$) was calculated as follows:

$$r_{NO_3^-} = F_{NO_3^-,in} - D \cdot C_{NO_3^-,out} \quad (\text{Equation 3})$$

With D (h^{-1}) as the dilution rate that is the result of the culture growth, and $C_{NO_3^-,out}$ (mmol L^{-1}) as the residual nitrate concentration in the reactor. Under N replete conditions, residual nitrate is present in the system, while under N limited conditions, all nitrate is consumed and $r_{NO_3^-}$ equals $F_{NO_3^-,in}$.

Two light absorption rates (Low light; LL and High light; HL) were combined with several nitrate supply rates, resulting in nine steady states ($[A]$ - $[I]$, see Table 3.2 and Figure 3.1).

Steady state was defined as a constant biomass concentration, cell characteristics, off gas concentrations and dilution rate during at least 6 consecutive days.

More information about photobioreactor specifications and operation, as well as a detailed description of culture media and conditions can be found in [27].

*Table 3.2: Combinations of several nitrate supply rates with two light absorption rates resulted in 9 steady states. The nitrogen replete reference conditions [A] and [E], for low light and high light respectively, are indicated in gray. *: Experimental conditions used in transcriptome analysis.*

Experiment	Low light conditions				High light conditions				
	A*	B	C*	D	E*	F	G	H	I*
PFD _{in} ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	193	193	193	193	460	460	460	460	460
PFD _{out} ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	18	16	16	17	69	70	70	71	70
$r\Upsilon$ ($\text{mmol L}^{-1} \text{h}^{-1}$)	32	32	32	32	70	70	70	70	70
FNO ₃ -.in ($\text{mmol L}^{-1} \text{h}^{-1}$)	0.36	0.10	0.06	0.04	0.61	0.21	0.16	0.11	0.06
rNO ₃ -. ($\text{mmol L}^{-1} \text{h}^{-1}$)	0.19	0.10	0.06	0.04	0.27	0.21	0.16	0.11	0.06

3.2.2 Biomass analysis and additional measurements

To monitor the culture, samples were taken every day at the same hour. These samples were analyzed for dry weight biomass concentration, optical density, cell concentration, cell size and residual nitrate in the supernatant, as described by [90].

The growth rate of the culture was equal to the steady state dilution rate, and was calculated by dividing the amount of overflow produced in 24h (V_{24} in L h^{-1}) by the reactor volume (V_{pbr} in L):

$$\mu = D = \frac{V_{24}}{V_{pbr}} \quad (\text{Equation 4})$$

The composition of the gas mixture leaving the system was analyzed using a prima dB mass spectrometer (ThermoFisher, USA). Additionally, a reference measurement was done by analyzing the composition of the in-going gas mixture. With these measurements, the volumetric oxygen production rate (r_{O_2} , $\text{mmol L}^{-1} \text{h}^{-1}$) and carbon dioxide consumption rate (r_{CO_2} , $\text{mmol L}^{-1} \text{h}^{-1}$) rates could be calculated:

$$r_{O_2} = F_{gas.out} \cdot x_{O_2.out} - F_{gas.in} \cdot x_{O_2.in} \quad (\text{Equation 5})$$

$$r_{CO_2} = F_{gas.out} \cdot x_{CO_2.out} - F_{gas.in} \cdot x_{CO_2.in} \quad (\text{Equation 6})$$

With $x_{O_2.out}$, $x_{O_2.in}$, $x_{CO_2.out}$ and $x_{CO_2.in}$ as the molar fractions of O_2 and CO_2 (-), and $F_{gas.in}$ and $F_{gas.out}$ as the total molar gas flow rates entering and leaving the system ($mmol L^{-1} h^{-1}$), respectively. To account for any difference between $F_{gas.in}$ and $F_{gas.out}$ due to CO_2 consumption and O_2 production inside the photobioreactor, the nitrogen gas feed flow ($F_{gas.in} \cdot x_{N_2.in}$) was used as a reference to determine $F_{gas.out}$:

$$F_{gas.out} = \frac{x_{N_2.in}}{x_{N_2.out}} \cdot F_{gas.in} \quad (\text{Equation 7})$$

When steady state was reached, biomass was collected on ice over a period of exactly 24h. The obtained biomass was centrifuged, washed and lyophilized as described by [90]. The supernatant of the 24h samples was evaluated for TOC (Total Organic Carbon), which was measured as the difference between total carbon (TC) and inorganic carbon (IC), using a TOC-VCPH/TOC-VCPN Total Organic Carbon Analyzer (Shimadzu corps, Kyoto, Japan). This determination of dissolved organic carbon, allowed for the correction of the measured biomass concentrations for any biomass lost due to cell lysis, assuming that all organic material in solution had the same elemental composition as suspended biomass.

The freeze-dried algae powder collected in each steady state was used for all biomass characterizations. The amino acid content per gram dry weigh was determined by Ansynth Service BV (Berkel en Roodenrijs, The Netherlands) for 20 individual amino acids, as described by [90]. The total amount of protein in biomass was assumed to be the sum of the water corrected mass of these 20 amino acids. The total fatty acid content and TAG content of dry biomass, and the fatty acid composition of each lipid pool was determined as described by [27]. Carbohydrates were measured by treating the freeze dried algae powder with a phenol solution and concentrated sulfuric acid, according to [93] and [94]. The absorbance of the resulting solution was measured at 483nm. Glucose was used as a standard. Starch was measured in the lyophilized biomass samples using a commercial kit (Total Starch (AA/AMG), Megazyme International, Bray, Ireland).

The carotenoid and porphyrin (i.e. chlorophyll) fractions were quantified using RP-UHPLC as previously described by [95], with some minor adaptations: Lutein (96%) was purchased from Carotenature (Lupsingen, Switzerland). Chlorophyll-a (90.0%) was purchased

from Wako Pure Chemical Industries (Osaka, Japan). RP-UHPLC identification and quantification of carotenoids and porphyrins was performed as previously described by Weesepeol et al. (2013) with minor adaptations. The elution program was started from 25% (v/v) A / 75% (v/v) B and then as follows: to 15 min – linear gradient to 100% (v/v) B; to 22.5 min – isocratic at 100% (v/v) B; to 29.5 min – linear gradient to 87.5% (v/v) B / 12.5% (v/v) C; to 31.5 min – linear gradient to 70% (v/v) B / 30% (v/v) C; to 41.5 min – linear gradient to 100% (v/v) C; to 42.5 min - isocratic at 100% (v/v) C. After 42.5 min, the eluent composition returned to its initial composition in 7.5 min, followed by an equilibration phase of 2.5 min. Detection wavelengths for UV-Vis were set at 450 nm (carotenoids c) and 660 nm (porphyrins). For quantification, lutein and chlorophyll-a were used. Approximately 1 mg of lutein was first dissolved in dichloromethane, and subsequently diluted with 4 volumes of ethanol. Chlorophyll-a was first dissolved in ethyl acetate, and subsequently diluted with 3 volumes of 90% (v/v) aqueous acetone. Standard concentrations were measured prior to RP-UHPLC analysis, using their respective absorption coefficients ($A_{1\%1\text{cm}}$): 2550 $\text{g L}^{-1} \text{cm}^{-1}$ (445 nm, 100% ethanol) for lutein and 887 $\text{g L}^{-1} \text{cm}^{-1}$ (664 nm, 90% (v/v) aqueous acetone) for chlorophyll-a (Roy et al. (2011)). PDA calibration was performed using five different concentrations of the standards injected in duplicate. For this calibration, the response of the all-trans and cis form of the carotenoid standards was considered equal for quantification. The detector was found to be linear for lutein between 0.08 and 8.04 $\mu\text{g mL}^{-1}$ with a R^2 of 0.996. For chlorophyll-a, the detector was linear between 0.06 and 6.01 $\mu\text{g mL}^{-1}$ with a R^2 of 0.968. Carotenoids were expressed as lutein equivalents using the lutein calibration curve. The responses were corrected using the following carotenoids $A_{1\%1\text{cm}}$: 2330 $\text{g L}^{-1} \text{cm}^{-1}$ (437 nm, 100% ethanol) for 9'-cis neoxanthin, 2550 $\text{g L}^{-1} \text{cm}^{-1}$ (433 nm, 100% ethanol) for violaxanthin, and 2500 $\text{g L}^{-1} \text{cm}^{-1}$ for antheraxanthin and unknown carotenoids were used. Similar procedure was used for porphyrins using the calibration curve of chlorophyll-a: 1270 $\text{g L}^{-1} \text{cm}^{-1}$ (664 nm, 90% (v/v) acetone) for chlorophyllide a, 742 $\text{g L}^{-1} \text{cm}^{-1}$ (667 nm, 90% (v/v) acetone) for pheophorbide a, 514 $\text{g L}^{-1} \text{cm}^{-1}$ (647 nm, 90% (v/v) acetone) for chlorophyll b and 512 $\text{g L}^{-1} \text{cm}^{-1}$ (667 nm, 90% (v/v) acetone) for pheophytin a (Roy et al. (2011), Britton et al. (1995)).

The ash content of biomass was determined as described by [27]. The nucleic acids DNA and RNA were not measured directly, but determined from cell number data. *N.*

oleoabundans has an estimated genome size of 68.2 Mb, with a CG content of 60% (internal communication). Assuming that each cell contains a single copy of DNA, it could be calculated that each cell contains 8.1×10^{-14} g DNA per cell. Transcriptome analysis revealed that RNA had an average GC content of 59%, and the RNA content was based on the observation that the cellular RNA:DNA ratio increase with increasing specific growth rate (μ , here in d^{-1}) of [96]. Their equation, derived for three nitrogen limited marine phytoplankton cultures was used in our calculations:

$$\mu = 0.255 \cdot \frac{RNA}{DNA} - 0.287 \quad (\text{Equation 8})$$

3.2.3 Flux balance analysis

The metabolism of any organism can be described by a set of reaction equations that define the stoichiometry of the conversion of substrates into products [97]. In this paper, a metabolic network is presented that describes the primary metabolism of the oleaginous green alga *N. oleoabundans*. To calculate the fluxes through this network, flux balance analysis was applied, using the measured production and consumption rates of the nitrogen replete and nitrogen limited steady states.

The stoichiometry matrix S of the CBM contains the stoichiometric coefficients of the substrates and products for the different reactions in the metabolic network, including the transport reactions over the membranes. Mass balances are written for all the intracellular metabolites present in the network. Assuming steady state and neglecting the net production of intermediates results in the next set of linear equations, which are used to calculate the fluxes in the network:

$$S \cdot x = 0 \quad (\text{Equation 9})$$

9)

In which S is the stoichiometry matrix and x is the vector which contains the reaction rates. The solution space of equation 9 was studied to find possible dead ends in the presented CBM, which were subsequently removed.

Three redundancy relations were extracted from the model stoichiometry being a nitrogen balance, a carbon balance and a redox balance. The measured production and consumption rates of functional biomass (i.e. biomass that does not contain starch, TAG and ash), starch,

TAG, nitrate, carbon dioxide and oxygen were checked for consistency using these redundancy relations and a Chi-Squared test. Subsequently, consistent rates were balanced according to [98].

Flux distributions were subsequently calculated using optimization of an objective function:

Objective function: $\max (c \cdot x)$

Constraints: $S \cdot x = 0$

$$\max (c \cdot x) \quad (\text{Equation 10})$$

In which c contains the objective function and LB and UB are the lower and upper boundary of reaction rate x . In this study, the objective functions ‘maximize ATP dissipation’ or ‘maximize product yield’ were used. Reactions that are irreversible were constrained to one direction by setting appropriate bounds. Constraints were set on transport fluxes depending on whether a compound was consumed or produced. Measured rates, rates were used to constraint the model of the corresponding reactions.

Due to underdetermined reaction sets, the obtained solutions for the flux values are not unique. In order to find the underdetermined reaction sets equation 9 is converted to:

$$S_c \cdot x_c = -S_m \cdot x_m \quad (\text{Equation 11})$$

Where x_c contains the unknown and x_m contains the measured rates. S_c and S_m are the corresponding parts of S . Underdetermined reaction sets were next found by studying the null space of S_c using singular value decomposition.

By studying the null-space of matrix S_c (Equation 11), it was revealed that the CBM contained 13 underdetermined parts. By constraining irreversible reactions and thermodynamically impossible combinations of reactions to the correct directions (Supplementary File 2), and by choosing the objective functions ‘maximize ATP dissipation’ and ‘maximize product yield’, unique values could be calculated for all but one underdetermined part. This underdeterminacy involved the anaplerotic routes between phosphoenolpyruvate, pyruvate, and oxaloacetate, and was solved by disabling reaction 55 (oxaloacetate carboxy-lyase), as described by [90].

Mathcad 14.0 (M020, Parametric Technology Corporation, USA) was used for network analysis and Matlab (version 6.0.0.88, release 12, The MathWorks Inc., USA) was used for *in silico* simulations.

3.2.4 RNA sampling and transcriptome assembly

Biomass for RNA extraction was obtained from 50 mL of culture volume that was centrifuged at 3900 RCF at 4 °C for 5 minutes. Resulting biomass pellets were quickly frozen and stored at -80 °C. Frozen pellets were ground using a mortar and pestle under liquid nitrogen. Subsequently, the total nucleic acid fraction of 0.5 g of ground and frozen pellet was extracted, using a preheated (65 °C) mixture of 1 mL liquefied phenol and 1 mL extraction buffer containing 1% SDS (v/v), 10 mM EDTA, 0.2 M NaAc (pH 5) and 2% β -mercaptoethanol (v/v). When samples were cooled to room temperature, they were mixed with a vortex for 15 seconds. Next, 1 mL of chloroform was added and samples were mixed again. The aqueous phase was collected after 15 minutes of centrifugation (13.000 RCF), and re-extracted with 1 mL chloroform. The last obtained aqueous phase was diluted 4 times with an 8 M LiCl solution containing 1% β -mercaptoethanol (v/v). The RNA was allowed to precipitate overnight at 4 °C and collected by centrifugation (15 min at 13,000 RCF). The RNA pellet was washed once with 2M LiCl and twice with 70% (v/v) ethanol, and finally dissolved in MilliQ water.

Sequencing libraries for the Illumina platform were generated from the isolated RNA and subsequently sequenced at BaseClear BV (Leiden, The Netherlands). Total RNA was first assessed for quality on a Bioanalyzer 2100 (Agilent Technologies Inc., Santa Clara, USA) and used as input for library preparation using the Illumina TruSeq RNA library preparation kit (Illumina Inc., San Diego, USA). Briefly, the mRNA fraction was purified from total RNA by polyA capture, fragmented and subjected to first-strand cDNA synthesis with random hexamers. After second-strand synthesis, barcoded DNA adapters were ligated to both ends of the double-stranded cDNA, the ligated product was size-selected and subjected to PCR amplification for 15 cycles. The resultant sequencing library was checked on a Bioanalyzer (Agilent Technologies Inc., Santa Clara, USA) and quantified. The libraries were multiplexed, clustered, and sequenced on an Illumina HiSeq 2000 with paired-end protocol for 50 cycles. The sequencing run was analyzed with the Illumina CASAVA pipeline (v1.8.2), with demultiplexing based on sample-specific barcodes. The raw sequencing data produced was processed removing the sequence reads which were of too low quality (only "passing filter" reads were selected) and discarding reads containing adaptor sequences or PhiX control. For all the samples, the reads are 51nt long and the median insert size is of 129bp.

To maximize the diversity and completeness of the *N. oleoabundans de novo* assembled transcripts, data from 16 RNAseq samples available in our lab were combined. Eight samples were generated from experimental conditions ([A],[C],[E],[F], and [I]) as described in [27] and

Figure 3.1, available at ArrayExpress (E-MTAB-3651) and at ENA (ERP010848). The other eight samples were generated from four experimental conditions, available at ArrayExpress (E-MTAB-3746) and at ENA (ERP0011032). The combination of these 16 RNAseq samples yielded a total of 496,158,724 reads and assembled with IDBA-UD v1.1.0 [100]. QUAST v2.3 [99] was used to estimate the quality of the assembly. CDS were extracted using QUAST and translated into protein sequences for functional annotation.

3.2.5 Transcriptome annotation

The protein sequences were annotated by aligning them using DELTA-BLAST 2.2.29+ (default parameters, E-value < 0.001) [91]. To this end, several databases were used sequentially: SwissProt, Chlorophyceae branch from SwissProt, Viridiplantae and Cyanobacteria branch from NCBI, Uniprot (SwissProt + Trembl) filtered for proteins with an annotated enzymatic reaction. Additionally, the proteins were annotated using InterproScan 5 [44] (with default parameters) for domain search analysis. Finally, Blast2GO V.2.7 [92] was used as the central tool to combine both analysis methods to assign GO terms to the protein sequences and to retrieve EC numbers.

Three tools were used to predict the subcellular location of the predicted enzymes: TargetP [100], PredAlgo [37] and WoLF PSORT [101]. Four common cellular locations were identified: chloroplast, mitochondrion, secretory pathway, and cytoplasmic. In the case of WoLF-PSORT, the localization defined as extracellular or plasma membrane were considered as secretory pathways.

3.2.6 Expression analysis

Read abundance estimations were done using the RSEM script from TrinityRNAseq pipeline [102, 103] with the default settings. Reads from each experimental condition were mapped onto the set of coding sequences generated with QUAST from the assembled transcriptome.

Data were normalized using Trimmed Mean of M-values as implemented in EdgeR package [104]. They were further normalized by CDS length to compute Fragments Per Kilo base of exon per Million fragments mapped (FPKM) using TrinityRNAseq TMM normalization script. CDS with read counts lower than 10 in all conditions were discarded and resulted in a reduction of ~18% of the transcripts. Finally, FPKM values corresponding to CDS annotated to

the same enzyme or transporter present in the CBM were added up to provide a single expression value for each reactions. Differential expression values per reaction were computed using the merged FPKM and the EdgeR package.

3.2.7 GO enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using the BINGO Cytoscape plugin [105] with default settings (hypergeometric test and Bejamini -Hochberg correction for multiple testing). The fully annotated transcriptome was used as background set. The test was performed for each pairwise comparison (Figure 3.1) and for three subsets of transcripts: down-regulated, up-regulated, and both up or down-regulated. Only terms with False Discovery Rate lower than 0.05 were retained for further analysis.

3.3 Results

3.3.1 Turbidostat experiments

Photobioreactors were operated in continuous mode as turbidostats, combining two light absorption rates with several nitrate supply rates resulting in 9 steady states. Experiments are presented in detail in the paper of Klok et al and part of the data presented there are used in this paper [27, 106]. The measured biomass compositions can be found in Table 3.1A. Functional biomass is defined as biomass without ash, starch and TAG. The nitrogen, carbon, and redox balance closed, according to the Chi-Squared test, for all experiments except [E], which was only just outside the 95% confidence interval (Additional File 3). In [E], nitrogen consumption rate was slightly lower than expected based on the measured functional biomass production rate, which could be explained by an inaccurate nitrate measurement. Therefore, for condition [E], the specific nitrate consumption rate was calculated from the nitrogen balance. Next for all conditions, except condition [E], the rates were balanced using the nitrogen, redox and carbon balances. For condition [E] the redox and carbon balances were used for balancing the rates. The balanced rates obtained for each steady state, expressed in $\text{mmol (g protein)}^{-1} \text{ h}^{-1}$, can be found in Table 3.1B. Rates and fluxes are expressed per gram of

protein, because protein is representative for functional biomass and contains the enzymes that carry the fluxes.

To study the effect of light intensity and nitrogen limitation, four steady states, indicated in Figure 3.1, were selected to compare flux and transcript levels. For high light conditions steady state E, which is nitrogen replete, was chosen together with steady state I, which is the steady state that is most nitrogen limited. For the low light intensity steady state, A, which is nitrogen replete and steady state C, which is nitrogen limited and has a nitrogen supply rate that is comparable to steady state I, were selected.

3.3.2 De novo transcriptome assembly and annotation

De novo assembled transcripts were used to predict coding sequences (CDS) that were further translated into 28266 putative protein sequences. Using InterProScan5 we identified conserved protein domains in 54% (15306) of these putative proteins, of which 7021 could be associated with Gene Ontology (GO) terms, of which 83% (5861) were unique. Additionally, 3249 putative proteins were associated with EC numbers, 834 of which unique.

Using three subcellular location prediction tools (TargetP, PredAlgo and WoLF-PSORT), four common cellular locations were identified for the putative proteins: chloroplast, mitochondrion, secretory pathway, and cytoplasm. It should be noted that protein localization prediction remains a challenge since signal peptides need to be further characterized. Currently, only PredAlgo is trained with algal sequences and solely with *C. reinhardtii*. Functional microalgae annotation would greatly benefit from curated training sets [38, 47].

3.3.3 Metabolic reconstruction

An initial metabolic network was constructed describing the primary metabolism of *N. oleoabundans* based on a previously published genome based reconstruction of *C. reinhardtii* [90]. The CBM comprises 373 reactions, distributed over two compartments: chloroplast and the cytosol. All reactions not located in the chloroplast were assigned to the cytosol. The reactions are described by a total of 227 unique EC numbers. This network of 373 reactions was further reduced to a more manageable size of 216 reactions and 219 metabolites by lumping linear pathways (Figure 3.2). In addition, complex processes such as photosynthesis and respiration were lumped into single reactions and not assigned to specific EC numbers.

Finally, transport reactions were included in the CBM, often as a result of gap filling to obtain a functional model, as many transporters have not been identified yet. EC numbers were cross-checked with the *de novo* assembled and annotated transcriptome (additional file 2) of *N. oleoabundans*, and the presence of genes coding for the enzymes catalyzing the modeled reactions was verified. The CBM contains only one enzymatic reaction for which the corresponding gene could not be identified: EC: 4.2.1.19, which is associated to the histidine biosynthesis pathway.

Stoichiometry of the functional biomass synthesis reaction was calculated for each steady state based on the measured composition (Additional File 3). As the amino acid composition of protein and fatty acid composition of TAG and membrane lipids did not exhibit large variations throughout the tested experimental conditions, the average composition of these molecules was used for determining the stoichiometry of their respective formation equations (Additional File 3). Based on their respective fractions, measured growth rates and biomass concentrations, the production rates of starch, TAG and functional biomass were calculated.

The non-growth associated maintenance was obtained from the *C. reinhardtii* CBM ($7.1 \text{ mmol. (g protein)}^{-1} \cdot \text{h}^{-1}$). The growth associated maintenance was estimated from experimental condition A assuming that in this steady state light was not dissipated. Thus, a value of $29 \text{ mmol. (g functional biomass)}^{-1}$ was obtained. More details about the metabolic reconstruction can be found in additional file 1.

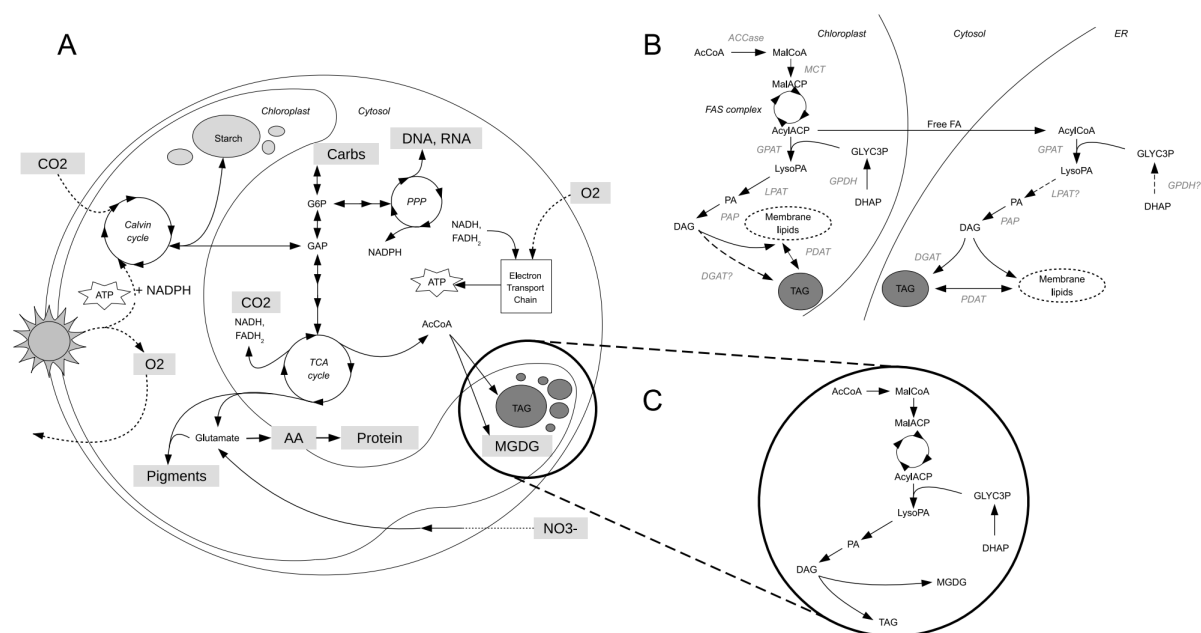


Figure 3.2: Overview of metabolic processes. A. Simplified overview of the metabolic network. B. Schematic overview of glycerolipid biosynthesis in microalgae; “?” denotes uncertainty about the existence of these enzymes. C. In the presented network, glycerolipid synthesis occurs only de novo and is located in the chloroplast.

3.3.4 Flux analysis

The balanced rates of Table 3.1B were used to calculate the metabolic flux distributions of experimental conditions [A]-[I]. Results for all steady states are given in Additional File 4. For the four selected steady states, these flux distributions were used to calculate the use of NADPH in the chloroplast, which represents the flow of electrons and energy in the chloroplast. The results are shown in Figure 3.3. The formation of NADPH occurs in the photosystems, where 10 photons are used to make 2 NADPH and 3 ATP. Thus, based on the total light absorption rate the potential rate of NADPH formation can be calculated, which represents the 100% value. However, part of the photons is dissipated as heat and fluorescence and are thus not used to make NADPH. This dissipation is represented in Figure 3.3A as fluorescence. It is assumed that in steady state A, which has a low light intensity and is nitrogen replete, all the light is 100% effectively used. Figure 3.3B shows the distribution of NADPH usage as a percentage of the NADPH that is actually formed. Under nitrogen replete conditions (A and E) 17% of the NADPH that is actually generated is used directly in the

chloroplast for nitrogen assimilation and fatty acid synthesis. However, by far the main fraction, 82%, is used to fix carbon. The electrons fixed in the C₃ carbon sugars have different fates. The majority (34%-38%) is respired again to generate energy for growth. The major remaining fates are protein (20%), starch (6-12%), membrane lipids (8-9%), TAG (3-5%) and other biomass components (1%) like carbohydrates other than starch, DNA and RNA. Thus, if the NADPH required to fix carbon for protein synthesis is included about 35% of the NADPH is involved in nitrogen assimilation and protein synthesis under replete conditions. This high NADPH demand for nitrogen assimilation and protein synthesis clearly illustrates the potential 'energy crisis' that a growing alga will be facing when it is exposed to sudden nitrogen depletion. When nitrogen fixation comes to an immediate halt, 14% of the generated NADPH has no direct sink anymore. Moreover, if no alternative is found for NADPH normally used in carbon fixation for protein synthesis, over one third of plastidic NADPH will become redundant. As a result, the photosynthetic electron transport chains will become over-reduced, which potentially triggers the redistribution of carbon towards highly reduced compounds such as TAG [107]. In Figure 3.3A, it can be seen that the main mechanism to deal with nitrogen limitation and the associated excess of energy is by reducing photosynthetic efficiency (53-75% of the energy is dissipated). When looking at the total energy supplied (Figure 3.3A) the increase in NADPH use for TAG formation upon nitrogen limitation is very small going from 2% (N-replete) to 4% (N-limitation) under low light and from 1% (N-replete) to 2% (N-limitation) under high light. However, when looking at the redistribution of NADPH that is actually formed (Figure 3.3B) the increase in NADPH use for TAG becomes more substantial when facing nitrogen depletion. For low light conditions, it increases from 2% (N-replete) to 8% (N-limitation), while for high light conditions it increases from 2% (N-replete) to 10% (N-limitation). However, this increase is mainly accounted for by the carbon fixation in the Calvin cycle and not by the fatty acid synthesis from the C₃ sugars. The evaluation of relative carbon and energy distributions only partially illustrates the effect of nitrogen limitation. Evaluation of absolute fluxes revealed that both NADPH and carbon use for all mentioned metabolites decreased under nitrogen limitation, with the exception of TAG (additional file 4). This indicates, that only TAG functioned as a sink in alleviating the NADPH imbalance in the plastid, as suggested by [107]. However, considering the large fraction of energy that is dissipated, it is more likely that the main function of TAG is storage of energy, carbon and electrons, while reduction of photosynthetic efficiency is the mechanism to deal with the energy excess.

The maximum theoretical yield of TAG on light for this model is $1.06 \text{ g (mol photons)}^{-1}$. TAG yields under nitrogen limited conditions are far from the theoretical maximum due to two factors being energy spent on other macromolecules, such as functional biomass and starch, and energy dissipation that progressively increases with nitrogen limitation.

To assess the impact of these effects, the TAG yield on light and the TAG productivity were estimated in three scenarios: starch-less mutant, conserved photosynthetic capacity in nitrogen limited condition, and both together. The results are shown in Figure 3.4. Clearly the most important factor in increasing the TAG yield on light is the photosynthetic efficiency. Under nitrogen limitation, maintaining photosynthetic efficiency at 100% would result in an increase of TAG yield on light to 0.68 for low light [C] and $0.89 \text{ g(mol photons)}^{-1}$ for high light [I]. Knock out of the starch pathway only leads to an increase of about $0.06 \text{ g.(mol photons)}^{-1}$ for low light [C] and $0.07 \text{ g.(mol photons)}^{-1}$ for high light [I]. Combination of both strategies leads to a yield of $0.94 \text{ g.(mol photons)}^{-1}$ for the high light condition I, which is close to the theoretical maximum.

For the specific productivities, (Figure 3.4B) more or less the same reasoning applies, with the difference that the productivities in condition C are much lower than in condition I due to the lower light intensity applied.

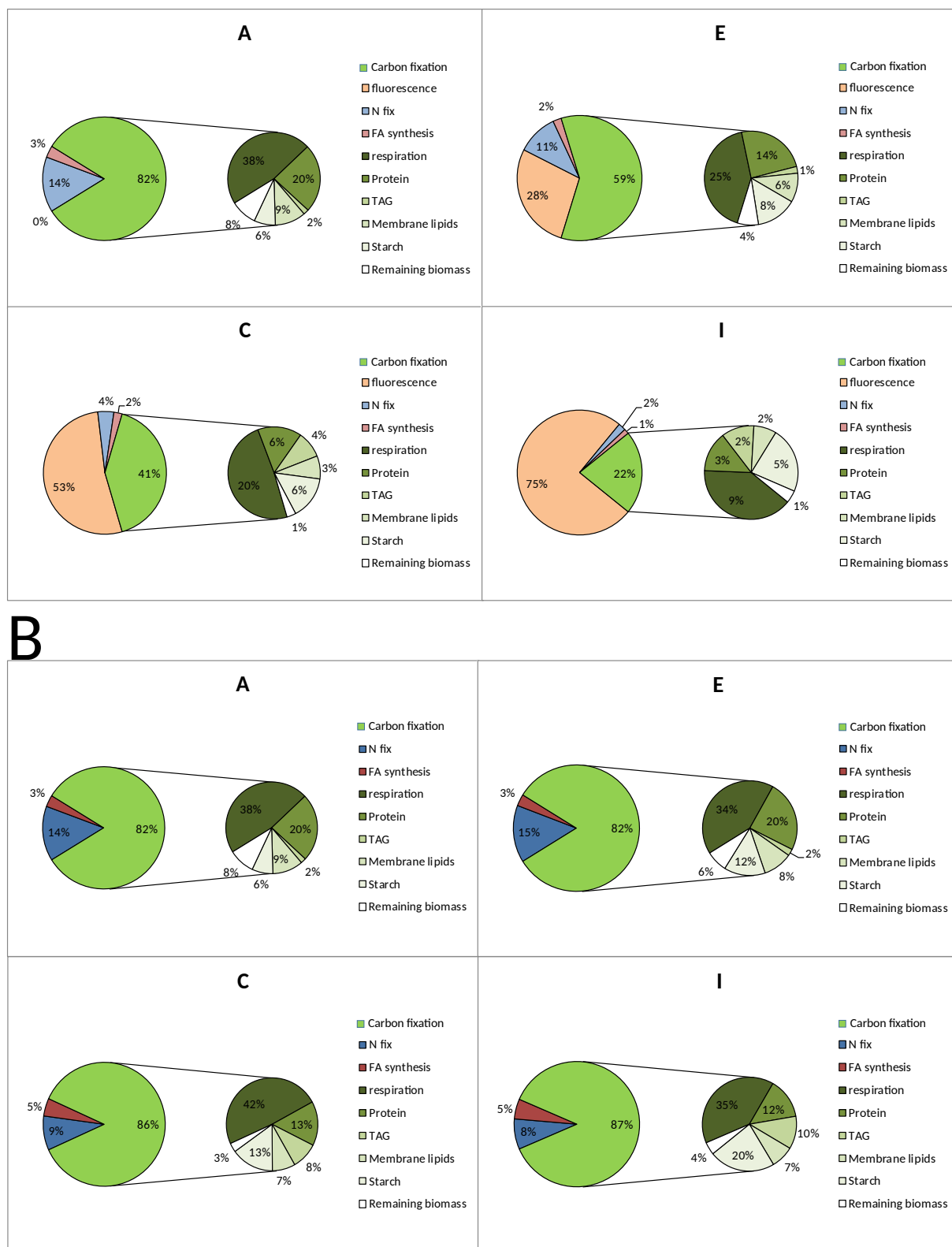


Figure 3.3: Partitioning of NADPH for condition A, C, E and I. A: relative to the total light energy received, where fluorescence represents NADPH that is not formed due to dissipation of light energy. B: relative to the actual amount of NADPH formed.

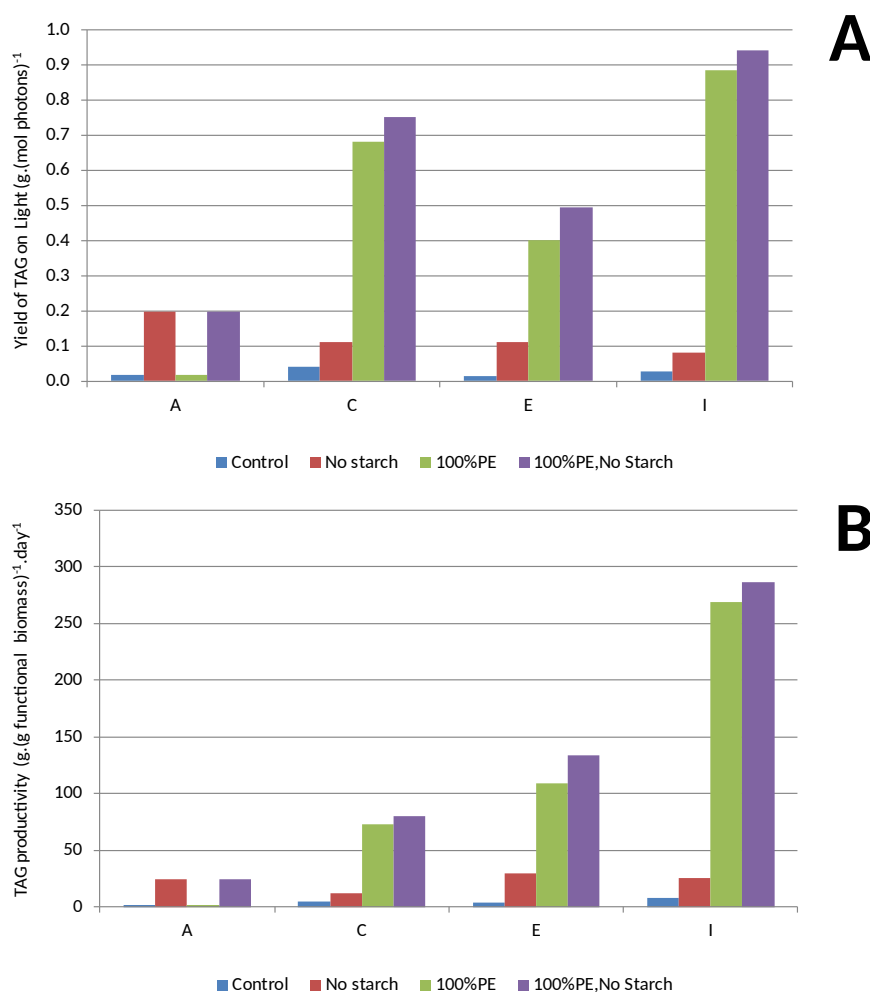


Figure 3.4: A: Yield of TAG on light in $\text{g TAG} \cdot (\text{mol photons})^{-1}$, B: specific TAG productivity in $\text{g TAG} \cdot (\text{g functional biomass})^{-1} \cdot \text{day}^{-1}$

3.3.5 Metabolic phenotype predictions constrained by differential expression

3.3.5.1 GO enrichment analysis

Fragments Per Kilo base of exon per Million fragments mapped (FPKM) were calculated by mapping RNAseq reads to the assembled transcriptome. Pairwise comparisons for the conditions presented in Figure 3.1 were performed. Up- and down- regulated transcripts were selected based on their fold change (FC), ($|\log_2\text{FC}| > 1$). GO enrichment analysis was performed in the selected sets (Table 3.3 and Additional file 5).

Comparisons associated to nitrogen limitation ([C][A], [I][E]) returned only few overrepresented GO terms, which were associated to up-regulation of fatty acid synthesis, cell wall biogenesis, transport activity and modifications in cell structure.

Comparisons associated to high light (HL) conditions ([E][A], [I][C]) resulted in a higher amount of overrepresented GO terms. Terms related to cell growth and division were found within the up-regulated transcripts, consistent with the high growth rate of [E]. Moreover, HL conditions cause down-regulation of transcripts enriched in terms associated to chloroplast membrane, thylakoid membrane, chlorophyll and carotenoids synthesis. This is clearly reflected in the reduced amount of pigments measured under HL (Table 3.1). During HL conditions, transcripts related to lipid synthesis and fatty-acids synthesis are down-regulated, while those related to the response to lipids term are up-regulated.

The comparison associated to combined induction of high light and nitrogen limitation ([I][A]) retains functions associated to HL induction, such as chlorophyll and carotenoids synthesis, which appear enriched in the down regulated gene set (Table 3.3) but none of the ones found in nitrogen limitation. This can at least in part be explained from opposite regulation for high light and nitrogen limitation. For example the down-regulation of fatty acid synthesis upon high light may be compensated by the up-regulation of these genes under N-limitation. Additional terms are found enriched in the combined induction. Those terms appear enriched by the six candidate proteins associated to the GOGAT being strongly up-regulated in the combined induction.

Table 3.3: Enrichment output for selected GO in up- and down-regulated transcripts. “+” indicates overrepresentation (FDR < 0.05) for one pairwise comparison, while “++” indicates that both pairwise comparisons display an over-representation of the GO term group.

Selection	HL condition		Nitrogen limitation		Combined induction	
	UP	DOWN	UP	DOWN	UP	DOWN
Cellular structure						
Cell wall biogenesis	+		++			
Microtubules		+	++			
Transport activity			++			
Lipid production						
Acetyl-CoA	+					
Phosphopantetheine binding			++			
Fatty acid synthesis		+	++			
Lipids synthesis		+				
Response to lipids	+					
Photo-acclimation						
Carotenoids		+				+
Chloroplast membrane		++				+
Thylakoid membrane		+				+
Magnesium chelatase		++				+
Chlorophyll synthesis		++				+
PPP & NADPH regen		+				+
Cellular reproduction						
Methylation	++					
DNA synthesis	+					
Cell replication	+					
Oxydoreductase	+					
UMP / ATP / AMP	+					
Cell AA derivative	+					
Nitrogen dependency						
Ammonium assimilation					+	
Resp. to ammonium					+	
Glutamate synthesis					+	

3.3.5.2 Comparing dynamics in flux and expression

Changes in flux and expression levels are shown in Figure 3.5. No obvious correlations between changes in flux and expression were found. Their correlations can be hampered by all kinds of post transcriptional regulatory events, by experimental set up, and by type and design of the model. In general, low correlations between transcript and protein measurements have been reported in different organisms [108]. Therefore, the lack of obvious correlations, at a

global level, is not surprising when comparing even more distant levels such as transcription levels and fluxes.

The flux values directly correlate with the flux through the light reaction meaning they increase with light intensity and decrease when nitrogen is limited. This is logical since this reaction determines the availability of energy and reducing equivalents.

For the change from LL to HL conditions, comparisons [E]/[A] and [I]/[C] (Figure 3.5, a and b), flux variations are limited in range, while relatively large variations in expression are observed for pathways such as carotenoid metabolism, chlorophyll metabolism, and purine and pyrimidine metabolism. For the comparison of nitrogen replete to nitrogen deplete: [C]/[A] and [I]/[E] (Figure 3.5, c and d), both flux and expression values show a broader range of variation. The pronounced flux changes under nitrogen limitation can be associated to drastic changes in growth rate and biomass composition (increased TAG content, decreased pigmentation) as shown in Table 3.1.

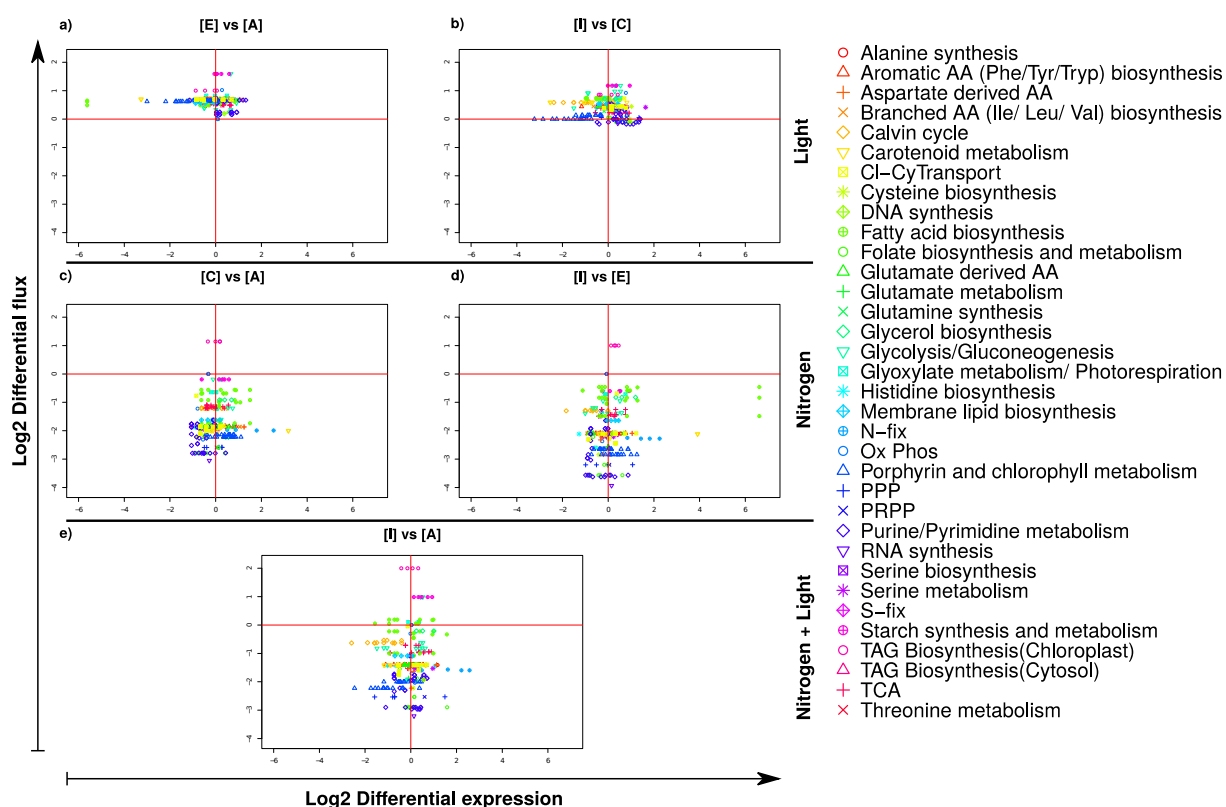


Figure 3.5: Flux and expression comparisons. Comparison of changes in expression and changes in fluxes for the 5 pairwise comparisons from Figure 3.1 (blue arrows), for each EC or transporter integrated in the CBM (colored by pathways and transport type). a) and b) correspond to an increase in light intensity; c) and d) correspond to a decrease in nitrogen concentration, and e) correspond to combined induction (HL induction and nitrogen limitation).

3.3.5.3 *Carotenoid and chlorophyll biosynthesis*

Figure 3.6 shows expression changes associated to transcripts annotated to enzymes in the carotene biosynthesis, chlorophyll biosynthesis and terpenoid backbone biosynthesis pathways. For HL induction (comparisons [E]/[A] and [I]/[C]), the transcripts are globally down-regulated in these pathways. The combined induction (comparison [I]/[A]) displays a behavior similar to HL induction and we conclude that HL has a stronger influence on the regulation of these three pathways than nitrogen limitation. Nitrogen limitation (comparisons [C]/[A] and [I]/[E]) affects these pathways in a different way. Terpenoid backbone biosynthesis and chlorophyll biosynthesis pathways are up-regulated especially at low light intensity. At the high light intensity the up regulation is present for fewer genes, which is possibly related to the dominant effect of light intensity on regulation in these pathways. The carotene biosynthesis appears slightly down-regulated except for EC1.3.5.5 that is strongly upregulated. However, this transcript has extremely low expression in all analyzed conditions and the observed upregulation may not be relevant with respect to changes in the flux through this pathway. Geranylgeranyl-diphosphate (GGPP) reductase (EC 1.3.1.83), is described as a reversible reaction and connects the three aforementioned pathways. Its pattern of change opposes that of these pathways with over expression upon high light conditions and reduced expression under nitrogen limitation, as can be seen in Figure 3.6. The over expression upon high light conditions may be related to the interconversion of chlorophyll to carotenes and the effect of photo damage, which occur at high light. Chlorophyllase (EC:3.1.1.14) exhibits a pattern of expression changes dissimilar from the rest of the enzymes in these pathways. Chlorophyllase is strongly down-regulated by HL with sufficient nitrogen supply ([E]). On the other hand, chlorophyllase is slightly up-regulated during HL with limited nitrogen supply ([I]), which is the condition with the lowest chlorophyll amount (Table 3.1). This might be linked to the inability of the organism to deal with the excess of energy under high light and low nitrogen condition thereby promoting active chlorophyll degradation to reduce this surplus.

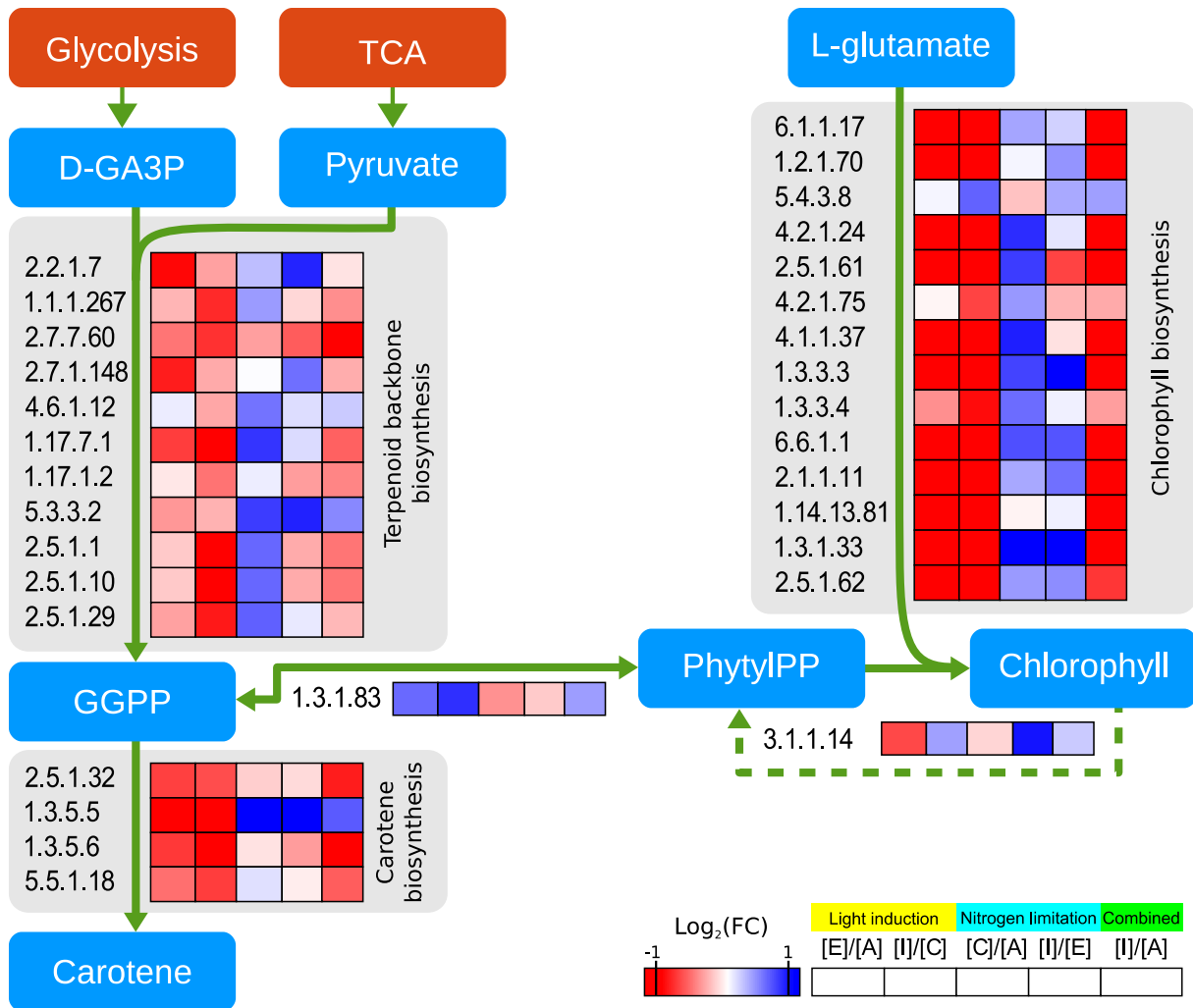


Figure 3.6: Crosstalk between Chlorophyll biosynthesis, terpenoid backbone biosynthesis and carotene biosynthesis pathways. This figure represents the reactions involved in the pathways of Chlorophyll biosynthesis, terpenoid backbone biosynthesis and carotene biosynthesis. All the reactions are ordered and named by their EC number. The differential expression of each reaction is given for the 5 pairwise comparisons from Figure 3.1 (blue arrows). The differential expression color representation is limited from -1 to +1 of the log₂(FC), any values respectively lower or higher than this cannot be distinguished here.

3.3.5.4 Nitrogen assimilation pathways

Differential expression of enzymes annotated to nitrogen assimilating reactions is summarized in Figure 3.7. During nitrogen limitation, transcripts associated to nitrate (EC:1.7.1.1) and nitrite (EC:1.7.7.1) reductase are strongly up-regulated. These reactions are the initial point of nitrogen assimilation in microalgae. Transcripts associated to GOGAT (EC:1.4.1.14) are strongly up-regulated both by nitrogen limitation and by HL induction. The GO enrichment

analysis for the nitrogen related terms is entirely due to the six candidate proteins associated to GOGAT being up-regulated. Furthermore, when both inductions are combined [I], we observe that the aforementioned transcripts have the strongest expression. The expression of Glutamate dehydrogenase (GDH, EC:1.4.1.3) associated transcripts remains stable, therefore, we hypothesize GOGAT to be the preferred path for nitrogen assimilation in *N. oleoabundans*. The expression of Glutamine synthase (GS, EC:6.3.1.2) does not show this strong upregulation under all conditions, which could be explained by the fact that the activity of this enzyme is regulated in another way than transcriptional.

In the case of nitrogen limitation the upregulation of nitrogen assimilation reflects an attempt to fixate the scarcely available nitrogen quickly. For the high light condition it reflects the higher growth rate under this condition, which also requires faster assimilation of nitrogen.

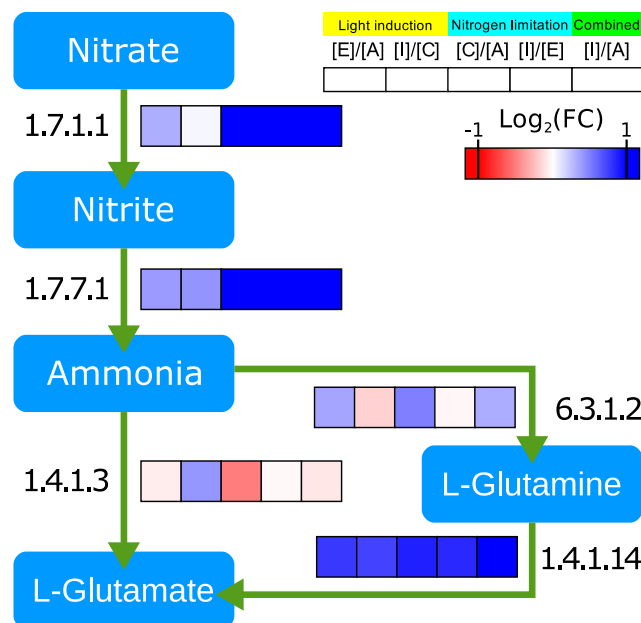


Figure 3.7: Nitrogen Assimilation. This figure summarizes the reactions involved in nitrogen assimilation. The differential expression of each reaction is given for the 5 pairwise comparisons from Figure 3.1 (blue arrows). The differential expression color representation is limited from -1 to +1 of the $\text{log}_2(\text{FC})$, any values respectively lower or higher than this cannot be distinguished here.

3.3.5.5 Starch biosynthesis

Higher relative levels of starch are found under HL conditions. (Table 3.1). This could be due to higher expression levels of starch synthesis genes or lower expression levels of starch degradation enzymes. For the HL conditions expression of both synthesis and degradation are slightly up-regulated (Figure 3.8), except 3.2.1.2 which is a starch degrading enzyme that is

down regulated under HL nitrogen replete conditions. The up-regulation of both synthesis and degradation suggests that under HL induction, starch turnover is increased, which agrees with its function as a transient carbon and energy storage. Furthermore, the increase in starch content is not necessary due to transcriptional regulation but could also be due to a higher synthesis flux caused by a higher availability of sugars under HL conditions due to the increased photosynthetic rate.

Nitrogen limitation did not affect the net flux towards starch. Yet, there is a strong over-expression of starch synthase (2.4.1.21), while the degradation of starch to single sugars appears down regulated. The down regulation of starch breakdown does not agree with the constant net flux to starch. Possibly the biosynthesis rate is reduced due to low availability of sugars under nitrogen limitation and the transcriptional regulation as described compensates for this to keep a constant starch synthesis flux.

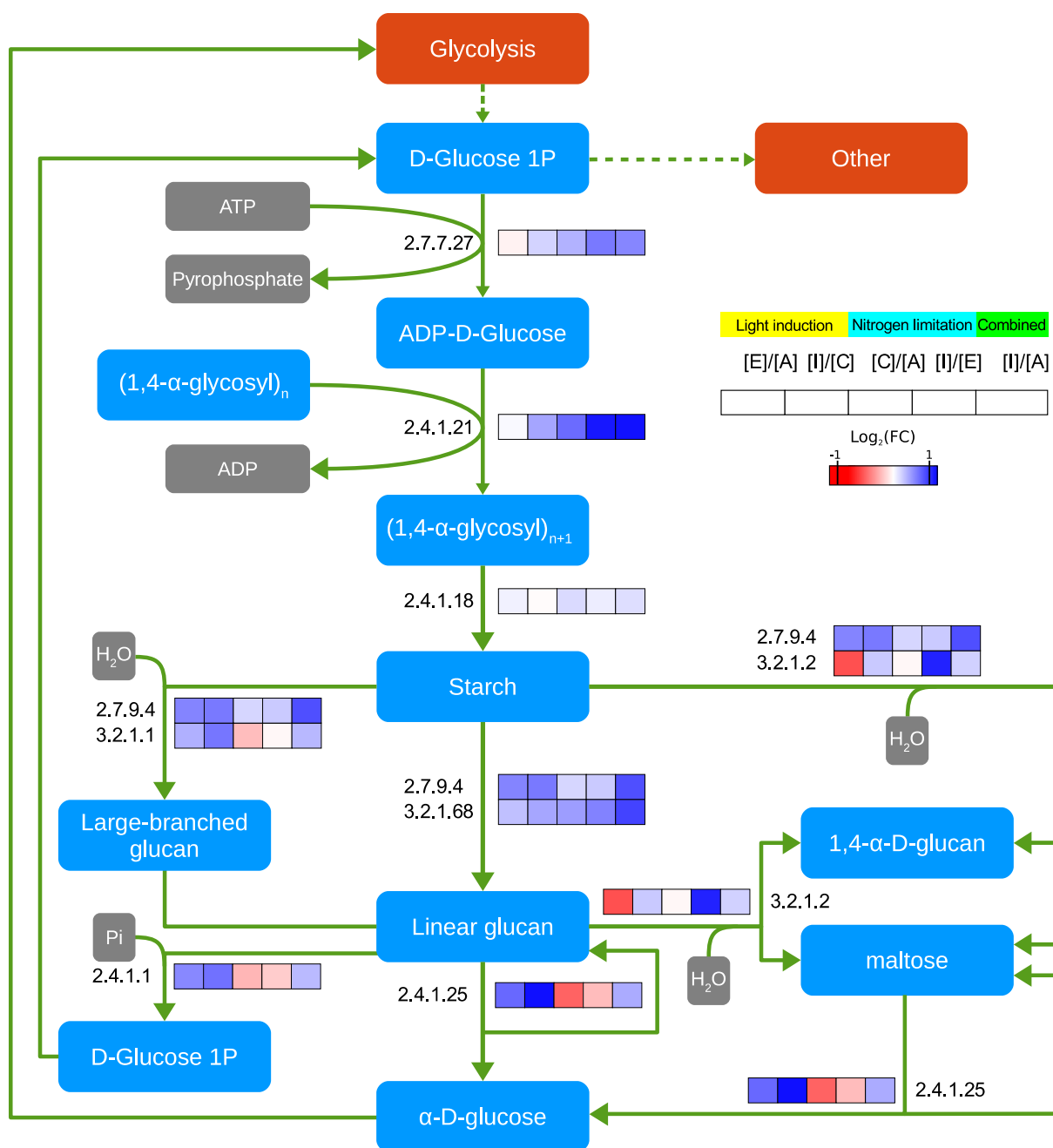


Figure 3.8: Starch biosynthesis. This figure summarizes the reactions involved in starch biosynthesis. The differential expression of each reaction is given for the 5 pairwise comparisons from Figure 3.1 (blue arrows). The differential expression color representation is limited from -1 to +1 of the $\log_2(FC)$, any values respectively lower or higher than this cannot be distinguished here.

3.3.5.6 Lipids and TAG biosynthesis

HL conditions leads to global down regulation of enzymes in fatty-acid and TAG synthesis pathways (Figure 3.9), which can be linked to the reduction in membrane lipid content under high light exposure (Table 3.1). On the other hand, nitrogen limitation induces a more diverse

pattern of expression changes with a few enzymes that show a distinctive pattern of overexpression being: acetyl-CoA carboxylase (6.4.1.2), biotin carboxylase (6.3.4.14), acyl-ACP synthase I (2.3.1.41), acyl-ACP reductase (1.1.1.100), acyl-ACP hydrolase (3.1.2.14), Glycerol-3P acyltransferase (2.3.1.15), the two acyl-CoA desaturases (1.14.19.1, 1.14.19.2), and the three Glycerol-3-Phosphate dehydrogenases (GPDH, 1.1.1.8, 1.1.1.94, 1.1.5.3). These are interesting candidates for genetic engineering, as their over expression correlates with an increased TAG content and thus, these might be key controllers of lipid biosynthesis under nitrogen limitation. Among the three GPDHs, two are NAD dependent reactions (EC:1.1.1.94 and EC:1.1.1.8) and are annotated to the same transcript, while the third (EC:1.1.5.3) is a FAD dependent enzyme. The transcript associated to this last enzyme is the one with the highest expression.

We have not been able to identify PDAT (EC: 2.3.1.158) nor MGDG synthase (EC: 2.4.1.46) among our transcripts, and as a result the role of membrane lipid turnover remains uncertain. However, expression of diacylglycerol acyltransferase (DGAT, EC:2.3.1.20) associated protein increased strongly under nitrogen limitation and, albeit to a lesser extent, upon HL. This expression pattern of DGAT suggests an important contribution of *de novo* synthesis to TAG production under nitrogen limitation. This agrees with previous suggestions in *C. reinhardtii* where PDAT-mediated membrane lipid turnover is strongly associated to healthy growth conditions and less prominent during stress or stationary conditions [109] and further justifies the decision to restrict the model to *de novo* lipid synthesis only.

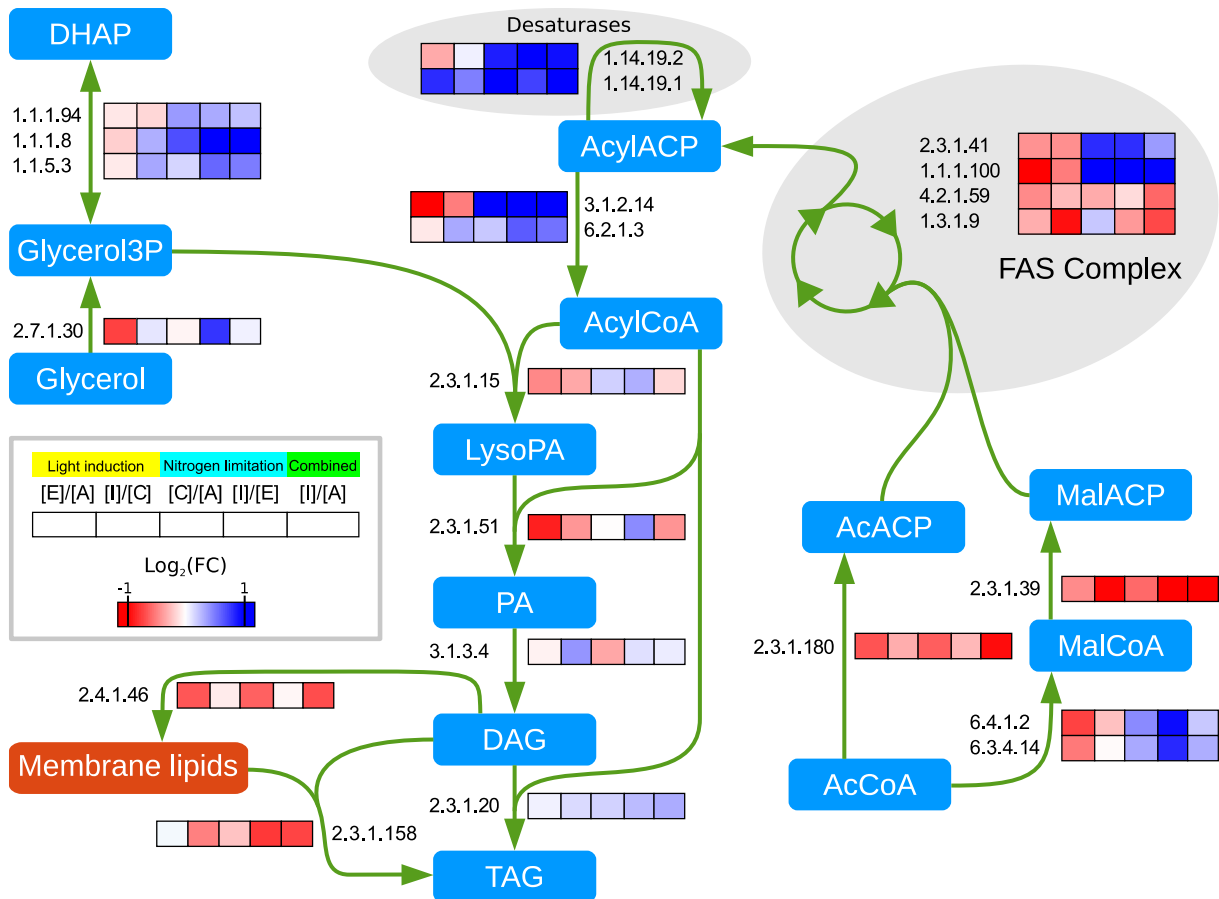


Figure 3.9: Lipids and TAG biosynthesis. This figure summarizes the reactions involved in lipids and TAG biosynthesis. The differential expression of each reaction is given for the 5 pairwise comparisons from Figure 3.1 (blue arrows). The differential expression color representation is limited from -1 to +1 of the $\text{log}_2(\text{FC})$, any values respectively lower or higher than this cannot be distinguished here.

3.4 Discussion

3.4.1 Adaptations upon high light induction and nitrogen limitation

Transcript level changes upon changes in light intensity and nitrogen concentration show distinct characteristics (Figure 3.5 and Table 3.3).

Nitrogen limitation induces a targeted response, with few differentially expressed transcripts, mostly related to cell structure, fatty-acids synthesis, transport activity and nitrogen assimilation. Such behavior has been described in other microalgae [110–112]. The response is most likely propagated through nitrogen-responsive regulators [110]. HL induces a wider regulatory response on transcripts related to cell growth, chloroplast, chlorophyll, carotenoids, and lipids synthesis. Excessive light energy seems to primarily affect the regulation of the photosystems. Similar behavior has been reported in *C. reinhardtii* and other photoautotrophic eukaryotes [113–115].

GO term enrichment analysis of differentially expressed genes uncovered processes directly related to nitrogen assimilation when both inductions are combined. This indicates a stronger response than separated inductions. This response suggests that while the cell gets more energy from the photosystems or oxidative stress, the lack of nitrogen becomes a bigger bottleneck and the system tries to compensate by forcing higher assimilation rates. The stronger impact from nitrogen limitation combined with HL stress can also be observed in the studied pathways shown in Figures 3.7, 3.8, and 3.9.

3.4.2 Potential regulation for photo-protection

HL induction is known to induce degradation of chlorophyll and photosystems in plants and green algae [116, 117]. Our experimental setup (Table 3.2) was set to keep a constant light absorption rate for each light condition, leaving only NPQ to alter the photochemical conversion rate. The interaction of carotenoids and chlorophylls is strongly dependent on the composition and structural arrangement within the antenna of the photosystem [118]. As depicted in Figure 3.6, the expression profiles of the synthesis pathways for terpenoids backbone, and chlorophyll, are very similar. High light conditions result in down regulation of these pathways, which is in agreement with the lower chlorophyll content at higher light intensities (photoadaptation). Also the carotenoid pathway is down regulated under high light conditions. *N. oleoabundans* is known not to accumulate secondary carotenoids for photoprotection and thus only contains carotenoids as accessory pigments that are a structural part of the photosystems and help in transferring light energy to the reaction centers. Consequently the down regulation of this pathway is in agreement with the reduction in photosystems due to photoadaptation. Under nitrogen limitation a global up-regulation of the terpenoid backbone and chlorophyll pathways is observed. Apparently, transcriptional

regulation is such as to compensate the loss of photosynthetic efficiency (Figure 3.3) and the limitation of nitrogen by increasing the chlorophyll content.

The regulation of GGPP reductase (EC:1.3.1.83), in which GGPP is desaturated to form PhytylPP, is opposite the regulation of the terpenoid and chlorophyll pathway. This reaction is needed for the synthesis of chlorophyll, but also forms the connection between carotenoid and chlorophyll pathways. The up-regulation under high light conditions could indicate and increased turnover of chlorophyll to carotenoids or a change in balance between chlorophyll and carotenoids. Some of the carotenoids that are structurally present in the photosystems are known to be involved in scavenging reactive oxygen species that are increasingly formed if the light intensity increases. Based on this, it is expected that also the breakdown of chlorophyll would be stimulated under high light. However, chlorophyllase (EC:3.1.1.14), which removes the phytol group from chlorophyll is only up-regulated under high light if nitrogen is limiting. Under replete conditions it is even down regulated if the light intensity goes up. This could indicate that under replete conditions sufficient enzyme is present and regulation is especially on the metabolite level, whereas under N-limitation the amount of enzyme is limiting. Under nitrogen limitation the GGPP reductase is down regulated while chlorophyll synthase (2.5.1.62) is up-regulated, and under high light the chlorophyllase (EC:3.1.1.14) is also up-regulated, which could point to increased turnover in combination with a reduced net synthesis.

With these results, we can only speculated about two possible reasons where GGPP reductase and chlorophyllase play a key role at the intersection of the three pathways depicted in Figure 3.6. The first is that the three pathways compete for the PhytylPP, which can be recovered from the chlorophyll by the chlorophyllase. It is also possible that the phytol recycled from chlorophyll degradation and can be reused for tocochromanols (vitamin E) and possibly participate in photo-protective activity [119–124]. The second hypothesis is that GGPP reductase take effect only after the chlorophyll synthase and acts as an activator of the chlorophyll [122, 125, 126]. Chlorophyll synthase can esterify chlorophyllide with either PhyPP or GGPP, respectively resulting in the “active” phytolated chlorophyll a (Chla_{phy}) or “inactive” geranylgeranylated chlorophyll a (Chla_{gg}) [127]. Hence, the microalgae can continue to synthesize Chl_{gg} during HL, and it can then rapidly activate its chlorophyll using the GGPP reductase when the HL stress is gone.

Overall, chlorophyll and carotenoid content is tightly regulated over three major points: the production by the chlorophyll and carotenoid synthesis reactions including the terpenoid backbone pathway, the degradation by the chlorophyllase, and the connection by GGPP reductase. The chlorophyll synthesis reactions and chlorophyllase regulate together the

turnover. The chlorophyll synthase, the chlorophyllase and the GGPP reductase possibly regulate together the distribution of GGPP and PhyPP between the carotenoids and chlorophyll.

3.4.3 Carbon partitioning

Under nitrogen limitation, the carbon flow towards proteins is decreased and is compensated by a higher flux towards TAG. This agrees with the up-regulation of a number of genes involved in the production of TAG. Furthermore, for each light condition the net flux to starch is maintained at nitrogen limitation due to a down regulation of the starch break down, which agrees with the decreased need for starch due to the lower specific growth rate.

Under HL with sufficient nitrogen supply [E], there is an increased carbon flux towards starch. This is not clearly correlated with up-regulation of the starch synthesis genes or down regulation of the starch break down genes. The only exception is the gene for glycogenase (Figure 3.8, EC:3.2.1.2), which debranches maltose units from polysaccharides and oligosaccharides. This enzyme that is part of the breakdown of starch is down regulated under the high light condition. By avoiding starch degradation into transportable sugar, the cell actively tries to keep it as a form of energy storage. The lack of correlation between gene expression and the starch synthesis may indicate that regulation of starch synthesis occurs on the metabolite level, for example by a concentration increase for the metabolites that form a substrate for starch synthesis. For example the metabolites D-GA3P, and pyruvate (also the source of acetyl-CoA) form the starting point of the pigment (Figure 3.6) and lipid (Figure 3.9) synthesis pathways and are also part of the route towards Starch. Thus down-regulation of these indirectly competing pathways may increase the availability of these compounds and thus facilitate starch synthesis.

3.5 Conclusion

A CBM was built describing the primary metabolism of *N. oleoabundans* and was used to evaluate carbon and energy partitioning under steady state, nitrogen limited conditions with varying light intensities. Simulations revealed that TAG yield on light can much better be

improved by optimizing photosynthetic conversion than by blocking competing pathways. Integration of expression data showed a direct effect of nitrogen limitation on nitrogen dependent pathways, while HL impacted a wider number of pathways, among which cell growth, chloroplast, chlorophyll, carotenoids, and lipids synthesis. Nevertheless, the response to the combined induction of HL and nitrogen limitations it seems that higher synthesis of glutamate is attempted, possibly to sink nitrogen towards protein synthesis. Our analysis highlighted a possible interplay between synthesis of chlorophyll, carotenoids and a possible accumulation of inactive geranylgeranylated chlorophyll a. Finally, our analysis pinpoints eight reactions, such as the biotin carboxylase and GPDH, as prime choice reactions to increase TAG synthesis during nitrogen limitation, and some others, such as other forms of GPDH and Glycerol-3P acyltransferase, that could redistribute carbons towards TAG synthesis.

Chapter 4

Neochloris oleoabundans is worth its salt:
Transcriptomic analysis under salt and nitrogen
stress

This chapter is published as:

Lenny de Jaeger[§], Benoit M. Carreres[§], Jan Springer, Peter J. Schaap, Gerrit Eggink, Vitor A. P.
Martins Dos Santos, Rene H. Wijffels, Dirk E. Martens.

Neochloris oleoabundans is worth its salt: Transcriptomic analysis under salt and nitrogen
stress.

PLoS One. 2018

§ Authors contributed equally

ABSTRACT

Neochloris oleoabundans is an oleaginous microalgal species that can be cultivated in fresh water as well as salt water. Using salt water gives the opportunity to reduce production costs and the fresh water footprint for large scale cultivation. Production of triacylglycerols (TAG) usually includes a biomass growth phase in nitrogen-replete conditions followed by a TAG accumulation phase under nitrogen-deplete conditions. This is the first report that provides insight in the saline resistance mechanism of a fresh water oleaginous microalgae. To better understand the osmoregulatory mechanism of *N. oleoabundans* during growth and TAG accumulating conditions, the transcriptome was sequenced under four different conditions: fresh water nitrogen-replete and -deplete conditions, and salt water (525 mM dissolved salts, 448mM extra NaCl) nitrogen-replete and -deplete conditions. In this study, several pathways are identified to be responsible for salt water adaptation of *N. oleoabundans* under both nitrogen-replete and -deplete conditions. Proline and the ascorbate-glutathione cycle seem to be of importance for successful osmoregulation in *N. oleoabundans*. Genes involved in Proline biosynthesis were found to be upregulated in salt water. This was supported by Nuclear magnetic resonance (NMR) spectroscopy, which indicated an increase in proline content in the salt water nitrogen-replete condition. Additionally, the lipid accumulation pathway was studied to gain insight in the gene regulation in the first 24 hours after nitrogen was depleted. Oil accumulation is increased under nitrogen-deplete conditions in a comparable way in both fresh and salt water. The mechanism behind the biosynthesis of compatible osmolytes can be used to improve *N. oleoabundans* and other industrially relevant microalgal strains to create a more robust and sustainable production platform for microalgae derived products in the future.

4.1 Introduction

Sustainable and renewable production of energy and food for an increasing world population is an enduring challenge in present-day research. This challenge must be addressed with urgency, because of the world's dependence on limited fossil fuels and the increase in living standards of emerging economies. Renewable energy platforms based on oleaginous agricultural crops such as rapeseed, palm oil, corn, and soybean are being studied. Although these crops are considered renewable and bio-based, they increase the competition for food, fresh water, the amount of available arable land and result in deforestation to create plantations [11, 128–131]. Ideally, we would use land that is not suitable for traditional agriculture such as salt contaminated land or very dry areas like deserts for this purpose. A promising alternative feedstock compared to traditional crops for the production of oil are microalgae [4, 11]. Microalgae can produce high amounts of neutral lipids, triacylglycerol (TAG), when exposed to unfavorable growth conditions such as nitrogen depletion. TAG can be easily converted in biodiesels by methylation, which results in a pure and clean fuel that can replace petroleum-derived fuels [11]. The TAG molecules can also be used directly in the food and feed industry as a sustainable vegetable oil replacement. For an acceptable sustainable production process with a reduced fresh water footprint the use of marine or salt tolerant microalgal species is essential.

Most organisms are not able to cope with a shift in osmotic pressure when the environment is changed from fresh water to salt water and their growth will be compromised. Some organisms are able to adapt to such changes. Plants have developed different strategies to deal with osmotic stress. In addition to the strategies that involve structural traits such as waxes and adaptation of flowering time to the right conditions and moment, plants can also regulate their osmotic homeostasis by actively excluding salts from the cell to maintain water absorption [132–135]. Another strategy involves the accumulation of certain compatible organic osmolytes. A few examples of these compatible osmolytes are proline and glycine betaine [136], cyclic polyols such as D-pinitol [137], and sugars such as sucrose hexoses and sugar alcohols [132, 138].

Prokaryotic microalgae (cyanobacterias) are known to accumulate sucrose or α -glucosylglycerol under salt stress conditions [139, 140]. In eukaryotic microalgae, there are

some strategies that can be found to overcome salt stress [141–145]. Some species of microalgae are known to be able to survive low levels of saline environments. Known mechanisms for salt tolerance are glycerol production [146], sucrose production [147], and amino acid accumulation [132, 148].

The oleaginous salt tolerant microalgae *Neochloris oleoabundans* is a very interesting candidate for lipid production [147, 149, 150]. In nitrogen-deplete conditions, *N. oleoabundans* can accumulate TAG at up to 44% of its dry weight resulting in a maximal productivity of 164 mg L⁻¹ day⁻¹ [13, 151]. The harsh desert conditions from which this oleaginous microalgae was isolated [152], forced *N. oleoabundans* to be a highly flexible species to deal with the daily salt, drought and temperature stresses during the hot days and cold nights. These properties, combined with the high growth rate of *N. oleoabundans* (μ 2.2 D⁻¹) [150], and its resistance to highly alkaline conditions (up to pH 10) [149], which enhances the CO₂ transfer and reduces risk of contamination, makes it a very interesting candidate for sustainable oil production. The aim of this study is to identify the mechanisms used by the green microalgae *N. oleoabundans* to cope with saline conditions under growth (nitrogen-replete) and TAG accumulating (nitrogen-deplete) conditions based on a transcriptomic approach. To obtain insight in the pathways and metabolic reactions involved in salt resistance and lipid accumulation differential gene expression was studied between four conditions being: fresh water nitrogen-replete and -deplete, and salt water nitrogen-replete and -deplete (Panel A in S1 Fig).

To our knowledge, *N. oleoabundans* is the first fresh water microalgae, studied on transcriptomic and metabolite level, which is able to alleviate the osmotic stress under salt water conditions. We will discuss and compare the different pathways that are involved in the saline and nitrogen stress response. These findings can be used to get a better understanding of these processes and to define targets for new strategies to enhance microalgal strains to increase lipid productivity in the future.

4.2 Material and methods

4.2.1 Strain, medium and pre-culture

Neochloris oleoabundans UTEX 1185 (University of Texas, Austin, USA) pre-cultures were maintained in 100 mL filter sterilized fresh or salt water medium in 250 mL Erlenmeyer shake flasks. The fresh water medium consisted of: KNO_3 50.5 mM; Na_2SO_4 4.6 mM; HEPES 100 mM; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 1 mM; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 0.5 mM; K_2HPO_4 4.1 mM; NaHCO_3 10 mM; NaFeEDTA 0.14 mM; $\text{Na}_2\text{EDTA} \cdot 2\text{H}_2\text{O}$ 0.4 mM; $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ 96 μM ; $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ 21 μM ; $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$ 6 μM ; $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ 6.6 μM ; $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ 0.5 μM ; biotin 0.2 μM ; vitamin B1 7.4 μM ; vitamin B12 0.2 μM . The same medium was used for salt water with the following modifications: NaCl 448 mM; $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ 5 mM; $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ 2.4 mM. The pH for both media was set to pH 7.5 using NaOH and the medium was filter sterilized (0.2 μm) prior to use. In nitrogen-deplete conditions were applied the KNO_3 was omitted and replaced by an equimolar amount of KCl. In the shake flasks, pH was buffered using HEPES, while in the bioreactor HEPES was omitted and the pH was controlled by CO_2 addition.

4.2.2 Reactor design

All experiments were performed in flat-panel, Algaemist airlift-loop photobioreactors [153]. The reactors, with a working volume of 0.38 L, were kept at 25°C and pH controlled using on demand CO_2 (pH 7.5). Light was supplied continuously and the incident light intensity was adjusted to maintain constant average light intensity (80 $\mu\text{mol m}^{-2} \text{s}^{-1}$). The cultures were inoculated at a biomass concentration of 0.15 g L^{-1} in either fresh or salt water medium. When the biomass concentration reached approximately 4.5 g L^{-1} , the cultures were harvested and washed with either salt or fresh water medium containing nitrogen or no nitrogen. The cells were cultivated for 24 hours before sampling. Samples for dry weight, total fatty acid, TAG, starch, Nuclear magnetic resonance (NMR) analysis, and RNA extraction were taken. All reactors were run in duplicate, resulting in 8 individual samples for four different conditions.

4.2.3 Determination of dry weight concentration

Dry weight concentrations were determined on biological replicates. Around 1.5 mg of biomass was filtered through pre-dried (100°C overnight) and pre-weight Whatman glass fiber filter paper (GF/F; Whatman International Ltd, Maidstone, UK). The filter was washed with 50 mL of filtered demineralized water supplemented with an equimolar concentration of NH_4HCO_2 to prevent osmotic shock and subsequently dried overnight at 100°C before weighing.

4.2.4 Starch analysis

The starch content was analyzed using the Total Starch assay (Megazyme International, Wicklow, Ireland) following the protocol described previously [33]. The method is based on enzymatic degradation of starch to glucose monomers by α -amylase and amyloglucosidase enzymes and measuring glucose monomers in a spectrophotometric-based assay for quantification against a D-glucose calibration control series at a wavelength of 510 nm.

4.2.5 Total fatty acid analysis

Total fatty acid (TFA) extraction and quantification were executed as described by Breuer *et al.* [154] with the following adjustments. Around 5 mg of pellet was transferred to bead beating tubes (Lysing Matrix E; MP Biomedicals, Santa Ana, CA, USA) and lyophilized overnight. Freeze-dried cells were disrupted by a 30-min bead beating step in the presence of a chloroform-methanol mixture (1:1.25) to extract the lipids from the biomass. Tripentadecanoin (T4257; Sigma-Aldrich, St Louis, MO, USA) internal standard was added to the extraction mixture to enable fatty acid quantification. For TFA analysis, samples were directly methylated (see below). For TAG analysis, directly after the TFA extraction, the chloroform methanol mixture was evaporated under N_2 gas and the TFA fraction was dissolved in 1 mL hexane and separated based on polarity using a Sep-Pak Vac silica cartridge (6 cc, 1,000 mg; Waters, Milford, MA, USA) prewashed with 6 mL of hexane. The neutral TAG fraction was eluted with 10 mL of hexane-diethyl ether (87:13% v/v). The polar lipid fraction containing the glycolipids and phospholipids remained in the silica cartridge. Methylation of the fatty acids to fatty acid

methyl esters (FAMES) and the quantification of the FAMES were performed as described by Breuer [154].

4.2.6 Bodipy staining

The presence of neutral lipid bodies in *N. oleoabundans* was measured by staining the cells with the fluorescent dye BODIPY 505/515 (4,4-difluoro-1,3,5,7-tetramethyl-4-bora-3a,4a-diazasindacene; Invitrogen Molecular Probes, Carlsbad, CA). An aliquot of 200 μ L of cells was incubated for 10 minutes with 4 μ L of a 40 μ M BODIPY stock solution (in 0.2% (v/v) DMSO) and subsequently studied using a confocal laser scanning microscope (LSM510; Carl Zeiss, Jena, Germany), using a 488 nm Argon Laser [27]. To increase the visualization, a color filter was applied to visualize the BODIPY fluorescence as yellow signal.

4.2.7 NMR analysis

NMR analysis was performed according to the protocol described by Kim et al. [155] with the following modifications. Freeze dried microalgal biomass (20 mg) was dissolved in 0.5 ml of 50% deuterated methanol in buffer (90 mM KH_2PO_4 in D_2O) containing 0.05% trimethyl silyl propionic acid sodium salt (TMSP, w/v) as internal standard. To effectively extract the metabolites, the cell suspension was vortexed and ultrasonicated for 10 min and centrifuged at 17,000xg for 5 min. From the supernatant, 300 μ l was used to perform NMR analysis. The NMR analysis and data analysis was carried out as described by Kim et al. [156]. The NMR plots can be found in S3 Fig.

4.2.8 RNA extraction

Samples for RNA isolation were immediately processed after sampling and kept on ice. Cells were collected by centrifugation at 4.000xg at 0°C for three minutes and immediately frozen in liquid nitrogen before storage at -80°C until further extraction. Cells were disrupted by grinding the pellet using mortar and pestle and liquid nitrogen. A 5-ml volume of heated (65°C) phenol-chloroform and 5 ml of extraction buffer (10 mM EDTA, 1% sodium dodecyl sulphate (SDS), 2% 2-mercaptoethanol and 200 mM sodium acetate, pH 5) was added to the ground

biomass. The RNA was precipitated by addition of 1/3 volume 8M lithium chloride (LiCl) enriched with 1% 2-mercaptoethanol. The RNA pellet was washed with 2M LiCl and twice with 70% ethanol. After evaporation of the last residues of ethanol, the pellet was resuspended in RNase free H₂O and sequencing and quality control was outsourced to BaseClear BV (Leiden, The Netherlands).

4.2.9 RNA sequencing and transcriptome assembly

RNA-seq library preparation and deep sequencing on Illumina HiSeq2500 instruments were carried out at BaseClear, (Leiden, The Netherlands). The Illumina TruSeq RNA sample preparation protocol was used to prepare libraries with a median fragment insert size of 230 bp. For all samples, 51 nt paired-end sequencing runs were carried out, data was delivered in Illumina format 1.8 and filtered for reads that did not pass the Illumina chastity filter, reads that aligned to the phiX control genome and reads that contained Illumina TruSeq adapters. The sequence data can be found at EBI ArrayExpress with accession number E-MTAB-3746.

To maximize the diversity and completeness of the *N. oleoabundans* *de novo* assembled transcripts, the data from 16 transcriptomes were combined yielding a total of 496,158,724 high quality reads and assembled with IDBA-UD v1.1.0 [157]. QUASt v2.3 [99] was used to estimate the quality of the assembly. The average number of read pairs per experiments was approximately 18 million, the GC content was 59%. Coding sequences (CDS) were extracted using QUASt and translated into protein sequences for functional annotation.

4.2.10 Transcriptome annotation

The protein sequences were annotated by aligning them against different databases by using DELTA-BLAST 2.2.29+ (default parameters, E-value < 0.001) [91] and by using InterproScan 5 (default parameters) for domain search analysis. Blast2GO V.2.7 [92] was used as the central tool to combine both analysis methods to assign GO terms to the protein sequences and to retrieve EC numbers. For DELTA-BLAST the following databases were sequentially used: SwissProt, Chlorophyceae branch from SwissProt, Viridiplantae and Cyanobacteria branch from NCBI, Uniprot (SwissProt + TrEMBL) filtered for proteins with an annotated enzymatic reaction.

Three tools were used to predict the subcellular location of the predicted enzymes: TargetP [100], PredAlgo [37] and WoLF PSORT [101]. Four common cellular locations were identified: chloroplast, mitochondrion, secretory pathway, and cytoplasmic. In the case of WoLF-PSORT, the localizations defined as extracellular or plasma membrane were considered as secretory pathways.

4.2.11 Expression analysis

Read abundance estimations were done using the RSEM script from TrinityRNAseq [102] with the default settings. Reads from each experimental condition were mapped onto the set of coding sequences generated with QUASt from the assembled transcriptome.

Data was normalized taking into account the library sizes using Trimmed Mean of M-values [104]. They were further normalized by the CDS length to compute Fragments Per Kilo base of exon per Million fragments mapped (FPKM) using TrinityRNAseq TMM normalization script. CDS with FPKM values lower than 10 in all conditions were discarded and resulted in a reduction of ~13%. Finally, FPKM values corresponding to CDS annotated to the same enzyme or transporter were added up to provide a single expression value for the reactions in the model.

4.3 Results and discussion

4.3.1 Biomass composition

N. oleoabundans was cultivated under four different conditions to assess the difference in gene expression during growth (nitrogen-replete) and TAG accumulation (nitrogen-deplete) in fresh water (A and B) and salt water (C and D), and the difference in gene expression for fresh and salt water at nitrogen-replete conditions (growth) (A and C) and nitrogen-deplete conditions (TAG accumulation) (B and D) (Panel A in S1 Fig). Samples were taken at a time point when Triacylglycerol (TAG) accumulation was induced but cells were not so stressed that apoptotic gene expression dominated the transcriptional profile (Panel B in S1 Fig). In Figure 4.1 (A-D), oil body formation in the *N. oleoabundans* cells is shown in the four different growth

conditions. The size of the cells grown in fresh water is much smaller with a diameter of approximately 4 μm , compared to the size of cells grown in salt water with a diameter of approximately 8 μm . The red fluorescence represents the autofluorescence of the chlorophylls and the neutral lipid bodies that are stained by BODIPY have a yellow fluorescence signal. Under nitrogen-deplete conditions, there is significantly more BODIPY fluorescence indicating accumulation of neutral lipids that mainly consist of TAG molecules. Not all cells show the same neutral lipid content under nitrogen-deplete conditions. This heterogeneity between cells has been observed previously in *N. oleoabundans* [27, 104]. This could be explained by the fact that, under constant light conditions, not every cell is in the same stage of the cell cycle when the TAG accumulation phase is induced. Individual differences in metabolic activity and internal nitrogen pool might determine whether cells start to produce TAG or not. The total fatty acid (TFA) content increased as a consequence of the increase in TAG when the cells are exposed to nitrogen-deplete conditions (Figure 4.1E). In the first 24 hours of nitrogen depletion, *N. oleoabundans* accumulates $12.1\% \pm 1.4$ and $10.3\% \pm 0.8$ of their dry weight as TAG molecules in fresh water and salt water conditions, respectively. The TFA and TAG content is therefore comparable for the salt water adapted cells compared to the fresh water culture, and the TAG fraction of the TFA content is the same under both conditions. In Figure 4.1F (or S1 Table) the fatty acid profiles are shown for both TFA and TAG relative to the total cell dry weight. No significant changes in fatty acid composition are observed when comparing fresh water with salt water conditions, indicating that the salt resistance of *N. oleoabundans* is unlikely to be a consequence of modifications of the plasma membrane lipid composition. This is different for *Dunaliella salina*, which induces fatty acid elongation and expression of their related genes under saline stress [158].

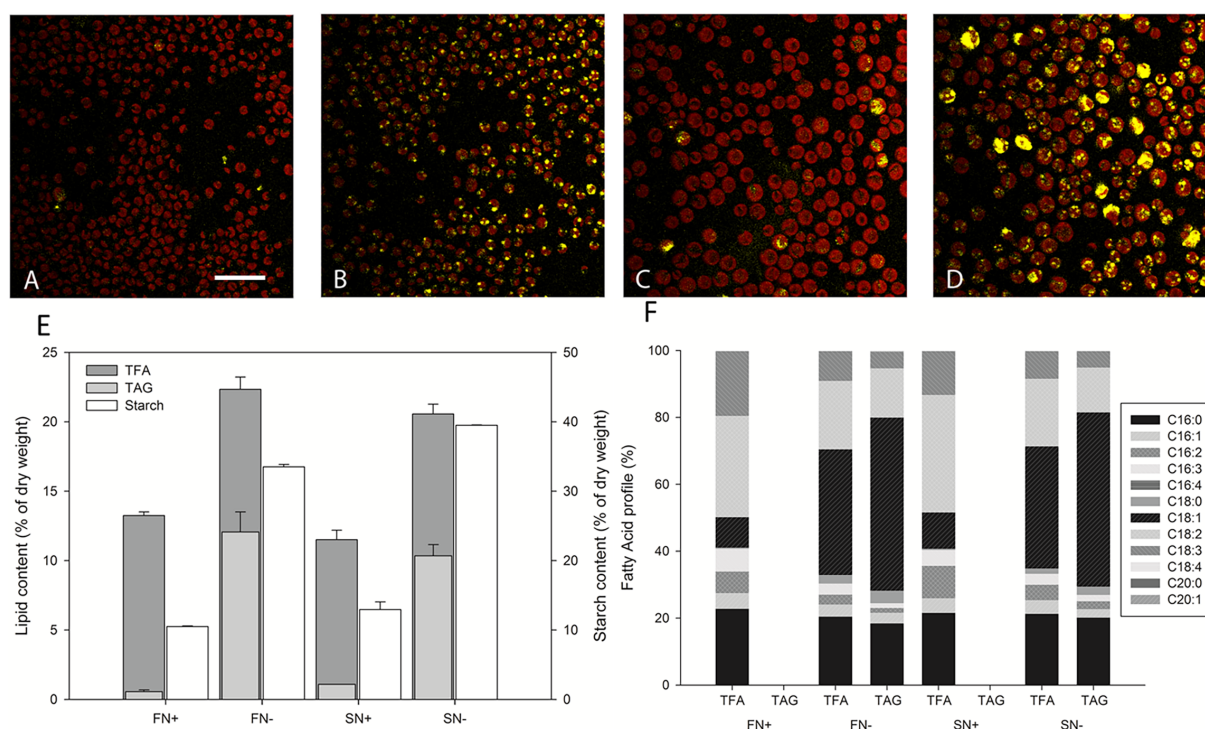


Figure 4.1: Biomass composition of *N. oleoabundans* in the four tested conditions. Upper panel: Confocal laser scanning microscope images of *N. oleoabundans* under four different cultivation conditions. (A) Nitrogen-replete fresh water (FN+). (B) Nitrogen-deplete fresh water conditions (FN-). (C) Nitrogen-replete salt water conditions (SN+). (D) Nitrogen-deplete salt water conditions (SN-). Chlorophyll autofluorescence is shown in red and the BODIPY stain is shown in yellow. The bar represents 20 μ m. Lower panel: (E) Lipid and starch content. TFA (dark grey), TAG (light grey), and starch (white) 24 hours after medium replacement. The height of the bars represents the average of the two independent measurements. Error bars represent distance of the sample values to the average value. TAG, TFA and starch are given as a percentage of total dry weight. (F) Fatty acid profile under the different conditions expressed as percentage of TFA.

4.3.2 De-novo transcriptome assembly and annotation

For the *de novo* assembly, RNAseq samples were pooled, amounting to 496,158,724 reads yielding 30489 contigs with a N50 of 2372 bp. Using a lower bound in contig size of 500 bp, 18097 contigs remained with a total length of 30,540,822 bp. QUAST [99] predicted these contigs to encode 32,136 protein coding sequences (CDS) with a total length of 20,517,587 bp. InterProScan5 [44] subsequently identified the presence of conserved protein domains in 54% (15,306) of these CDS. Additionally, 7021 putative proteins could be associated with Gene

Ontology (GO) terms, of which 83% (5861) were unique. Finally, 3249 putative proteins were associated with EC numbers, 834 of which were unique. All sequences of genes, proteins, and the mentioned annotation results can be found in S1 File. At this time, there are two published studies of *N. oleoabundans* grown under nitrogen-deplete conditions, one with transcriptomics data and the other with proteomics [159, 160]. The experimental conditions are fundamentally different from our set-up, which makes it difficult to compare the results directly.

4.3.3 Compatible solutes

Compatible solutes are highly soluble low molecular weight molecules that can be accumulated to high concentrations without being toxic to the cell. They can protect cells against drought or saline stress by regenerating cellular osmotic homeostasis, relieving oxidative stress caused by Reactive Oxygen Species (ROS), and protecting membrane integrity and stabilization of enzymes or proteins [161]. Some examples of compatible solutes are, free amino acids, sugars, polyols and quaternary ammonium compounds (QAC). We used the transcriptomic landscape under stress (nitrogen-deplete) conditions in comparison to the unstressed conditions as proxy to identify compounds that may act as a compatible solute or otherwise might be involved in protecting the algae against salt stress. For some of these compounds identified with RNAseq data we were able to confirm their role by measuring the intracellular concentrations.

4.3.3.1 Sugars

Several sugars are known to have a protective role in organisms experiencing different stress conditions [132]. Examples of osmoprotectant sugars are trehalose and sucrose. In most marine and freshwater cyanobacteria sucrose is synthesized from fructose-6-phosphate and UDP-glucose by the enzymes sucrose phosphate synthase (SucPS, EC:2.4.1.14) and sucrose-phosphate phosphatase (SucPP, EC:3.1.3.24) [162]. *Anabaena sp.* synthesizes sucrose in one step converting fructose and ADP or UDP glucose using sucrose synthase (SucST, EC:2.4.1.13) [163]. In *N. oleoabundans*, transcription of several genes in the sucrose biosynthesis pathway are up-regulated under saline growth conditions. Transcripts are considered differentially expressed when the FPKM values are at least log₂ fold change (LFC) 0.59 (FC 1.5) compared to the reference condition. In this pathway, D-glucose-1P derived from glycolysis is converted into UDP glucose by the enzyme G1PUT which is marginally overexpressed under saline growth conditions. The conversion from UDP-glucose to sucrose is catalyzed by the enzyme

sucrose synthase (SucST) and is strongly overexpressed under-nitrogen-deplete conditions (Figure 4.2).

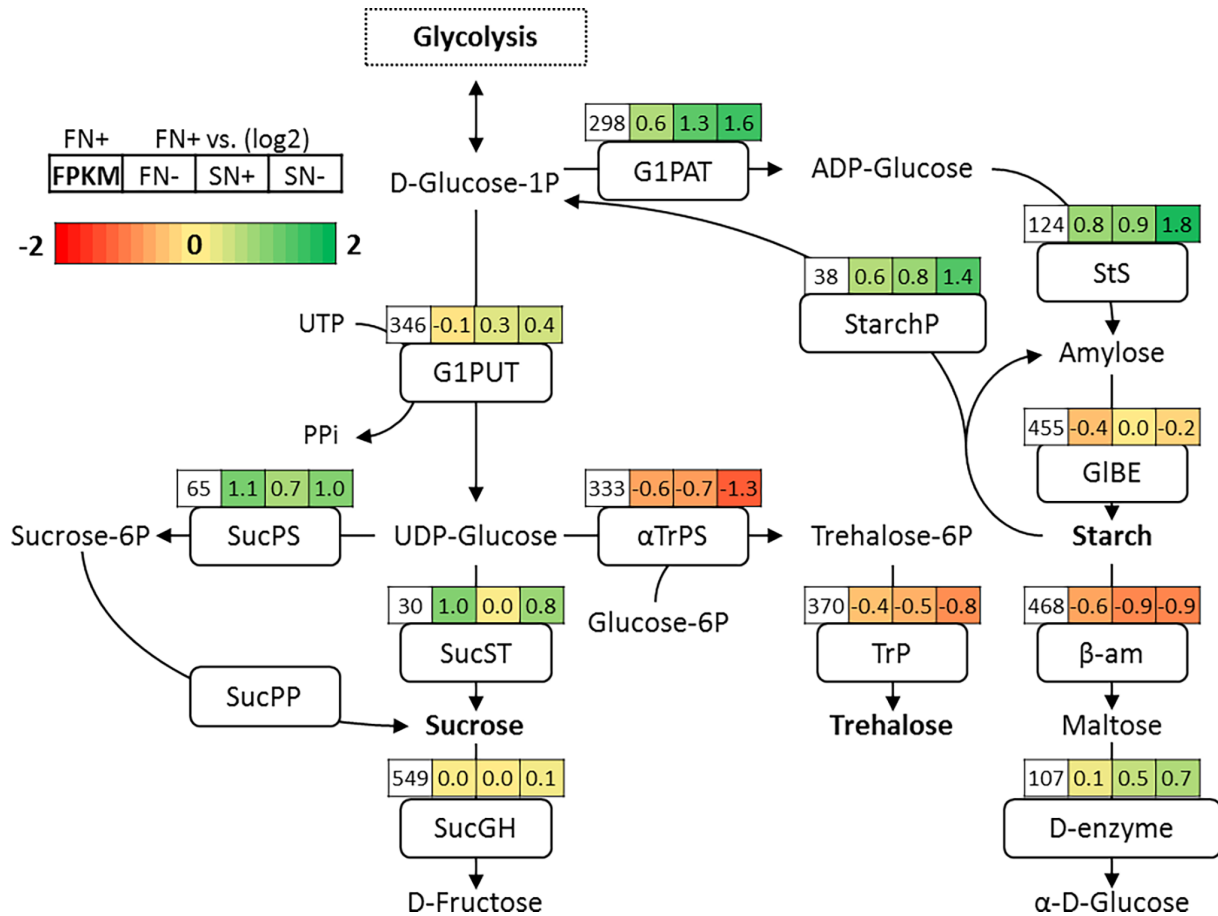


Figure 4.2: Biosynthesis pathway of sucrose and starch. The values that are shown in the following figures refer to the Fragments Per Kilobase of transcript per Million mapped reads (FPKM) for each condition. The white left-most box represents the FPKM value of the respective gene in the fresh water nitrogen-replete reference condition. This is followed by three colored boxes that represent the log2 fold change (LFC) of the other conditions compared to the reference condition. The order of the remaining three boxes are from left to right, FN-, SN+, SN- respectively. Abbreviations: G1PUT: glucose-1-phosphate uridylyltransferase; UTP: Uridine triphosphate. SucPS: sucrose-phosphate synthase. SucST: sucrose synthase. G1PAT: glucose-1-phosphate adenylyltransferase. αTrPS: alpha,alpha-trehalose-phosphate synthase. αTrS: alpha,alpha-trehalose synthase. TrP: trehalose-phosphatase. StS: starch synthase. GIBE: glycogen branching enzyme. SucPP: sucrose-phosphate phosphatase. SucGH: sucrose glucohydrolase. β-am: β-amylase. D-enzyme: 4-alpha-glucanotransferase. StarchP: Starch phosphorylase.

The first step in the pathway that goes from UDP-glucose to sucrose via sucrose-6P catalyzed by SucPS is up-regulated under both salt water and nitrogen stressed cells. The second step catalyzed by sucrose-phosphate phosphatase (SucPP) is absent from the annotation. Sucrose

can be degraded by sucrose glycohydrolase (SucGH) into D-fructose. Band et al. found that the main soluble carbohydrate that is accumulated in *N. oleoabundans* experiencing a salt osmotic up-shock is sucrose [147]. Based on the sucrose levels measured and the expression levels of the genes involved, sucrose might be functioning as quick response to saline stress [147], but does not seem to be responsible for salt resistance in long term salt adapted cells. The NMR spectroscopy analysis revealed that sucrose concentrations may be increased under saline and nitrogen stress conditions (S2 Fig). Sucrose accumulation seems to be more of an overflow mechanism that results from nitrogen depletion.

Another sugar that is often found to be involved in salt resistance is trehalose. In high-temperature stressed yeast cells, trehalose concentrations are increased to protect enzymes from the elevated temperatures [164]. Trehalose is found in different microalgal species under saline stress conditions [165, 166], but does not seem to be an osmoprotectant in *N. oleoabundans*, since key enzymes in this pathway are not transcriptionally up-regulated under saline conditions (Figure 4.2). This observation does however not completely exclude trehalose as a candidate for osmoprotection, because trehalose concentration was not analyzed directly and it could be that the flux toward this compound is not controlled at transcript level but on a metabolic level.

4.3.3.2 Proline

Proline was already found to have many positive effects to cope for osmotic stress in bacteria, plants and algae [132, 167–171]. We summarize the effects and pathway regulations for proline as found in diverse studies in S1 Text.

In *N. oleoabundans*, many genes upstream of proline starting from the tricarboxylic acid (TCA) cycle are up-regulated under saline growth conditions. Based on the transcript levels it appears that the major route from the TCA cycle intermediate 2-oxo-glutarate towards L-glutamate is catalyzed via the enzyme glutamate dehydrogenase (GDH). There are two enzymes known that can catalyze this reaction: EC:1.4.1.3 and EC:1.4.1.4 using different co-factors. Both enzymes are up-regulated under saline growth conditions, the first by LFC 0.7 in SN+ and by LFC 0.8 in SN- and the second by much stronger LFC 7.6 (FC ~200) in SN+ and LFC 8.5 (FC ~350) in SN-(Figure 4.3). The enzyme complex P5CS1 and P5CS2 converts L-glutamate into L-glutamyl-P and GSA respectively. Both enzymes are strongly up-regulated under saline growth conditions. For the P5CS1 gene this is LFC 0.9 times and LFC 0.6 times for nitrogen-replete and -deplete saline growth conditions respectively. For the P5CS2 enzyme, the up-regulation under these conditions are LFC 1.5 and LFC 1.4 respectively. The alternative way

to produce GSA by OAT from ornithine seems to be less important and only active under nitrogen-replete conditions and with slightly higher expression under salt water conditions (Figure 4.3). In contrast, the diatom *Fragilariopsis cylindrus* is primarily using the ornithine pathway to generate proline as a response to salt stress [172]. The final conversion from P5C to proline through the P5CR reaction, is markedly up-regulated under saline conditions. However, this overexpression is stronger in SN+ than SN- (LFC 2.3 and LFC 0.7, respectively) Furthermore, catabolism of proline is done by proline dehydrogenase (PRODH, EC:1.5.99.8) and is upregulated in all conditions in comparison to FN+. Most importantly, PRODH overexpression was stronger in SN- than SN+ (LFC 2.0 and LFC 1.4 respectively). Thus, based on this proline synthesis would be more upregulated and breakdown less upregulated under the SN+ as compared to the SN- condition.

Proline content was then studied for all four conditions by NMR spectroscopy, results are shown in S2 Fig. Although the differences between the duplicate measurements are large, still an increase in proline content can be observed for the SN+ condition, which is in agreement with the transcriptome results. The cause for the large difference between the duplicate measurements is likely due to the difficulty to disrupt the cell wall of *N. oleoabundans*. Furthermore, no increase in proline content is measured for the SN- condition. This could be due to the upregulated breakdown pathways under the SN- condition and the fact that in the SN- condition nitrogen, which is needed for proline synthesis, is absent.

To conclude, since proline levels are increased in salt water adapted cells and all genes involved in the proline biosynthesis are up-regulated, proline is most likely to be the primary mechanism in the saline resistance of *N. oleoabundans*. Furthermore, under nitrogen-deplete conditions, PRODH is up-regulated possibly to enable the recycling of nitrogen to use the nitrogen molecules from proline in other nitrogen requiring processes in the cell.

More detailed descriptions and analysis of Proline and other potential osmoregulatory mechanisms can be found in S1 Text.

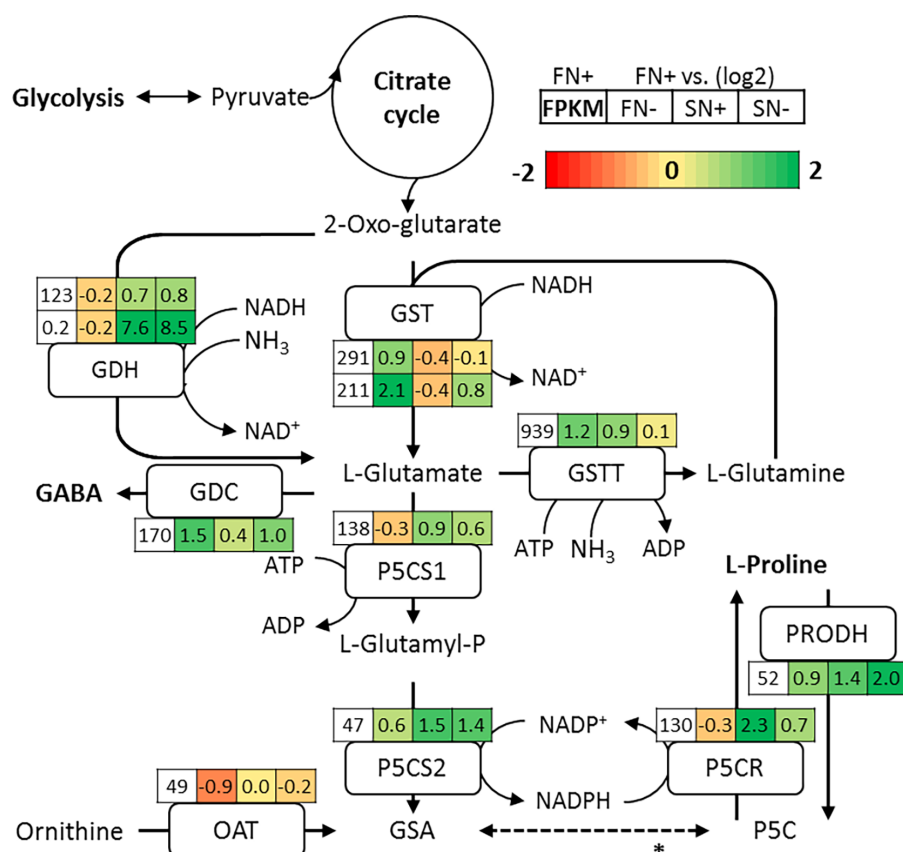


Figure 4.3: Proline and GABA biosynthesis pathway. The figure legend refers to Figure 4.2
 Abbreviations: GSA: glutamate-semialdehyde. P5C: L-1-Pyrroline-5-carboxylate. GABA: 4-aminobutanoate. P5CR: P5C reductase. P5CS1: glutamate-5-kinase. P5CS2: GSA dehydrogenase. GDH: glutamate dehydrogenase (EC:1.4.1.3/4). GST: glutamate synthase (EC:1.4.1.13/14). GSTT: glutamine synthetase. GDC: glutamate decarboxylase. OAT: ornithine aminotransferase. PRODH: Proline dehydrogenase. *This reaction occurs non-enzymatically.

4.3.4 Oxidative stress

Under stress conditions ROS can be formed which need to be removed to prevent damage to the photosystems and other cellular equipment [173]. Glutathione (GSH) is a pivotal compound for many plant species and microalgae to scavenge ROS, such as superoxide hydrogen peroxide and lipid hydroperoxides, which can be accumulated under environmental and oxidative stress [174]. In addition to direct scavenging of ROS, GSH can also function as the reductant in the glutathione-ascorbate cycle which can alleviate ROS build up and provide protein protection. The tripeptide glutathione (GSH) is synthesized from glutamate and

cysteine into γ -glutamylcysteine by γ -glutamylcysteine synthetase (γ -GCSTT EC:6.3.2.2) at the expense of one ATP. In the next step, another ATP molecule is needed to convert γ -glutamylcysteine and glycine into glutathione, by glutathione synthetase (GSHSTT EC:6.3.2.3) (Figure 4.4).

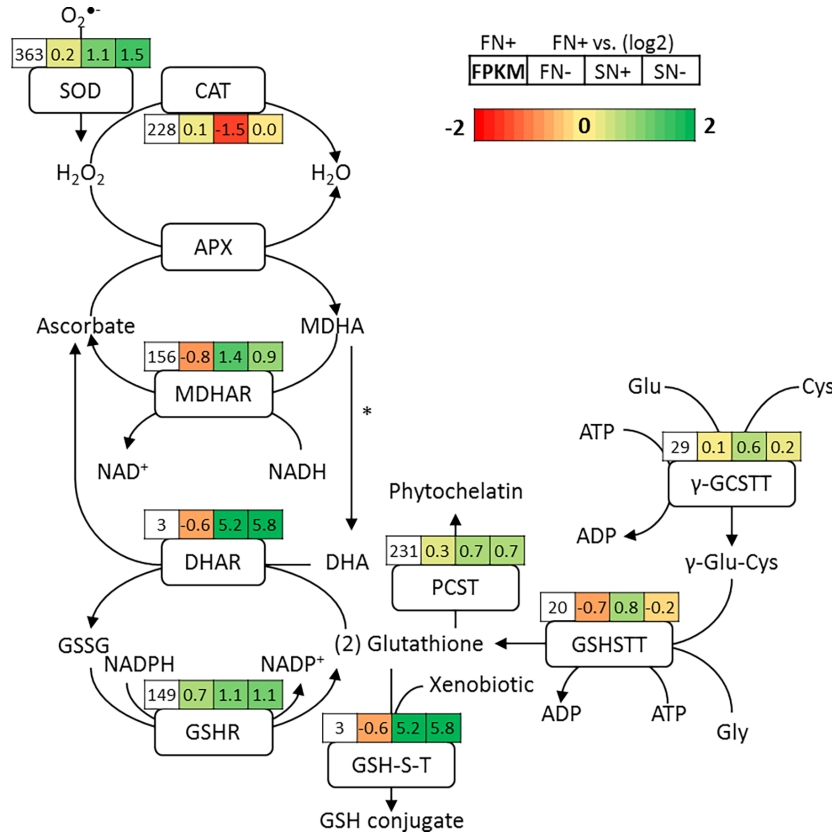


Figure 4.4: Glutathione biosynthesis pathway and glutathione-ascorbate cycle. For the figure legend refers to Figure 4.2. Abbreviations: SOD Superoxide dismutase; CAT Catalase; APX Ascorbate peroxidase (not annotated); MDHA monodehydroascorbate; MDHAR monodehydroascorbate reductase; DHA dehydroascorbate; DHAR dehydroascorbate reductase; GSH (reduced) Glutathione; GSSG oxidised glutathione; GSHR Glutathione reductase; Glu Glutamate; Cys Cysteine; Gly Glycine; γ -GCSTT γ -glutamylcystein Synthetase; γ -Glu-Cys γ -glutamylcystein; GSHSTT Glutathione Synthetase; GSH-S-T glutathione S-transferase; PCST Phytochelatin synthase. * This reaction can occur enzymatically or non-enzymatically.

The oxidative stress can be relieved by antioxidant enzymes. First, superoxide can be converted into hydrogen peroxide by the superoxide dismutase (SOD; EC:1.15.1.1). SOD is highly overexpressed under saline conditions LFC 1.1 and LFC 1.5 in nitrogen-replete and depleted conditions respectively (Figure 4.4). Second, the toxic hydrogen peroxide can be converted to H₂O by catalase (CAT, EC 1.11.1.6) or by ascorbate peroxidase (APX EC:1.11.1.11). CAT

expression is stable under nitrogen-deplete conditions and strongly down regulated under nitrogen-replete saline conditions. Such a regulation of CAT suggests that APX is the preferred way to relieve the harmful radicals.

Unfortunately, APX could not be annotated in *N. oleoabundans* from the mRNA sequences. MDHA, which is produced by APX, needs to be converted back to ascorbate via two different ways. Enzymatically, by monodehydroascorbate reductase (MDHAR EC:1.6.5.4), or non-enzymatically by spontaneous disproportionation of two MDHA molecules resulting in ascorbate and dehydroascorbate (DHA) [175]. DHA will subsequently be converted back to ascorbate by dehydroascorbate reductase (DHAR EC:1.8.5.1) coupling two GSH molecules into glutathione disulfide (GSSG). Glutathione reductase (GSHR EC:1.8.1.7) can catalyze the reduction of GSSG back to two GSH molecules. The MDHA reductase gene is LFC 1.4 increased under saline replete conditions and LFC 0.9 increased under nitrogen-deplete conditions. DHA reductase was strongly up-regulated under salt water conditions to regenerate the ascorbate from DHA. DHAR was LFC 5.2 (FC 37) and LFC 5.8 (FC 56) up-regulated under nitrogen-replete and -deplete salt water growth conditions respectively. To recycle the GSH molecules, GSHR is up-regulated under salt water nitrogen-replete and -deplete conditions as well, namely LFC 1.1 and LFC 1.1 times respectively.

The efficiency of GSH is dependent on its concentration in the cell, which in turn depends on the activity of the GSH reductase enzyme, which determines the ratio between GSH and its oxidized form GSSG. As shown, all genes involved in the ascorbate-GSH cycle were up-regulated under salt water growth conditions, including the biosynthesis of GSH (Figure 4.4).

Another function of GSH is the detoxification of xenobiotics, compounds that have no significant nutritional function in cell metabolism, but do affect cellular homeostasis in too high concentrations, this in contrast to compatible osmolytes that do not interfere with metabolism when present in high concentrations. The xenobiotic molecules can be conjugated to GSH, by GSH-S-transferase (GSHT, EC:2.5.1.18), and transported to vacuoles to be detoxified. Under saline conditions this enzyme is strongly up-regulated, LFC 5.2 and LFC 5.8 under nitrogen-replete and -deplete saline conditions respectively. This indicates the likeliness of GSH conjugate formation and exclusion to relieve xenobiotic pressure under saline conditions (Figure 4.4). It is known in plants that GSHT is induced under salt and drought stress to reduce the ROS in plants and GSHT could also be involved in reducing the harmful byproducts of oxidative stress such as lipid peroxidation [176, 177].

A third protecting feature of GSH is the formation of phytochelatin (PC) by the enzyme phytochelatin synthase (PCST EC:2.3.2.15). PCs are oligomers of GSH and are able to detoxify

heavy metals by chelation of toxic ions [178, 179]. The PC-ion complexes can be compartmentalized in the vacuole of plants where they can do no further harm. PC has also been found to be involved in stress responses other than to heavy metals. In the cyanobacterium *Anabaena doliolum* PCs are produced in response to UV-B radiation [180]. The cloning of the PC synthase gene of *Anabaena sp.* into *E. coli* increased the resistance of this bacterium to heat, metals, UV-B, salt, and herbicides [180]. Some plants have been shown to produce PCs in response to heat or salt stress [181, 182]. The *N. oleoabundans* PCST gene is up-regulated in salt water medium under both nitrogen conditions (LFC 0.7), indicating that PCs are likely to be formed to protect *N. oleoabundans* against osmotic stress.

Based on the general increase in expression of the genes involved in the ascorbate glutathione cycle, it is very likely that *N. oleoabundans* is using this cycle to alleviate the pressure of ROS that arise under saline growth conditions and that the GSH derived conjugates and phytochelatin GSH oligomers are likely to be involved as well. This mechanism is not used to alleviate the ROS pressure under nitrogen stress, at least not under the tested conditions.

4.3.5 Starch and triacylglycerol accumulation

4.3.5.1 Starch pathway

Starch is known to be transiently accumulated under nitrogen-deplete conditions in the beginning of the stress phase [33, 34, 183]. In this study, the sample time was 24 hours after nitrogen stress induction and it is expected that the genes involved in starch biosynthesis would be up-regulated in this early phase of nitrogen stress. The starch levels under the four different conditions are shown to be strongly increased in the first 24 hours after nitrogen depletion. Under fresh water conditions, the starch content increased from $10.5\% \pm 0.1$ to $33.5\% \pm 0.4$ when switched to nitrogen depletion and from $12.9\% \pm 0.1$ to $39.5\% \pm 0.0$ under nitrogen depletion in salt water conditions (Figure 4.1E). Interestingly, salt water adapted cells had a 1.2 times higher starch content compared to the fresh water adapted cells under both nitrogen-replete and -deplete conditions.

The genes encoding glucose-1-phosphate adenylyltransferase (G1PAT, EC:2.7.7.27) and starch synthase (StS, EC:2.4.1.21) are strongly up-regulated under nitrogen-deplete conditions and seems to be up-regulated by salt water as well (Figure 4.2). The G1PAT gene is LFC 0.6 times up-regulated in FN-, while this gene is up-regulated by LFC 1.3 in SN+ and by LFC 1.6 times

SN-. Starch accumulation upon nitrogen depletion seems to be facilitated by transcriptional up-regulation of starch biosynthesis genes with the exception of the final step catalyzed by glycogen branching enzyme (GIBE, EC:2.4.1.18). Starch catabolism can be facilitated by different enzymes. β -amylase (β -am, EC:3.2.1.2) degrades starch by removing maltose units from the non-reducing ends of the chains. β -am was shown to be down-regulated under salt water and nitrogen-deplete conditions (Figure 4.2). Those results correlate with another nitrogen deplete transcriptomics study of *N. oleoabundans* [159] in which starch is also accumulated, the glucose branching enzyme (GIBE) is slightly down-regulated, and β -amylase is down-regulated. However, that study observed down-regulation of starch building enzymes G1PAT and StS, while in this study, we observed up-regulation of those genes. Unexpectedly, starch phosphorylase (StarchP, EC:2.4.1.1) and α -amylase (α -am, EC:3.2.1.1), two other starch degrading enzymes, are up-regulated under nitrogen-deplete conditions when there is accumulation of starch (S1 Table). While the change of expression is rather strong for StarchP, the change of expression of α -am is very small and cannot be considered significant. A possible explanation could be that during nitrogen depletion, starch turnover is accelerated with a faster anabolism than catabolism, while the starch degradation products are being redirected towards other sugars and TAG. Salt stress does seem to increase the starch content, which is possibly a result of reduced growth creating a surplus of energy that needs to be channeled from the photosystems [184].

4.3.5.2 *Triacylglycerol pathway*

When oleaginous microalgae are exposed to unfavorable growth conditions such as nitrogen-deplete conditions, cells start to accumulate triacylglycerol (TAG). These neutral glycerolipids are composed of a glycerol backbone with three fatty acid molecules attached. The fatty acid biosynthesis takes place in the chloroplasts and starts downstream of the glycolysis with the conversion of acetyl-CoA to malonyl-CoA by the heteromeric enzyme complex acetyl-CoA carboxylase (ACCase). The acetyl-CoA pool can be supplied in two different ways. The first way is via pyruvate and the pyruvate dehydrogenase complex, in which the first two steps are up-regulated under nitrogen-deplete conditions in *N. oleoabundans* (EC:1.2.4.1 and EC:1.8.1.4, see Figure 4.5). Another way is via the TCA cycle in which citrate can be converted into acetyl-CoA and oxaloacetate by ATP citrate lyase (EC:2.3.3.8) [185]. The *N. oleoabundans* ATP citrate lyase is up-regulated under nitrogen-deplete conditions and it is likely that this pathway is involved in supplying part of the acetyl-CoA.

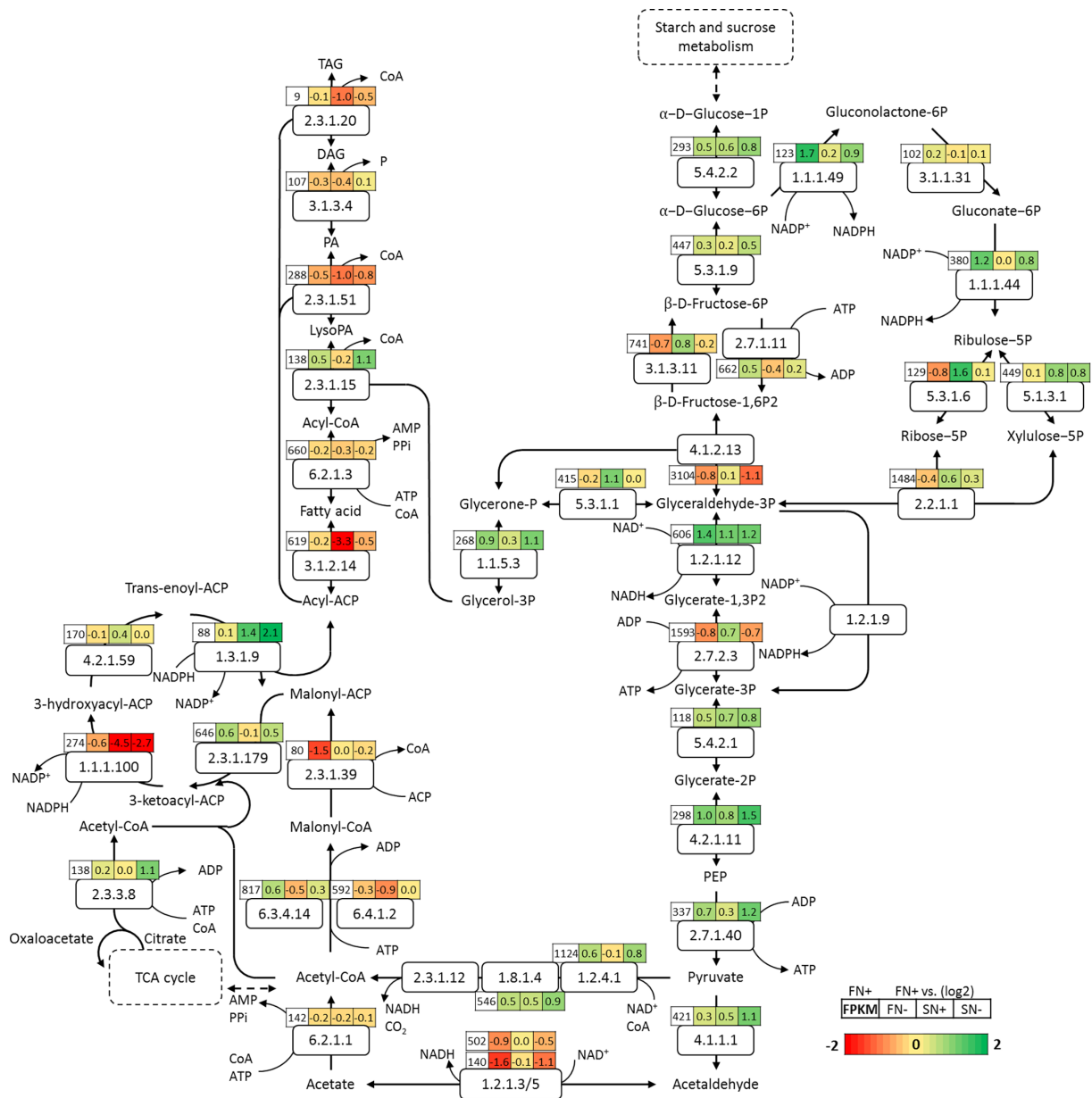


Figure 4.5: Carbon metabolism in *N. oleoabundans*. For the figure legend refers to Figure 4.2. Codes in the white boxes represent the corresponding EC numbers. Abbreviations: PEP phosphoenol pyruvate; LysoPA Lysophosphatidic acid; PA Phosphatidic acid; DAG 1,2-Diacylglycerol; TAG Triacylglycerol. EC:2.3.1.12 could not be annotated from the transcriptome of *N. oleoabundans*.

ACCase catalyzes an important step in fatty acid biosynthesis and overexpression of this enzyme resulted in increased lipid contents in *Arabidopsis* [186], but not necessarily in microalgae [187]. Over expression of the ACCase complex is challenging because of the multigene-encoded enzyme complex and post-translational modifications [188, 189]. Next, an acyl-carrier protein (ACP) is exchanged for the CoA moiety by malonyl-CoA:ACP transacylase giving rise to a malonyl-CoA molecule. Malonyl-CoA enters the heteromultimeric fatty acid

synthesis (FAS) cycle where it is extended with two carbon atoms per repetitive cycle to usually 16 or 18 carbon long acyl-ACP groups.

In this study, two subunits of the enzyme complex ACCase could be identified, biotin carboxylase (EC: 6.3.4.14) and biotin carboxyl carrier (EC:6.4.1.2). The biotin carboxylase subunit is up-regulated under nitrogen-deplete conditions, and the biotin carboxyl carrier has little (fresh water) to no (salt water) down-regulation under nitrogen-deplete conditions (Figure 4.5). These results thus show that the biotin carboxylase is the main regulator for ACCase complex which is in agreement with the results from both aforementioned *N. oleabundans* studies [159, 160].

Our results show that the third reaction of the FAS cycle did not display any changes in expression (Figure 4.5), while the first (EC:2.3.1.179) and last (EC:1.3.1.9) step of the FAS cycle displayed up-regulation: the first step seems to be up-regulated due to nitrogen depletion, whereas the last step is strongly up-regulation due to salt water conditions. In contrast, the second reaction of the FAS cycle displayed light down-regulation due to nitrogen depletion and displayed very strong down-regulation due to salt water. In agreement with the other transcriptome study of *N. oleabundans* [159], Acyl-ACP desaturase (AAD) was found to be significantly up-regulated, which also correlates with the very similar change in fatty acid profile in that study being mainly an increase in C18:1. Since the TAG content is strongly increased in the first 24 hours of nitrogen starvation (Figure 4.1), it was expected that transcription of the genes encoding the enzymes involved in the TAG synthesis pathway would be up-regulated under nitrogen-deplete conditions. This was not the case with the exception of the glycerol 3-phosphate acyltransferase (EC:2.3.1.15) gene, which showed a LFC 0.5 and LFC 1.1-under fresh water and salt water nitrogen-deplete conditions respectively (Figure 4.5). This enzyme links glycolysis and the TAG synthesis pathway by attaching the first acyl-ACP molecule to sn-Glycerol 3-phosphate resulting in lysophosphatidic acid. Glycerol-3-phosphate can be formed from glycerol via glycerol kinase (EC:2.7.1.30) or from dihydroxyacetone phosphate via glycerol-3-phosphate dehydrogenase (EC:1.1.5.3). The latter seems to be the case in *N. oleabundans* since transcription of the gene for this enzyme is strongly up-regulated under nitrogen-deplete conditions. The overexpression of a yeast glycerol-3-phosphate dehydrogenase in rape seed, resulted in a 40% increase of the seed oil content, indicating that the available glycerol backbones are a limitation in TAG accumulation in rape seed [190]. To ensure sufficient supply of G3P, transcription of triose-phosphate isomerase (EC:5.3.1.1), the enzyme that converts glycero-P (Dihydroxyacetone phosphate, DHAP) to glyceraldehyde-3-P, is reduced under nitrogen-deplete conditions (Figure 4.5). Chen

et al. described the increase of DHAP, G3P and glycerol in *Arabidopsis* when triose-phosphate isomerase was knocked out [191]. A similar regulatory mechanism was found in microalgae [51].

The last step from diacylglycerol (DAG) to triacylglycerol (TAG) is catalyzed by diacylglycerol O-acyltransferase (DGAT, EC:2.3.1.20). This enzyme is down regulated in *N. oleoabundans* under all tested conditions compared to the fresh water replete conditions. DGAT has been extensively studied since it is regarded as one of the most important rate limiting steps in TAG biosynthesis. Over expression of this gene or several subunits have resulted in very different outputs. In *Phaeodactylum tricornutum* there have been very promising results, resulting in higher levels of TAG synthesis, but in the green alga *C. reinhardtii* there have not been consistent increases in TAG content in combination with DGAT over expression [192–195]. The fact that DGAT does not seem to be over expressed under nitrogen-deplete conditions is surprising, since TAG accumulation is occurring in *N. oleoabundans* under these circumstances. In this study only one DGAT gene could be identified and it is known from other species that there are several genes that encode for DGAT enzymes and that they have very different expression profiles as a response to nitrogen deprivation [184, 196, 197]. More research needs to be done to understand and identify the different DGAT enzymes in *N. oleoabundans*.

As can be seen in Figure 4.5, most of the genes involved in energy and carbon metabolism are up-regulated under nitrogen-deplete conditions. Under salt conditions, transcription of genes involved in glycolysis, TCA cycle, and pentose phosphate pathway (PPP) were increased. This can be explained by the need for sufficient carbon precursors such as acetyl-CoA and ATP. Furthermore, it was already hypothesized [34] that PPP might be used to redirect the starch degradation products towards TAG biosynthesis. In the PPP, the transcripts for the NADPH generating enzymes are strongly up-regulated to generate the required NADPH that is needed for TAG biosynthesis and to scavenge ROS.

To maintain high levels of TAG molecules, the rate of TAG catabolism should be low [198]. The gene triacylglycerol lipase (EC:3.1.1.3) is highly expressed under nitrogen-replete fresh water conditions (FPKM: 1734). The expression is down regulated under nitrogen-deplete and salt water conditions, LFC -0.3, LFC -0.9, and LFC -0.6 for fresh water nitrogen-depleted, salt water nitrogen-replete and depleted conditions respectively (S1 Table).

4.4 Conclusion

To assess the metabolic and transcriptomic response to nitrogen and salt stress, *N. oleoabundans* was cultured under four different conditions being fresh water and saline water both under nitrogen-replete and nitrogen-deplete conditions. The most likely compatible solute in *N. oleoabundans* is proline because under saline conditions, transcripts involved in proline biosynthesis are up-regulated and NMR indicated that proline is accumulated in the cells. Next to this transcriptome analysis shows that anti-oxidant pathways, like the ascorbate GSH cycle, GSH-conjugation, PC formation and GSH itself, are upregulated under saline conditions, probably protecting the cells from oxidative stress occurring under this condition. Unlike other microalgae, the plasma membrane lipid composition of *N. oleoabundans* was not adjusted in salt water adapted cells. Under Nitrogen-deplete conditions both starch and TAG were accumulated at both fresh water and salt water conditions. With the overall gene expression, we could explain how starch is accumulated. However, no strong correlation was found between accumulation and gene expression in the TAG pathway. Nevertheless, we identified two genes as promising targets to enhance TAG production: glycerol-3-phosphate acyltransferase and glycerol-3-phosphate dehydrogenase. Not only because they link the glycolysis to the TAG biosynthesis pathway, but also because of their expression profile. Overall, the results of this study can be used to develop strategies to enhance salt resistance of *N. oleoabundans* and other industrially relevant microalgal strains and ultimately help to develop a competitive feasible large-scale production of TAG for feed and bio-fuels, while reducing the dependence on precious fresh water resources and arable land.

Chapter 5

Draft Genome Sequence of the Oleaginous Green
Alga *Tetradismus obliquus* UTEX 393

This chapter is published as:

B. M. Carreres, L. de Jaeger, J. Springer, M. J. Barbosa, G. Breuer, E. J. van den End, D. M. M. Kleinegris, I. Schäffers, E. J. H. Wolbert, H. Zhang, P. P. Lamers, R. B. Draaisma, V. A. P. Martins dos Santos, R. H. Wijffels, G. Eggink,^b P. J. Schaap, D. E. Martens.
Draft Genome Sequence of the Oleaginous Green Alga *Tetradismus obliquus* UTEX 393.

Genome announcements, 2017

ABSTRACT

The microalgae *Tetrademus obliquus* is able to maintain a high photosynthetic efficiency under nitrogen limitation and is considered a promising green microalgae for sustainable production of diverse compounds, including biofuels. Here, we report the first draft whole-genome shotgun sequencing of *T. obliquus*. The final assembly comprises 108,715,903 bp with over 1,368 scaffolds.

5.1 Genome Announcement

Microalgae are promising photosynthetic microorganisms for the sustainable production of many compounds ranging from bulk chemicals such as biofuels, to high-value compounds such as pigments and polyunsaturated fatty acids [4]. The freshwater microalga *Tetradesmus obliquus* (previously known as *Scenedesmus obliquus* and reclassified as *Acutodesmus obliquus* [199]) is a member of the *Chlorophyceae* and has four chromosomes [200]. *T. obliquus* UTEX 393 was identified as a promising candidate for industrial applications, because it can accumulate triacylglycerol (TAG) up to 40% of its dry weight. Furthermore, *T. obliquus* is able to maintain a high photosynthetic efficiency under nitrogen limitation for a relatively long period, which results in a high yield of lipids on light and thus high volumetric and areal productivities [13, 34, 201, 202].

The genome was sequenced using Illumina HiSeq2000, yielding nearly 200 million 101-bp reads, with an estimated average insert size of 248 bp. Additionally, more than 8 million 51-bp mate pairs were sequenced, with an estimated average insert size of 4,139 bp. Two methods were used in parallel for assembling and scaffolding, and the resulting scaffold sets were merged. In the first method, contigs were assembled using IDBA-UD version 1.1.1 [157], producing a 96,197,736-bp assembly in 9,191 contigs, and scaffolded using SOPRA version 1.4.6 [203], as previously recommended [204], to obtain 2,509 scaffolds. In the second method, contigs were assembled using CLC Genomics Workbench version 5.1, producing a 93,347,907-bp assembly in 10,609 contigs, and scaffolded using SSPACE to obtain 2,768 scaffolds [205]. The two scaffold sets were merged using *quickmerge* [206] with suitable settings. Additionally, the scaffolds matching 96% of their length over another scaffold were taken out. Between each scaffolding or merging iteration, the gaps were closed using GapFiller [207]. The result was 1,368 scaffolds with a total length of 107,715,903 bp. Using QUAST version 4.3 [99], we compared our assembly to the complete chloroplast genome sequence of UTEX-393 [208], covering 99.97% (~50-bp gap) of the reference with few overlapping scaffolds.

Accession number(s). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers [FNXT01000001](#) to [FNXT01001368](#). The versions described are the first versions, FNXT01000001 to FNXT01001368.

Chapter 6

The diurnal transcriptional landscape of the
microalga *Tetradesmus obliquus*

This chapter is published as:

Benoit M. Carreres, G. Mitsue León-Saiki, Peter J. Schaap, Ilse M. Remmers, Douwe van der Veen, Vitor A. P. Martins dos Santos, René H. Wijffels, Dirk E. Martens, Maria Suarez-Diez

The diurnal transcriptional landscape of the microalga *Tetradesmus obliquus*.

Algal Research, 2019

ABSTRACT

Tetrademus obliquus is a promising oleaginous microalga. We functionally annotated its genome and characterized the transcriptional landscape of *T. obliquus* adapted to 16:8h light dark (LD) cycles in turbidostat culture conditions at very high temporal resolution (1 h intervals). Revealing a cycle of cellular events, six distinct expression profiles were obtained, each with transcriptional phenotypes correlating with measurements of biochemical composition.

The impact of starch deficiency was studied using the starchless mutant *slm1*. Significant changes in the transcriptional landscape were observed. Starch deficiency resulted in incapacity to supply energy during the dark period, resulting in a shift of energy demanding processes to an earlier or later time point. Our study provides new perspectives on the role of starch and the adaptation to LD cycles of oleaginous microalgae.

6.1 Introduction

Microalgae are a promising source of compounds of interest (lipids, proteins, and pigments) for the production of food, feed, chemicals, and fuels [4, 209, 210]. Large scale microalgal production will be primarily done outdoors under natural diurnal light/dark (LD) cycles [9, 211]. Diurnal cycles are ubiquitous and photosynthetic organisms synchronize their metabolic activities to anticipate light changes in the environment and schedule specific tasks during the night or day [24, 183, 212, 213]. Synchronization in photosynthetic organisms involves the regulation of photosynthesis to maximize carbon fixation and use of light during the day and to schedule light sensitive processes (such as DNA synthesis and cell division) at night [183, 214–216].

Tetrademus obliquus is a microalga recognized as an industrially relevant strain for food and fuel production [13, 28]. *T. obliquus* can reach a maximum triacylglycerides (TAG) content of $0.45 \text{ g}\cdot\text{gDW}^{-1}$ and a maximum TAG yield on light of $0.14 \text{ g}\cdot(\text{mol photons})^{-1}$ under batch nitrogen starvation and continuous light conditions [153]. For further improvement of TAG yield on light, de Jaeger et al. developed the starchless mutant *slm1* [33], which cannot synthesize starch due to a missense mutation in the small subunit of ADP-glucose pyrophosphorylase, the committed step of starch biosynthesis [35]. Under culture conditions of continuous light and batch nitrogen starvation, *slm1* showed a higher maximum TAG yield on light ($0.217 \text{ g}\cdot(\text{mol photons})^{-1}$) and maximum TAG content ($0.57 \text{ g}\cdot\text{gDW}^{-1}$) compared to the wild-type (WT) without a decrease in photosynthetic efficiency [153]. Comparable results were obtained under light-dark cycles and batch nitrogen starvation [28].

LD cycles give an advantage over continuous light, leading to higher energy conversion efficiency and higher yield of biomass on light. The physiological behavior of *T. obliquus* WT and of the starchless mutant (*slm1*) was also studied under 16:8h LD cycles in turbidostat controlled systems [24, 214]. Under this light regime and nitrogen replete conditions, *T. obliquus* showed synchronization and diurnal patterns of its metabolism, which among others, suggested that starch was used as a temporary energy storage. This is, starch was accumulated during the light period and was used during the dark period. Cell division was also synchronized and occurred mainly during the night. The starchless mutant *slm1* showed a lower energy conversion efficiency (11–24% lower) and biomass yield on light (13–39% lower) compared to the WT under different photoperiods [214]. Furthermore, for the *slm1* mutant cell

division still occurred mainly during the night, but at a slower rate and no diurnal oscillations on any of the other measured compounds were found [214]. Unlike for the WT, diurnal LD cycles did not provide an advantage for the *slm1* mutant compared to continuous light. On light, biomass yield as well as the energy conversion efficiency were similar [24]. Starch may therefore play an important role in harvesting additional light energy during LD cycles.

While the biochemical analysis allowed us to draw some conclusions, we still lack information on how cellular processes are regulated during the diurnal cycle and how the synthesis and use of starch is connected to these processes. Furthermore, while the inability to make starch had an impact on energy conversion efficiency, it is not known how this will affect the timing and regulation of the different cellular processes. LD cycles and the subsequent synchronization of the microalgal population make it of paramount importance to unravel the timing of cellular and subcellular events. Understanding the timing at which metabolic changes take place is essential to understand the phenotype exhibited by the WT and *slm1* strains and to optimally design experiments characterizing mutant strains.

Time resolved transcriptome analysis can give useful insights into the timing and regulation of the different physiological stages and on the succession of processes in the cell. Thus it allows the association of the cellular processes with biochemical properties. Algal diurnal transcriptional regulation is not well known. To our knowledge, only a few reports on diurnal oscillations under light/dark cycles in microalgae have been published [217–222]. The first one was on the eukaryotic red alga *Cyanidioschyzon merolae* [223] grown under 12:12 h LD cycles, where they studied the transcriptional changes in intervals of 2 h. Furthermore, the diurnal transcriptional changes were studied in the diatom *Phaeodactylum tricornutum* [219] under 16:8 h LD cycles. The authors studied changes in biochemical composition (carbohydrates and lipids) in 8 different unequally distributed time points during a period of 26.5 h (5 points during the light period and 3 during the dark period). Finally, transcriptional analysis on *Nannochloropsis oceanica* [220] and *Chlamydomonas reinhardtii* [221] under a 12:12 h LD cycle was published. For *N. oceanica* the authors studied growth, changes in biomass composition (lipids and glucose) and in gene expression in intervals of 3 h, while for *C. reinhardtii* a more detailed study was done with intervals of 1 h during most of the cycle and every 30 min for some time points, ending up with a total of 28 points distributed over the 24 h cycle. Our study, however, is the first one to look into the transcriptional changes in a diurnal cycle of an oleaginous green algae with a high sampling frequency (intervals of 1 h). Additionally, this is the first study reporting on the transcriptome changes over a diurnal cycle for a starchless mutant, which could give insights into the role of starch in microalgae.

To fully characterize the changes induced by LD cycles and the role of starch metabolism, we analyzed and compared the transcriptional landscape of *T. obliquus* WT and *slm1* under diurnal 16:8h LD cycles. For this, we used the same samples collected from our previously published article where also the measurements of the biochemical composition are published [24]. Each strain was cultivated in two separate photobioreactors that were operated in a continuous turbidostat controlled mode under a 16:8 LD cycle resulting in an oscillating steady state that was synchronized to the light regime. Samples from all four photobioreactors were taken in intervals of 1h for WT, and intervals of 3h for *slm1*. Finally, changes in the transcriptional landscape were related with their previously published [24] measurements of biochemical composition.

6.2 Materials and methods

6.2.1 Experimental setup and sampling

Wild-type (WT) *Tetrademus obliquus* UTEX 393 (formerly known as *Acutodesmus obliquus* and *Scenedesmus obliquus* [224]) was obtained from the Culture Collection of Algae, University of Texas. The starchless mutant of *T. obliquus* (*slm1*) was generated as described by de Jaeger et al. [33]. *T. obliquus* was continuously cultivated in a sterile flat panel airlift-loop reactor with a 1.7L working volume (Labfors 5 Lux, Infors HT, Switzerland). Culture conditions and reactor set-up (27.5 °C, pH7.0 and gas flow rate of 1 L·min⁻¹ air enriched with 2% CO₂) were controlled as previously described in León-Saiki et al., where also the measurements of the biochemical composition are published [24]. Light was provided in a 16:8h light/dark (LD) block at an incident photon flux density of 500 μmol·m⁻²·s⁻¹ (warm white spectrum 450–620 nm). Cultivations were turbidostat controlled, where fresh medium was fed to the cultures when the light intensity at the rear of the reactor dropped below the setpoint (10 μmol·m⁻²·s⁻¹, OD750). Feeding of medium was stopped during the dark period. After steady state was reached, 8 mL samples were taken for RNA extraction (approximately 10 mgDW). Cells were immediately collected by centrifugation (4255 ×g, 0 °C for 5 min), supernatant was discarded and pellets were frozen in liquid nitrogen and stored at -80 °C until

further extraction. Samples for RNA extraction were taken in intervals of 1 h for WT or every 3 h for *slm1*. Due to restrictions on working hours of the laboratory, the samples were collected in two successive time settings to allow sampling the dark period during the day. After collecting samples of the first half of the cycle, light settings were shifted and the culture was then allowed to reach oscillating steady-state before collecting samples for the second half of the cycle. The first and the last samples of each time settings are overlapping samples for control. Therefore, four RNA samples are present at these time points. Overall 72 samples were taken for RNA extraction.

6.2.2 RNA isolation and quality control

RNA extraction was performed using the Maxwell® 16 LEV simplyRNA Tissue kit (Promega). Frozen algae pellet ($\approx 200\ \mu\text{L}$) were submerged in $400\ \mu\text{L}$ of homogenizing buffer supplemented with $8\ \mu\text{L}$ 1-thioglycerol in a 2 mL Lysing matrix C tube (MP), prefilled with a mix of glass beads. Samples were disrupted using a FastPrep-24 instrument (MP). After disruption, all liquid was transferred to a LEV RNA Cartridge. $300\ \mu\text{L}$ lysis buffer (from the kit) were added and the rest of the extraction was performed using a Maxwell MDx AS3000 machine (Promega) following manufacturer's instructions. RNA integrity and quantity were assessed with an Experion system (Bio-Rad), and only high quality samples (RIN value ≥ 7) were selected. Total RNA was sent for whole transcriptome sequencing to Novogene Bioinformatics Technology Co. Ltd. (HongKong, China).

The corresponding data have been submitted to EBI ArrayExpress and can be found under the accession number E-MTAB-7009.

6.2.3 Genome structural annotation

Using the available genome sequence of *T. obliquus* [225], we performed an RNA-Seq-based genome annotation using BRAKER1 [226]. RNA reads from 38 samples of both WT and *slm1*, supplemented with an additional sample from each strain under nitrogen limited condition were given as additional BRAKER1 input. The annotation information was processed using our semantic framework pipeline [39] and the information was stored according to our integrated ontology [227] which respects the FAIR data principles [228]. The annotation

framework and the ontology were extended with the necessary tools [226, 229, 230] and ontology terms for the purpose of this analysis.

The genome feature annotation has been used to update the original genome ENA project with accession number PRJEB15865.

6.2.4 Genome functional annotation and pathway mapping

Proteins were annotated to GO terms using InterProScan5 and Argot2, with default parameters [44, 61]. Proteins were annotated to enzyme commission (EC) number using EnzDP [231]. Results with a complete EC number and a likelihood-score of at least 0.2 were used for subsequent analysis. To choose a threshold for the likelihood-score that results in a good trade-off between true positive and false positives, we visually compared the completeness of KEGG metabolic maps while avoiding dispersion of each reaction into different expression clusters. The functional annotation can be found in Supplementary file S1 that includes EC numbers and GO terms.

6.2.5 RNA-seq normalization and expression calculation

Using all the transcripts found from the genome annotation, we aligned the reads from each sample and calculated the FPKM values using Cufflinks [230]. As a pre-filter to remove unexpressed genes and false positives from gene annotation, we selected the transcripts with a coverage of at least 1 and a FPKM of at least 0.1 in at least 10 of the 113 analyzed samples. These samples include the 72 of this study, and the rest were taken under LD cycles and nitrogen limited conditions (not studied in this manuscript). Additionally, we selected the transcripts having an expression of at least 10 samples with a value higher or equal to the 0.15 quantile. Following the advice of maSigPro [232], the FPKM of all samples were normalized using the scaling normalization method TMM [233] using the R functions from edgeR package “calcNormFactors” [104].

6.2.6 Gene clustering

To identify genes with significant expression profile changes over time, maSigPro was used [232, 234], with the modified parameters: regression model was set to a maximum of 23 degrees, the parameter “counts” was set to true, “nvar.correction” was also set to true, the “step method” set as backward, and R-squared was set to 0.7. The R-squared was chosen based on the relevance and the balance in the number of enriched pathways identified in each cluster (see below). To better estimate the number of clusters to group the gene expression, we evaluated combinations of significance level (Q: 0.005 to 0.05) and R-squared (0.85 to 0.5). We decided to keep the standard-strict values (Q=0.05, R-squared=0.7), which would also give a good balance in pathway enrichment and number of genes in each group.

Hierarchical clustering with agglomerative linkage was performed using the R stats package (hclust function). Number of clusters from 3 to 25 were evaluated using the R function “cluster.stats” from package “fpc” [235]. For each set of clusters, the following indexes were computed: average silhouette widths, normalized gamma, two Dunn indexes, average within and average between ratio, and Calinski and Harabasz index. Additionally, each set of clusters were compared one on one with the Rand index.

6.2.7 Enrichment analyses

Enrichment analyses were performed using the hypergeometric function to model the probability density using the “phyper” function from the R package stats [236]. Two types of analysis were performed: pathway and GO term enrichment. Pathway enrichment required associating annotated Enzyme Commission (EC) numbers to metabolic maps. We used the online available resource from KEGG pathway maps [237, 238]. The KEGG pathways fitting the following requirements were kept for further analysis: 60% coverage if 3 to 6 EC numbers annotated, and 50% coverage if 6 to 10 EC numbers annotated, and 25% of coverage if >10 EC numbers were annotated. For the hypergeometric test we considered the universe size, N , to be the total number of EC numbers in all pathways in the genome, m is the number of successes in this universe and is defined as the number of EC numbers in the corresponding pathway in the genome, k and x are the sample size and the number of successes in the sample (or considered gene subset) respectively. Enrichments with a p-value lower than 0.05 were considered significant. Similarly, for the GO enrichment, N is the total number of genes

annotated to any GO terms in the genome, m is defined by the number of genes annotated to the considered GO term in the genome, and k and x refer to the considered subset of genes. Multiple test correction for the GO enrichment was performed using the Benjamini–Hochberg procedure. Enrichments with $FDR < 0.05$ were considered significant. To handle the GO information such as ontology, ancestor, and offspring, we used “GO.db” database from Bioconductor [239]. Additionally, to reduce the number of GO terms and conserve the most specific, only terms not having any offspring in the selection were retained. We additionally provide the pathway coverage, and the x and m values to better understand the reason of each significant enrichment.

6.3 Results

6.3.1 Genome annotation

The genome of *T. obliquus* UTEX 393 is about 100 million base pairs in size [225]. BRAKER1 annotation revealed 19,795 genes, which transcribed 21,493 coding sequences and translates into 19,723 protein sequences.

To analyze the changes in transcription over a diurnal cycle of *T. obliquus* WT and *slm1* strains, we sequenced RNA from biological duplicate runs. WT was sampled every hour in duplicate resulting in 52 samples. *slm1* was sampled in intervals of 3 h resulting in 20 samples. For both strains, time points 0 h and 13 h were overlapping samples and therefore sequenced in quadruplicate (see material and methods for details). Not all the transcripts were found to be expressed and filtering was performed to identify genes with very low expression that were not further considered. From the 19,723 proteins, 16,810 proteins remained after filtering for subsequent analysis. GO term annotation yielded 13,687 proteins annotated to 2394 unique GO terms. EC annotation yielded 3559 proteins annotated to 1315 EC numbers. The functional annotation can be found in Supplementary file S1.

6.3.2 Light-dark 16:8 h cycle induces systemic transcriptional changes following a circular pattern

4686 genes were found to have a significant change of expression during the diurnal cycle. A principal component (PC) analysis (PCA) for these selected genes is shown in Figure 6.1A. The PC plot provides a global overview of changes in gene expression over time. The time point samples form an ordered and circular pattern over the two first PCs. These two first PCs together explain 87% of the variation in expression. To demonstrate that this is not a consequence of gene selection, we rendered another PCA considering the whole set of expressed genes (Supplementary file S2). Within this set, the overall pattern is noisier, but the two first components still explain 78% of the variation in expression. The main changes in gene expression occur along PC1 from 1 h to about 5 h and back again from 8 h to about 13 h. The second important change occurs along PC2 from 6 h to 9 h and back again from about 22 h to 0 h. During the period from 14 h to 22 h there is very little change in gene expression. Overall, genes associated to light harvesting complex (LHC) were found to be the main contributors of PC1. The two highest contributing genes to PC2 are glyceraldehyde-3-phosphate dehydrogenase (GA3PDH) and glucose-6-phosphate isomerase (GPI). LHC genes also contributed strongly to PC2.

The heatmap representation in Figure 6.1B offers a complementary view of these multi-dimensional data. It allows to evaluate differences and similarities in expression between the time points. Two samples corresponding to time points 13 h and 15 h are separated in the dendrogram from the neighboring samples and from the biological duplicate samples. However, in the heatmap it can be seen that these samples still show relatively high correlation with their corresponding biological duplicates and samples from the surrounding hours. Three main groups of time points can be identified in the dendrograms in Figure 6.1B, which we will refer to as time phases I, II, and III. Additionally, each phase can be subdivided in two sub-phases, a and b, which represent subtler expression changes. The occurrence of the phases over the 16:8 h LD cycle is shown in the Figure 6.1E for both strains. The phases were named according to their order of appearance from 0 h to 23 h. These time phases are also apparent and can be associated to changes along the two first PC. Similarly, time phases were also associated to the *slm1* strain based on the PCA plots. For *slm1*, we noticed a change in phase timing that represents a significant change of expression occurring one to 2 h before each dark-light shift.

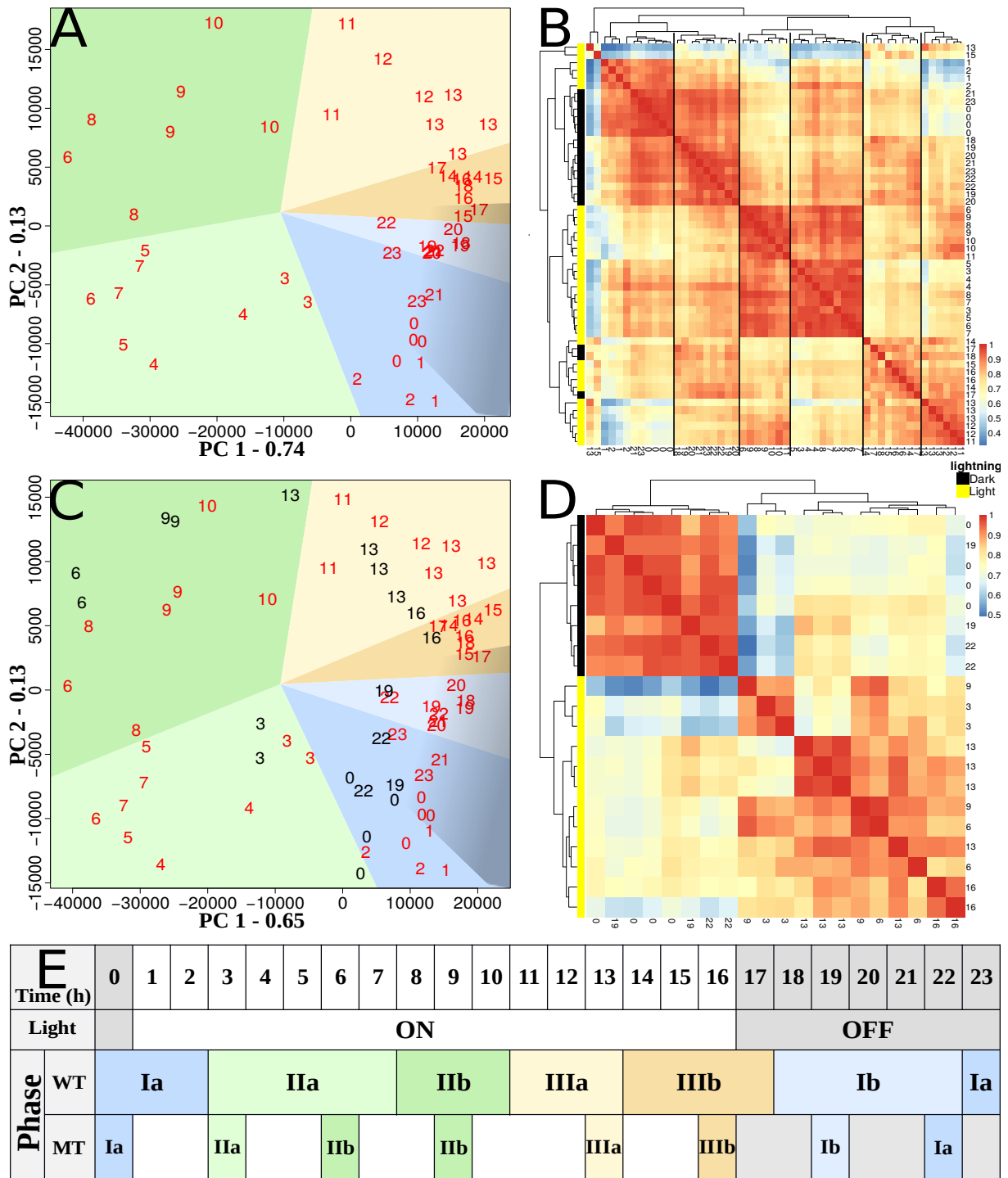


Figure 6.1: Principal Component Analysis (PCA) (A,C) of gene expression data for *T. obliquus* wild-type (WT) (A) and both WT and starchless mutant (*slm1*) (C) during 16:8h light-dark cycle. Numbers represent the time points of the samples, with red for the WT samples and black for the *slm1* samples. Background colors in both PCA plots refer to the colors given to the six time sub-phases described in (E), and the dark shade refers to the dark period. (B, D) Heatmaps showing similarity of gene expression data among samples for WT (B) and *slm1* (D) strains, samples are identified by their time points. Dark vertical lines across the heatmap in (B) indicate different time phases. (E) Overview of the time phases.

6.3.3 Gene expression clusters and their time profiles

6.3.3.1 Cluster profiles

We clustered the 4686 genes that showed significant changes over time in different numbers of clusters ranging from 3 to 25. Next we evaluated the cluster separation using seven well established indexes to assess similarities within clusters and differences between clusters. These results are depicted in the Supplementary file S3. Some of the selected metrics (such as the first Dunn index) gave no clear indication on the optimal number of clusters. However other metrics such as average silhouette width, second Dunn index, normalized gamma or Calinski-Harabasz index, show a local maximum (or minimum in the case of the average within and average between ratio) at 4–6 clusters. Inspection of cluster similarities based on the Rand index (Supplementary file S3) led to select 6 clusters (over 4 or 5) as it would result in a new and distinct cluster. Based on these results, we considered separating the genes in six clusters to be optimal.

The resulting six cluster profiles were summarized using for each time point the median value of gene expression of each gene present in the cluster and are depicted in Figure 6.2. Cluster 1 to 6 (WT#1-6) contain respectively 829, 364, 1020, 929, 952 and 592 genes. All the cluster profiles are different, but there are certain similarities between clusters 1 and 6, and between 3, 4, and 5. Interestingly, the two peaks of expression observed in cluster 3, 4, and 5 are all 5–6 h apart.

The temporal profile of each cluster peaks at different moments of the diurnal cycle. Cluster 1 peaks the earliest, 1 h after light goes on. Cluster 2 peaks at 3 h, cluster 3 peaks around 11 h and shows a smaller secondary peak at 17 h. Cluster 4, with a similar profile, shows the first peak at 13 h and the second at 18 h. Cluster 5 has a rather gradual increase of expression and peaks around 14 h followed by a smaller peak at 20 h. Finally, cluster 6 has very gradual expression changes throughout the day, and shows a peak at 23 h.

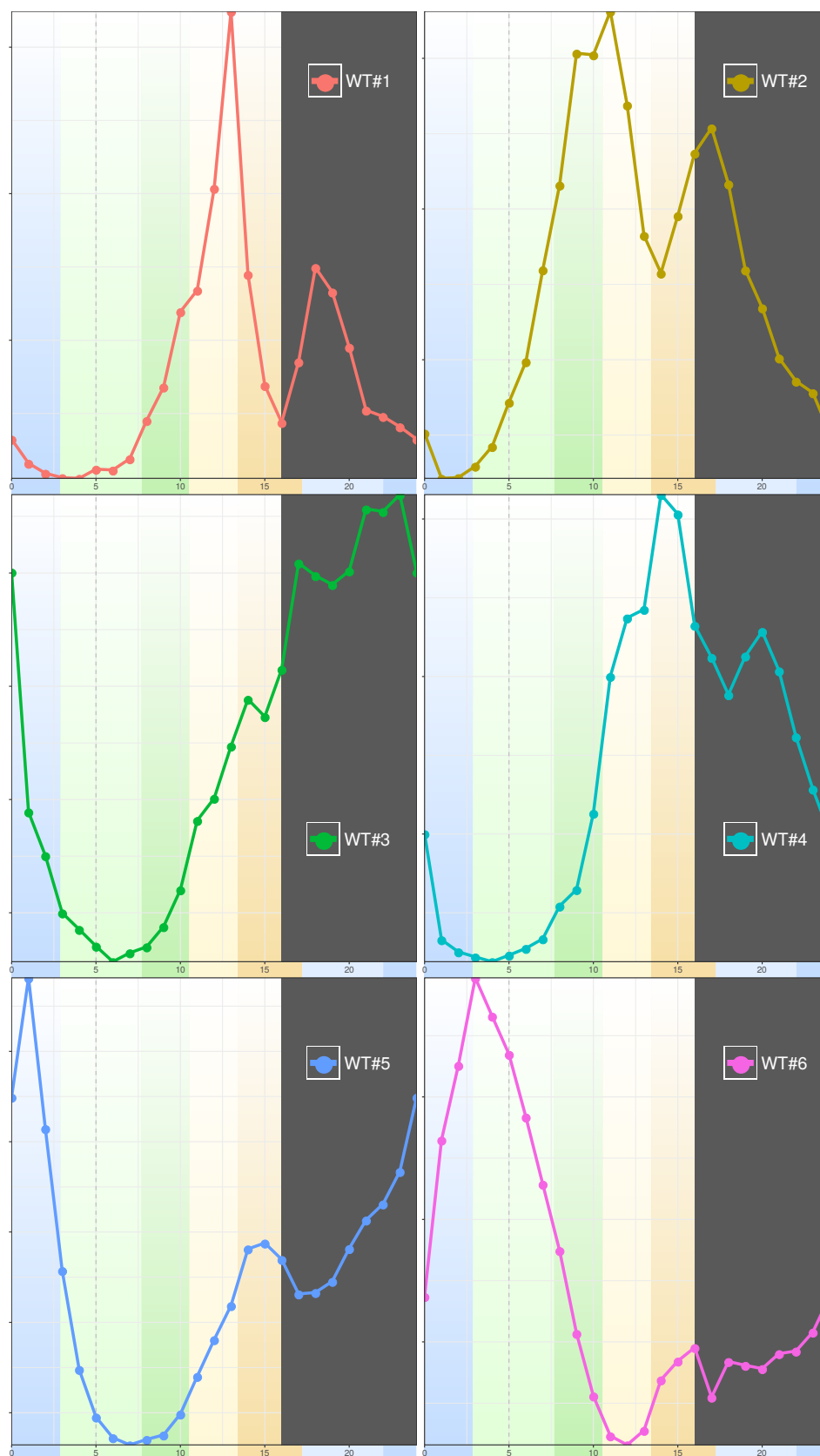


Figure 6.2: Expression profile of the 6 gene clusters for *T. obliquus* under 16:8 h light-dark cycles. The plots show the median profile of gene expression in the indicated clusters. The background colors correspond to the colors given to the time phases in Figure 6.1. The vertical gray dashed line represents the time point of maximum dilution rate taken from [24]. Dark area represents the dark period.

6.3.3.2 Cluster functional analyses

To understand the functions of the genes in the clusters, we performed two sets of complementary enrichment analyses. Metabolic pathway enrichment analysis provides direct information on gene metabolic functions. The functional annotation identified genes associated to EC numbers. Mapping these EC numbers to KEGG pathway maps revealed 67 pathways for which enough genes can be associated to the corresponding reactions to warrant further analysis, as detailed in Materials and methods. GO enrichment analyses were performed on the three ontologies: biological process (BP), molecular function (MF) and cellular component (CC). These analyses provide a wider overview not restricted to metabolism. The enrichment results are summarized in Table 6.1. Each cluster median time profiles and the full set of significant (p -value < 0.05) enrichments are presented in the Supplementary file S4.

Cluster 1 shows enrichment in pathways related to amino acids (AA), and riboflavin metabolism. GO terms enrichment is in agreement and contains terms such as tRNA modification, aminoacyl-tRNA biosynthesis, ribosome, alpha-amino acid biosynthetic process and translation. All these pathways and terms indicate protein synthesis and the associated strong demand for AA. Riboflavin metabolism, leading to FAD synthesis, and ubiquinone and other terpenoid-quinone biosynthesis appears enriched in cluster 1. Inspection of this last pathway (Supplementary file S5) indicates that ubiquinone is the most probable product in this time frame. Ubiquinone is an important step in the oxidative phosphorylation by oxidating FADH₂ back into FAD and it indirectly plays a role in ATP synthesis.

Cluster 2 mainly shows enrichment in the processes related to pigments synthesis, which include carotenoids, chlorophylls, and their precursor (Phytyl-diphosphate) from the terpenoids backbone synthesis. Those are needed for building the photosystems and starting the carbon fixation. Three of the four reactions involved in the starch synthesis were found in cluster 2, this can be observed from the pathway map (Supplementary file S5). Cluster 2, also shows enrichment in genes related to transcription, amino acid (AA) and protein synthesis, but to a lower extent than cluster 1. The carbon fixation pathway is found enriched in cluster 2, which goes in line with enrichment in the pigments and in starch synthesis. The GO terms in biological process are in agreement with the pathway enrichments. The GO terms in molecular function provide extra information on the nature of chemical reactions performed by these genes. We also found that this cluster contains genes associated to the LHC that were found to

strongly contribute to PC1 and PC2 and among the proteins associated to carbon fixation in this cluster GAPDH was found to be a main contributor in PC2. Interestingly, while all the other amino-acid synthesis pathways were found enriched in cluster 1, the pathway of “Phenylalanine, tyrosine and tryptophan biosynthesis” is found enriched in cluster 2. This result indicates that these aromatic amino-acids are probably synthesized later than the other, more simple, amino-acids.

In cluster 3, the GO enrichment provides valuable information that cannot be covered by the pathway enrichment. The GO term enrichments display clear terms such as “DNA replication”, “organelle fission”, “chromosome organization”, “microtubule”, “cytoskeletonDNA-directed”, “DNA polymerase activity”. This cluster clearly groups all processes related to the full cell cycle. The pathway enrichment reveal processes related to AA pathways: “Lysine degradation” and “Phenylalanine metabolism”. Additionally, the enrichment in “N-Glycan biosynthesis” and “Various types of N-glycan biosynthesis” together with the enriched GO terms “protein complex assembly” and “vesicle-mediated transport” indicates important protein maturation processes leading to complex proteins and some transport of proteins to membranes.

Cluster 4 is mostly enriched in pathways associated to different kind of carbohydrates metabolism. The first pathways enriched are “Galactose metabolism”, “Starch and sucrose metabolism”, “Other glycan degradation”, and “Amino sugar and nucleotide sugar metabolism”. With the GO terms displaying “polysaccharide binding”, “hydrolase activity”, and “hydrolyzing O-glycosyl”, we observed the same trend towards starch degradation. The GO term for “microtubule binding” is the only one of this kind in this cluster, but it comes right after cluster 3 where a lot of cytoskeletal and microtubule terms were found enriched. Finally, there is enrichment in Cytochrome P450 which hints on repair mechanisms potentially related to photo-damage.

Cluster 5 shows enrichment in a very diverse set of metabolic pathways, covering glycolysis, pyruvate metabolism, glutathione metabolism, co-enzyme-A, starch and other carbohydrates polymers. More importantly, the cluster is also enriched in genes related to nitrogen metabolism and glutamate associated genes. ATP synthase is found in this cluster too, but the gene is highly expressed throughout the whole light period with highest expression before dark. Among the genes associated to the chlorophyllase (EC 3.1.1.14), the best candidate (g19660.t1) is found in cluster 5. This also agrees with the observed enrichment of Cytochrome

P450, which was recently found to take major part in the breakdown of chlorophyll in *Arabidopsis thaliana* [240–242].

Cluster 6 shows enrichment in fatty acids degradation and folate biosynthesis and contains the reactions from GTP until the precursor of dihydrofolate (DHF).

Table 6.1: Summary of the results of the enrichment analyses. The first column contains the cluster identifier. The second contains the enriched pathways (p-value < 0.05). The last three columns are the results of the GO enrichment analyses (FDR < 0.05) for each of the GO ontologies: biological process (BP), cellular component (CC) and molecular function (MF). Full set of GO enrichment results are available in the Supplementary file S4.

Cluster	Pathways	BP	CC	MF
1	Ubiquinone and other terpenoid-quinone biosynthesis Purine metabolism Glycine, serine and threonine metabolism Cysteine and methionine metabolism Valine, leucine and isoleucine biosynthesis Lysine biosynthesis Arginine and proline metabolism Riboflavin metabolism Aminoacyl-tRNA biosynthesis	RNA methylation tRNA modification translation nucleobase-containing small molecule metabolic process alpha-amino acid biosynthetic process	ribosome	RNA binding structural constituent of ribosome GTPase activity GTP binding S-adenosylmethionine-dependent methyltransferase activity ligase activity, forming carbon-nitrogen bonds
2	Glycine, serine and threonine metabolism Phenylalanine, tyrosine and tryptophan biosynthesis Carbon fixation in photosynthetic organisms Porphyrin and chlorophyll metabolism Terpenoid backbone biosynthesis Carotenoid biosynthesis Aminoacyl-tRNA biosynthesis	tRNA aminoacylation for protein translation steroid biosynthetic process coenzyme metabolic process porphyrin-containing compound biosynthetic process photosynthesis pigment biosynthetic process oxidation-reduction process	photosystem	nucleotide binding 3-beta-hydroxy-delta5-steroid dehydrogenase activity aminoacyl-tRNA ligase activity iron ion binding calcium ion binding oxidoreductase activity, acting on a sulfur group of donors oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen transferase activity, transferring alkyl or aryl (other than methyl) groups carbon-carbon lyase activity hydro-lyase activity heme binding precorrin-2 dehydrogenase activity coenzyme binding 2 iron, 2 sulfur cluster binding
3	Lysine degradation Phenylalanine metabolism Glutathione metabolism N-Glycan biosynthesis Various types of N-glycan biosynthesis Propanoate metabolism	DNA replication DNA repair DNA recombination protein complex assembly vesicle-mediated transport cell cycle process regulation of cellular metabolic process proteasome-mediated ubiquitin-dependent protein catabolic process	microtubule cytoskeleton protein complex cytoskeletal part	DNA binding DNA-directed DNA polymerase activity ATP binding microtubule binding transcription factor binding four-way junction helicase activity ATPase activity, coupled

		organelle fission chromosome organization regulation of primary metabolic process carbohydrate derivative biosynthetic process single-organism organelle organization		
4	Galactose metabolism Starch and sucrose metabolism Other glycan degradation Amino sugar and nucleotide sugar metabolism Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450	carbohydrate metabolic process		hydrolase activity, hydrolyzing O-glycosyl compounds microtubule binding polysaccharide binding
5	Glycolysis / Gluconeogenesis Glutathione metabolism Starch and sucrose metabolism Sphingolipid metabolism Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Thiamine metabolism Pantothenate and CoA biosynthesis Nitrogen metabolism Metabolism of xenobiotics by cytochrome P450 Drug metabolism - cytochrome P450	carbohydrate metabolic process oxidation-reduction process		oxidoreductase activity, acting on CH-OH group of donors
6	Fatty acid degradation Pyrimidine metabolism Butanoate metabolism Folate biosynthesis			

6.3.3.3 *Transcriptional landscape of Tetradismus obliquus slm1*

We compared changes in gene expression between the starchless mutant (*slm1*) and WT to understand the differences resulting from the lack of starch accumulation. An overview of gene expression data in *slm1* in comparison to WT is shown in Figure 6.1C. The PCA shows very similar regulation over time in both strains. While many time points display similar expression patterns, there are also clear differences related to the separation in time phases. At the time points from 16 h until 3 h (phase IIIa to phase Ia), all *slm1* samples along PC1 are shifted left compared to WT, but they remain at the same level along PC2. This reflects a state of expression of WT that is never reached by *slm1*, rather than a diurnal dysregulation. On the contrary, the samples at 6 h and 9 h along PC2 are shifted up for *slm1* as compared to WT, but remain at the same level along PC1. This reflects an early state of expression, especially for *slm1* samples at 6 h, which fits into the phase IIb. Another smaller time dysregulation is a small shift of time point 22 h that is shifted down for *slm1* as compared to WT. This reflects an early state of expression that fits into to the time phase Ia. The relative time phases of *slm1* are also

shown in Figure 6.1E. Overall, two main trends are observed: earlier changes in expression of processes shortly after light and a delay in change of expression of the processes before the dark period.

The PCA considering the whole set of expressed genes (Supplementary file S2) displays a very similar expression between WT and *slm1*. While there is a general overlap, the expression of *slm1* seems noisier, especially for the time points between 19 h and 0 h.

6.3.3.4 Comparison of the gene expression dynamics in WT and *slm1*

Using the regression based approach of maSigPro, we identified genes with differences in expression profile between *slm1* and WT. Our experimental design included sequencing samples of WT every hour and of *slm1* every 3 h. However, maSigPro is designed to allow uneven distribution of time points. 784 genes showed no significant differences in expression profile between WT and *slm1* strains. 3902 genes showed significant differences in expression profile between WT and *slm1*. Additionally, 40 genes were found to have a time profile in the *slm1*, while no time profile was detected in WT for these genes. Apart from having no time profile it could also be that these genes are too noisy or are not expressed in WT.

6.3.3.5 Differences in time profiles

Clustering was done for the gene expression data obtained from the mutant *slm1*. Analysis of clustering performance indicators again resulted in 6 clusters. As previously stated, 784 genes did not show any significant difference in time dynamics between the two strains, therefore, these genes were used to guide the identification of the mutant clusters in relation to the wild type clusters. The cluster assignments of these 784 genes in both strains are shown in Figure 6.3A. The genes in WT clusters 3 and 4 are all found in the same *slm1* cluster, consequently named 3&4. Likewise, the WT genes in clusters 1 and 6 are all found in the same *slm1* cluster, consequently named 1&6. The WT genes in cluster 2 and 5 ended up in separate *slm1* clusters and consequently kept the same cluster name. Finally, the two remaining clusters, named A and B, do not contain genes with conserved expression. Those clusters have a time profile that does not fit any of the cluster profiles in WT.

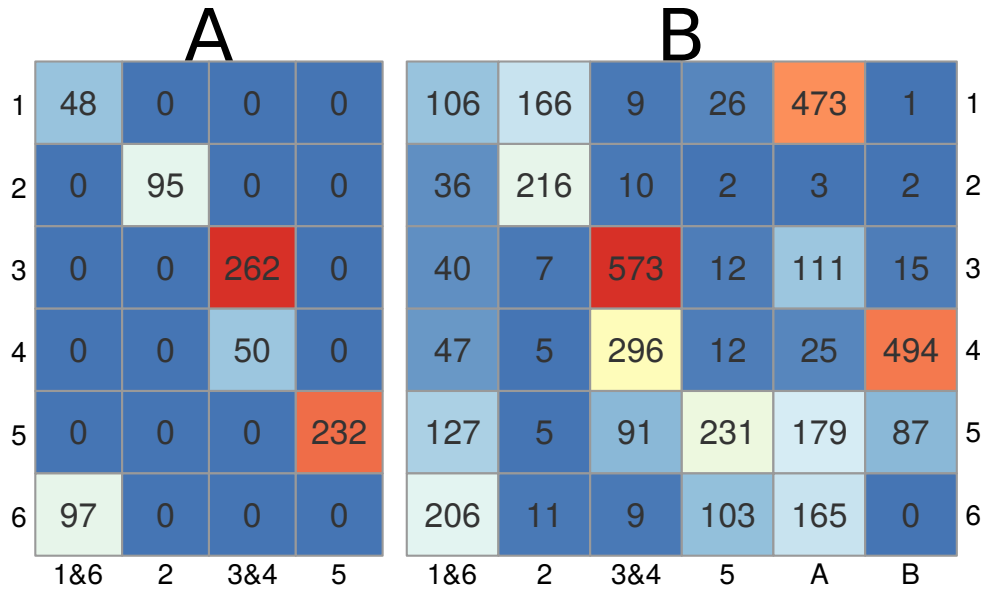


Figure 6.3: Distribution of genes between time profile clusters of *T. obliquus* wild-type (WT) (rows) and *slm1* (columns). A: 784 genes with the same time expression profile in *slm1* and WT. B: Genes with significantly different time profiles between *slm1* and WT, as identified by *maSigPro* (3902 genes). Background color ranges from blue, white, to red, with red being the highest value and blue the lowest.

Figure 6.3B shows the cluster assignments of the 3902 genes with significant differences in their time profiles when comparing WT and *slm1*. From these genes with altered expression (Figure 6.3B), there is a general trend of genes to remain clustered together. This means that most genes are transcriptionally controlled in large groups. The genes in WT cluster 3 are almost all found in the associated *slm1* gene cluster 3&4. Again, almost all genes in WT cluster 2 are found in the associated *slm1* cluster 2. The genes in WT cluster 5 are distributed over five *slm1* clusters, with the majority in the associated *slm1* cluster 5. The genes in WT cluster 6 are distributed over three clusters, with the majority in the associated *slm1* cluster 1&6. Interestingly, a minority of genes in WT cluster 4 is found in the associated *slm1* cluster 3&4, but the majority is found in *slm1* cluster B. Similarly, a minority of WT genes in cluster 1 is distributed over three *slm1* gene clusters, with a small portion in the associated *slm1* cluster 1&6, a bigger portion in *slm1* cluster 1, and the majority in *slm1* cluster A. Thus the new cluster A mainly receives genes from the WT clusters 1, and to lesser extent from cluster 3, 5 and 6. On the other side, the new cluster B receives genes mainly from cluster 4 and to a lesser extent from cluster 5.

slm1 cluster median profiles are depicted in Figure 6.4, together with the associated WT clusters. From cluster 1&6 to cluster B each *slm1* cluster grouped respectively 735, 507, 1304,

621, 958, and 600 genes. Overall, the time profiles of *slm1* clusters show strong similarities with the associated WT clusters. *slm1* clusters A and B are displayed together with WT clusters 1 and 4, respectively, due to the large number of genes found in common. The new cluster A is characterized by low expression at dark and the first half of the light period, with a single expression peak during the second half of the light period. The new cluster B is characterized by a shift in the timing of the two peaks and a change in amplitude. A large number of genes were detected to be expressed differently between the two strains, but are associated to clusters with similar profile. These genes are more affected on their amplitude, and not necessarily on the time regulation.

91 genes that were found in WT cluster 5, are now found in *slm1* cluster 3&4. This represents a shift in time that is reflected in an earlier expression of these genes. This correlates with the observed early expression of genes along PC2 in phase II. A substantial number of genes from WT cluster 1 are found in *slm1* cluster 2, suggesting that these genes are shifted towards later expression in time and are not expressed during the dark period anymore. Again, the transfers of genes from WT clusters 4 and 5 to *slm1* cluster B reflect an earlier change of expression. Opposite to the previously mentioned changes, which all indicate decreased expression during dark, the genes in *slm1* cluster 1&6 are now expressed exclusively during the dark period.

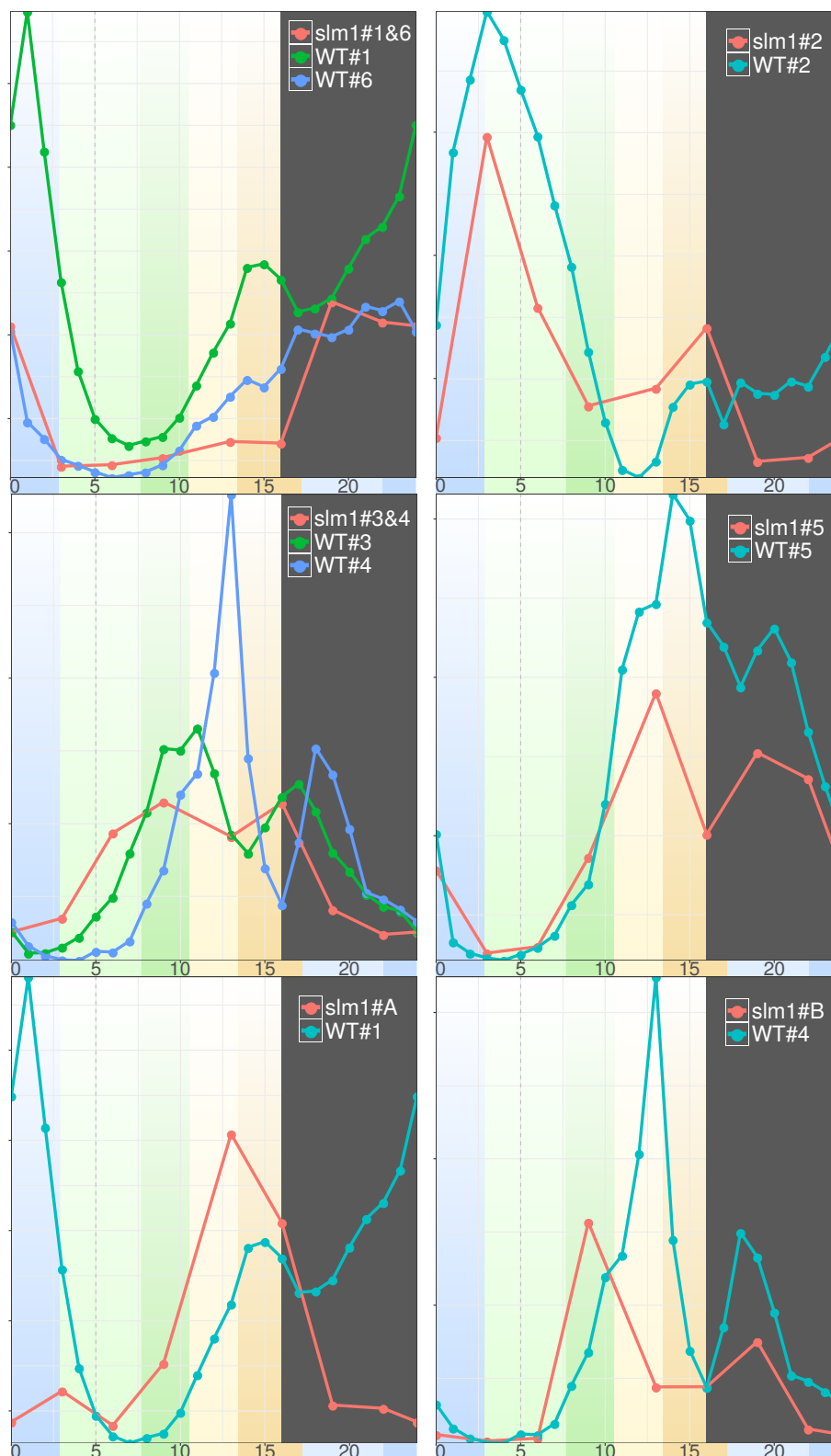


Figure 6.4: Expression profiles of clusters obtained with slm1 samples and their associated profiles in the WT. WT clusters 1 and 6, slm1 cluster 1&6; WT and slm1 cluster 2; WT clusters 3 and 4, slm1 cluster 3&4; WT and slm1 cluster 5; cluster A; WT cluster 4 and slm1 cluster B. slm1 clusters are plotted with the related WT clusters according to the common conserved genes and the redistribution of genes as in Figure 6.3. The background colors correspond to the colors given to the time phases in Figure 6.1. The gray area represents the dark period.

6.3.3.6 *Functional differences between gene clusters in *slm1* and WT*

As observed by the similarities and changes of expression between the two strains (Figure 6.3B), genes have related expression in large groups. As a result, the enrichments are consistent with the large groups of genes changing expression or not. The detailed enrichments are available in Supplementary file S4. Enrichments of *slm1* clusters A and B are respectively similar to WT clusters 1 and 4. *slm1* cluster 3&4 is not enriched for the two AA pathways found in WT cluster 3 (lysine degradation and phenylalanine metabolism). This cluster also remains enriched in cellular structure and DNA replication, but without any biological process GO terms. Unlike for WT cluster 6, *slm1* cluster 1&6 does not display enrichment for other fatty-acids related pathways. AA synthesis and protein translation activities are now enriched in *slm1* cluster A, but were enriched in WT cluster 1. In *slm1*, nitrogen metabolism is no longer found enriched in cluster 5. This is due to three EC numbers (out of the five) that are associated to *slm1* cluster A. Globally, the enrichments are following the expectations drawn from the gene transfers described in Figure 6.3B.

6.4 Discussion

This study analyzed gene expression dynamics in a 16:8h light-dark (LD) cycle synchronized culture of *T. obliquus* wild-type (WT). Hourly sample collection and RNA sequencing allowed us to observe in detail the changes of gene expression during a diurnal cycle. A starchless mutant of the same species (*slm1*) was also studied. *slm1* was cultivated in the same conditions, but RNA samples were taken every 3h. The succession of cellular events were then compared between WT and *slm1*, allowing us to examine the role and importance of starch as a transient energy storage compound in this microalga.

In this work, RNA measurements were done in duplicate, where the two samples were taken from separate photobioreactors (biological duplicates). Due to restrictions on working hours of the laboratory, the samples were collected in two successive light settings in such a way that sampling could always be done during working hours. First, samples were collected on the first half of the cycle, then light settings were shifted 12h and the culture was then allowed to reach an oscillating steady-state before collecting samples for the second half of the cycle. The first

and the last samples of each time settings are overlapping samples for control (corresponding to time points 0 h and 13 h). Figure 6.1A and C show the high similarity in expression between these samples which indicates the robustness of these cultures and the validity of the approach.

6.4.1 The diurnal rhythm of *T. obliquus* wild-type is driven by six transcriptional phases

6.4.1.1 *Transcriptional phases*

In this experimental setup, *T. obliquus* cells are synchronized to the diurnal light-dark (LD) cycles [24]. Many organisms synchronize their metabolism to anticipate the changes in the environment [213, 243, 244]. For photosynthetic microorganisms, this synchronization provides a benefit as they can capture sunlight efficiently during the day and perform light sensitive processes at night [183, 216]. In *T. obliquus*, synchronization was observed in growth and cell division, as well as in changes in biomass composition [24]. The overall analysis of the changes in gene expression in both the WT and *slm1* strain shows a circular pattern in the PCA plot over two strong principal components. In WT, changes in expression appeared to occur sequentially, back and forth along each PC. In the studied process, the two main effects impacting gene expression are light availability and time itself.

By analyzing similarities between samples (Figure 6.1) and correlating genes expression profiles (clusters), we identified six time phases and six clusters of genes, respectively. The agreement between these two independent analyses indicates that they are related to each other. In fact, the time phase changes correspond to jumps in expression between the time points, when either a large increase or large decrease in overall gene expression occurs. The gene clusters correspond to synchronized peaks of expression. Thus, the identified phases are associated to peaks in expression for each clusters. Accordingly, cluster 1 peaks during phase Ia; cluster 2 during IIa; cluster 3 during IIb, cluster 4 during IIIa, cluster 5 during IIIb, and, finally, cluster 6 during Ib. Because the clusters regroup genes from the same biological processes, the six transcriptional phases correspond to a logical succession of biological processes.

6.4.1.2 Temporal succession of cellular events as described by the gene clusters

An overview of the changes in the transcriptional landscape and experimental measurements of WT through the diurnal LD cycle is presented in Figure 6.5.

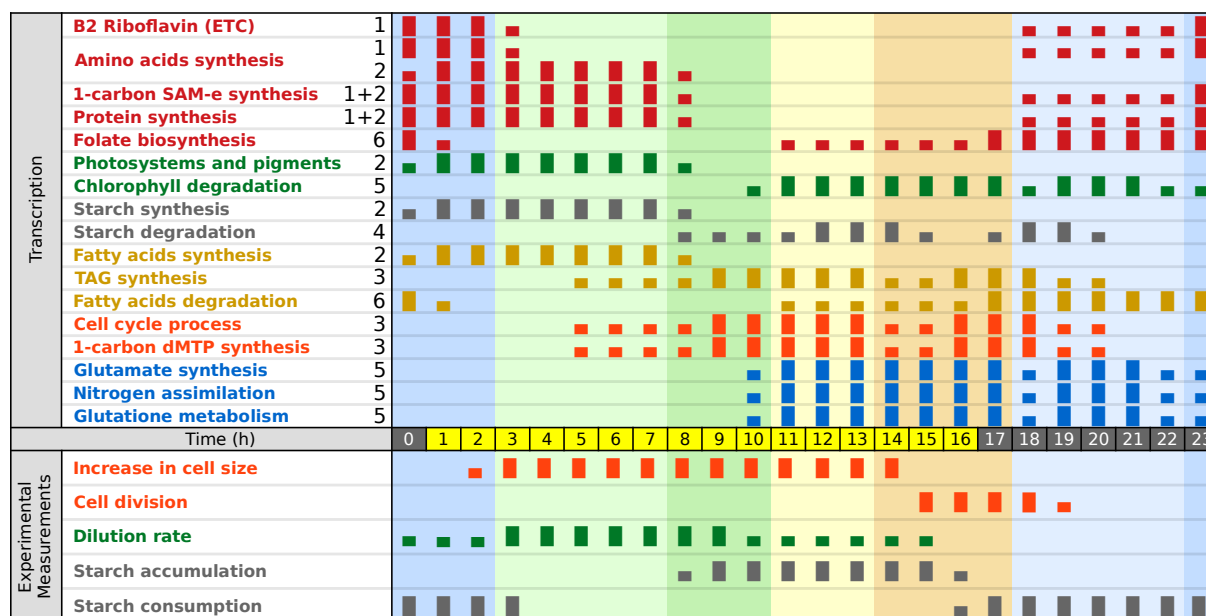


Figure 6.5: Overview of the diurnal changes in the transcriptional landscape and experimental measurements of *T. obliquus* wild-type under 16:8 h light/dark cycles. Small and big rectangles indicate that the process is occurring at relatively low or high level, respectively. The cluster ID associated to the expression profile is indicated by the black number in the first column. Groups of related processes are indicated in the same color. The background colors correspond to the colors given to the time phases in Figure 6.1. Amino-acids biosynthesis two lines reflect the synthesis of different amino-acid fitting different cluster profiles.

The expression of genes in cluster 1 starts to increase during the dark period and peaks the first after the light is switched on. The enrichment of cluster 1 reflects amino acids (AA) and protein synthesis and suggests increased ATP synthesis. These transcriptional changes point to increased catabolism for the production of energy (ATP) and agree with the observe starch consumption during the night and at the beginning of the day. The energy and intermediates of catabolism are apparently used to synthesize amino acids for proteins synthesis as part of cell growth that starts at the start of the day.

Cluster 2 mainly contains genes involved in pigment and starch synthesis, besides carbon fixation and AA synthesis. The expression profile of this cluster matches the dilution rate

profile, although gene expression precedes the dilution rate with about 3 h [24], as can be seen in Figure 6.5. The expression profile of processes related to starch synthesis precedes starch accumulation by 8 h while starch degradation at transcriptomic level appears correlated to the measured starch accumulation (Figure 6.5). Starch is a transient energy source and is subject to simultaneous synthesis and degradation [24, 245], which results in daily fluctuations between net production during the day and consumption during the night and early day. Between 16 h and 0 h, degradation occurs presumably to provide energy for the night processes, since light is not available. Between 0 and 4 h, starch is still consumed and it appears that it is used to supply additional energy and carbon for the build up of new photo systems, which is also up regulated at this time. Between 4 and 8 h, starch synthesis increases following the up regulation of the starch synthesis genes in cluster 2. The starch synthesis machinery is getting fully operational and energy demanding processes, such as AA and protein synthesis, are down regulated, resulting in net starch synthesis.

Genes in cluster 3 show a peak in expression at 13 h. GO enrichments point towards cell division which agrees with the observed cell division that starts just before the night and continues during the early night. Shortly after the genes in cluster 3 peak, genes associated to cluster 4, related to starch degradation, increase in expression. This agrees with the observed starch degradation during the night.

The next cluster for which the gene expression shows a peak is cluster 5, enriched in carbohydrates pathways, glycolysis and contains the ATP synthase. These point to starch degradation and glycolysis as the means to generate the needed energy during the night and to generate precursors, for example for glutamate metabolism. Glutamate associated genes are also found in cluster 5. Glutamate synthesis, which is at the center of nitrogen assimilation and amino-acid metabolism, seems to be upregulated before and during the night as anticipation for the synthesis of AA (cluster 1) and pigment (chlorophyll and carotenoids, cluster 2) during the night and early light period. It also suggests that *T. obliquus* continues to assimilate nitrogen during the night, which is in agreement with the observed nitrate consumption during the night by WT (data not shown). Notably, *slm1* does not consume nitrate in the night and nitrogen metabolism is not enriched for *slm1*, which suggests that pathways in nitrogen metabolism are regulated by the energy status of the cell. Cluster 5 also contains genes associated to glutathione reductive cycle, indicating some form of oxidative stress. This stress

could be a side effect of the whole day exposure to light leading to accumulation of reactive oxygen species and damaged photosystems.

Finally, cluster 6 is enriched in fatty acid degradation, which appears to occur during the night and early morning. However, fatty acid degradation was not observed experimentally. Possibly, these enzymes are involved in remodeling of membranes that has to occur after cell division, or maybe are involved in recycling unsaturated fatty-acids damaged from oxidative stress.

Folate (vitamin B9) is a vital cofactor, notably in thymidylate (dTMP) synthesis and in S-adenosylmethionine (SAM-e) synthesis. The gene expression in the pathway “one carbon pool by folate” of WT (Supplementary file S5) is clearly decoupled for these two processes. While dTMP related reactions are associated to cluster 3, SAM-e reactions are associated to clusters 1 and 2. Figure 6.5 shows their respective synchronization with cellular division and protein synthesis. Considering the temporal succession of these events, it seems that folate is being accumulated before the light period in order to sustain the high levels of protein synthesis at dawn.

Models of a multiple fission cell cycle have been established for *C. reinhardtii* and *Scenedesmus quadricauda* where [246, 247], the latter from the same family as *T. obliquus*. Comparison to the established cell cycle model, suggests that processes leading to the cell growth (G1), DNA replication (pS), mitosis (G2), and cellular division (G3), are associated to genes in clusters 2, 3, 4, and 5, respectively and correlate to phases IIa, IIb, IIIa and IIIb, respectively. In our experimental condition, the doubling time was measured to be 0.67 day in WT and 0.75 day in *slm1*, resulting in 1.5 and 1.33 divisions per day respectively. This can be explained by a double fission for a portion of the cells, 50% for WT and 33% for *slm1*. Time profiles of clusters 3, 4 and 5 show a double peak of expression 4 to 5 h apart. This second and lower peak could be associated to that portion of cells performing a double fission.

In summary, the annotation and enrichment analyses uncovered the general function of the genes in each cluster, thus showing a clear succession of cellular events: RNA transcription, gene expression, protein synthesis, pigments synthesis, starch synthesis, fatty acids synthesis, protein glycosylation, cellular growth and division, starch degradation and sugars inter-conversions, riboflavin synthesis, folate synthesis, and, finally, fatty acids degradation (Figure

6.5). This succession indicates that *T. obliquus* does not simply adapt its cellular phenotype to direct signals from the light switch, but it also anticipates the changes in light conditions.

6.4.2 Impact of starch deficiency on the diurnal rhythm

When analyzing data from *slm1*, significant changes in the temporal profiles were identified for the vast majority of the genes compared to the WT (3902 genes). In most cases, the genes remained clustered together, as for the WT, and most differences can be explained by a time shift or a change in amplitude. Time changes of expression in phase II (and to a lesser extent IIIa) contain most of the overall differences between WT and *slm1*. However, this could be due to the less frequent sampling of *slm1* (every 3h). From the six clusters of *slm1*, cluster A showed a really novel expression profile not seen in WT. In WT the profile of these genes more or less matches the observed net consumption of starch that occurs during the night and early day. In other words, impossibility to consume starch during this period in *slm1* may be related to the low expression of the genes occurring in cluster A.

The biochemical data for *slm1* also match the transcriptome results, but to a lesser extent than for WT. Data obtained from the starchless mutant *slm1* shows that the phases from the WT are to some extent preserved, and starch deficiency seems to result in rescheduling biological processes from the dark to the light period. It is possible that these processes could not happen at night due to the lack of energy originally provided by starch degradation. Processes required for expression and translation were transferred from cluster 1 to cluster A in *slm1*. Thus, expression of these genes seems to depend on the presence of starch and is possibly related to the energy status of the cell. In *slm1*, the expression of these processes is shifted to the time when starch accumulates in the WT during which it may be assumed that sufficient energy is available.

Genes in *slm1* cluster 1&6 are mainly expressed during the dark period. This indicates that the scheduling of processes that occur in the dark period and that do not depend on energy sources only suffers small changes. Notably, genes associated to fatty acid degradation in *slm1*, are in this cluster, although no significant reduction in TAG content was observed.

The reason of the incapacity of *slm1* to synthesize starch has already been studied [35]. The gene coding for the small subunit of ADP-glucose pyrophosphorylase was found to contain a non-sense mutation. However *slm1* has been obtained with UV, that is known to produce multiple mutations in genomes, therefore we cannot rule out differences in gene regulation due to other mutation not related to starch deficiency.

6.4.3 Selected reactions and pathways

In addition to the described time shifts caused by starch deficiency, a number of processes show differences in expression between *sml1* and WT that seem more complex. These are discussed in the following paragraphs.

6.4.3.1 Starch synthesis

The diurnal expression of genes related to the reactions analyzed in this section is depicted in Figure 6.6. Three of the four enzymes necessary for starch synthesis are found in cluster 2, being the ADP-Glucose pyrophosphorylase (EC:2.7.7.27), the starch synthase (EC:2.4.1.21) and the granule-bound starch synthase (EC:2.4.1.242). The fourth, starch branching enzyme, was not found to be regulated in time. Genes associated to the starch synthase (*g7327.t3* and *g2865.t1*), show complementary profiles in WT, each being highly expressed during one half of the light period. However, their expression appears stretched during the light period in *slm1*. This could be the response to the lack of ADP glucose. Similarly the granule-bound starch synthase expression peaks in the middle of the light period for the WT, but appears stretched over the light period in *slm1*. The difference in time regulation of the two types of starch synthase could be explained by their associated processes on starch branching nature [248]. The amylose isomerase (EC:2.4.1.18) was associated to two genes, with very similar profiles, but without apparent time regulation for either strain.

The reason of the incapacity of *slm1* to synthesize starch has already been studied [35]. The gene coding for the small subunit of ADP-glucose pyrophosphorylase was found to contain a non-sense mutation. In conditions of continuous light, the gene coding for the small subunit of ADP-glucose pyrophosphorylase (known to be mutated in *slm1*) was found to be strongly down-regulated in comparison to WT (approx. 5 folds). In this study, we found that this gene (*g788.t1*) is even more strongly down-regulated, meaning that the changes become even more

prominent in presence of a diurnal cycle. Additionally, we found that the time regulation is preserved between the two strains.

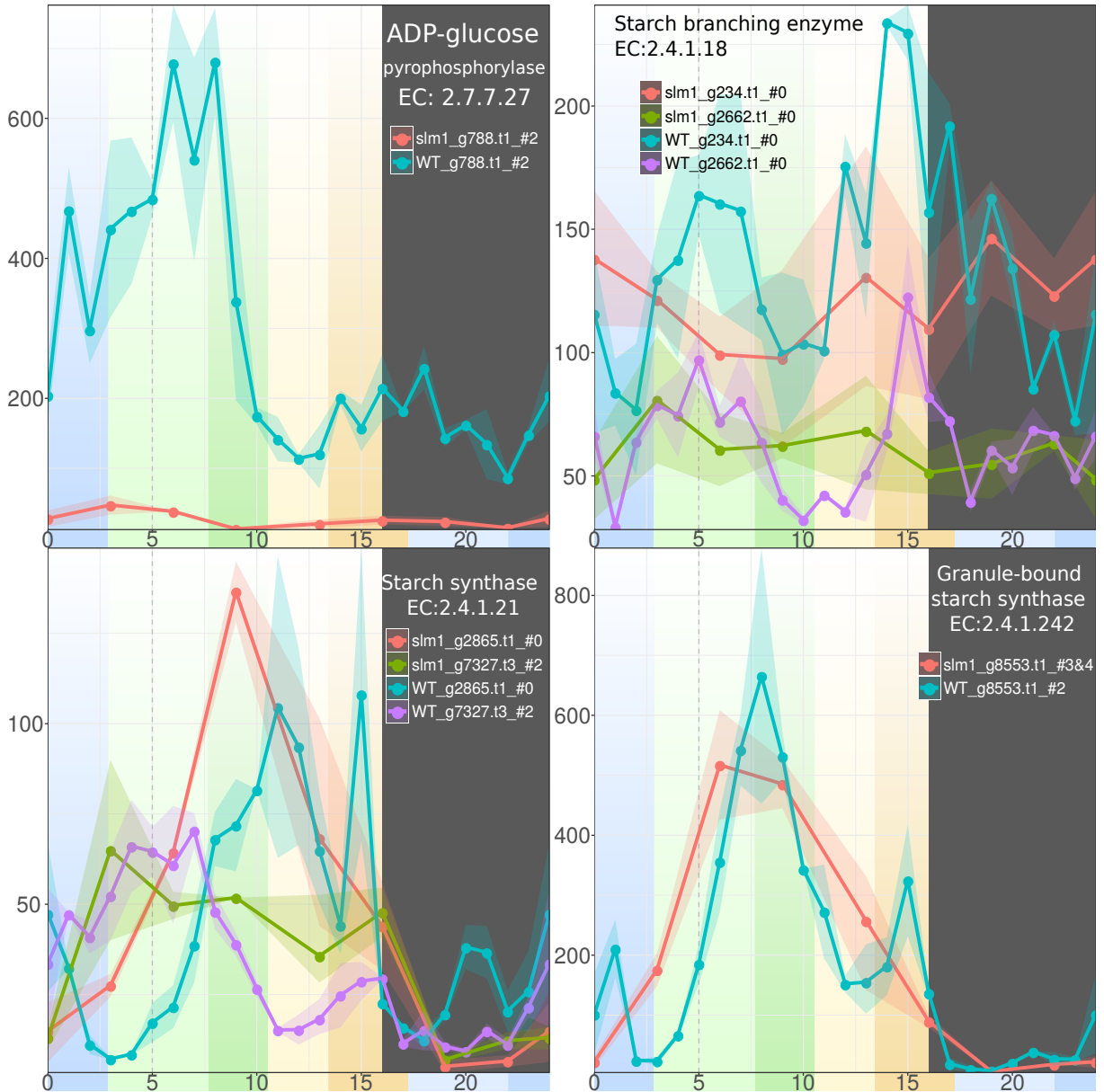


Figure 6.6: Diurnal expression of genes related to carbon fixation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the six plots varies. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combination. For each gene-strain combination, the letter or the number behind the # symbol, indicate the cluster with which the gene is associated. The background color corresponds to the time phases from Figure 6.1. The dark gray area corresponds to the dark period.

6.4.3.2 *Carbon fixation*

The diurnal expression of the genes related to the reactions analyzed in this section is depicted in Figure 6.7. The expression of RuBisCO (EC:4.1.1.39) is noisy, although it is generally lower in *slm1* than in WT. The logical explanation is that less carbon can be fixed due to starch deficiency. Similarly, the bisphosphatase profile is noisy but with a clear pattern of higher expression between 4 h and 9 h in WT that is not observed in *slm1*. Other reactions associated to carbon fixation are fitting the expression pattern of clusters 1 or 2 in both WT and *slm1*. Among these, the two enzymes preceding in the carbon fixation pathway (phosphoriboisomerase and phosphopentokinase) follow an expected pattern with carbon fixation during the light with a strong peak at the beginning of the light period.

The RuBisCO activase is activated by light and helps to release the substrate from the RuBisCO active site, accelerating the reaction [249, 250]. Three candidate genes were identified among which only *g6221.t1* displays high expression and a detected time profile, shown in Figure 6.7. Overall, the expression of the RuBisCO activase resembles the expression of the RuBisCO, but with an higher amplitude and a few hours earlier.

In complement to the RuBisCO activase, the RuBisCO inhibitor 2-Carboxy-D-arabinitol 1-phosphate (CA1P) is inactivated by the 2-carboxy-D-arabinitol-1-phosphatase (CA1Pase EC: 3.1.3.63). CA1P normally binds to RuBisCO under dark conditions, which prevents it from performing chemical reactions. In WT the gene *g18093.t1*, associated to the CA1Pase, displays a rather constant expression through both light and dark periods, except for a dip in expression around 1–2 h. However, in *slm1*, CA1Pase has a higher expression during the dark period and lower during the day.

All the aforementioned enzymes are co-localized in the pyrenoid with the RuBisCO [251]. This sub-cellular micro-compartment plays a major role in carbon fixation and it is logically localized in the chloroplast, along the thylakoid membrane. It is found in green algae (chlorophyta), and therefore found in *T. obliquus* [252]. Pyrenoids are not delimited by a membrane, but they accumulate starch at their periphery in the form of a sheath. It is believed that starch is a barrier that limits the transfer of CO₂ to the rest of the chloroplast and outer compartments. It was demonstrated that the pyrenoid plays a role in the CCM, allowing higher levels of carbon fixation in *C. reinhardtii* [58], however analysis of a *C. reinhardtii* starchless

mutant showed that the starch sheath surrounding the pyrenoid is not involved in the CCM [253]. Therefore, the starchless mutant, *slm1*, should not suffer from a lessened CCM in comparison to the wild-type. The lower yield on light should be due to the lack of a transient energy storage allowing for harvesting more light energy during the day and using this in the night.

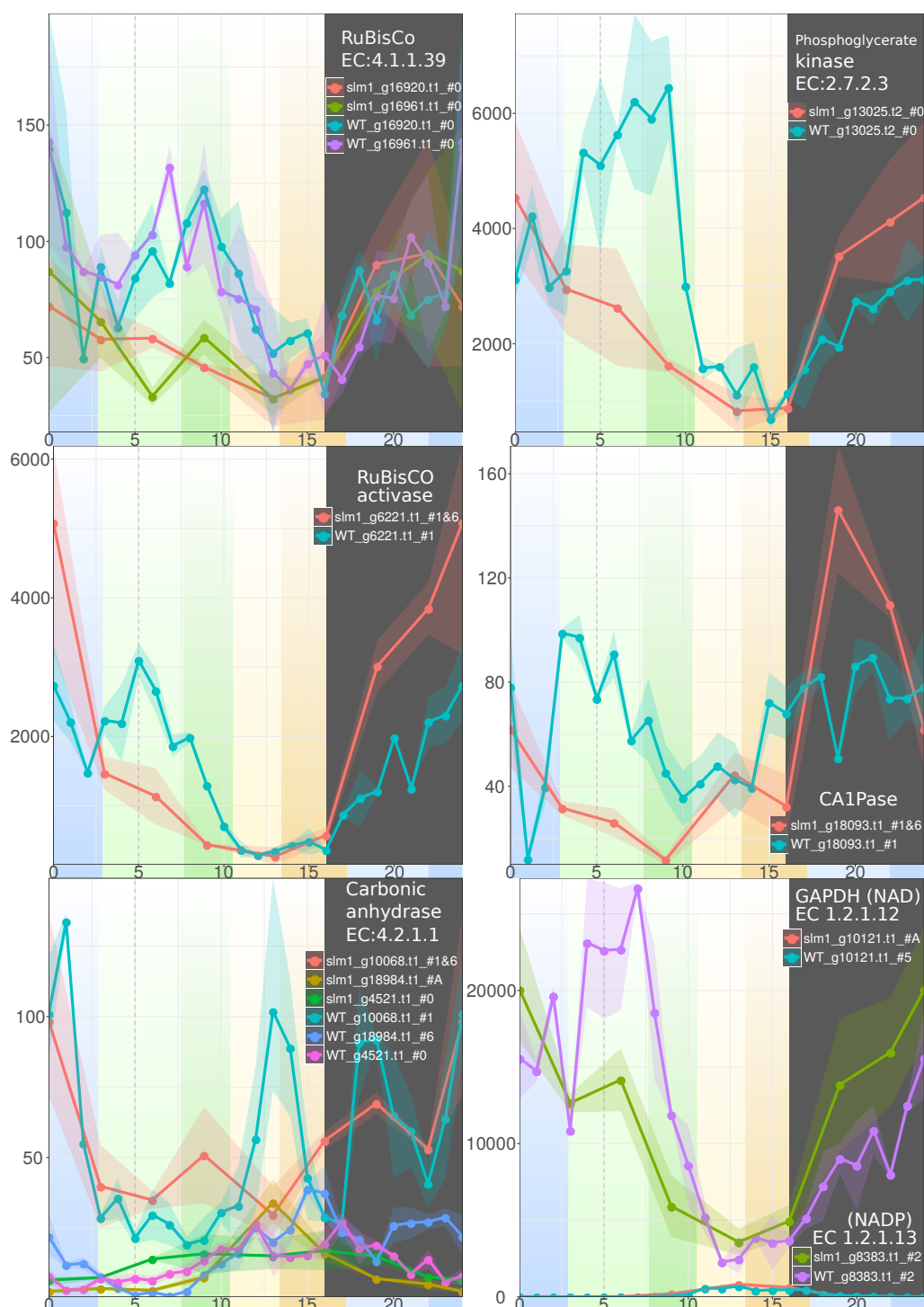


Figure 6.7: Diurnal expression of genes related to carbon fixation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the six plots vary. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each strain-gene-cluster combination. For each gene-strain combination, the letter or the number behind the #symbol, indicates the cluster in which the gene was found. The background color corresponds to the time phases from Figure 6.1. The dark gray area corresponds to the dark period.

6.4.3.3 ***Nitrogen metabolism***

The diurnal expression of the genes related to the reactions analyzed in this section is depicted in the Figure 6.8. The first enzymes of the nitrogen assimilation pathways are nitrate and nitrite reductases converting nitrate to ammonium. These enzymes are known to be strongly regulated by the light and nitrogen condition of the cell [254]. We were not able to identify the nitrate reductase gene in the transcriptome. The nitrite reductase gene displayed extremely low expression and no identifiable diurnal profile. Ammonium can be transformed into glutamate via two pathways, either by the glutamate dehydrogenase (GDH, EC 1.4.1.3 and 1.4.1.4), or via the glutamate synthase cycle that consists of glutamine synthase (GS, EC 6.3.1.2) followed by glutamate synthase (EC 1.4.1.14 and 1.4.7.1). GS is known as the main contributor for nitrogen assimilation [255], whereas GDH is generally involved in balancing nitrogen between metabolites by recycling ammonium for GS.

The candidate gene for GS (g10416.t1) was located in the chloroplast [37, 255] and showed higher expression at night and early day with similar profiles in WT and *slm1*. Expression during the dark period seems to peak at the same time as the GDH (g547.t1). For the glutamate synthase, we have three candidates, two using NAD as a cofactor, and one using ferredoxin as a cofactor. The two transcripts associated to the NAD-dependent reaction peak at the end of the day and have a similar profile in WT and *slm1*. WT expression of the ferredoxin-dependent reaction is very similar to GS but with a small time-shift (approximately 2h) towards later expression. Additionally, we also notice a small peak of expression right before the dark period that comes right after the peak of the two genes associated to the NAD-dependent glutamate synthase. Additionally, the ferredoxin-dependent glutamine synthase also peaks right after the GDH during the late light period, but also several hours after the GDH peak during the dark period. These observations in WT match the observed behavior of these enzymes in *C. reinhardtii* [256], meaning that the ferredoxin-dependent enzyme cannot only perform the assimilation during the light period, but also contributes to the assimilation during the dark period. Expression of the ferredoxin-dependent enzymes in *slm1* follows a similar temporal profile but is strongly reduced.

Two genes were associated to GDH EC 1.4.1.3, namely g547.t1 and g12033.t1. The second one, g12033.t1, was also associated to EC 1.4.1.4, although with a lower score. Both genes have similar WT expression profiles. In *slm1*, g547.t1 exhibits a similar profile whereas g12033.t1 in

slm1 appears to have lost sharp peaks with very gradual variation. The opposing expression of GDH and GS matches previous observations in *C. reinhardtii* [256], suggesting GDH to be an exclusively catabolic enzyme. These profiles suggest nitrogen recycling at the end of the light period and beginning of dark. Differences in the GDH are probably related to differences in cofactor utilization [254]. Nitrogen concentration measurements (data not shown) revealed that *slm1* does not assimilate nitrogen during the dark period, which would suggest EC 1.4.1.4 as the main route for nitrogen fixation during the dark period.

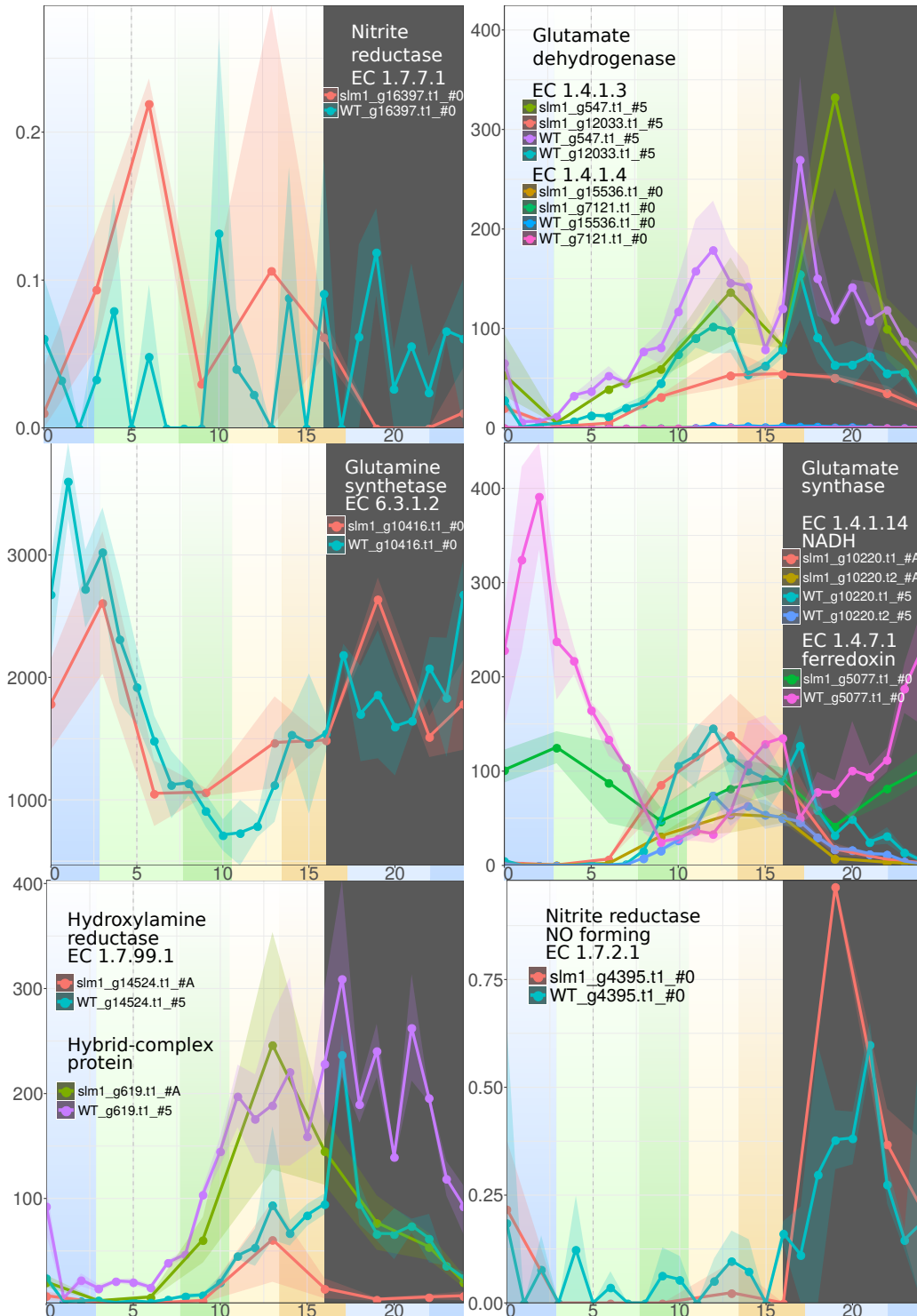


Figure 6.8: Diurnal expression of genes related to nitrogen assimilation. The points represent the mean for each time points, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together. All genes in the same plot are using the same y axis scale (not shown), but the scale between each of the five plots vary. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each strain-gene-cluster combination. For each gene-strain combination, the letter or the number behind the # symbol, indicate the cluster in which the gene was found. The background color corresponds to the time phases from Figure 6.1. The gray area corresponds to the dark period.

6.5 Conclusions

Oleaginous microalgae are a promising source of biofuels. Among the oleaginous microalgae, *Tetrademus obliquus* is a promising candidate that can reach a high maximum TAG yield on light and TAG content under nitrogen starvation. However, in order to develop strategies to enhance lipid productivity for commercial production, a system-level understanding of metabolism is essential. Large scale microalgal production will be done outdoors under natural LD cycles. Therefore, the impact of LD cycles has to be carefully considered when characterizing the behavior of microalgae.

In this work, we show that LD cycles induce systems level transcriptional changes in 4686 genes showing a clear expression pattern and indicating a strict succession of cellular events. These cellular events were found to be in accordance with the biochemical measurements. While some regulations seem a direct response to light, other regulations reflect an anticipation to the switch from light to dark and vice-versa. This observed anticipation indicates an inner time-keeping system.

Additionally, we studied the diurnal transcriptional changes in the starchless mutant of *T. obliquus* *slm1*. Pathways directly associated to energy storage, such as carbon fixation, and other processes such as nitrogen metabolism are strongly affected in *slm1*. The large changes in activity of these processes are attributed to starch deficiency, which seems to be the only transient energy storage compound in *T. obliquus*. Genes associated to TAG and lipid degradation are highly expressed during the dark period in *slm1*. This suggests TAG as a transient energy storage, however, no significant changes in TAG content were measured. As a result of the lack of an energy source during the dark, some of the energy related processes shift to the light period and the algae are less well prepared to start the next cycle at the beginning of the day. More subtly, processes related to the cellular cycle were detected to start earlier in *slm1* than in WT.

Overall, we provide for the first time a diurnal transcriptional landscape under LD cycles with a high resolution (1h intervals) of an oleaginous green microalgae that produces both starch and TAG under nitrogen starvation. This is also the first time a diurnal transcriptional landscape is described and compared for a starchless mutant of green algae. The diverse set of insights revealed in this analysis of the diurnal cycle of *T. obliquus* is very valuable to develop

strategies to increase yields. We suggest that the presented transcriptional landscape should be carefully considered when designing future experiments with LD cycles, including metabolic engineering approaches.

Chapter 7

Evaluation of diurnal responses of *Tetradesmus*
obliquus under nitrogen limitation

Benoit M. Carreres[§], G. Mitsue León-Saiki[§], Ilse M. Remmers, René H. Wijffels, Vitor A.P. Martins dos santos, Douwe van der Veen, Peter J. Schaap, Maria Suarez-Diez, Dirk E. Martens

§ Authors contributed equally

ABSTRACT

Tetrademus obliquus is an oleaginous microalga with high potential for triacylglycerides production. We characterized the biochemical composition and the transcriptional landscape of *T. obliquus* wild-type and the starchless mutant (*slm1*), adapted to 16:8h light dark (LD) cycles under nitrogen limitation. In comparison to the nitrogen replete conditions, the diurnal RNA samples from both strains also displayed a cyclic pattern, but with much less variation which could be related to a reduced transcription activity in at least the usually highly active processes. During nitrogen limitation, the wild-type continued to use starch as the preferred storage compound to store energy and carbon. Starch was accumulated to an average content of $0.25 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$, which is higher than the maximum observed under nitrogen replete conditions. Small oscillations were observed, indicating that starch was being used as a diurnal energy storage compound, but to a lesser extent than under nitrogen replete conditions. For the *slm1* mutant, TAG content was higher than for the wild-type (average steady state value was $0.26 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ for *slm1* compared to $0.06 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ for the wild-type). Despite the higher TAG content in the *slm1*, the conversion efficiency of photons into biomass components for the *slm1* was only half of the one obtained for the wild-type. This is related to the observed decrease in biomass productivity (from $1.29 \text{ g}_{\text{DW}} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the wild-type to $0.52 \text{ g}_{\text{DW}} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the *slm1*). While the transcriptome of *slm1* displayed clear signs of energy generation by degrading TAG and amino-acids during the dark period, no significant variation of the metabolites could be measured. When looking through the diurnal cycle, the photosynthetic efficiency was lower for the *slm1* mutant compared to the wild-type especially during the second half of the light period, where starch accumulation occurred in the wild-type.

7.1 Introduction

Microalgae are considered as one of the most promising renewable sources for the production of feed, fuels and chemicals [40, 106, 210]. However, in order to make large scale production of algal biodiesel economically feasible, high triacylglycerides (TAG) productivity are needed [210, 257].

Tetrademus obliquus (formerly known as *Scenedesmus obliquus* [199]) is a microalga of interest due to its high TAG content and the fact that it can retain a high photosynthetic efficiency under nitrogen starvation [24, 258–260]. Besides, de Jaeger et al. [261] developed the starchless mutant *slm1* that showed a higher maximum TAG yield on light compared to the wild-type ($0.22 \text{ g}_{\text{TAG}} \cdot \text{mol}_{\text{ph}}^{-1}$ compared to $0.14 \text{ g}_{\text{TAG}} \cdot \text{mol}_{\text{ph}}^{-1}$), as well as a higher maximum TAG content ($0.57 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ compared to $0.45 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$) under batch nitrogen starvation and continuous illumination [34].

In large scale production, microalgal biomass first will be grown outdoors under favorable nitrogen replete conditions and light/dark cycles (LD). After biomass has been produced, the lipid production phase can be done under nitrogen deplete conditions. The behavior of *T. obliquus* under nitrogen replete conditions and diurnal 16:8 h light/dark (LD) cycles has been previously studied [24, 262]. Under such conditions, *T. obliquus* wild-type and starchless mutant *slm1* showed synchronized cell division and growth. *T. obliquus* wild-type showed diurnal oscillations in biomass composition, with an accumulation of starch during the light period that was consumed during the dark period and beginning of next light period. For the *slm1*, no such oscillations in biomass composition were found, which shows that this microalga does not need starch as a temporary energy storage compound to survive dark periods up to 12h [214]. The lack of starch, however, did result in a reduction of the conversion of energy (photons) to biomass [24, 214].

Under nitrogen deplete conditions, many microalgal species accumulate starch or TAG, which allows the capture and storage of energy and carbon that can be rapidly used when nitrogen becomes available again [263]. However, under deplete conditions growth stops, the photosynthetic efficiency rapidly decreases and also TAG production stops after a certain time. Since production of TAG occurs also under nitrogen limitation in combination with LD cycles, this may result in continuous production and in the end a higher yield of TAG on light. Consequently, the behavior of *T. obliquus* under these conditions is of interest. The average

steady state behavior under nitrogen limitation and LD cycles of both *T. obliquus* wild type and *slm1* has been previously reported [28]. However, the detailed diurnal behavior during a single LD cycle under nitrogen limitation has not been studied in these strains. Therefore, the aim of this paper is to obtain a better understanding of the diurnal behavior of *T. obliquus* wild-type and starchless mutant *slm1* under nitrogen limitation, with focus on the diurnal changes in starch and TAG content, and on energy efficiency. Next to studying the biochemical composition and light use, gene expression was examined from RNA sequencing data to obtain more insight in the precise regulation in the affected pathways. These results will also be compared to the previously observed changes without nitrogen limitation [262].

7.2 Materials and methods

7.2.1 Strains, pre-culture conditions and cultivation medium

Wild-type *Tetradismus obliquus* UTEX 393 was obtained from the Culture Collection of Algae, University of Texas. The starchless mutant of *T. obliquus* (*slm1*) was generated by de Jaeger et al [261]. Pre-cultures were grown as described by León-Saiki et al. [24] in defined medium described by Breuer et al. [258].

7.2.2 Reactor set-up and experimental conditions

T. obliquus was continuously cultivated in a sterilized flat panel airlift-loop photobioreactor with a working volume of 1.7 L and a 0.02 m light path (Labfors 5 Lux, Infors HT, Switzerland). Temperature was maintained at 27.5 °C and pH was controlled at 7.0 by the automatic addition of 2.5% (v/v) H₂SO₄. The reactor was continuously sparged with 1 L·min⁻¹ air enriched with 2% CO₂. Light was provided by a light panel with 260 LEDs with a warm white spectrum at an incident photon flux density of 500 μmol·m⁻²·s⁻¹ in a 16:8 h light/dark (LD) block cycle. The reactor was inoculated to an optical density (OD₇₅₀) of 0.1.

Duplicate turbidostat cultivations were done, where the light intensity at the rear of the reactor was kept constant at $10 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ by addition of medium. Dilution medium [258] was prepared without KNO_3 and KNO_3 was separately fed at $0.075 \text{ g}_\text{N}\cdot\text{L}^{-1}\text{day}^{-1}$ for the wild-type and $0.052 \text{ g}_\text{N}\cdot\text{L}^{-1}\text{day}^{-1}$ for the *slm1*. This corresponds to a nitrogen limitation of 30% of the nitrogen consumption rate observed under nitrogen replete and light limitation [28]. Dilution medium was only added during the light period and switched off during the night. This type of system allows the study of the interaction between growth, nitrogen consumption and lipid accumulation [88].

Cultures were allowed to reach steady state, which was defined as a constant biomass concentration and daily dilution rate for a period of at least 3 days. After steady state was reached, liquid samples were freshly taken from the reactor and either immediately used for dry weight measurements or centrifuged for 5 min at $2360 \times g$ for biochemical analysis (approximately 10-12 mg for proteins, 5 mg for starch, 7 mg for triacylglycerides (TAG) and 5 mg for total carbohydrates). In addition, at least three daily overflow samples were collected for each strain. Due to restrictions on working hours of the laboratory, the samples were collected in two successive time settings to allow sampling the dark period during the day. After collecting samples of the first half of the cycle, light settings were shifted and the culture was then allowed to reach oscillating steady-state before collecting samples for the second half of the cycle. The first and the last samples of each time settings are overlapping samples for control.

7.2.3 Dilution rate

Dilution rate was calculated by measuring the feed of the dilution medium (without KNO_3) and the feed of the KNO_3 medium over intervals of 10 min. Dilution rates were then calculated as the total medium added over 30 min divided by the reactor volume. After that, a moving average of 60 min was done.

7.2.4 Measurements

Dry weight (DW) concentration was determined as described by Kliphuis et al. [90]. Total carbohydrates were determined according to DuBois et al. [264] and Hebert et al. [94] using

glucose as a standard. Starch content was measured as described by de Jaeger et al. [261] using a total starch kit (Megazyme, Ireland). Triacylglycerol (TAG) content was quantified as described by Remmers et al. [28] using glyceryl trinonadecanoate (T4632; Sigma Aldrich) and 1,2-dipentadecanoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) (840446, Avanti Polar Lipids Inc.) as internal standards. Total protein concentration was measured done using a colorimetric assay (Bio-Rad DC protein assay) as described by Postma et al. [265].

7.2.5 Conversion of photons into biomass

The theoretical amount of photons converted into biomass was calculated based on the theoretical photon requirements for the biomass components ($1.33 \text{ g}\cdot\text{mol}_{\text{ph}}^{-1}$ for TAG, $3.24 \text{ g}\cdot\text{mol}_{\text{ph}}^{-1}$ for starch, and $1.62 \text{ g}\cdot\text{mol}_{\text{ph}}^{-1}$ for functional biomass [90, 202] and the photon absorption rate ($1.36 \text{ mol}_{\text{ph}}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$) [24]. Samples were taken in intervals of three hours. For the calculation of the hourly energy conversion efficiency the additional points for biomass composition were estimated assuming a proportional change between the measured points. Biomass composition of the overflow samples (collected during 24 h on ice) was used to calculate the average steady state values.

7.2.6 Starch and TAG productivities

The starch productivity ($r_{\text{starch},t}$ in $\text{g}\cdot\text{L}^{-1}\cdot\text{h}^{-1}$) was calculated using a balance for starch over short time intervals (Eq. 1), considering the change in starch content over three hour time intervals (

$\frac{dC_{\text{starch}}}{dt}$), the dilution rate (D_t in h^{-1}) and the starch content ($C_{\text{starch},t}$ in $\text{g}\cdot\text{L}^{-1}$):

$$\frac{dC_{\text{starch}}}{dt} = -D_t \cdot C_{\text{starch},t} + r_{\text{starch},t} \quad (\text{Eq. 1}).$$

The TAG productivity was calculated in a similar way where the changes in TAG content over intervals of three hours ($\frac{dC_{\text{TAG}}}{dt}$), the dilution rate (D_t in h^{-1}), and the TAG content ($C_{\text{TAG},t}$ in $\text{g}\cdot\text{L}^{-1}$) were used. During the dark period the dilution rate is zero and the productivity is determined by the accumulation term only.

7.2.7 RNA sampling, extraction, and sequencing

After steady state was reached, 8 mL samples were taken for RNA extraction (approximately 10 mg_{DW}). Cells were immediately collected by centrifugation (4255 x g, 0 °C for 5 min), supernatant was discarded and pellets were frozen in liquid nitrogen and stored at -80 °C until further extraction. Samples for RNA extraction were taken in intervals of 3 h for both the wild-type and the *slm1* mutant. Sampling strategy and RNA extraction were performed as described in our previous study [262].

7.2.8 Analysis of RNA expression

The RNA-seq samples were treated and analyzed with the same method as our previous study [262]. To summarize, the expression was calculated by aligning the read samples over the available genome of *T. obliquus* UTEX-393 [225]. From the aligned samples, Fragments Per Kilobase of transcripts per Million reads mapped (FPKM) values were computed and used in the following steps. Genes with significant changes of expression over time were identified using maSigPro [232]. These genes were then separated using hierarchical clustering and Pearson's correlation. The optimal number for cluster separation was determined by combining the results from seven indexes commonly used for this purpose. These indexes were calculated using the functions of "clusters.stats" from the "fpc" R package [266]. Functional characterization of genes in each cluster was done using the functional annotation presented in [262]. Again, the same methods were applied to determine the enrichment of clusters in GO terms and in KEGG pathways [262]. In brief, enrichment analyses were performed using the hypergeometric function to model the background probability density. Enzyme commission (EC) numbers were associated to metabolic pathways using KEGG pathways maps. Pathways fitting the following requirements were kept for further analysis: 60% coverage if 3 to 6 EC numbers annotated, and 50% coverage if 6 to 10 EC numbers annotated, and 25% of coverage if more than 10 EC numbers were annotated. Enrichments with a p-value lower than 0.05 were considered significant. For the GO enrichment analysis multiple test correction was performed using the Benjamini-Hochberg procedure. Enrichments with FDR<0.05 were considered significant.

The annotation was curated manually for specific reactions related to starch, lipids and TAG synthesis and degradation pathways. To do so, the combination of top score and p-value from

Enz-DP were considered, as well as the relative difference between the scores, and as a final verification, the proteins were aligned using DELTA-Blast and HMMER3.

7.3 Results and discussion

7.3.1 Growth of *Tetrademus obliquus* wild-type and *slm1*

Our experimental set-up using turbidostat controlled systems imposes a fixed light uptake. We created an energy imbalance by decreasing the nitrogen supply rate to 30% of the value that would be needed to have nitrogen replete growth at the fixed light uptake rate.

Nitrogen limitation resulted in a reduced growth rate (Figure 7.1). Nitrogen limited cultures showed a diurnal pattern in dilution rate (D), with both the wild-type and the starchless mutant *slm1* showing a similar pattern. Dilution patterns were, however, different from those under nitrogen replete conditions, specifically with respect to the start of dilution after the light went on and the moment where the maximum dilution rate was reached. Dilution started about 4 h after start of the light period under nitrogen limitation, while under nitrogen replete conditions dilution started immediately after the beginning of the light period. As a result, the time at which the maximum dilution rate was reached was also shifted and was reached at 7h. A possible explanation for this shift may be that, as nitrogen was not added during the night, nitrogen content was equal to zero (starvation). This may have resulted in cells being unable to prepare for the next day, or cells switching to a nitrogen starvation mode, breaking down proteins and pigments, from which they had to recover. With respect to the first case, it is possible that cells prepare during the night for the next day so they can start growing quickly. If nitrogen is needed for this preparation, cells would have to first fixate nitrogen when the light is on, which would explain the delay. With respect to the second case, cells switch to a nitrogen starvation mode, stopping pigment synthesis and breaking down pigments and proteins, which would result in a decreased pigment content at the beginning of the light period. This last point is supported by the fact that the light out for the wild-type was $1 \mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ above set-point, indicating breakdown of light absorbing material. Consequently, first a certain amount of pigment has to be synthesized before the light out drops below the

set-point and dilution starts. For *slm1*, an oscillatory pattern was observed (Figure 7.1B), which was also observed in our previous study in nitrogen replete conditions [262].

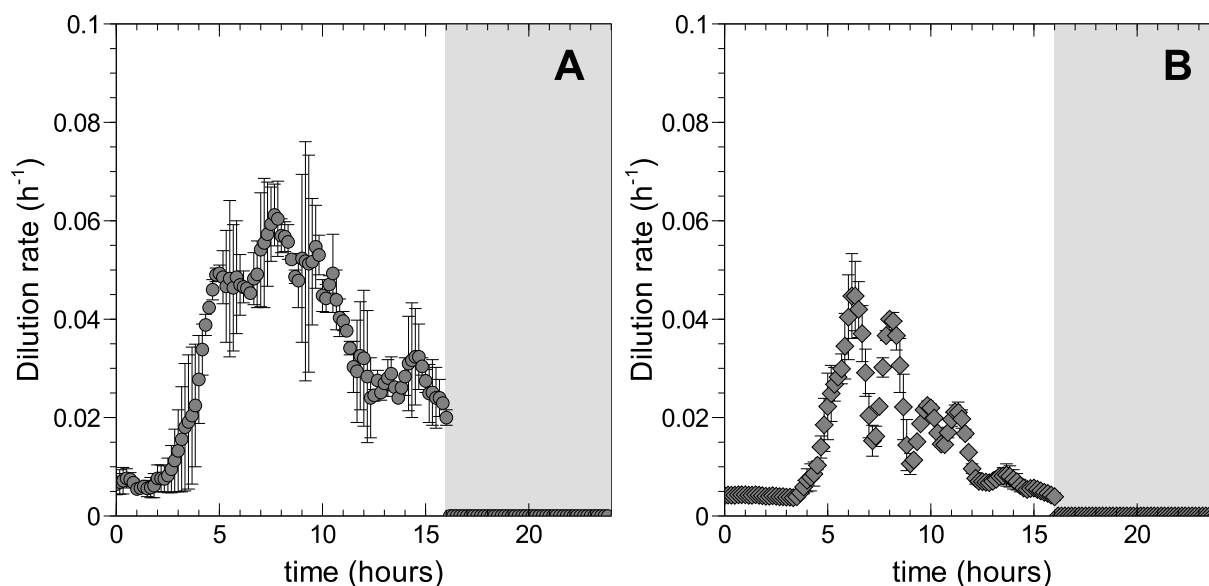


Figure 7.1: Changes in dilution rate for *Tetrademus obliquus* wild-type (A) and *slm1* (B) under nitrogen limitation. Values represent the average of replicate reactors and error bars represent minimum and maximum values. The x axis shows hours after start of light period. Shaded areas indicate the dark period.

Under favorable nitrogen replete conditions, oleaginous microalgae produce only small amounts of TAG, while under unfavorable nitrogen starvation conditions, TAG accumulation is induced and contents of up to $0.40 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ can be reached [267, 268]. However, these unfavorable environmental conditions come with the disadvantage of a complete stop of cell division [28]. Nitrogen limitation processes were suggested as an attempt to overcome the disadvantages observed under nitrogen starvation [88]. As observed for *T. obliquus*, growth still occurs under nitrogen limitation, but this is reduced to approximately half of the one observed under nitrogen replete conditions (from $1.12 \pm 0.01 \text{ day}^{-1}$ under nitrogen replete [24] to $0.55 \pm 0.07 \text{ day}^{-1}$ under nitrogen limitation). This was also observed for other microalgae, such as *Neochloris oleoabundans* [257]. For the starchless mutant *slm1* the 24 h average daily dilution rate, i.e. growth rate, was reduced even more (from $0.90 \pm 0.01 \text{ day}^{-1}$ under nitrogen replete to $0.22 \pm 0.02 \text{ day}^{-1}$ under nitrogen limitation), showing again the reduction in growth and lower photosynthetic efficiency when blocking starch production.

7.3.2 Diurnal changes of storage compounds: Starch and TAG.

During nitrogen limitation, starch continued to be the preferred storage compound to store energy and carbon for the wild-type (Figure 7.2A). Starch was accumulated to an average content of $0.25 \text{ g}\cdot\text{g}_{\text{DW}}^{-1}$, with a maximum content of approximately $0.29 \text{ g}\cdot\text{g}_{\text{DW}}^{-1}$, which is higher than the maximum observed under nitrogen replete conditions ($0.20 \text{ g}\cdot\text{g}_{\text{DW}}^{-1}$) [24]. The overall preference of *T. obliquus* wild-type towards storing starch has been previously observed [28, 34]. When looking into the starch content during the diurnal cycle, a small oscillation can be observed (Figure 7.2A), where starch content slightly increases between 9-12 h and decreases during the dark period. This can also be seen in the starch productivity shown in Figure 7.2C, that was calculated according to equation 1. Starch productivity remained positive just above zero during the first 9 h, then increased reaching the maximum value on $t=12 \text{ h}$ after which the starch productivity decreased and became negative at $t=21 \text{ h}$, indicating consumption.

Under nitrogen replete conditions, TAG is not accumulated, whereas under nitrogen limitation or starvation substantial amounts of TAG can be accumulated in *T. obliquus*, as observed by Remmers et al. [28]. However, no information is known on the diurnal role of this compound under nitrogen limitation in *T. obliquus*. Therefore, we also measured the TAG content during the diurnal cycle in intervals of 3 h. The wild-type showed a constant content of approximately $0.07 \text{ g}\cdot\text{g}_{\text{DW}}^{-1}$ for all measured time points (Figure 7.2A). This rather low content of TAG can be explained by the fact that carbon and energy are mainly stored as starch. TAG is only formed if the carbon and energy supply exceeds the storage capacity in starch, as also mentioned in literature [28, 34, 269]. For *T. obliquus*, Remmers et al. [28] found that TAG is only accumulated after starch reaches its cellular maximum of $0.40 \text{ g}\cdot\text{g}_{\text{DW}}^{-1}$. Under this nitrogen limited condition, this maximum content is not reached, which explains the low content of TAG in the wild-type. The preference for starch cannot be explained from the energy density, since more energy can be stored in fatty acids as compared to starch (yield of complete oxidation of fatty acids is about $9 \text{ kcal}\cdot\text{g}^{-1}$ compared to $4 \text{ kcal}\cdot\text{g}^{-1}$ for carbohydrates [270]). A possible explanation for the preference for starch could be that the synthesis and degradation of TAGs requires more cellular resources and enzymes [271]. Another reason might be that when cells grow photoautotrophically, fatty acids synthesis represent a significant loss of carbon as conversion of pyruvate to acetyl-CoA by the pyruvate dehydrogenase complex involves the loss of one of the fixed carbons. This renders starch a more efficient carbon storage compound regarding carbon fixation [272].

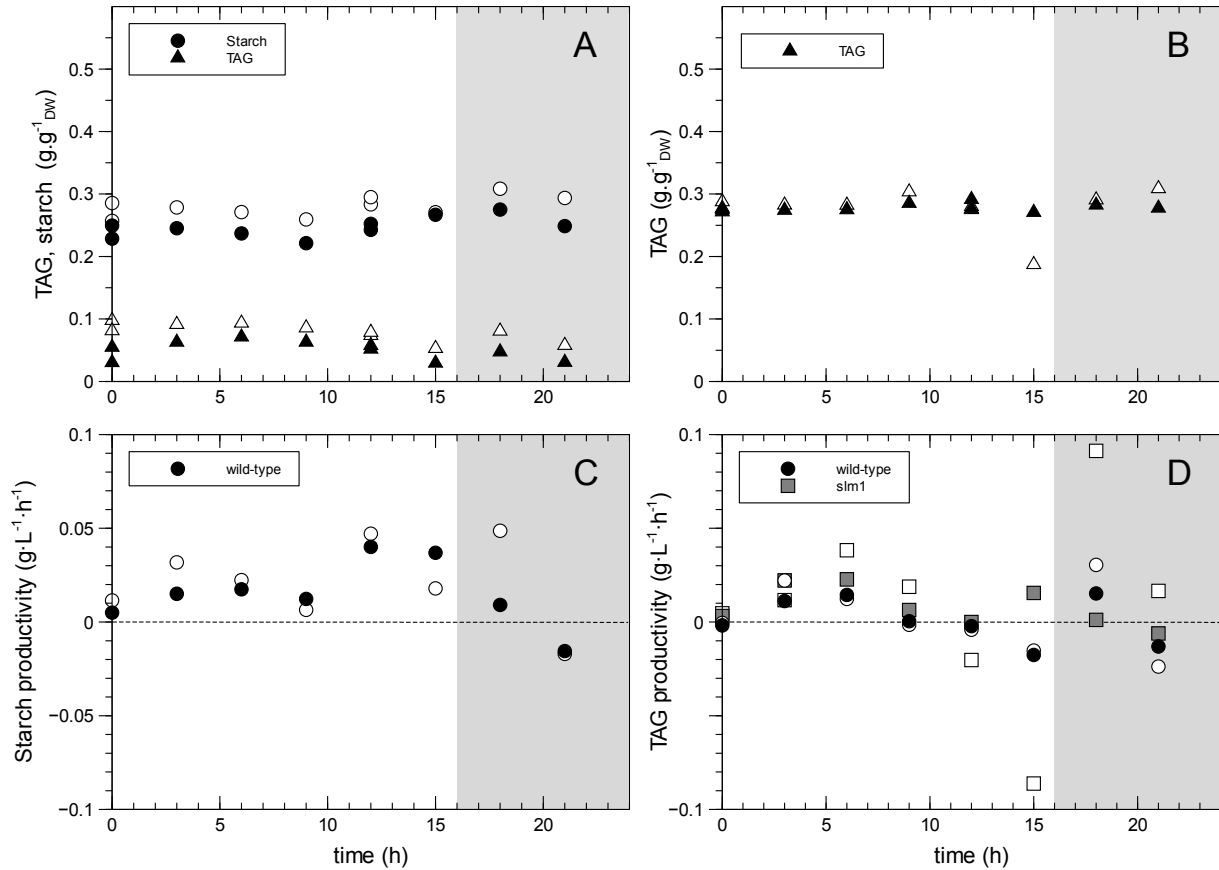


Figure 7.2: Diurnal changes under nitrogen limitation in starch and triacylglycerides (TAG) content for *Tetradesmus obliquus* wild-type (A) and *slm1* (B) and in starch productivity for the wild-type (C) and TAG productivity for wild-type and *slm1* (D). Open and closed symbols represent replicate cultures. The x axis shows hours after the start of the light period. Shaded areas indicate the dark period.

For the *slm1* mutant the TAG content (approximately 0.27 g·g_{DW}⁻¹) was about 4 times higher than for the wild-type during the diurnal cycle (Figure 7.2B). However, TAG showed no diurnal oscillations under nitrogen limitation for either strain (Figure 7.2A&B). When looking at the TAG productivity, diurnal variations were expected for the *slm1* as the dilution rate is changing. However, no clear diurnal variations were observed (Figure 7.2D). This is because

the values for dilution rate are similar or smaller than the accumulation term ($\frac{dC_{TAG}}{dt}$). The accumulation term is in turn highly sensitive for small changes in the measured TAG content. Therefore TAG productivity calculations are highly affected by measurement errors in TAG content. For the wild-type, the values for the dilution rate are higher making the calculation of the TAG productivity less sensitive to small errors in TAG content, which in turn makes the differences between the time points more significant.

The other measured biomass components, proteins and non-starch carbohydrates did not show oscillations through the diurnal cycle for either strain (Supplementary Figure S1).

7.3.3 Energy conversion efficiency

Under nitrogen replete conditions, the energy conversion efficiency into biomass of the wild-type is higher than that of the *slm1* mutant [24, 214]. We therefore investigated if this was also the case for nitrogen limitation conditions where TAG accumulation occurs. For this, we calculated the minimal number of photons needed for the production of functional biomass, TAG, and starch (for the wild-type), based on the biomass composition of the overflow samples (average steady state values). Results are displayed as a percentage of used photons from the total amount supplied in Error: Reference source not foundA. As can be seen, also for nitrogen limited conditions the overall efficiency is lower for the *slm1* mutant than for the wild-type (approximately 50% for the wild-type and 25% for the *slm1*). However, the TAG content on the *slm1* is higher than for the wild-type resulting in a slightly higher TAG volumetric productivity for the *slm1* as compared to the wild-type ($0.09 \text{ g}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$ for the wild-type and $0.14 \text{ g}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$ for the *slm1*). The lower overall efficiency is related to the fact that the biomass productivity in the mutant is reduced to less than half of that of the wild-type ($1.29\pm 0.11 \text{ g}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$ for the wild-type and $0.52\pm 0.06 \text{ g}\cdot\text{L}^{-1}\cdot\text{day}^{-1}$ for the *slm1*) (Figure S1). Also, both strains are less efficient under nitrogen limited conditions than under nitrogen replete conditions ($62.77\pm 0.08\%$ for the wild-type and $49.75\pm 0.07\%$ for the *slm1* [214]). When looking into the conversion efficiency during the diurnal cycle in intervals of one hour (Error: Reference source not foundB&C), we observe that during the first half of the day, both strains behave similarly, with the wild-type showing an increase in energy efficiency about one hour before the *slm1*. However, in the second half of the light period (approximately from $t=7 \text{ h}$), the *slm1* mutant is much less efficient than the wild-type. Notably the second half of the light period is the moment when starch accumulation occurs in the wild-type, probably explaining the extra energy fixated.

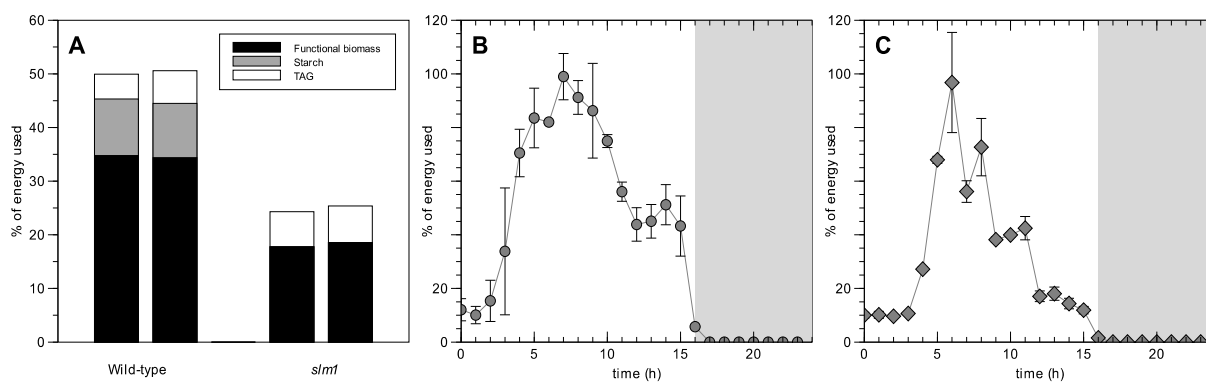


Figure 7.3. The percentage of photons (as a fraction of the total supplied) minimally needed to make the different biomass components as an average over 1 day (A) and over each hour of the day for *Tetrademus obliquus* wild-type (B) and *slm1* (C). For (A), values represent the average values of at least 3 overflow samples. For the hourly energy conversion efficiency (B and C) the x axis represents hours after the light was switched on. Values represent average of replicate reactors and error bars represent minimum and maximum values. Shaded area indicates the dark period.

Under nitrogen starvation both starch and TAG can act as an overflow sink for electrons [268]. For starchless mutants, it has been suggested that the difference in energy efficiency and biomass yield on light might be due to an increased rate of energy dissipation compared to the wild-type [28, 273] since electrons can no longer be channeled to starch. However, under nitrogen limitation growth still occurs and starch still seems to have a role as a diurnal energy storage compound, such as under nitrogen replete conditions [24], which provides an extra benefit for the wild-type compared to the *slm1*.

7.3.4 Transcriptional landscape

Previous transcriptome analysis on *T. obliquus* under the same LD cycles but in nitrogen replete conditions uncovered that genes following a diurnal expression pattern were contributing to the vast majority of expression changes [262]. Principal component analysis (PCA) of nitrogen replete samples and nitrogen limited samples (this study) under LD cycles, shows that the two first components explain 83% of the total variation in gene expression (Figure 7.4A). Samples from both strains (wild-type and *slm1*) grown in the same condition with respect to nitrogen availability and from the same time points overlap. However, the samples taken from the nitrogen replete condition are completely separated from those taken from the nitrogen limited condition. This indicates the limited nitrogen supply has a much

greater impact on gene expression than the changes caused by the mutation. In both conditions (nitrogen replete and deplete), there is a clear succession of time points in the PCA. Under nitrogen replete conditions, time points are arranged in a circular pattern (Figure 7.4A). However, under nitrogen deplete conditions, the circular pattern seems lost and samples at 3h overlap with the samples at 12h. Yet, when selecting the genes with a detected diurnal expression, the circular pattern is recovered (Figure 7.4B). This noticeable difference was not observed in the nitrogen replete condition, which indicates that nitrogen limitation results in an overall reduction of diurnal changes in expression. The time point samples of each strain in nitrogen limited condition (Figure 7.4B) draw an overlapping elliptic pattern, but the variation of expression in *slm1* is clearly lower than in the wild-type. Moreover, the samples of *slm1* during the dark period are all extremely close to each other, indicating very little transcriptional changes. Therefore, starch deficiency has a greater impact on gene expression in nitrogen limited condition than it has in nitrogen replete condition, with *slm1* being particularly inactive during the dark period. In complement, the heatmaps showing similarity of gene expression between time points is available in the supplementary Figure S2.

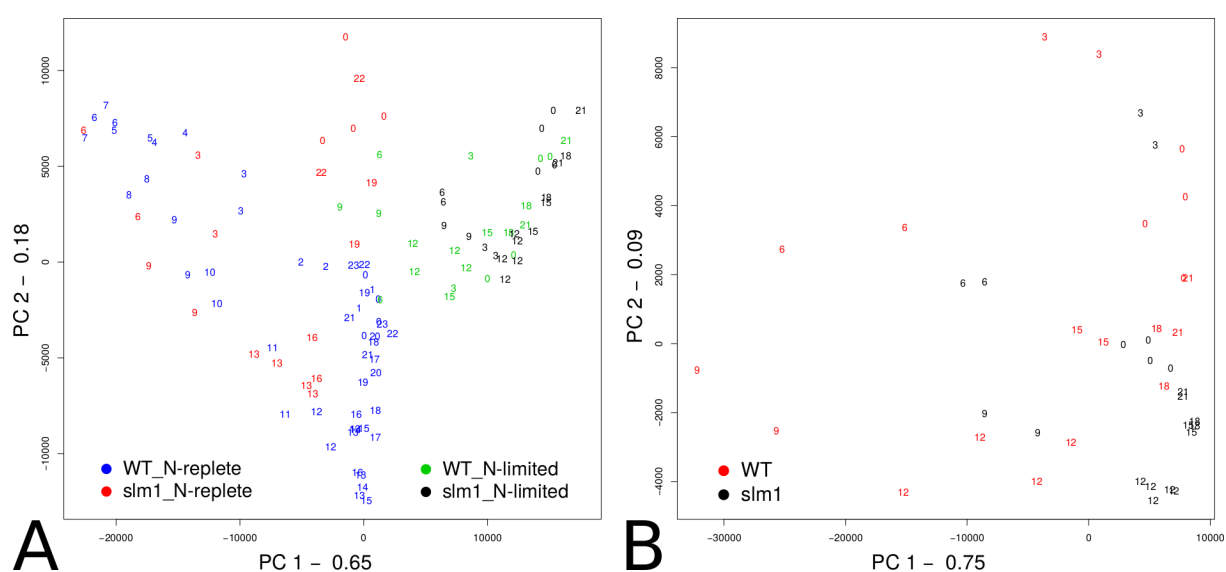


Figure 7.4: Principal component analyses of time point samples. (A) Comparison between samples considering all genes (no selection) in nitrogen replete [262] and nitrogen limited conditions. (B) Comparison between samples considering only selected genes (detected time profile) in nitrogen limited conditions. The variance explained by the first two components is indicated on axes labels.

7.3.5 Diurnal gene regulation and functions

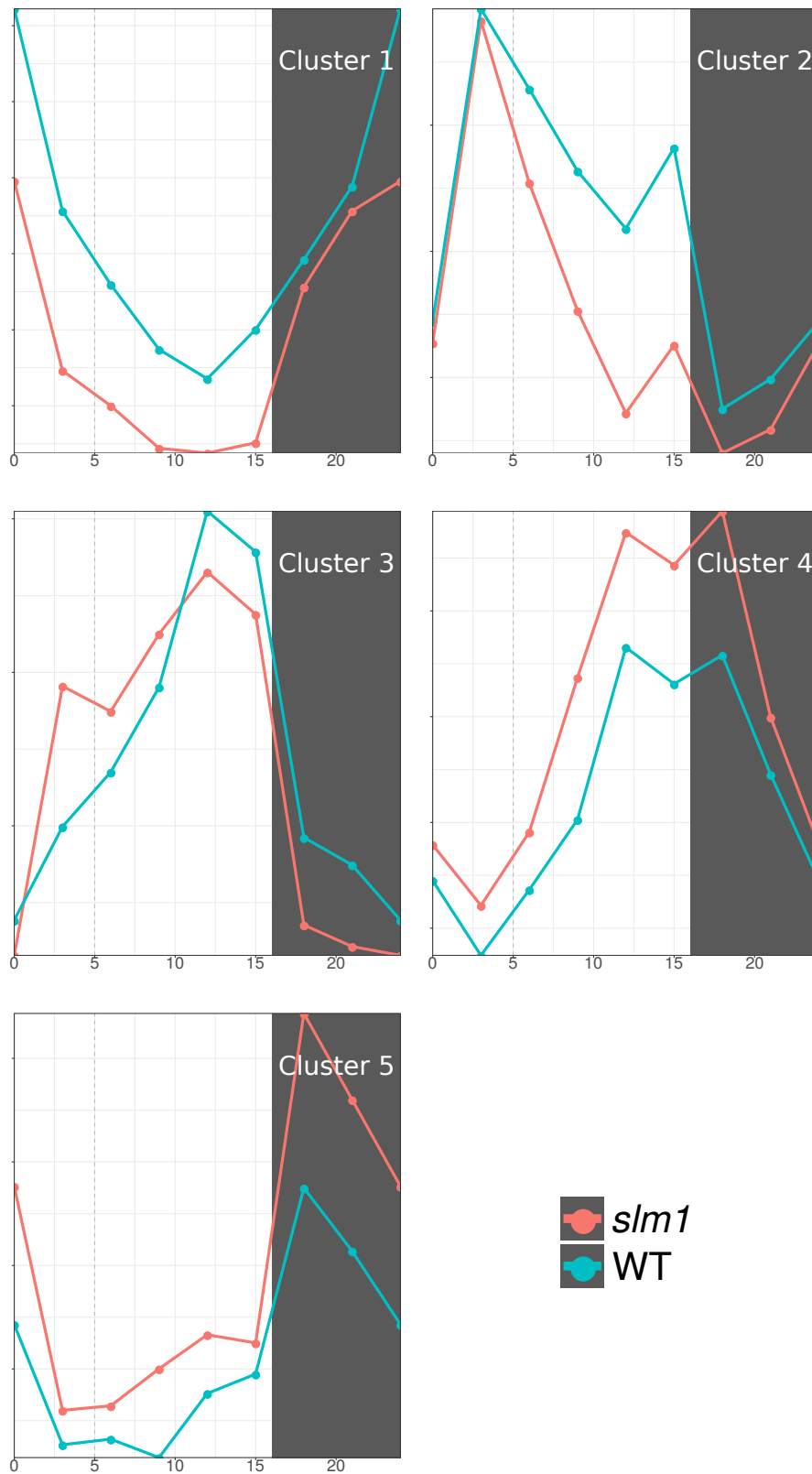


Figure 7.5: Expression profile over 16:8 light-dark cycle of the 5 genes clusters identified in both strains. The plots show the median profile of gene expression in the indicated clusters. Dark area represents the dark period.

Genes with a detected diurnal regulation were clustered using hierarchical clustering based on Pearson's correlation in different numbers of clusters ranging from 3 to 25. Similarities within and differences between clusters were evaluated with well established indexes, depicted in the supplementary file S3. Some of the selected index (first Dunn, Calinski-Harabasz and the ratio between the average within and average between clusters) gave no indication on the optimal number of clusters. However other metrics such as average silhouette width, second Dunn index or normalized gamma, show a local maximum (or minimum) at 4-6 clusters. Inspection of cluster similarities based on the Rand index (supplementary file S3) led to select 5 clusters (over 4 or 6). Based on these results, we considered separating the genes in 5 clusters to be optimal. Similar results are obtained for both strains. The median profile of these 5 clusters is displayed in Figure 7.5. Between the two strains, the clusters with the nearest median profile are also the clusters with the most genes in common (supplementary Figure S4). For those reasons, clusters were named the same.

As observed in our previous study under nitrogen replete conditions [262], the clusters reflect a succession of peaking expressions. In the wild-type, cluster 1 expression peaks at 0h, progressively decreases until 12h and next progressively increases. Cluster 2 expression peaks at 3h, decreases until 12h before peaking again at 15h, and finally reaches its lowest expression during dark period. Cluster 3 expression progressively increases from its lowest point at 0h until 12h, maintains high expression at 15, and its expression drops during the dark period. Cluster 4 expression is lowest at 3h and increases progressively until 12h, maintains high expression during the beginning of the dark period before decreasing at 18h. Finally, cluster 5 displays high expression during the dark period only with a peak of expression at 18h. The profile of the clusters in the *slm1* strain are similar to those of the wild-type, with noticeable differences of shape of the first three clusters. *Slm1* cluster 1 displays a higher expression increase right after the dark period. *Slm1* cluster 2 displays a much lower secondary peak at 15h, meaning that these genes are now mainly expressed during the first half of the day. The median profile of *slm1* cluster 3 seems exclusively expressed during the light period with a stronger expression in the early light period (time point 3, 6 and 9h). *Slm1* clusters 4 and 5 are the same as the wild-type, besides a simple upper shift on the y axis. Interestingly, in both strains, the expression profile of cluster 2 is opposite to the profile of cluster 4, and the profile of cluster 3 opposes that of clusters 1 and 5.

Table 7.1: Results of enrichment analyses. Summary of the results of the enrichment analyses. The first column contains the cluster identifier. The second and third contains the enriched pathways (p-value < 0.05). GO terms relate to biological processes (FDR<0.05) for the wild-type. Fourth and fifth column present the same information for *slm1*. Full set of enrichment results are available in the supplementary file S5.

Cluster	wild-type		<i>slm1</i>	
	Pathway	Biological Processes	Pathways	Biological Process
1	Fatty acid biosynthesis Glycine, serine and threonine metabolism Starch and sucrose metabolism Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Carbon fixation in photosynthetic organisms	protein phosphorylation fatty acid biosynthetic process oxidation-reduction process	Purine metabolism Geraniol degradation	protein phosphorylation movement of cell or subcellular component
2	Ubiquinone and other terpenoid-quinone biosynthesis Photosynthesis Glycine, serine and threonine metabolism Phenylalanine, tyrosine and tryptophan biosynthesis Methane metabolism Carbon fixation in photosynthetic organisms Porphyryn and chlorophyll metabolism Terpenoid backbone biosynthesis Carotenoid biosynthesis Aminoacyl-tRNA biosynthesis	translation lysine metabolic process steroid biosynthetic process coenzyme metabolic process aromatic amino acid family metabolic process photosynthesis tetrapyrrole metabolic process dicarboxylic acid metabolic process cofactor biosynthetic process oxidation-reduction process alpha-amino acid biosynthetic process	Glycolysis / Gluconeogenesis Fatty acid biosynthesis Ubiquinone and other terpenoid-quinone biosynthesis Pyruvate metabolism Carbon fixation in photosynthetic organisms Porphyryn and chlorophyll metabolism Aminoacyl-tRNA biosynthesis	cellular amino acid metabolic process fatty acid metabolic process photosynthesis tetrapyrrole metabolic process pigment metabolic process carboxylic acid biosynthetic process cofactor biosynthetic process oxidation-reduction process organonitrogen compound biosynthetic process
3	Oxidative phosphorylation Arginine biosynthesis Purine metabolism Pyrimidine metabolism Valine, leucine and isoleucine biosynthesis Lysine biosynthesis Arginine and proline metabolism Histidine metabolism Glutathione metabolism Glyoxylate and dicarboxylate metabolism Aminoacyl-tRNA biosynthesis	RNA methylation tRNA threonylcarbamoyladenosine modification rRNA processing protein folding purine ribonucleotide biosynthetic process purine ribonucleoside triphosphate biosynthetic process monocarboxylic acid metabolic process cellular amide metabolic process ATP metabolic process alpha-amino acid biosynthetic process	Photosynthesis Arginine biosynthesis Purine metabolism Pyrimidine metabolism Alanine, aspartate and glutamate metabolism Lysine biosynthesis	cellular amino acid biosynthetic process monocarboxylic acid metabolic process amide biosynthetic process nucleobase-containing small molecule metabolic process oxidation-reduction process
4	Citrate cycle (TCA cycle) Fatty acid biosynthesis Fatty acid elongation Fatty acid degradation Arginine biosynthesis Alanine, aspartate and glutamate metabolism Glycine, serine and threonine metabolism Arginine and proline metabolism Tyrosine metabolism Pyruvate metabolism Propanoate metabolism Butanoate metabolism Carbon fixation pathways in prokaryotes Thiamine metabolism Nitrogen metabolism Drug metabolism - cytochrome P450	glycolytic process cellular biogenic amine metabolic process tRNA processing energy derivation by oxidation of organic compounds	Glycolysis / Gluconeogenesis Citrate cycle (TCA cycle) Alanine, aspartate and glutamate metabolism Cysteine and methionine metabolism Glyoxylate and dicarboxylate metabolism Carbon fixation in photosynthetic organisms	
5	Glycolysis / Gluconeogenesis Galactose metabolism Fatty acid biosynthesis Fatty acid degradation Purine metabolism Pyrimidine metabolism Valine, leucine and isoleucine degradation Phenylalanine metabolism Tryptophan metabolism beta-Alanine metabolism Pyruvate metabolism Propanoate metabolism Methane metabolism	protein phosphorylation movement of cell or subcellular component cell cycle process	Fatty acid biosynthesis Fatty acid elongation Fatty acid degradation Valine, leucine and isoleucine degradation Geraniol degradation Phenylalanine metabolism Tryptophan metabolism beta-Alanine metabolism alpha-Linolenic acid metabolism Pyruvate metabolism Glyoxylate and dicarboxylate metabolism Propanoate metabolism Butanoate metabolism Carbon fixation pathways in prokaryotes Biotin metabolism Biosynthesis of unsaturated fatty acids	generation of precursor metabolites and energy

The wild-type strain enrichment analyses (Table 7.1) revealed amino acids related processes in all five clusters. On the other hand, *slm1* pathway enrichment in amino acids related processes is more scarce, and GO enrichments clearly reveal that amino acids synthesis (found in

clusters 2 and 3) occurs now exclusively during the light period. In contrast, the wild-type shows some clear signs of amino acid synthesis during the dark period. This means that *slm1* lacks the needed energy to start synthesizing amino acids during the dark period, and amino acid synthesis is delayed until light is available. Besides this delay, the enrichment results indicate that for both strains, protein synthesis occurs throughout the whole day, whereas protein degradation occurs more during the night. Unlike in nitrogen replete condition [262], under nitrogen limitation both strains show signs of amino acids degradation during the dark period. This degradation is a logical consequence of the limited nitrogen supply. During the dark period there is no nitrogen supply and thus likely the nitrogen concentration is zero and nitrogen starvation occurs. Notably, the pathway “Valine, leucine and isoleucine degradation”, which produces acetyl-CoA, is found enriched in cluster 5 which is confirmed by the mapping to the metabolic maps (supplementary file S6). Enrichment of “protein phosphorylation” appears in clusters 5 and 1 (43 and 31 genes respectively) in the wild-type. In the mutant *slm1*, the “protein phosphorylation” appears only in cluster 1 (30 genes). Additionally, “movement of cell or subcellular component” is enriched in wild-type cluster 5 and in *slm1* cluster 1. This change reflects a possible delay of the cell cycle that might not be finished before light is on. This correlates with the longer duration of cell division that was measured in the *slm1* as compared to the wild-type. The GO term “generation of precursor metabolites and energy” was found enriched in *slm1* cluster 5. This term indicates the formation of precursor metabolites from which energy is derived. Those precursors can be formed via different pathways including photosynthesis, glycolysis, TCA, pentose phosphate pathway, glyoxylate cycle, and fermentation. Except for the last two, none of these pathways were found enriched (Table 7.1) because not many reactions of those are associated to cluster 5 (supplementary file S6). The glyoxylate cycle is complementary to the TCA for which it pre-processes degradation products originating from nucleic-acids, amino-acids and fatty-acids. Enrichment for degradation processes of the last two is also found in the *slm1* cluster 5. Within the fermentation processes, we can find the enriched butanoate metabolism and amino acid catabolic processes. Since more pathways related to amino acid metabolism are found enriched in the *slm1* cluster 5 than in wild-type, it is most probable that more intense amino acid catabolism occurs during the dark period.

The gene clusters of wild-type display a similar succession of transcriptional events as observed in nitrogen replete [262], which consist of protein synthesis, photosystem and pigments synthesis, carbon fixation, cellular replication, glycolysis, cell division, and fatty acid degradation. Compared to the nitrogen replete condition, the nitrogen limited condition

displays a more prominent role for lipid metabolism, notably prior to and during the dark period. Fatty acids related processes are found enriched in clusters 1, 4 and 5 for wild-type, and in clusters 2 and 5 for *slm1*. A vast majority of genes are found in the cluster 3 for both strains. Looking at the change in profile for *slm1* cluster 3, it is safe to assume that these genes (mostly amino-acids synthesis) have shifted their expression to start as soon as light is available and result in light-exclusive expression due to an even lower expression at dark. Furthermore, the large transfer of genes between wild-type cluster 2 and *slm1* cluster 3 indicates that these genes kept an expression pattern that was similar but without an increased expression at the end of the light period (Figure S4). Likewise, the genes in cluster 2 have shifted their expression to the earlier part of the light period. The enrichments confirm that the photosynthesis is maintained throughout the whole light period. The genes in cluster 3 participate in sinking photosynthetic energy, fixating carbon, and synthesizing the amino-acids. Detailed comparison after manual curation is discussed in a dedicated section below. Starch and sucrose metabolism is found enriched only in wild-type clusters. This indicates that in *slm1* genes in these pathways are affected in such a way that they do not show a shared diurnal pattern and do not appear associated to any cluster. This suggests some change in the regulation (dysregulation) of these genes. No dysregulation was observed between the two strains in nitrogen replete condition [262].

7.3.6 Selected processes and pathways

Genes associated to carbon fixation, synthesis of the photosystems, its pigments (chlorophylls and carotenoids) and their precursor terpenoid backbone, are all found enriched in the cluster 2 for both strains. Detailed inspection of these pathways (supplementary File S6) shows all the genes annotated to these consecutive reactions are associated to cluster 2. Also for the nitrogen replete conditions the expression pattern of these genes is the same for both strains and identical to the limiting conditions. This indicates a strong and synchronized regulation of these processes with little to no difference between the strains. We can therefore conclude that these processes are not affected by either the starch deficiency or by the nitrogen limitation. The genes associated to starch synthesis, fatty-acids biosynthesis, carbon fixation, pyruvate metabolism are also tightly regulated within the same time frame. From the detailed pathways overview (supplementary File S6), we can see that the carbon fixation remains an early process in *slm1*, but the other subsequent pathways and reactions are delayed in comparison to the

wild-type strain. These delays reflect the higher difficulty to synthesize the complex machinery required to store energy, due to the nitrogen limitation and the very limited energy available before light is turned on. Since expression of genes associated to carbon fixation and photosynthesis machinery are not altered unlike what happens to other key systems, we deduce that their role is the most essential, if not critical, and their maintenance is given the highest priority. Possibly, these processes are directly controlled by the light availability. Even if the expression is unchanged, it is possible that the processes are being stretched (or simply delayed) over time by the lack of available nitrogen and starch, effectively delaying the subsequent processes as we observed.

7.3.6.1 Starch metabolism

As previously observed in two studies [35, 262], the lack of starch synthesis in *slm1* is due to a nonsense mutation in the ADP-pyrophosphorylase small subunit. The associated gene was found to be expressed at much lower level in *slm1* than the original gene in wild-type but with a conserved temporal regulation. The reads mapping over the mutant gene revealed the nonsense mutation and more importantly, a homogeneous distribution over the whole gene length. The reduced expression (see Figure 7.6) can be explained by a lower transcription or post-transcriptional regulation that would degrade the transcript. Since the diurnal profile of the mutated gene is maintained, it is more likely that the transcript is being degraded instead. Furthermore, the nonsense mutation suggests strongly that this post-transcriptional degradation is done by nonsense-mediated mRNA decay (NMD) [274, 275]. Since this quality-control mechanism is common in eukaryotes, it is very likely that *Tetradesmus obliquus*, and possibly other green algae are subjected to NMD. Nevertheless, it should be noted that *slm1* has been obtained with UV therefore we can not rule out other mutations that could affect regulation.

As previously observed in nitrogen replete condition, the similarity of expression between the two strains means that there is little response from the lack of starch. The strongest observed difference is for the gene g2865.t1, annotated to the starch synthase (EC 2.4.1.21, Figure 7.6). Its expression is effectively doubled in the *slm1* as compared to in the wild-type. Another noticeable difference relates to gene g234.t1, associated to the starch branching enzyme (EC 2.4.1.18). This gene is overexpressed in *slm1*, but with a very similar time profile to that in wild-type (Figure 7.6). Though the starch synthesis is operated by few enzymes, we did not observe a strong response in the mutant that could be attributed to a transcription regulatory mechanism to counter the incapacity to synthesize ADP-Glucose and starch. Since the

transcript g2865.t1 coding for the starch synthase (EC 2.4.1.21) is the only one that displays a change in diurnal expression profile, it is likely to be a regulatory point for synthesizing starch in the microalga *T. obliquus*. Furthermore, the same transcript had also different expression between strains in nitrogen replete condition. Unlike here, the change in nitrogen replete was an early shift in expression. However, we can not rule out other post-translational regulatory mechanisms.

Similarly to starch synthesis, the expression of genes associated to starch degradation does not seem to be affected by the incapacity to synthesize starch. We identified genes associated to five enzymatic reactions. The profiles of these genes are shown in Figure 7.7 except for the glucoamylase (EC 3.2.1.3) which did not have a detected time profile. Isoamylase (EC 3.2.1.68) was found in the same cluster for both strains and showed identical profiles for both associated genes (Figure 7.7). The alpha-amylase (EC 3.2.1.1) has several candidate genes, but significant changes were observed only for g17712.t1. This gene was found in wild-type cluster 4 and in *slm1* cluster 3, with an apparent higher expression during the light period and lower expression during the dark period. As for the beta-amylase (EC 3.2.1.2), the associated gene (g13560.t1) displays a significant difference between the two strains, with a generally higher expression and a strong peak right after the switch to dark at 18h (Figure 7.7) for *slm1*. Finally, the glycogen phosphorylase (EC 2.4.1.1) was associated to two genes and both were affected in a similar way, since they were found in wild-type cluster 4 and in *slm1* cluster 3. While the beta-amylase and the isoamylase are more expressed during the dark period, the alpha-amylase and the glycogen phosphorylase are more expressed at the end of the light period. The first two are degrading beta glycosidic linkage, while the later two are degrading alpha glycosidic linkage. The later two are also displaying a stronger decrease right after dark in *slm1*. These results suggest that the genes associated to alpha and beta linkage lysis are not only expressed differently over time, but are also regulated differently by starch deficiency.

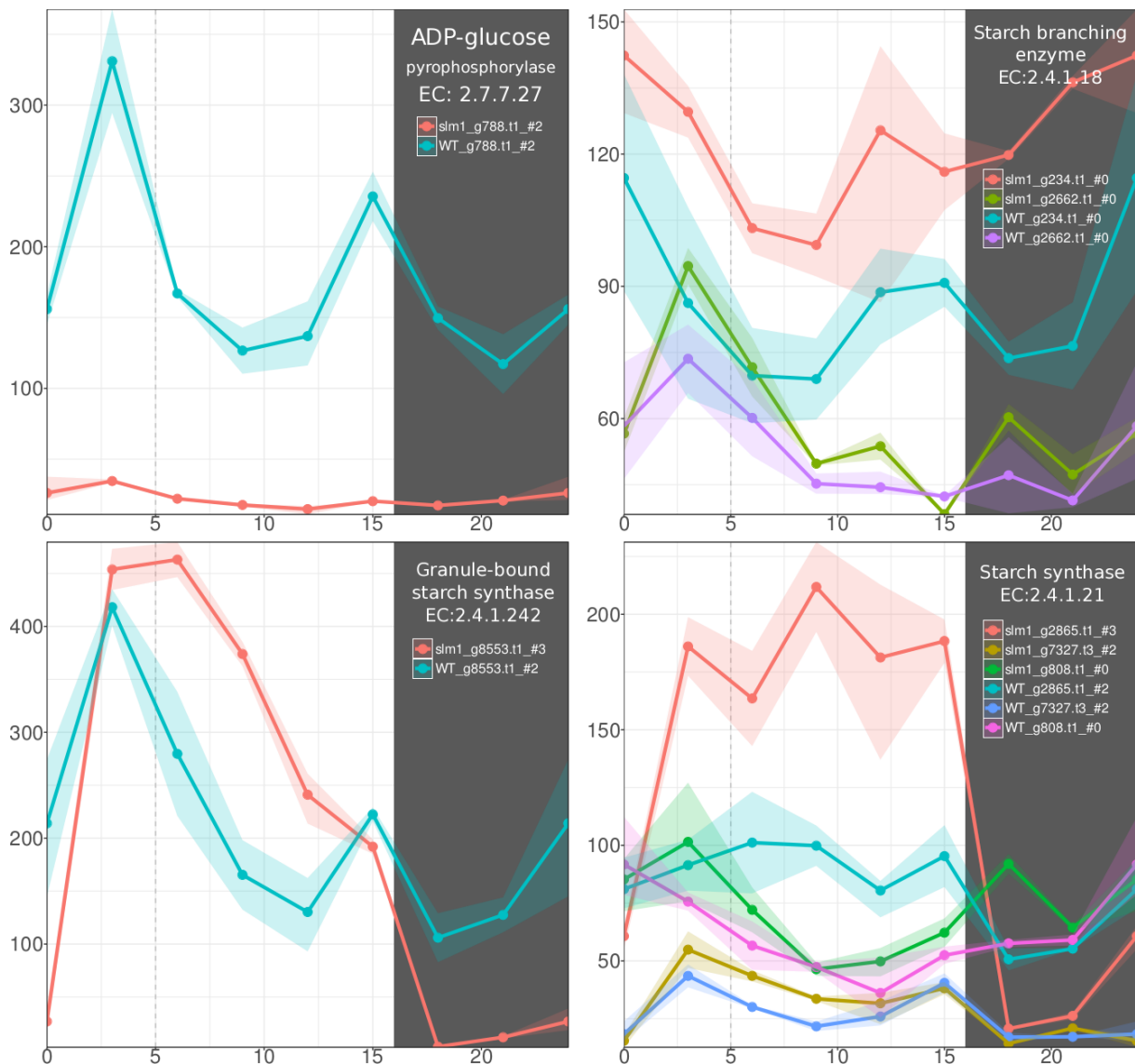


Figure 7.6: Diurnal expression of genes associated to starch synthesis. Points represent mean values for each time point, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together with the same scale. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combinations. For each gene-strain combination, the number after the # symbol, indicates the associated cluster and #0 indicates that the gene showed not significant changes in expression over time. The dark area corresponds to the dark period.

Overall, the expression of the genes associated to the starch synthesis and degradation agrees with the biochemical measurements. For the wild type, the genes associated to starch synthesis show diurnal regulation with higher expression during the first half of the light period, which agrees with the accumulation of starch later on. Likewise, the genes involved in degradation are upregulated either just before or during the dark period, which agrees with the observed starch degradation. With respect to *slm1*, the genes associated to the starch synthesis are

strongly upregulated during the day, which may be a physiological response to the inability to shuttle an excess of energy and carbon to starch. Likewise, a few of the degradation enzymes are up regulated during late light period or during the dark period, which may be again, a response to the lack of transient energy due to the absence of starch.

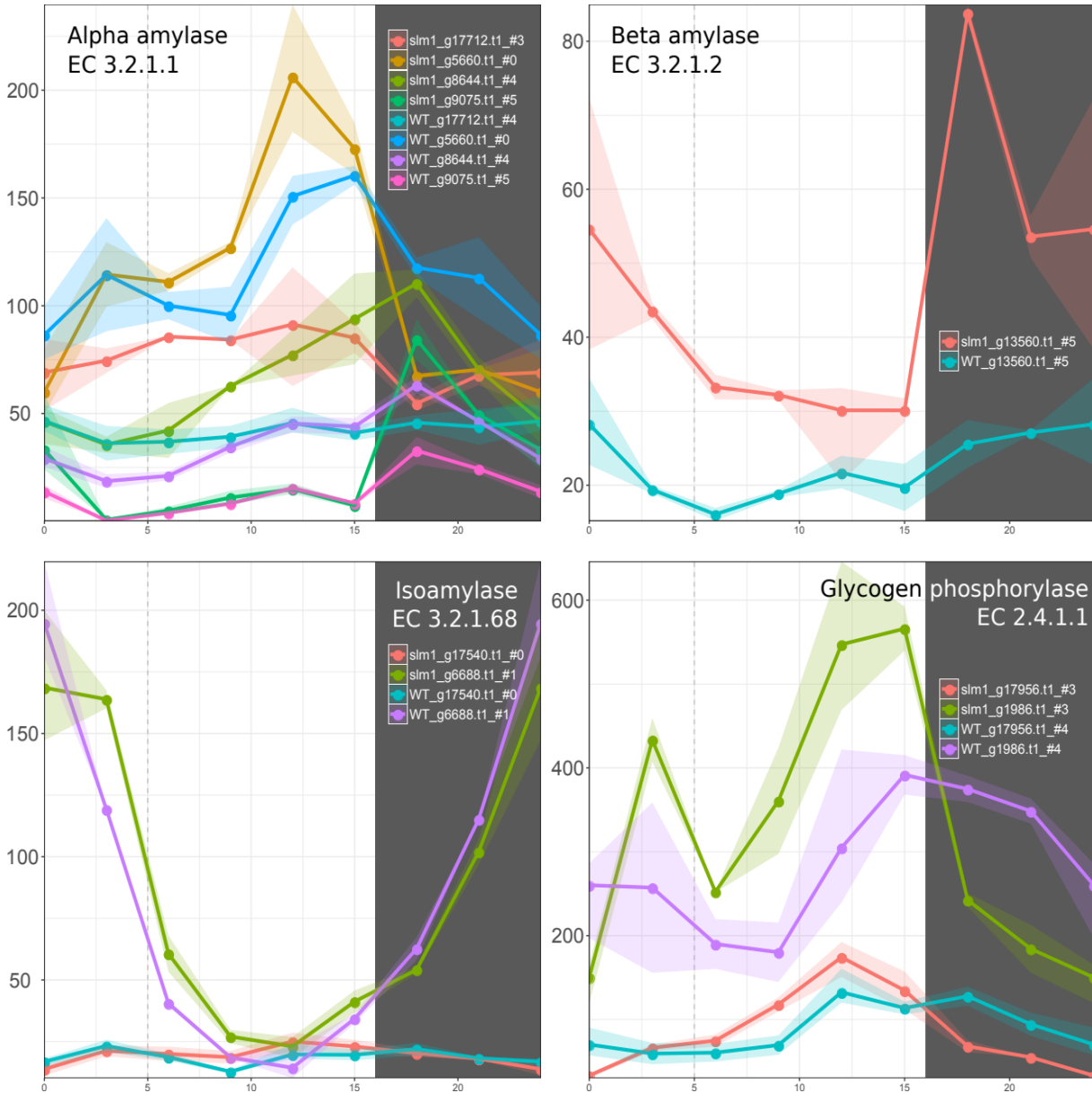


Figure 7.7: Diurnal expression of genes associated to starch degradation. Points represent mean values for each time point, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together with the same scale Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combinations. For each gene-strain combination, the number after the # symbol, indicates the associated cluster and #0 indicates that the gene showed not significant changes in expression over time. The dark area corresponds to the dark period.

7.3.6.2 *Metabolism of lipids and TAGs*

The wild-type expression of genes associated to lipid synthesis fits a general trend: highest expression is reached right at the beginning of the light period, lowest expression is reached half way of the light period, increase of expression right before the dark period, and slight decrease during the dark period. This trend is very similar to the TAG productivity measured in Figure 7.1D, and the first peak corresponds to the increased TAG productivity at 6h (Figure 7.1), but with a short delay. Remarkably, the same genes in *slm1* display larger variations in expression with the lowest expression during the dark and highest expression during the day. The increased expression at the peaks is roughly 1.5 to 2 times higher in *slm1* than in the wild-type. This difference suggests that the lack of the ability to synthesize starch is actively compensated by increased accumulation of lipids and TAG during the light period. Furthermore, the presence of starch correlates with a maintained expression of the lipid synthesis enzymes during the night, whereas in the absence of starch these enzymes are no longer expressed in the night. This suggests that starch is possibly used to synthesize lipids and TAG when there is no photosynthesis.

TAG and lipid synthesis

In the pyruvate metabolism, pyruvate dehydrogenase subunits synthesizing acetyl-CoA (EC 1.2.4.1, 2.3.1.12, 1.8.1.4) appear in wild-type cluster 1 and *slm1* cluster 2 (File S6), which could be related to the fact that *slm1* cannot synthesize acetyl-CoA from starch during the dark because starch is not present. Acetyl-CoA is involved in many processes, but most importantly, it is the building block to start the lipid synthesis by the Acetyl-CoA carboxylase (ACC, EC 6.4.1.2). As seen in Figure 7.8, the best candidate genes for the ACC are all found in wild-type cluster 1 and *slm1* cluster 2 and 3. All these time profiles are very similar to each other, which is synchronized with the described synthesis of acetyl-CoA. The following reaction in the lipids synthesis is the Beta-ketoacyl-ACP synthase I (KASI, EC 2.3.1.41). Two genes were associated to this reaction, g16858.t1 and g14707.t1, where g16858.t1 is identified with a trans-membrane domain and g14707.t1 is not. The two genes are expressed with near identical profiles fitting into wild-type cluster 1 and *slm1* cluster 2. Overall, the same pattern is observed for genes associated to reactions in the lipids biosynthesis (supplementary Figure S7). The Acyl-CoA synthetase (EC 6.2.1.3) is associated to two genes, displaying very different patterns. This is because it is a reversible reaction that is occurring also in the beta oxidation pathway. Our results and the manual curation of the annotation confirm that g1817.t1 performs the forward

reaction, towards synthesis, while g4377.t1 performs the reverse reaction. The forward reaction is in phase with the other synthesizing genes.

The synthesis of TAG can be performed by two different enzymatic reactions, known as diacylglyceride acyltransferase (DGAT, EC 2.3.1.20) and phospholipid diacylglycerol acyltransferase (PDAT, EC 2.3.1.158). While several DGAT genes have been identified in green algae, our functional annotation revealed two genes with high confidence [276]. The expression of these two genes is noticeably different, with g2587 being the only one with a temporal regulation. In wild-type, the gene g2587 is highly expressed during the dark period. In *slm1*, its profile is very similar to the wild-type but with overall higher expression and apparently more exclusive to the dark period. In *slm1*, PDAT is displaying a somewhat opposite expression to DGAT, meaning that fatty acids from phospholipids are transferred onto diacylglycerides (DAG) to form TAG during the light period and free fatty acids are added to DAG during the dark period. This seems to agree with the findings in *Chlamydomonas reinhardtii*, where PDAT mediated membrane lipid turnover is helping to recycle phospholipids and synthesize TAG [277]. Interestingly, the monogalactosyldiacylglycerol (MGDG) synthase (EC 2.4.1.46) is highly expressed during the light period with a similar pattern in both strains. This expression is not synchronized with any of the other lipids synthesis reactions, but because MGDG is the main lipid form of the chloroplast membrane, the expression of the MGDG synthase is logically correlated with cellular growth and chlorophyll synthesis.

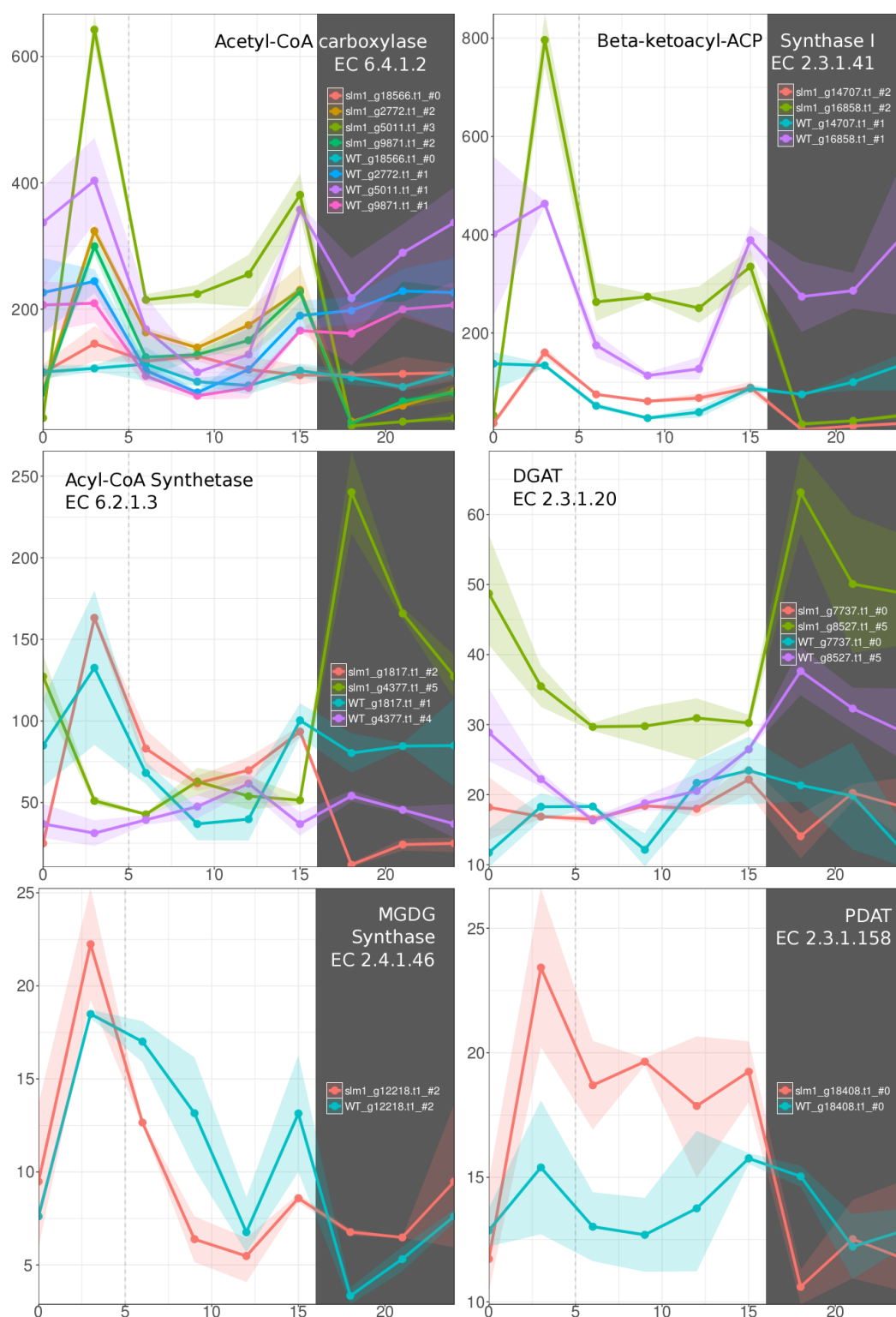


Figure 7.8: Diurnal expression of genes associated to lipid and TAG synthesis. Points represent mean values for each time point, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together with the same scale. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combinations. For each gene-strain combination, the number after the # symbol, indicates the associated cluster and #0 indicates that the gene showed not significant changes in expression over time. The dark area corresponds to the dark period.

TAG and lipid degradation

In the wild-type strain, the diurnal expression of genes associated to TAG and lipid degradation shows a temporal regulation very similar to that of genes associated to synthesis, but with a lower expression. These genes are expressed higher during the day period and lowest at night. This transcriptional regulation is coherent with the measured accumulation of TAG and lipids during the day and night, and with a very limited turnover. However, the *slm1* strain is regulated in a complete opposite way, strongly over-expressing the genes associated to TAG and lipid degradation during the dark period (Figure 7.9). Though the profile of *slm1* under N-limitation is very similar to that in the previously analyzed replete condition, the amplitude of the up-regulation at dark is much more important, with fold changes ranging from 1.5 to 2.5 folds. Interestingly, also the acyl-CoA oxidase (EC 1.3.3.6) is now displaying a strong increase in expression in *slm1* compared to the wild-type.

The first and second step to degrade TAG is performed by the same enzyme, the triacylglycerol lipase (EC 3.1.1.3). Its highest expression is measured at 18h, the first time point after dark, and its expression reduces progressively until 6h (Figure 7.9). Next, the last fatty acid chain is removed from the glycerol by the monoacylglycerol lipase (EC 3.1.1.23). Out of the four genes annotated to this reaction, only two displayed some expression variation, although these appear lowly expressed and do not display any temporal regulation (Figure 7.9). The following step is the Acyl-CoA synthetase, the reversible reaction described in the synthesis section (Figure 7.8), for which we have identified separate candidate proteins for each direction. The degrading candidate (g4377.t1) is highly expressed during the dark period, which is in phase with the other degrading candidates. The four following reactions are repeated reactions and are performed by three enzymes: Acyl-CoA oxydase (EC 1.3.3.6), Multifunctional MPF-a (EC 1.1.1.35, 4.2.1.17), Acetyl-CoA acyltransferase (EC 2.3.1.16), Acetyl-CoA C-acetyltransferase (EC 2.3.1.9). All the reactions are highly expressed in *slm1* during the dark period. Only the Acetyl-CoA C-acetyltransferase is expressed with a similar time profile in the wild-type strain but at a lower level. Additionally, the enzyme representing the beta oxidation of unsaturated fatty-acids, enoyl-CoA isomerase (EC 5.3.3.8), is found to be expressed in the same cluster as the other degrading enzymes (*slm1* cluster 5, supplementary file S6).

The acetyl-CoA molecules produced from the described reactions can then be utilized by the organism. For example, the acetyl-CoA can be broken down in the TCA cycle to generate energy or can be converted into succinate and malate through the glyoxylate cycle. Next

succinate and malate can be converted into oxaloacetate through the TCA cycle in the mitochondria, and finally converted into hexoses or sucrose through the glyconeogenesis in the cytosol [278]. Overall, all these points strongly suggest that a global regulatory mechanism, as observed in *A. thaliana* [279], allows the organism to use TAG as a form of transient energy storage, which accumulates during the light period and is degraded during the dark period.

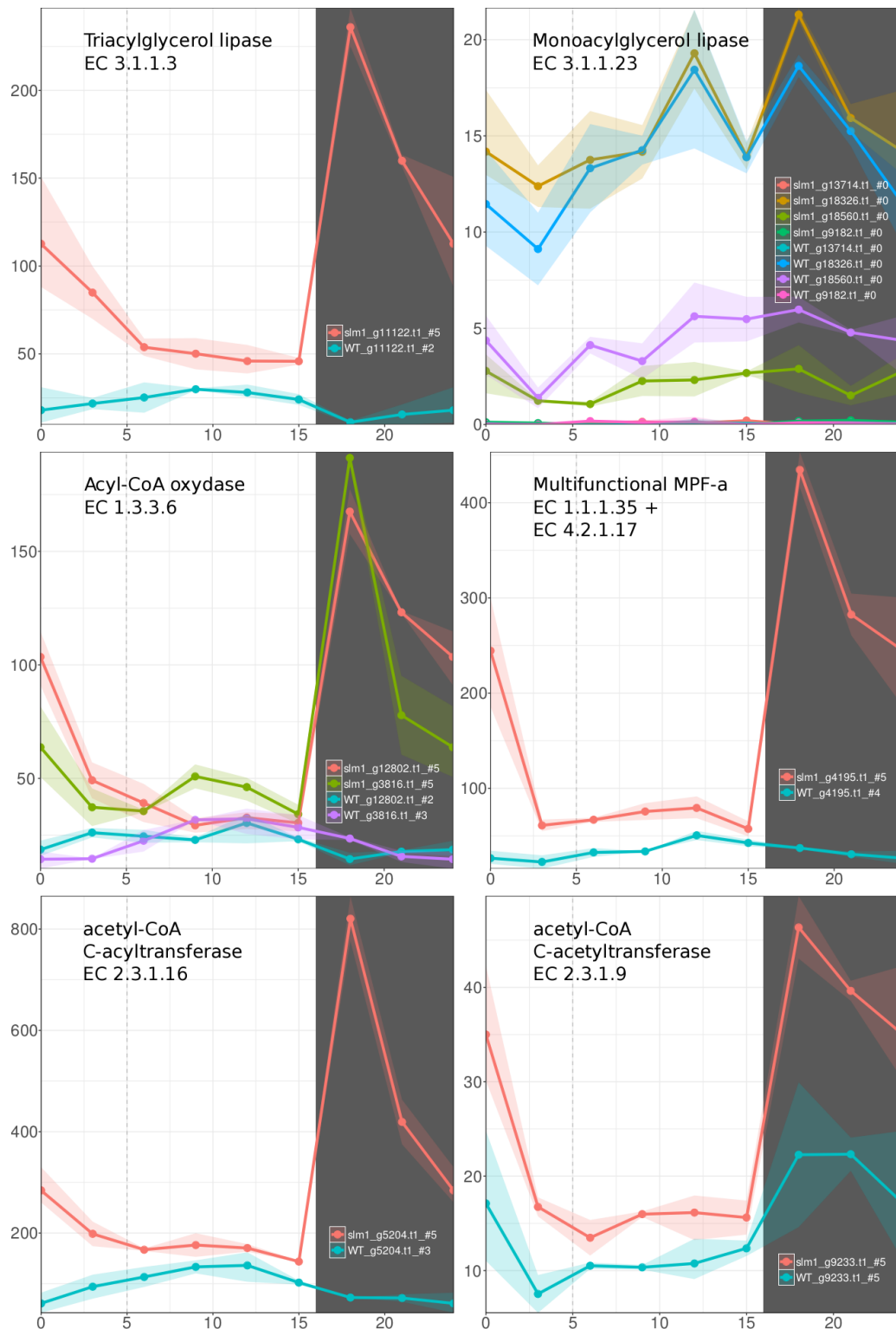


Figure 7.9: Diurnal expression of genes associated to lipid and TAG degradation. Points represent mean values for each time point, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together on the same scale. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combination. For each gene-strain combination, the number after the # symbol, indicates the associated cluster and #0 indicates that the gene showed not significant changes in expression over time. The dark area corresponds to the dark period.

7.3.6.3 *Nitrogen metabolism*

The expression profiles of the genes associated to nitrogen metabolism are displayed in Figure 7.10. Very little difference is noticeable between the strains at the first sight as most of the profiles are overlapping and only one gene stands out. Five genes from three of the six annotated reactions are in cluster 4, resulting in an enrichment of this pathway. These strong similarities suggest that nitrogen metabolism in both strains is regulated in a very similar way. In other words, nitrogen metabolism is mostly affected by nitrogen limitation and not from starch availability.

In the previous paper on N-replete conditions we could not find a protein for nitrate reductase. In this work we were able to identify a candidate protein for nitrate reductase [262]. The EnzDP score for this protein came second. However, its expression is sufficient, unlike the top scored protein that was not found expressed in both the nitrogen limited condition as well as the nitrogen replete condition. Both strains display the exact same pattern and level of expression for this protein, fitting the associated cluster 4 pattern but with only the peak of expression before the dark period (Figure 7.10). This diurnal expression pattern indicates that for the wild type and *slm1* under nitrogen limitation, nitrogen is assimilated gradually throughout the whole light period and gradually reduced right before the dark period until the very early light period. In the nitrogen replete condition, the expression of the transcript for *sml1* is identical to the one observed here in the nitrogen limited condition, and is part of the *slm1* cluster A as defined in the previous N-replete paper [262]. Besides, the wild-type the expression is very different and fits the pattern of cluster 5 as defined in the N replete paper [262], with a progressive expression throughout the light period similar to the *slm1* pattern in this study, but with a continued high expression during the late day and early night, where the expression of the *slm1* dropped. Its expression reached its highest value early in the dark period after which it decreased. This seems to indicate that nitrogen is probably metabolized during the dark period. Since this reaction requires the presence of nitrogen and energy in form of NADH, it can only occur in nitrogen replete conditions for strains with available transient energy storage like starch.

The genes associated to the nitrite reductase and the nitrite reductase NO-forming are too lowly expressed to detect any time pattern. Hence, it is unlikely that these are used by *Tetrademus obliquus* to perform the reduction of nitrite. Besides, two genes were associated to two forms of hydroxylamine reductase (EC 1.7.99.1, Figure 7.10). These two are expressed differently throughout the diurnal cycle, their expression is not affected by the mutation of

slm1, and the expression of the hybrid cluster protein (HCP) is much higher (~10 folds). More interestingly, the diurnal expression of the transcript associated to the HCP is nearly identical to the one observed from the nitrate reductase. In addition, the same expression is also observed between the two transcripts in nitrogen replete condition. Since the gene identifiers are given by their order of appearance in the genome, it is clear that the nitrate reductase and the HCP are neighbor genes, further indicating that these genes have a special bound. Additionally, the hydroxylamine reductase reaction is fairly similar to the nitrite reductase and was even found to perform the NO forming nitrite reduction [280]. We are pushed to believe that the proteins translated from these transcripts (g618.t1 and g619.t1), though different than known enzymes, are performing the successive reactions of nitrate and nitrite reduction. It would then be valuable to confirm the functionality of these two proteins.

The glutamine synthetase (GS, EC 6.3.1.2, g10416.t1) is highly expressed throughout the whole light period, and lowly expressed during the dark period, effectively fitting clusters 2 and 3 of wild-type and *slm1* respectively. In both strains the expression is very similar. Compared to our previous observations in nitrogen replete conditions, its maximum expression in nitrogen limiting conditions is about 3 times lower and its expression profile is very different. In the nitrogen limited condition, its expression is high throughout the whole light period, while in nitrogen replete, its expression was high during the early light period and the dark period. Among the glutamate synthase (GltS, EC 1.4.1.14) in nitrogen replete condition, only the ferredoxin-dependent glutamate synthase (fd-GltS) was displaying significant differences between the strains and it was down-regulated in *slm1*. However, in this nitrogen limited condition, fd-GltS only displays a marginal difference at 3h between the strains with a slight overexpression. The levels of expression are comparable to *slm1* in nitrogen replete condition. Most importantly, while the two NADH dependent GltS are displaying a similar profile between strains, their expression is roughly 2 folds higher in *slm1* than in the wild-type. For glutamate synthase (GltS, EC 1.4.1.14) both genes conserved a similar expression pattern, but with an expression twice higher in *slm1* than in wild-type. Overall, it appears that when *T. obliquus* is exposed to nitrogen limitation, it relies more on the NADH dependent GltS than the fd-GltS, and with an even stronger expression for *slm1*. It would seem that with increasing stress, GltS is more prominent over fd-GltS, possibly due to stopped ferredoxin oxidation by the photosystem I.

Green algae are known to synthesize glutamate with GS, GltS, and fd-GltS, while ammonia is recycled by the glutamate dehydrogenase (GDH, EC 1.4.1.3) from metabolites such as amino acids [254]. Deamination by GDH suggests that amino acids are used as source of energy and

possible ammonium for protein recycling during the heterotrophic condition present in the dark period. Proteins are recycled for different purposes: to regulate enzyme activities, to remove abnormal proteins, to generate energy and to adapt to the changing environment [281]. Protein turnover is believed to be an adaptive process and also to be one of the most energy demanding processes during dark respiration [282–284]. Similarly to our previous findings in nitrogen replete conditions, among the genes associated to GDH, only *g547.t1* displays a difference between the strains with for *slm1* a lower peak of expression just before dark and a higher peak of expression just after the start of the dark period. Even if the difference between the strains is not large and the peak before dark has overlapping values, the values in the second peak are significantly different. In fact, the replicate values are much closer to each other, and the minimum difference between the strains is ~1.5 folds higher for *slm1*. More subtly, after being subjected to nitrogen limitation, only *slm1* kept the high peak of expression after dark. The stronger GDH activity, the enrichments for “generation of precursor metabolites and energy”, enrichment for diverse amino acids pathways (including degrading ones), transaminases candidate proteins, and other enzymes involved in the amino acids degradation, indicate that amino acids are actively degraded at the end of the light period and during the dark period, with higher *slm1* activity in the dark period.

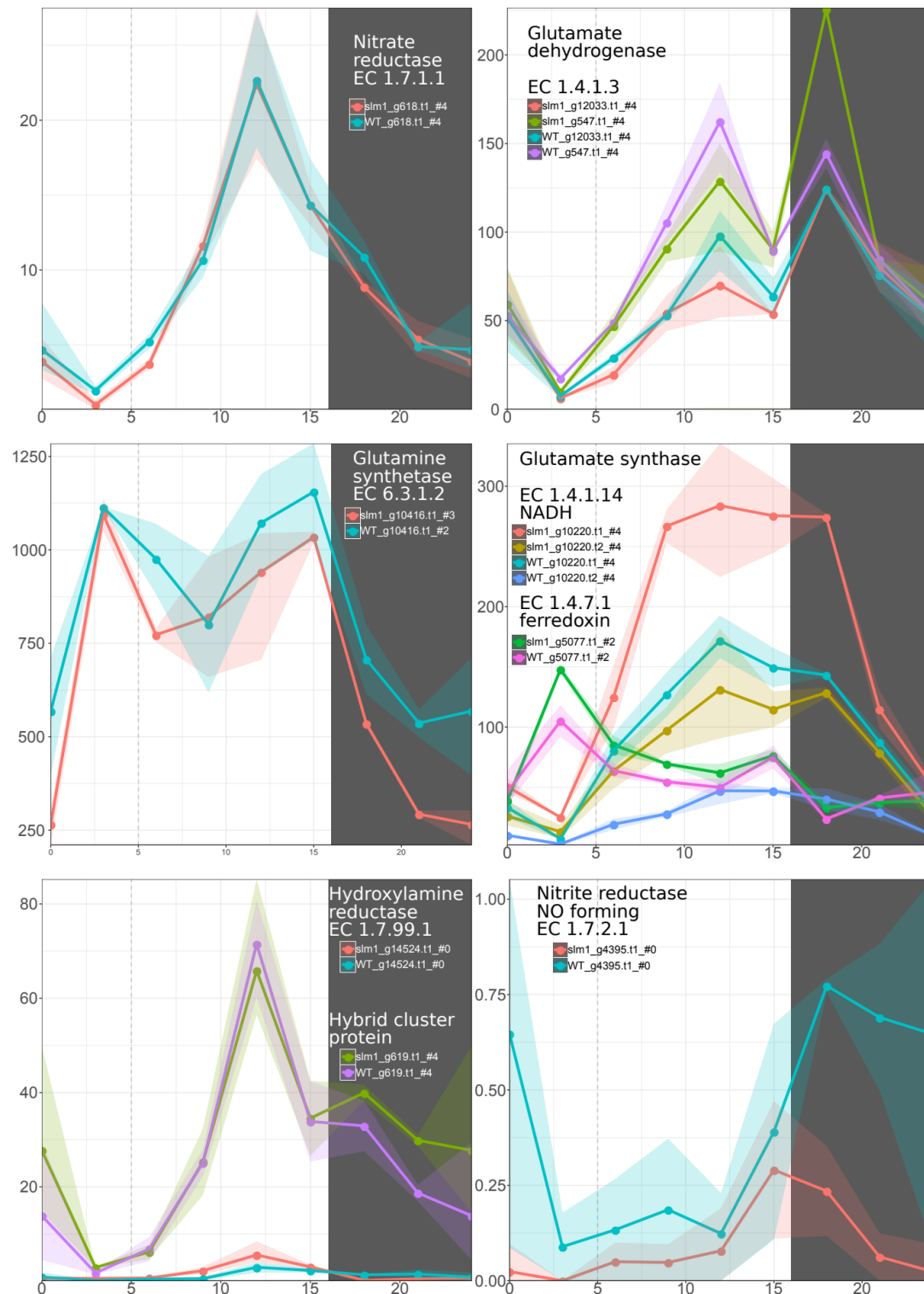


Figure 7.10: Diurnal expression of genes associated to nitrogen metabolism. Points represent mean values for each time point, the ribbon covers the minimum and maximum values for each time points. The genes associated to the same EC number are plotted together on the same scale. Each plot is labeled with the reaction name, the corresponding EC number, and the color legend for each gene-strain combination. For each gene-strain combination, the number after the # symbol, indicates the associated cluster and #0 indicates that the gene showed not significant changes in expression over time. The dark area corresponds to the dark period.

7.3.6.4 *Glyoxylate cycle*

The “Glyoxylate and dicarboxylate metabolism” that is enriched in wild-type cluster 1 and 3, and is enriched in *slm1* cluster 4 and 5. The glyoxylate cycle serves as an intermediary step that accepts the products from degradation pathways of non glycolytic metabolites such as amino- fatty- and nucleic-acids. The resulting pyruvate can then be transported to the mitochondrion to generate energy through the TCA cycle. This is a sign that the three type of metabolites cited above can be degraded to generate energy during the dark period in *slm1*. In the previous sections, we have shown that during the dark period, *slm1* display clear signs of energy generation and degradation of amino-acids, fatty-acids and TAG. However, *slm1* do not seem to degrade them during the dark period (supplementary Table S8).

The glyoxylate cycle converts two acetyl-coA molecules to one malate molecule and contains two specific enzymes that are not found in the TCA cycle. The transcripts associated to these key enzymes, cycle malate synthase (MS, g18100.t1) and isocitrate lyase (ICL, g13192.t1), are both found in wild-type cluster 3, while for *slm1* they are in cluster 5. Their expression profile in *slm1* is identical to the lipid degrading enzymes (I.g. ACC, Figure 7.9), and most importantly, orders of magnitude higher than the wild-type. We have also analyzed the other reactions involved in the glyoxylate cycle, and we have identified candidate proteins that agree with *slm1* dark period activity (Supplementary File S9). These results suggest that the glyoxylate cycle is only active in *slm1* and exclusively during the dark period. This confirms that in *slm1* fatty acids are broken down to acetyl-CoA to directly generate energy from. At the same time the glyoxylate pathway is used to convert acetyl-CoA to intermediates of the TCA cycle to keep this cycle running, or use them to make various biomass components. Finally also some amino acids are broken down to glyoxylate or acetyl-CoA and can be used in this way.

7.4 Conclusions

Under continuous nitrogen limitation, both *T. obliquus* wild-type and *slm1* showed a repeated diurnal dilution pattern, with the *slm1* showing a lower 24 h dilution rate (growth rate) compared to the wild-type. The transcriptome analysis revealed a similar diurnal pattern, with a reduced amplitude of the changes in expression for *slm1*. Under nitrogen limitation, *T. obliquus* wild-type continued to prefer starch to store energy and carbon. Starch was

accumulated to an average content of approximately $0.25 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$. Additionally, starch showed small diurnal oscillations, confirming its role as a transient energy storage compound, but to a lower extent than under nitrogen replete conditions. Analysis of expression changes during the cycle of genes associated to starch synthesis and degradation correlate with the measured productivity. The possibility to synthesize starch provides an advantage for the wild-type as compared to *slm1* as it has a higher energy conversion efficiency into the biomass components. With respect to the diurnal cycle this higher efficiency of the wild-type is reached especially during the last hours of light, where starch is accumulated. The energy conversion efficiency of photons into biomass components for the *slm1* was only half of the one obtained for the wild-type, which resulted in a decrease in biomass productivity in *slm1* (from $1.29 \text{ g}_{\text{DW}} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the wild-type to $0.52 \text{ g}_{\text{DW}} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the *slm1*). However, the TAG content in the *slm1* was higher than for the wild-type (average steady state values of approximately $0.27 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ for *slm1* compared to $0.07 \text{ g} \cdot \text{g}_{\text{DW}}^{-1}$ for the wild-type), which resulted in a slightly higher TAG volumetric productivity for the *slm1* as compared to the wild-type ($0.09 \text{ g} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the wild-type and $0.14 \text{ g} \cdot \text{L}^{-1} \cdot \text{day}^{-1}$ for the *slm1*).

The transcriptome analysis revealed breakdown of lipids and proteins in *slm1* during the dark period, which are probably used for the generation of energy in the absence of starch. In agreement with this, the enzymes in the glyoxylate cycle are upregulated in the *slm1* mutant. This cycle is used to convert the acetyl-CoA from amino acid and fatty acid breakdown to intermediates of the citric acid cycle to either keep this cycle running or to be converted to other biomass components. Our measurements do not show significant diurnal variation in DW proteins and TAG content, which indicates that substantially less energy is generated in the mutant during the dark period.

Chapter 8

Discussion

8.1 Rationale for the work carried out and background on algae and transcriptomics

Microalgae are unicellular photoautotrophic organisms that played and are still playing a very important role on the planet. Today's global energy needs still largely rely on fossil fuel extraction (~66%), while almost all renewable energy sources are based on electricity [7]. Efforts are being made to transition to sustainable and carbon-neutral fuel production. The first generation of biofuels were produced using plants, but the competition with food production became quickly an issue. In this regard, microalgae that accumulate large amounts of lipids precursors of biofuel molecules are a good alternative for sustainable biofuel production since they have many advantages over plants [16, 26, 106]. Still, production costs are currently too high and the technical limitations too strong for a microalgae based process to become economically feasible at large industrial scale. To become a viable alternative at industrial scale, the total solar-to-product conversion efficiency needs to be improved. This can be done by improving the design of the used bioreactors and bioprocesses, and by improving the microalgal strains through for example metabolic engineering. To be able to do rational strain improvement through metabolic engineering, detailed knowledge of algal metabolism in relation to production conditions is required.

The aim of this thesis was to obtain a better understanding of algal physiology by analyzing the transcriptional landscape of lipid-producing microalgae under conditions relevant for large industrial scale production. Thus, we studied the effect of light-dark cycles, salt water, intense light, and nitrogen limited or depleted conditions.

The internal phenotype can be studied by analyzing the different omics layers that compose it. As displayed in the Figure 8.1, the layers composing the internal phenotype range from the genome and epigenome through the transcriptome and proteome till the metabolome. The metabolome, in turn, participates in making the external phenotype. The transcriptome is an intermediary layer and it is possible to extrapolate information from the other layers by analyzing its composition. While the transcriptome is the direct result from the epigenetic and transcriptional regulation, interpreting the possible outcome from the subsequent layers is more difficult. This is where the absolute expression of genes alone is not sufficient, since it mainly informs if each gene is expressed or not. While comparing the expression between genes can be informative, their regulation can be very different. Furthermore, comparing

transcriptomes under different conditions allows to calculate the relative composition of the transcriptome and informs on the importance of genes under each condition. Whether a relatively high level of mRNA will result in higher protein levels depends on the synthesis and degradation rate of this protein. Although it not impossible, it seems unlikely that for the majority of proteins synthesis and degradation are up and down regulated in the same amount. Often, mRNA abundances can explain ~40% of the variations in protein concentration, while the other ~60% is generally explained by post-transcriptional regulation and measurement noise [285]. Thus, a relative increase in mRNA likely results in a higher protein level. However, the relation between proteome and metabolome is difficult to predict. The flux depends on the activity of the different proteins, which is partly determined by the metabolome itself.

Despite these limitations, even if transcriptomics alone cannot give a perfect snapshot of the internal phenotype, it is an efficient high throughput method to globally estimate how the organism adapts to its environment. While observations and hypotheses can be confirmed by the transcriptome landscape, unknown or unexpected ones should be validated by means of other omics or dedicated and specific biochemical measurements.

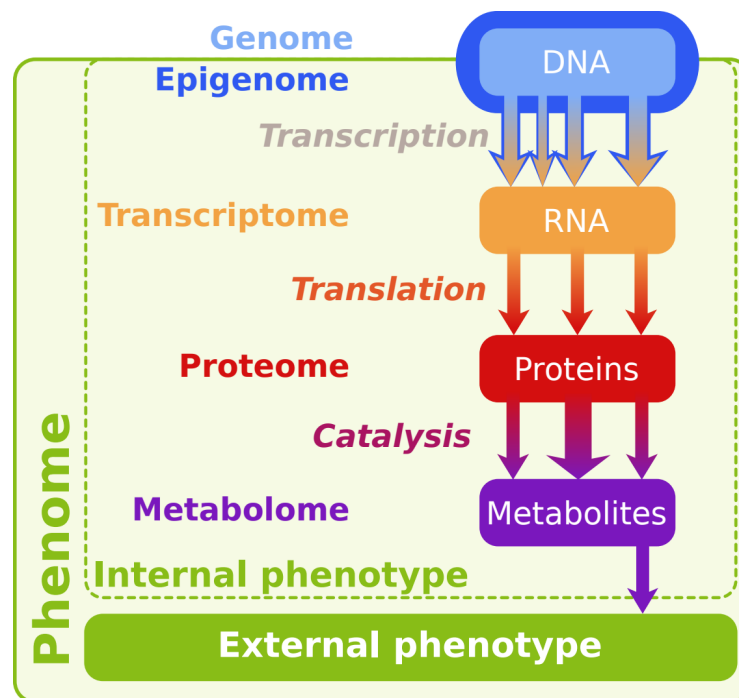


Figure 8.1: Omics layers from Epigenomics to Metabolomics are building the internal phenotype of the organism. The internal phenotype bridges the gap between the genotype and the external phenotype.

8.2 Annotation

8.2.1 *Ettlia oleoabundans*: de novo assembly

The initial annotation of *Ettlia oleoabundans* (*Neochloris oleoabundans*) was performed using common, albeit largely uncurated and automatic methods, which rely on plant genome annotations [38, 47]. As a consequence, the proportion of wrong annotations is certainly high, but impossible to determine to which extent. Another reason for the poor annotation was because the *de novo* assembly from RNA-seq data, which is in principle possible, was difficult in itself (chapters Error: Reference source not found and Error: Reference source not found). All RNA-seq specific tools resulted in more than four times the expected ~20000 transcripts, with many small contigs. This appeared to be due to the high quantity of intronic sequences present in the sequencing reads. These reads originate from rare cases of sequenced pre-mRNA. Even full intron sequences could be recovered due to the high sequencing depth, which was made possible by combining 16 samples from the different conditions present in Chapters Error: Reference source not found and Error: Reference source not found. These lead to more sequences being partially assembled and more sequences detected as either different genes or splice variants. Even though it was designed for genome assemblies, IDBA-UD performed better with our samples because it is designed for highly uneven sequencing depth [157]. While this method would allow very low sequencing depth regions such as intronic regions, it could allow to assemble the full transcript instead of generating multiple contigs. This way, a minority of the contigs contained intronic sequences which possibly led to frame-shifting, rendering their annotation either wrong or impossible. Due to the lack of algae data in the databases and because their genetic background is much more different than those of other eukaryotes, standard alignment methods (e.g. those relying on blastP) were not effective and reliable enough, as described in the Error: Reference source not found. The following annotation method developed for this purpose was laborious as it combined different methods and different sources of sequences for alignment. The method was based and performed in Blast2GO [92]. The sequence similarity was performed using DELTA-BLAST, which improves BLAST alignment by increasing the score in conserved regions (domains) [91]. The alignment was done stepwise, starting with closely related species and relying on curated databases. Only the transcripts without matches under a strict threshold were used for the following steps.

Besides, the transcripts were also annotated using InterProScan [44], to identify the domains, protein families, and their associated GO terms. The GO terms from both methods were then combined and the corresponding Enzyme Commission numbers (EC) were retrieved. This method yielded satisfactory results because only one enzyme from the metabolic model of Error: Reference source not found was not identified.

8.2.2 ***Tetrademus obliquus*: genome based read mapping**

For *Tetrademus obliquus* the situation with respect to annotation was very different as the genome was previously sequenced and assembled (Error: Reference source not found), which served as a solid base to perform the feature annotation that was needed for the expression analysis done in Chapters Error: Reference source not found and Error: Reference source not found. Using BRAKER1 [226] and 38 RNA-seq samples, we were able to identify the genes with their accurate exons and introns positions. The resulting protein sequences were then annotated to EC numbers using EnzDP [231]. This method was published in 2015, after the annotation for *E. oleoabundans* was done in Chapters Error: Reference source not found and Error: Reference source not found, and was unknown to us until early 2017. Unlike this previously used method, EnzDP is a bottom-up method that makes the direct link between sequence structure and enzymatic function. For this reason, the result is more accurate and has a much larger number of identifiable EC numbers. While the scoring method in EnzDP is very efficient, using a simple threshold, it is still difficult to effectively reduce false positive matches without risking to generate false negatives. To improve the result, the best scores (both likelihood-score and max-bitscore) were selected with a restriction on the maximum score gap between two proteins (Figure 8.2). Using this approach, each EC number had much less proteins associated to them (EC → Proteins, Figure 8.2 A), while keeping the best possible matches within a reasonable range. Then, the same score-gap filter was applied but now for each protein regarding their association to EC numbers (Protein → EC number, Figure 8.2 B). This resulted in each protein having a more limited number of associated EC numbers and thus reactions. However, after discovering the polyketide synthase of *T. obliquus*, which contains the functions of multiple enzymes, it was clear that this score distribution method described above would discriminate this type of proteins. To avoid missing potentially useful annotation, it was decided to use a standard likelihood-score threshold at the risk to require manual curation later on. Later on, manual correction was necessary to find the correct candidate

proteins for the reactions in few pathways (e.g. lipids and starch metabolism in Chapter Error: Reference source not found and Error: Reference source not found). The scores for each proteins and EC numbers were evaluated, all the while aligning the protein sequences using online tools [44, 91, 276]. This manual selection was on par with the aforementioned automated score distribution based filtering method. This means that the automated method could replace the manual labor while yielding comparable quality. The automated method seems to be a very good trade-off between removing false positive matches and limiting the removal of true positive. Besides, the case of proteins such as the polyketide synthase is rather rare and could be easily dealt with manual treatment. The method could be further improved using a dynamically selection by measuring the euclidean distances between the scores of each proteins. The distances from both score types could then be used for simple hierarchical clustering. Naturally, the clusters would contain the most similarly scored proteins, and the top cluster would contain the best annotations.

Protein ID	EC number	likelihood-score	max-bitscore
A			
ProtA	EC:1.1.1.1	1	691.6
ProtA	EC:1.1.1.2	0.91	400.2
ProtA	EC:1.1.1.3	0.098	274.7
ProtA	EC:1.1.1.4	0.75	752.8
B			
ProtB	EC:2.1.1.1	1	489
ProtC	EC:2.1.1.1	1	752.9
ProtD	EC:2.1.1.1	0.3	687.3
ProtA	EC:2.1.1.1	0.1	120

Figure 8.2: Hypothetical annotation results from EnzDP and selection of most suitable candidates. (A) case of one protein annotated to few EC numbers. (B) case of one EC number annotated to few proteins. The grayed lines represent the best protein ↔ EC number annotation. The red background values are too distant from the closest higher value. The green background values are within a short distance of the closest higher value.

The annotation method of EnzDP, used for *Tetrademus obliquus*, is then very competent at identifying protein functions of microalgae. The filtering method (Figure 8.2) should yield much better ratio between true positive and false positive protein-EC number associations. Furthermore, this method relies on building models directly from manually curated proteins in Swiss-Prot. With each update of Swiss-Prot, more proteins from various organisms enrich the

database and will ultimately help the method in defining better models for each enzymatic reactions.

8.2.3 Current state of annotation

Despite that our current annotation state still needs further curation, a substantial number of associations over a very broad range of functionalities was attained and is shown in Table 8.1, which contains basic information for the number of proteins and associated EC numbers. There is a large difference in total proteins between the displayed species. This is because other organisms contain multiple projects with multiple copies of the same genes, proteins and annotations. There are more proteins stored in UniProtKB for the human species than for the whole Chlorophyceae class. This reflects the fact that the quantity of curated data for microalgae severely lacks behind (Error: Reference source not found) [38, 47]. Both *T. obliquus* and *C. reinhardtii* count for a large part of the proteins in their common phylum and class. For well studied species like Humans and *A. thaliana*, the difference between the number of proteins (prot total) and proteins annotated to an EC number (prot-ec total) is much smaller than for the microalgae species in this table. As expected from the statements in the previous paragraph, our automated annotation contains significantly larger amount of EC numbers annotated per protein. Despite this easily fixable over-fitting, it is clear that thanks to the bottom-up approach of EnzDP, we were able to identify a very large number of functions. The number of unique EC numbers identified in our genome is comparable to the number of EC numbers found in the whole UniProtKB for human proteins, although there still is probably a not negligible number of false positives. It is also important to stress that the annotation of *T. obliquus* was done two years before. With the steady growth of SwissProt, today's annotation should potentially yield more EC numbers and possibly less false positives. Among the 18768 proteins of *T. obliquus* available in UniProtKB, 18538 proteins (95%) are from our contribution and are labeled by "BQ4739_LOCUSxxx". Yet, key reactions were still missing for our analysis. This automated annotation should be a useful complementary information from the already provided annotation with the published article (Error: Reference source not found).

Table 8.1: Number of proteins and EC annotations found in UniProtKB and Swiss-Prot, in comparison with annotations done for *T. obliquus*. The column UniProtKB regroups the databases of Swiss-Prot and TrEMBL, respectively the computationally annotated and manually annotated records. The columns “Prot” report the number of proteins. The column “Prot-EC total” represents the total number of times an EC number is annotated to a protein. The column “Prot-EC unique” represents the number of proteins annotated to an EC number. The column “EC unique” represents the number of unique EC numbers annotated. These numbers were collected in May 2019. Data obtained for *T. obliquus* are from Chapters Error: Reference source not found-Error: Reference source not found

	UniProtKB				Swiss-Prot			
	Prot total	Prot-EC total	Prot unique	EC unique	Prot total	Prot-EC total	Prot unique	EC unique
<i>H. sapiens</i>	171063	12780	12279	1393	20421	3800	3384	1373
<i>A. thaliana</i>	89205	6932	6605	1106	15856	4455	4169	1070
Chlorophyta (phylum)	314813	20078	19811	873	1168	311	305	80
Chlorophyceae (class)	149368	8479	8348	835	614	149	148	58
<i>C. reinhardtii</i>	31244	1146	1101	386	331	73	73	49
<i>T. obliquus</i>	18768	577	557	230	69	18	18	8
					This thesis			
<i>T. obliquus</i>					19723	4543	3911	1315

8.2.4 Microalgae-specific protein domains for improved annotation

More advance alignment methods use forms of position specific score matrices, giving higher scores for the regions and amino-acids that are evolutionarily conserved. The tools used to annotate *Ettlia oleoabundans* use protein domains to identify the conserved regions and functional groups. Each species has its own repertoire of protein domain (domainome). While many domains are known to be shared between species, they can also be kingdom specific [286]. Profile Hidden Markov Models are very powerful to deal with distant organisms, but the detection of domains is still limited at increased genetic distances. This is what we have shown in Error: Reference source not found, where we demonstrate that proteins of microalgae have low sequence similarity with the domains in the protein family database (PFAM). Even though

the known domainomes are diverse, they are clearly lacking domains discovered from microalgae. Using all the proteins available in UniProtKB for the Chlorophyta (Table 8.1), it would be possible identify the domainome of algae. The newly found domains could then be compared to existing ones, functionally annotated, and ultimately, they would greatly help in the difficult task of protein alignment and annotation for microalgae.

8.2.5 The importance of protein localization

Although we annotated a large number of proteins to enzymatic reactions, we do not have the information on where the reactions occur. While some reactions occur in only one organelle, many can occur in more than one compartment and are subject to different regulation. Clearly, information on localization of proteins and metabolic processes is very important for understanding the algal metabolism and for the development of genome scale metabolic models. This raises three questions: is the gene originating from the nuclear genome or an organelle genome? In the case of the nuclear genome, what are its final destination(s), which organelle genome the gene is originating from, which organelle the protein is being transported to, and how the necessary metabolites are transported between the organelle compartments. The first question can be easily answered is if the organelles and nuclear genomes are known like it is the case in this thesis for *Tetradismus obliquus* (Chapters Error: Reference source not found, Error: Reference source not found, Error: Reference source not found). In the case of *de novo* assembly, only mapping the contigs over other microalgae could give this information. However, the protocol used for illumina whole-transcriptome sequencing filters for polyadenylated transcripts. Since mitochondria and chloroplast have a low abundance of these, in addition to a low abundance of RNA in general, no expression from these organelles was detected in Chapter Error: Reference source not found and Error: Reference source not found despite the large data-sets. Again, today's UniProt's automatic pipeline has identified the subcellular location for some of our *T. obliquus* proteins. The annotation method used for *T. obliquus* could be improved by integrating the information about the protein location and generate separate models for each sub-cellular compartment. To this day [June 2019], Swiss-Prot counts 16,459 entries associated to mitochondrion, 11,623 on chloroplasts. Considering the numbers provided in the Table 8.1, these numbers are not negligible and could provide useful addition to the method of EnzDP.

8.2.6 Genetic code

The difficulties encountered when annotating microalgae proteins [38, 47], we suspected that it could be due to a difference in their genetic code. The impact of a single change in the genetic code should have major impacts on the resulting proteins, but that could explain the strong difficulties to annotate the proteins of certain microalgae like *Botryococcus braunii*. Since no information was clearly stated for the microalgae we were working on, we decided to use a prediction tool called FACIL [287]. To use it, some gene sequences should be provided, but any form of genome or transcriptome sequencing reads or contigs works as well. However, the results were not very clear using the tool as it displayed a rather high background noise. We hypothesized that this was because the tool aligns the entire domains. Domains correspond to small regions of proteins that are conserved. However not all amino-acids are equally conserved. In fact, only a minority of them have a high conservation score. We therefore decided to not simply align the conserved regions of the proteins (domain), but only the conserved amino-acids within each domain. Reusing the original code of FACIL, we added the conservation value and a threshold to count matching amino-acids aligned onto the domain. This simple improvement reduced greatly the background noise and resulted in more exclusive association between codons and amino-acids. The tool was left with a simple output in form of a matrix with codons on the rows and amino-acids on the columns, with the counts of matches. The result output could be further improved by displaying Genetic Code Logo similar to the one in the original tool. Unlike the original tool, the user does not need to provide the expected genetic code to improve the results, also meaning that no successive runs are necessary. With this version, we were able to show that none of the several microalgae tested, including the ones studied in this thesis, were using a different genetic code. However, *T. obliquus* has a different genetic code for its mitochondrial genome [287], and we were able to find the correct codon to amino-acid association. However, since this method is much more selective than the original one, more genetic material is necessary. Providing the mitochondrial genome sequence of *T. obliquus* was just enough to associate all the codons using a not too permissive threshold (lesser conserved amino-acids within the domains). Additionally, we were able to validate the results from recent articles describing mutations in the genetic code of several yeasts (data not shown). Today, more non-standard mitochondrial genetic codes are found in microalgae [288], suggesting that it may be necessary to systematically verify the mitochondrial genetic code of a microalgae before studying its mitochondrial proteins.

8.3 Transcriptome analysis

In this thesis, transcriptome analyses were done for different purposes. In Error: Reference source not found, steady state turbidostat cultures were operated at nitrogen replete and nitrogen limiting conditions at two light intensities under continuous red light. The aim of the transcriptome analysis was to get a better understanding of light stress, nitrogen limitation, and to ascertain the relation between metabolic flux and mRNA levels. In Error: Reference source not found, transcriptome analysis was used to obtain insights in adaptation mechanisms to salt water in combination with TAG production under nitrogen depletion. In Error: Reference source not found, we studied the changes in the transcriptional landscape during day night cycles under nitrogen replete conditions. This was then used as a reference to assess the effect of a mutation that inhibited starch formation on the gene transcription, thereby obtaining more insight in the physiological role of starch and the consequences of a starchless phenotype for growth. Likewise, in Error: Reference source not found, the effect of nitrogen limitation on the gene expression profile in the same wild type and starchless mutant were studied.

8.3.1 Transcription to flux correlation

As shown in the Figure 8.1, the transcriptome is a layer between the epigenome and the proteome. Part of the translated proteins participate in enzymatic reactions that transform metabolites and take part into the molecule turnover in metabolic pathways. This turnover rate of metabolites, in units of concentration per unit of time, is called metabolic flux. However, the relative change in a transcript does not need to result in a comparable change in flux through the reaction. Before translation, post-transcriptional events can occur and affect the effective translation level thus decreasing the correlation between transcript level and protein level. After translation and post-translational modifications, the metabolic flux is not only determined by the amount of enzyme. The metabolite concentrations determine the direction of the reaction, and together with the enzyme level the flux through the reaction. All these factors make it difficult to identify correlation between transcription and metabolic flux. In the Error: Reference source not found, we compared the relative transcription and the relative flux calculated from the constraint based metabolic flux model. Linear pathways are lumped and modeled as linear blocks that share one single flux value, whereas the transcripts represent the

different steps in the pathway and each have their own independent expression. While the results were not totally random, there was no clear correlation between the fluxes and mRNA content, as it has been often shown [289]. Since not all enzymes are equally driving the metabolic fluxes, a possible approach to find better correlation would be to select a number of flux regulating enzymes from the linear pathways. More forcefully, each enzymes could be tested for higher correlation, which would allow to identify possible regulatory enzymes. Another approach could be to consider the global trend and use the mean expression of the enzymes in the same linear pathway. Lastly, the model had some form of compartmentalization with reactions occurring in the chloroplast and in the mitochondria. However, the annotation of the protein localization was not sufficiently reliable for us to integrate this information and improve our results.

8.3.2 Single cell RNA sequencing

While the results of these diurnal cycle analysis were very insightful (Chapters Error: Reference source not found and Error: Reference source not found), there are few important points that rendered the interpretation of the data difficult if not ambiguous. These are cell homogeneity, sampling resolution, and choice of sampling time. The homogeneity of the cells in the sample is an important factor since we have sequenced bulk RNA and thus measured the average expression of all the cells. Still, it is well established that noise is unavoidable in all biological systems at all levels [290]. Homogeneity of cells is affected by two factors, the synchronization of the cells, and the presence of sub-populations of cells with specialized functionality. Bulk sample sequencing of desynchronized cells should result in weaker and wider peaks of transcriptional expression, also observed in our results. It is probable that not all the cells are stressed equally and certainly with more than a simple bi-polar state. Cells subjected to specialized functionality could be analogous to cell types. If those exist within a sample, they would have different expression than the others and logically increase the noise of a bulk sample.

As observed in our data, the more stress applied to the population, the less transcriptional variability is observed throughout the diurnal cycle. Nitrogen limitation results in strong stress that interferes with protein synthesis and cell growth. It is likely that it can also affect the synchronicity of the microalgae cells in the culture by not affecting all the cells with the same level of stress. If only a portion of the cells are stressed, it means that others are not and those

do not accumulate TAG. For best yield, all cells should be stressed at the same wanted level so TAG can be accumulated most efficiently. While this is an issue for our measurement, it is also an issue for the industrial applications. In fact, a loss in synchronicity is an important factor of variability in the culture that would systematically result in lower achievable yield upon harvest. Regardless the heterogeneity of the culture is the result of different stress state or desynchronization, and in addition to the reduced photo-efficiency, nitrogen limitation may not be the best choice for TAG accumulation.

Tetradismus obliquus (Chapters Error: Reference source not found to Error: Reference source not found) is known to be a colonial green algae in clusters of 2 to 8 cells encapsulated within a cellulosic wall [291, 292]. It is possible that the cells in these colonies have different phenotypes. Furthermore, the known sexual cycle of *T. obliquus* is also a potential source of cell differentiation [293, 294]. Analyzing these species with single-cell RNA sequencing should provide enough information to answer these suppositions. Even though most gene functionality is unknown, simply identifying large transcriptional differences within a tightly controlled environment would be a significant step. The smaller number of clusters and less extreme peaks of expression could also be due to the sampling frequency. While hourly sampling proved to be precise enough, sampling every three hours was not sufficient to be able to identify the quick responses that can occur. While selecting the samples, we were certain to which extent the sampling frequency would be sufficiently precise. We then decided to take frequent samples (1h) for the wild-type and less frequent samples (3h) for the starchless mutant and align their expression. However, the high frequency sample revealed strong and short burst of expression that could be entirely missed within the time frame of 3 hours. Besides, it is important to have precise measurements around the light switches in order to identify whether the transcriptional change is an anticipation or possibly a consequence of the switch. The best example is in Error: Reference source not found, where the measurements are surrounding the switch to dark. In Error: Reference source not found, sample was taken at the time of light switch, which allowed us to make the distinction between adaptation and anticipation. Sampling every two hours in combination with every 1 hour around the light switch could thus be a good compromise.

8.3.3 Characterization of the starchless mutant (*slm1*) gene

In Chapters Error: Reference source not found and Error: Reference source not found, we used transcriptomics data to obtain insights into how the mutation in the small subunit of the ADP-glucose pyrophosphorylase affected expression of genes in the starch metabolism and how this mutation caused the starchless phenotype. It was already demonstrated that the starchless mutant *slm1* cannot synthesize molecules of ADP-glucose and any subsequent products including starch or products of starch degradation [33, 34]. It was also shown that the gene was affected by an early stop codon (non-sense mutation), and the transcription of the gene was also greatly reduced [35]. The mRNA data from the Chapters Error: Reference source not found and Error: Reference source not found confirmed that the expression level was much lower in *slm1* than in the wild-type (>5-10 folds). Most importantly, we found that the diurnal regulation of the ADP-glucose pyrophosphorylase small subunit was the same between the wild-type and *slm1*. The reads mapping displayed in Figure 8.3 clearly shows a rather even distribution on the whole gene with a nearly identical distribution for the two strains, but with much less reads overall for *slm1*. The lower abundant reads in combination with a conserved diurnal regulation suggest some form of post-transcriptional regulation. Since the gene contains a nonsense mutation, it is logical to think that the transcripts are degraded by the Nonsense-Mediated mRNA decay (NMD) pathway [274, 275]. NMD is a post-transcriptional and translation-dependent mechanism that actively degrades from both ends the abnormal mRNA coding for shortened proteins. While this protective mechanism leads to lower mRNA content for the mutated gene, it is unlikely that any translation occurs for the same reasons. Quantifying the protein content will confirm if NMD is the cause of lower mRNA content in the starchless mutant of *Tetradismus obliquus*.

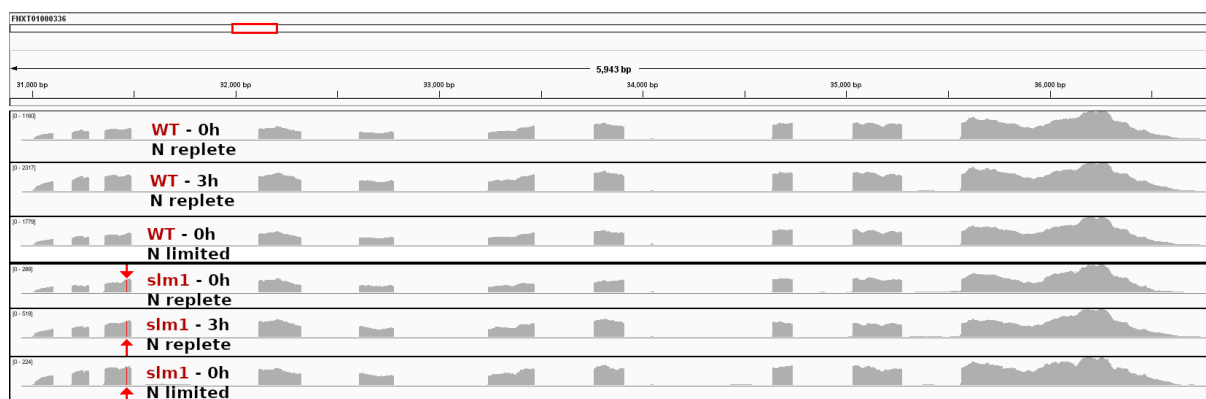


Figure 8.3: RNA sequencing reads mapping onto the genome of *Tetrademus obliquus* (Error: Reference source not found). Using Integrative Genomics Viewer (IGV), scaffold FNXT01000336, region ~31000-37000, location of the gene *g788* coding for the ADP-glucose pyrophosphorylase small subunit (Uniprot gene id BQ4739_LOCUS4183). Reads from three samples of each strains: UTEX393 (WT); starchless mutant (*slm1*). The samples were taken at the same time points for each strains. Samples were also taken from each growth conditions: N replete (Error: Reference source not found) and N limited (Error: Reference source not found). The red arrows were added to indicate the red vertical bar indicating the consistent mutation that is the nonsense mutation responsible for incapacity of *slm1* to synthesize starch. The gray barplots represent the reads mapping density over that area. On the top left of each sample rectangles, a range of number is written to indicate the y-axis scale of the respective sample.

8.4 Experimental condition and bioreactors control systems

The photobioreactors are the state-of-the-art reactors that allow growing microalgae with tight control. These reactors are perfectly suited to make sure that only the factor of study is varied and all other factors remain constant. Thus, confounding factors can be excluded and noise on the measured parameters can be reduced. Still, the experimental conditions were not all perfect, due to either the reactor itself or due to how it can control certain factors. The limited volume of the reactors does not allow to extract too large volume if it is necessary to keep the batch unperturbed after sampling. In Chapters Error: Reference source not found and Error: Reference source not found, large duplicate samples needed to be collected at each sampling time, which imposed the diurnal cycle sampling to take place over weeks instead of a single day. The reactor is set to control the culture in a stable diurnal cycle over the whole sampling time, and the transcriptome data show very similar expression between biological duplicates.

Still, there are few time points where the differences between duplicates are rather large, and these tend to be at times of large transcriptional changes. These conditions were more than sufficient for the objective of these studies with this level of precision, but it must be improved if more frequent sample timing is used or if more precise measurements are wanted.

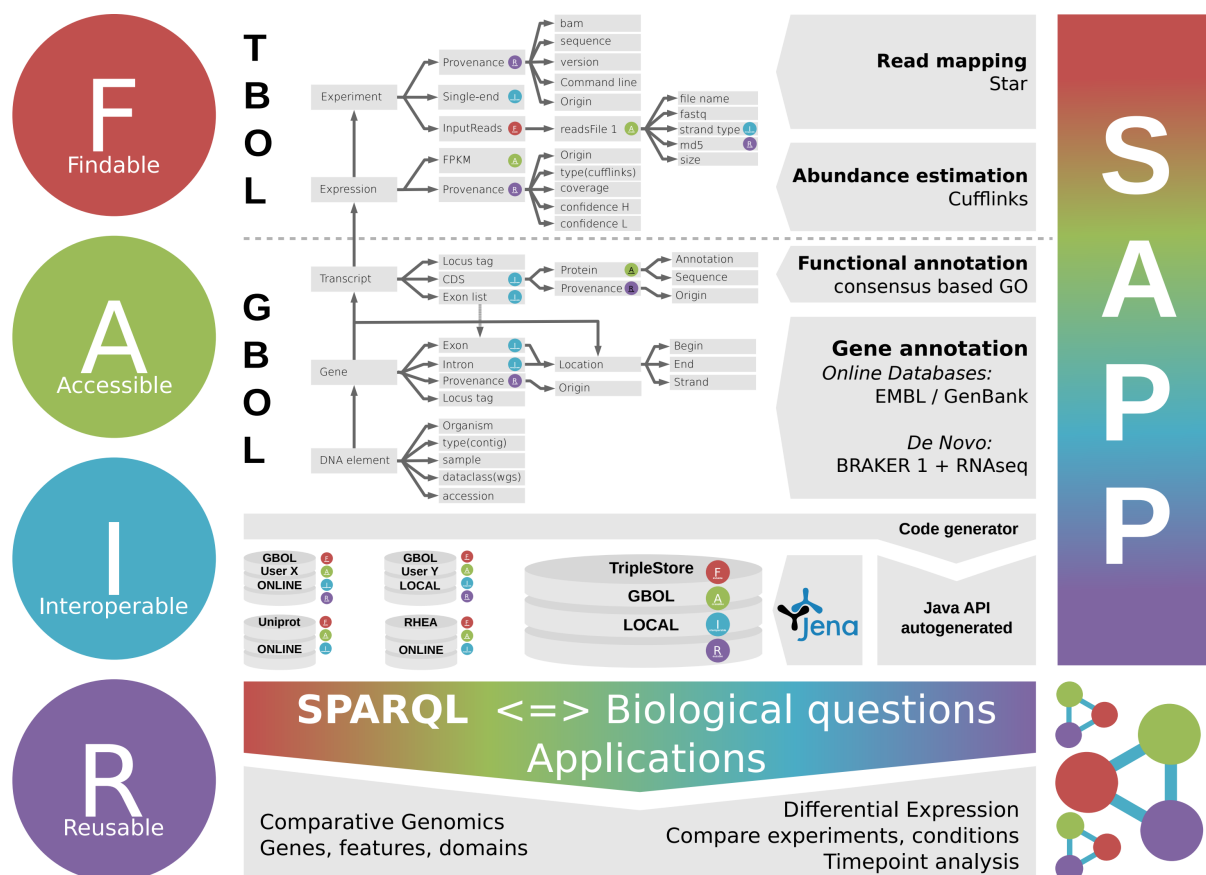
Although our experiments were carried in 16-8h light-dark cycles, the conditions are still not representative of what happens outdoors. The photobioreactors lighting system is binary, with no intensity changes, and most importantly, no light spectrum changes. It is well known that both factors play a role in setting the circadian clock. Our results show that particular light spectrum or changes in light intensity are not necessary for the system to display a diurnal rhythm. Nevertheless, we do not know how these missing factors could affect the temporal regulation of biological processes described in this thesis. A previous study showed that light exposure prior to dark period affects the cellular respiration during the dark period, and that if microalgae are still in growth phase right before the dark period, more respiration will be necessary [295]. This means that having progressive light switch should help reduce cellular stress by buffering diverse processes and allow smoother transcriptional changes. Microalgae are found in a large variety of environments, and so is the variety of circadian rhythms in behavioral and physiological processes [243]. Aside, the change in temperature and light can regulate circadian rhythm cycles in plants, and these changes are also likely to happen in outdoor conditions [296]. It would be valuable to study and compare the effect of diverse progressive light switches and the effect of different light spectrum between light switches. By the progressive light or change in blue light and far-red light, should be possible to identify some key genes and regulatory events that would be triggered by these different conditions. Using single cell transcriptome analysis, it would also be possible to observe how these factors affect the synchronicity of the algae population in the culture in comparison to our not so natural experimental condition. Such information is crucial in the case of outdoor industrial scale production, as it would allow to optimize yield by harvesting in prevision of weather changes.

8.5 FAIR data management of transcriptomics data

RNA-sequencing is a powerful method with a wide variety of applications, notably to understand, explain, and predict biological systems. However, this variety of applications multiplies to quantity of tools, pipelines and practices [297]. Furthermore, these are often not interoperable, static, lacks provenance, and therefore, hardly reproducible. In some cases, combining the tools outputs and predictions could help in assessing the quality and gain the best insights [38]. The best way to improve the quality of biological data is to follow the FAIR guiding principles: Findability, Accessibility, Interoperability, and Reusability [228]. To tackle biological data management bottlenecks, our group started to develop the semantic Genetic Biology Ontology Language (GBOL) [227]. This ontology is the foundation of a harmonized consistent description of biological data by taking advantage of available semantic technologies. GBOL is tightly integrated to other ontologies, semantic data management, and bioinformatics software, into a toolbox called SAPP (Semantic Annotation Pipeline with Provenance) [39]. While SAPP and GBOL were mainly oriented to genomics data, particularly for the microalgal work we started developing the tools and the ontology for transcriptomics data. We created a new semantic ontology called Transcript Biology Ontology Language (TBOL). It provides, on the top of GBOL's consistent and detailed genome structure and annotation, RNA sequencing specific terminology that includes transcripts expression. We implemented the first set of tools with the semantic data integration with provenance following terminology and structure provided by the TBOL ontology. The information about TBOL is summarized in the Figure 8.4, including information such as the type data stored in the ontology, their relevance with the FAIR principles and the tools integrated.

GBOL can store the information as found in the standard formats, but in a non ambiguous format and with data provenance. TBOL currently extends it with RNA sequencing information, which ranges from details of input files (e.g: file type, sequencing type, strand pairing, strand orientations), to the expression of the transcripts, while providing the means to keep information about the method for best reproducibility (data provenance). We focused our effort in implementing a genome based strategy for which we used the genome of *T. obliquus* from Error: Reference source not found and analyzed more than 100 samples of RNA-seq for Chapters Error: Reference source not found and Error: Reference source not found. The results of this analysis are a proof that a semantic framework such as SAPP can analyze RNAseq and deliver properly integrated FAIR data using semantic technologies. Later on, SAPP was extended to analyze RNA-seq data with InterproScan and EnzDP for protein family annotation

and EC number annotation, respectively. More tools should be deployed to cover the best practices of RNA-seq data analysis.



- F** All data are relevantly assigned with globally **unique** and **persistent identifiers**. **Meta-data is a priority**, and used for unique identification.
- A** Our software and ontologies are **free**, **open source** and **available online**. Users are welcome to review the code, make propositions.
- I** **TBOL** and **GBOL** are our semantic data description language, allowing precise description of information with cross database access.
- R** **SAPP** ensures a proper use of data access and creation, while keeping **process meta-data** (provenance) for the future. Data can be shared, compared, improved, reanalysed, etc.

Figure 8.4: Framework for RNA-seq data analysis according to the FAIR data principles. SAPP is a semantic genome annotation tool tightly integrated with the Genetic Biology Ontology Language (GBOL). To add transcriptomics data analysis, SAPP was extended with new set of bioinformatics tools that use the Transcript Biology Ontology Language (TBOL), also designed as an extension to GBOL. The F,A,I,R bubbles informs whether the data or the type of database complies with one of the principles.

8.6 A look back at transcriptional landscapes to move forward with strain improvement

8.6.1 Looking at the circadian clock

In the Chapters Error: Reference source not found and Error: Reference source not found, we have analyzed the diurnal transcriptional landscape of *T. obliquus*, describing the succession of biological processes throughout the whole diurnal cycle. However, we were not able to identify the circadian clock genes, nor any of the processes directly regulated by them. Microalgae are believed to possess a simpler circadian clock system with only two plant homologous genes that are suspected to function as feedback loop [243, 298–300]. Most of the information could already be inferred from the model organism *C. reinhardtii*. Since *T. obliquus* is a rather distant organism (class) than the studied ones, a complementary analysis could be useful to identify and compare the homologs with other microalgae and plant [301]. To this date, no *T. obliquus* genes were automatically annotated by the UniProt pipeline. A dedicated analysis of the circadian rhythm system in *T. obliquus* could be very valuable in improving metabolic model prediction of diurnal adaptation in response to commonly occurring stimuli.

8.6.2 Use of diurnal transcriptional landscape

In Error: Reference source not found, the hourly sampling allowed us to identify succession of biological processes throughout the diurnal cycle in detail. The flat panels photobioreactors used to mimic the cycle provided precise control to fine tune many parameters which in turn allowed us to measure tightly regulated transcriptional changes. We also observed that lack of starch in the starchless mutant *slm1* resulted in dysregulation of the diurnal cycle leading to time shifts of processes and a small decrease in amplitude. In Error: Reference source not found, nitrogen limitation resulted in much decreased amplitude of transcriptional changes over the diurnal cycle and this effect was further amplified in *slm1*. Dysregulation of the diurnal cycle may lead to a less synchronized population, which would result in reduced complexity of diurnal profiles with a reduced number of clusters. This is what we observed

when comparing the clusters in Chapters Error: Reference source not found and Error: Reference source not found.

The aforementioned clusters were found to group genes from same pathways or biological processes, which allowed us to describe the succession of metabolic events throughout the whole diurnal cycle. The diurnal profiles of these clusters can now serve as reference to describe the diurnal expression of genes. Analyzing and comparing the diurnal transcriptional landscape of different light-dark cycle setups could give insights about the light period stress and the respective recovery during the dark period. Furthermore, it would of great value to better predict the metabolic adjustments of the different clusters and their associated processes. For example, the early-light period peak is probably a consequence of recycled photosystems throughout the dark period. Similar patterns can be found in other processes, for which we can expect stabilization until the dark switch. These processes are dependent on nitrogen for amino-acids synthesis and energy supply. Logically, they will take longer with limited supply. Without actual data to compare with, we can already predict some transcriptional changes in response of different light-dark cycle setups. Longer light period should, if anything, not result in a stretched peak activity but simply stretched base level. Cell division processes would probably be extended till the end of the light period, leading to more successive cell fission at dark. It is also possible that more stress would be put on the cell to synthesize amino-acids before light, delaying early light period processes like we observed with lack of transient energy in *slm1*. *A contrario*, shorter light period should result in less damaged photosystem, requiring less intense peak in early day. The maintenance of other proteins as part of the early light period could result in lower peak and reach stable level earlier. With shorter light period, more starch should be accumulated, and with longer dark period, it is possible that more starch is being consumed. Cell growth processes should start only after the earlier demanding processes (mostly photosystem synthesis) stop sinking carbon and energy. Additionally, the observed double peak of expression could be associated to two successive of cell growth and results into fission of four cells. Changes in light period duration should then result in respective increase or decrease in the number of cell division. These hypotheses are consistent with the observation in the 12:12h diurnal cycle RNA sequencing project of *C. reinhardtii* [221]. Their higher number of division per diurnal cycle (8-16 daughter cells) must be due to the faster growth rate of *C. reinhardtii* in comparison to *T. obliquus* [302]. Predicting when the processes occur and when the metabolites reach their highest concentration is very valuable in bioprocessing application. In addition, predicting how the organism adapts with changes in light is important for decision making, regarding harvesting or culture adjustments and help

maintaining higher yield. More specifically, this could help in adapting large-scale cultures to predictable but uncontrollable outdoor weather.

8.6.3 Improving the nitrogen limitation stress phase: Starch synthesis

Starch is the main metabolic competitor to TAG, and if not synthesized, more energy and carbons can be redirected to the synthesis of TAG. This idea was tested using the metabolic model of *Ettlia oleoabundans* in Error: Reference source not found, and the results were very positive promising. However, the model also showed that the potential gain is far inferior compared to what can be reached by optimizing light energy usage. In continuous light condition, the lack of starch increases the yield of TAG on light by 51% [34], but in light-dark cycle condition, the lack of starch results in 20% reduced energy conversion efficiency [24]. Furthermore, in nitrogen limited LD cycle conditions, even though the TAG content in *slm1* is higher than in the wild-type (0.23 vs 0.07 g·g_{DW}⁻¹), the overall energy conversion efficiency from photons to biomass components in *slm1* was only half that of the wild-type. Thus, the lack of starch results in a considerable loss in the efficiency with which light is used. This is because starch is being used a transient energy storage that provides the needed energy for respiration during the dark period. We have shown using transcriptomics that a number of energy demanding metabolic processes occur during the dark period. These processes are associated to the cell maintenance, cell division, amino-acid synthesis and protein turnover. This allows the wild-type to kick-start the synthesis of pigments for well-functioning photosystems at the start of the day. However, in *slm1* these dark period processes are delayed to the early light period when energy becomes available again. This delays all the early light period processes. An important consequence is that the photosystems are not synthesized in time, and photon energy is effectively wasted. This issue is even more dramatic in condition of limited nitrogen. We observed that limited supply of nitrogen seems to impose the organism to degrade more proteins and amino-acids during the dark period. Finally the loss of photosynthetic efficiency in the starchless mutant occurs especially during the second half of the day when in the WT starch synthesis occurs. Apparently, the ability to synthesize starch allows a more efficient use of the light during this period.

Based on this, instead of relying on a fixed starchless phenotype, it could be better to develop a microalgae strain in which starch synthesis can be controlled over time and be limited to the

right amounts. Thus, during the growth phase, starch synthesis should be kept fully functional to allow for fast growth and optimal light use. Next during the TAG accumulating stress phase the accumulation of starch should be restricted to the second half of the day to an amount that is just enough to supply sufficient energy for the processes that should occur during the night. Thus only the necessary amount of starch should be synthesized, to minimize the competition with TAG synthesis. As the measurements show in the Error: Reference source not found, the starch content in nitrogen limited condition is approximately 25% of the dry weight, and the daily variation seems to be fluctuating by approximately 5% of the dry weight and 1/5th of total starch content. This means that in this well controlled LD cycle and nitrogen limited condition, there is 5 times too much starch accumulated in the microalgal cell. Ultimately, controlling starch synthesis intensity on-demand depending for example on light intensity and temperature would allow for the best results. Yet, this level of synthetic organism engineering with such deep control is far from being possible at the moment, and plenty of information still needs to be collected before anyone are capable to build models that would control such a complex system efficiently.

8.6.4 Improving the nitrogen limitation stress phase: Nitrogen supply

Microalgae found in ocean move upwards during the day for photosynthesis, and move downwards in the depths of the ocean at night, where more nutrients and notably nitrogen, are available [243, 303]. Mimicking this condition could be beneficial for TAG accumulation. Nitrogen limitation over depletion was an answer to allow the microalgae to survive for a longer period of time in continuous batches while limiting cell growth. However, while light-dark cycle normally results in a gain of efficiency, it is lost with limited nitrogen supply. We have demonstrated that *T. obliquus* synthesizes amino-acids during the dark period and early light period, also recycling proteins in preparation of a new day cycle. Providing the needed nitrogen during the dark period could help the microalgae to sustain its daily protein synthesis and may result in less proteins degradation during the night. It should then be possible to recover diurnal cycle gain of efficiency by providing nitrogen during the dark period. While still limiting the nitrogen concentration during the day, cell growth should remain inhibited. Further limiting nitrogen during the light period, or even depleting it, could further increase the TAG yield on light. It is also described that nitrogen can affect all canonical properties of

circadian rhythm, namely the amplitude, phase and period [304, 305]. Therefore, supplying nitrogen only during the dark period has the potential to reschedule diurnal processes and result in stronger changes than nitrogen limitation throughout the whole day. It would be interesting to observe whether this cyclic nitrogen stress will impact only nitrogen dependent processes or if it will influence in the circadian rhythm and impact a wider range of processes. Finally, this approach could be combined with the above-mentioned control of starch synthesis, which would guarantee that just the right amount of energy is available during the night for nitrogen assimilation.

8.6.5 Salt water adaptation

In the Error: Reference source not found, we have analyzed the transcriptional adaptation of *E. oleoabundans* when grown in salt water. We have identified several pathways that could be responsible for the adaptation in this osmotic stressful condition. Proline synthesis and ascorbate-glutathione cycle seem to be the most important ones, but only proline content was experimentally verified using nuclear magnetic resonance (NMR) where a higher concentration was measured in salt water condition. Even though the replicate measurements were very different, the lowest measured value was still multiple times higher than the highest measurement in fresh water. We also discussed about the possibility that sugars such as sucrose or trehalose could play a role in alleviating the osmotic stress. We only found differences in transcription for sucrose synthesis, but it seems to be only responsible for a quick response and not for long term adaptation. It is possible that the proteins involved in the many analyzed pathways were wrongly annotated or not annotated at all. Comparatively, many less EC numbers were identified in *E. oleoabundans* than in *T. obliquus*, furthermore, the accuracy of the annotation was also much improved in the case of *T. obliquus*. Thus, it would be valuable to re-perform the transcriptome analysis with current state knowledge and technologies, using the method applied to *T. obliquus* in Chapters Error: Reference source not found and Error: Reference source not found with the improvements described in this discussion. The annotation could also be improved with a high quality genome, but due to consistent contamination of samples, the genome has still not been assembled well enough to be accepted in databases. Besides verifying the statements already made on proline and sucrose synthesis, the improved protein annotation should help identifying whether the other osmo-protecting processes or metabolites described in Error: Reference source not found can take

place or can be synthesized. For example, the biosynthesis glycine betaine was missing the important choline monooxygenase, if it was to be found after reannotation, the glycine betaine would then be a potential factor for the protection of salt stress. Besides, as more transcription does not always result in more proteins or flux, it would be valuable to verify the metabolite and protein content. With sufficient information, it would be possible to draw a map of the salt stress protective mechanisms of *E. oleoabundans* or any other suitable microalgae. Such a map would allow to understand the importance of each individual mechanism. Two approach could be undertaken: introducing the salt water resistance to high TAG yield producers, or manipulate TAG production in salt water friendly strain. This would require to introduce multiple genes and would raise many issues regarding epigenetics, and regulation of transcripts and proteins. To undertake this non trivial experiment, the general knowledge in green microalgae should be greatly improved together with microalgal genetic modification tools. These synthetic manipulation could then be combined with forced evolution in salt water.

8.6.6 Metabolic engineering to improve lipids and TAG accumulation.

In the Chapters Error: Reference source not found and Error: Reference source not found, we have observed strong increase of transcription for the genes involved in lipids and TAG degradation. Actual degradation of lipids during the night could not be confirmed on the biochemical level as the changes were not large enough to be detectable by our measurement method. While lipid catabolism helps in maintaining a healthy turnover of membrane lipids, it will also decrease the yield of lipids on light. With respect to the last point it would be logical to avoid TAG degradation by suppressing the two enzymes that are responsible for the debranching of the fatty-acid chains from the glycerol backbone: the triacylglycerol lipase and the monoacylglycerol lipase. Our functional annotation found candidate proteins for both enzymes, and after verifying their function experimentally, the right proteins could be targeted for removal. Besides, while membrane phospholipids are degraded as part of the turnover, some of them can be directly used to synthesize novel TAG by the help of the phospholipid diacylglycerol acyltransferase (PDAT). While blocking forms of phospholipase could be beneficial for TAG accumulation, it could also lead to troubles in the cell membrane remodeling. Our annotation of *Tetradesmus obliquus* has 15 candidate proteins for the

phosphatidylcholine-sterol acyltransferase (LCAT, EC 2.3.1.43) and four for the phospholipase D (EC 3.1.4.4). Two interesting candidate proteins associated to LCAT, which originate from the same gene (g18408), are also associated to PDAT. Aside, to ensure that all the lipids destined for degradation end up stored in form of TAG, important step in lipid degradation must be avoided. The most logical approach would be to avoid the first step in lipid degradation to happen. This reaction can be performed by a multitude of enzymes available in the mitochondria by the acyl-CoA dehydrogenases or in the peroxisomes by the acyl-CoA oxidases. The different acyl-CoA dehydrogenases enzymes (EC 1.3.8.7-9) have preferences for certain lengths of the acyl chain, but none found with our annotation. However, we have two candidate proteins for the acyl-CoA oxidases (EC 1.3.3.6, g12802.t1 and g3816.t1). Even though we cannot be sure without experimental validation, it is possible that these candidate proteins are the only starting point of the beta-oxidation as it was found in other microalgae [306]. However, before these initial long chain degradation steps occur, the fatty-acid must be bound to a Coenzyme A (CoA) with the help of the acyl-CoA synthetase (ACS, EC 6.2.1.3), and this reaction occurs in both anabolic and catabolic processes. Our manual annotation and analysis revealed two main candidate proteins that are transcriptionally regulated like the other synthesizing or degrading genes. This led to the conclusion that one is responsible for the synthesis and the other one, ACS (g4377.t1), for the degradation and thus could be a candidate for knock out to suppress degradation of fatty acids. Again, experimental validation should be done to confirm this hypothesis.

8.7 Conclusion

The work in this thesis is the result of the integration of two disciplines, Bioprocess Engineering and Bioinformatics. It demonstrates how such a multidisciplinary approach can lead to more knowledge and at the same time bring scientific findings closer to application.

The most important challenge of this thesis was to functionally annotate proteins. The annotation method using EnzDP was a great improvement over the more conventional DELTA-Blast coupled with Gene Ontology EC number inference, used for *Ettlia oleoabundans*. Yet, EnzDP could greatly benefit from a better selection and protein location within the cell organelles. At the end of this thesis, the challenging protein annotation in high-throughput way is already yielding better results. Still, until many more microalgae proteins are experimentally identified and described in databases, thus increasing the accuracy and number

of correctly annotated proteins, the biological interpretation will always be filled with uncertainty. In preparation for the *Tetrademus obliquus* projects, the genome sequence was assembled with a high coverage. The genes were accurately identified within the genome and served as a solid base for accurate protein annotation. State-of-the-art bioinformatics tools and specially developed ontology and software base were developed to fulfill the FAIR data principles.

Our analyzes focused on comparing adaptation of microalgae to industrially relevant growth conditions: diurnal cycle, light stress, salt stress, and nitrogen limitation and depletion. We have shown that intense light stress and salt water result in oxidative stress in the cell. We were the first to report insights in the saline resistance mechanism of a fresh water oleaginous microalgae. The mechanism behind the biosynthesis of compatible osmolytes can be used to improve *E. oleoabundans* and other industrially relevant microalgal strains to create a more robust and sustainable production platform for microalgae derived products in the future. Using a metabolic model, we have shown that TAG yield on light can be more efficiently improved by optimizing photosynthetic conversion than by blocking competing pathways. We have identified as a possible regulator for photosynthetic capability in response light intensity changes.

We built a solid knowledge base of how a microalgae transcriptional landscape changes in a diurnal cycle. We described that there is a clear succession of metabolic processes and careful scheduling of events during the various phases of the cycle. Such knowledge is important for future experimental design, metabolic engineering strategies, and it serves as a reference case to study other experimental conditions. Starch seemed to be the only transient energy storage that is responsible for yield gain in a diurnal cycle. When missing, insufficient energy is available during the dark period, resulting in a loss of synchronization of a number of diurnal regulated processes.

Nitrogen limitation significantly reduces cellular activity. An insufficient supply of nitrogen for cellular maintenance seems to result in increased protein degradation just before and during the dark period. On top of that, the lack of starch in this condition results in much delayed activity in the early light period, and possibly more protein degradation during the dark period to provide some energy. Furthermore, we observed that the lack of starch triggered lipids and TAG degradation during the dark period, which would provide energy to the cell, but no significant variations were measured within this experimental setup.

Finally, specific points related to the analyses were discussed, such as the correlation between expression and fluxes, homogeneity of the synchronized cells, and the experimental conditions.

Then, based on the knowledge acquired in this thesis, clear metabolic engineering strategies were given to improve total solar-to-TAG conversion efficiency of microalgae under industrially relevant growth conditions. Increasing and maintaining light conversion efficiency of microalgae at industrial scale will improve step by step, by correcting each of the individual bottlenecks that are already known and described in this thesis. A crucial challenge to solve before reaching competitive industrial production of biofuels will be to allow secretion of TAG out of oleaginous microalgal cells. TAG secretion would allow long-term production with a single batch, avoid wasting photons in cell growth, and ultimately, it would greatly reduce the production costs. Hopefully, this drop in the ocean of knowledge will contribute for the future of our planet and help alleviating human society's dependence on fossil fuels.

Summary

Microalgae are unicellular microscopic and photosynthetic organisms. They are found all over the planet and in all sorts of environments. Their role has been and is still very important for the planet, most notably that they are currently accounted for producing half of the atmospheric oxygen. While they used to be studied because of their capabilities to depollute water, the interests have shifted towards oleaginous microalgae and their high level of fatty-acids accumulation. Fatty-acids such as Triacylglycerides (TAG) are of particular interest for their easy chemical treatment to produce clean biodiesel. Even if microalgae have higher energy conversion efficiency than plants, do not need arable lands to grow on, and do not compete with feed production, the optimal conditions of production are still too costly to compete with fossil fuel pricing. To decrease production costs, growth conditions and the physiological efficiency of the microalgae needs to be optimized. This requires a deep understanding of the microalgae phenotype in the relevant growth conditions. A phenotype represents a set of behavioral traits of an organism. In this thesis, the internal phenotype, the transcriptional landscape of two oleaginous microalgae species was studied using different growth conditions. The RNA content was chosen because it provides a dynamic system-wide view, can be done in high-throughput and, as proxy, can inform about the cellular and metabolic activities in response to a changing environment.

To analyze the transcriptome, it is necessary to know the functions of the transcripts. In contrast to model organisms like human or *Arabidopsis thaliana* whose genomes have been deeply annotated and studied, microalgae are far from that state. For both organisms studied in this thesis, *Neochloris oleoabundans* and *Tetrademus obliquus*, it was necessary to annotate the genes and transcripts since it was never done before. To functionally annotate a gene, most methods rely on sequence similarities to identify the closest gene in known organisms. Green algae are a difficult case due to the large genetic distance between them and the greater distance to reference organisms from land plants. Error: Reference source not found treats about the particular problem of protein annotation in microalgae. I analyzed the general state of data availability in microalgae, and we discussed about several annotation methods that are better suited than sequence similarity and discussed the limitations of using domain-based recognition methods. Besides, we also discussed the identification of the protein localization within the different subcellular compartment in microalgae. Finally, we stressed the importance of a large-scale wet-lab efforts for a few selected microalgae in order to provide a solid foundation for computation based methods.

To obtain more insight into the metabolism of *Ettlia oleoabundans* in Error: Reference source not found, a constraint-based metabolic model of *Ettlia oleoabundans* was built around the central carbon metabolism. This model was built based on the knowledge of central metabolism of algae at that moment and was cross checked with the *de novo* assembled annotated transcriptome. Experiments in controlled turbidostat were conducted in different combinations of light intensity and nitrogen supply. The measurements from the experimental conditions were used as constraints on the inputs and outputs of the model, effectively allowing us to estimate the metabolic flux distributions. In addition, RNA samples from the different experimental conditions were sequenced and analyzed. These data were used to validate the model structure as stated before, correlate expression levels with flux distributions and get a better understanding of the effect of light and nutrient conditions on algal physiology. The metabolic model calculates a maximum TAG yield of *N. oleoabundans* on light of $1.06 \text{ g (mol photons)}^{-1}$, more than 3 times the current experimental yield under optimal conditions. The model also shows that TAG yield on light can be more efficiently improved by optimizing photosynthetic conversion than by blocking competing pathways. Geranylgeranyl diphosphate reductase was identified as a potential regulator for photosynthetic capability that complements the fine-tuning of chlorophyll levels from synthesis and degradation. Finally, we identified some key reactions that could be targeted to improve TAG yield, by not only specifically increasing the flux within the lipids and TAG pathways, but also potentially redirect carbons from other pathways.

Water is a precious resource, and using fresh drinkable water to grow plants or microalgae could be considered non sustainable. However, salt-water is abundantly available and would be cheaper to supply. Therefore it is important to understand how algae deal with a salt water environment under growth and production conditions. This allowed us in Error: Reference source not found to study how algae deal with high salinity conditions under nitrogen replete (growth) and nitrogen deplete (TAG accumulation) conditions using a transcriptomics approach. The oleaginous microalgae, *Ettlia oleoabundans* (formally known as *Neochloris oleoabundans*) was chosen as a model algae, since it can accumulate large amounts of TAG and can grow in both fresh and salt-water. For this algae experiments were done in fresh water and salt water in combination with nitrogen replete and nitrogen deplete conditions. In addition to the transcriptome, we analyzed the biomass composition including TAG and starch accumulation and used the data to look into different salt resistance mechanisms. We found that Proline and the ascorbate-glutathione cycle seem to be of importance for successful

osmoregulation in *N. oleoabundans*. Genes involved in Proline biosynthesis were found to be upregulated in salt water, which is supported by Nuclear magnetic resonance (NMR) spectroscopy. Oil accumulation is increased under nitrogen-deplete conditions in a comparable way in both fresh and salt water. The mechanism behind the biosynthesis of compatible osmolytes can be used to improve *N. oleoabundans* and other industrially relevant microalgal strains to create a more robust and sustainable production platform for microalgae derived products in the future.

Although the TAG content that can be reached in *Ettlia oleoabundans* is high, the volumetric TAG productivity in *Tetradesmus obliquus* was evaluated to be clearly higher, while reaching the same TAG content. This was mainly due to the ability of *T. obliquus* to maintain photosynthetic efficiency for a longer time longer during the nitrogen depletion phase. Therefore, it was decided to switch to *T. obliquus* as a model organism. To obtain an idea of the capabilities of *T. obliquus* and to make transcriptome experiments easier to analyze, the genome of *T. obliquus* was sequenced. In Error: Reference source not found, the sequencing of the genome of *T. obliquus* is presented. The assembly approach was unconventional by combining two different methods and was able to combine the higher coverage from one method with the precision of the other method. In this way, the coverage and the accuracy of the assembly was maximized.

Production using microalgae will in many cases occur outdoors using sunlight. Consequently the algae will be exposed to the naturally occurring day night cycle. To better understand the effect of these day night cycles, in Chapters Error: Reference source not found and Error: Reference source not found, the transcriptional response of algae to diurnal cycles was studied under nitrogen replete conditions and nitrogen limiting conditions for the wild type and a mutant that can not synthesize starch. In Error: Reference source not found, hourly samples of RNA of *Tetradesmus obliquus* UTEX 393 were taken from a turbidostat culture operated under nitrogen replete conditions over a diurnal cycle of 16 hour light and 8 hours dark, to obtain more insight in in the transcriptional response towards diurnal cycles. In addition, to understand the effects of a lack of starch, the major transient energy storage, we sequenced samples of the starchless mutant *slm1* that were collected every three hours under the same conditions. At the same time, samples were collected and measurements of the biochemical composition of biomass and the specific light absorption rate were performed. These data are presented in a previously published article [24]. The genome features were annotated using

more than 38 RNA-seq samples from this study, using a specially developed extension of the FAIR principle based framework called SAPP. The work done to extend this framework for transcriptome analysis is described in the discussion Error: Reference source not found. We described the succession of metabolic events that occurred during the diurnal cycle, which are in agreement with the biochemical measurements. Comparing the wild-type with the starchless mutant *slm1*, we found a few temporal shifts in expression that reflect transcriptional adaptation to the lack of a transient energy storage compound during the dark period. Our study provides new perspectives on the role of starch and the adaptation to LD cycles of oleaginous microalgae.

In Error: Reference source not found a similar experimental approach was taken, where samples were taken for biochemical and transcriptome analysis every 3 hours from a turbidostat culture operated at the same diurnal cycle of 16 hours light and 8 hours dark, but this time in nitrogen limiting conditions, resulting in TAG accumulation. Again this was done for the wild type and the starchless mutant. The transcriptional landscape and biochemical data are compared to the nitrogen replete condition in Error: Reference source not found, to evaluate the effect of nitrogen limitation in general and study how the lack of starch is affecting TAG accumulation under nitrogen limiting conditions. We observed that the cycling diurnal effect is greatly reduced in comparison to nitrogen replete condition. The wild-type accumulated more starch than in nitrogen replete condition (Error: Reference source not found), and small oscillation was observed, indicating that it is being used as transient energy storage during the dark period. While the biochemical analysis did not reveal any oscillation in total lipids content in either strain, *slm1* over-expresses transcripts associated to TAG and lipid degradation during the dark period. Besides, while *slm1* accumulated more TAG than the wild-type, its conversion efficiency was only half of the wild-type. It appears also that the organism recycles more proteins during the dark period to supply nitrogen for the strong increase in amino acid synthesis right after light is turned on.

In Error: Reference source not found, the results of this thesis are discussed. The interest in analyzing the transcriptome of microalgae is explained. Then, the annotation methods are explained to show the large improvements between them, and could it still be improved. Basic numbers of the later annotation results are compared to the current status in UniProtKB. Particular points from the transcriptomics data are discussed, notably the expression of the mutant gene responsible for *slm1* phenotype, and the interests from using single-cell

technologies. Suggestions are then made to improve the experimental conditions and the photobioreactors setups. The efforts made in the thesis to generate and store the data according to the FAIR principles are explained. Finally, using the knowledge acquired during this thesis, suggestions are made to improve the growth conditions and to improve TAG production with divers metabolic engineering strategies. The work of this thesis contributes for the future of sustainable production of biofuels, which ultimately will help alleviating human society's dependence on fossil fuels.

Citations

1. Metting FB. Biodiversity and application of microalgae. *J Ind Microbiol Biotechnol.* 1996;17:477–89. doi:10.1007/BF01574779.
2. Falkowski P, Raven J. Aquatic photosynthesis. 2013. https://books.google.com/books?hl=en&lr=&id=kUCrAQAQAQBAJ&oi=fnd&pg=PP1&ots=bp-tfbSq_A&sig=aUeYBM0e1dA-xTD-iUFuetci7-o. Accessed 24 Feb 2019.
3. Chapman RL. Algae: the world's most important “plants”—an introduction. *Mitig Adapt Strateg Glob Chang.* 2013;18:5–12. doi:10.1007/s11027-010-9255-9.
4. Wijffels RH, Barbosa MJ, Oswald WJ, Golueke CG, Usui N, Ikenouchi M, et al. An outlook on microalgal biofuels. *Science.* 2010;329:796–9. doi:10.1126/science.1189003.
5. Stephens E, Ross IL, Mussgnug JH, Wagner LD, Borowitzka MA, Posten C, et al. Future prospects of microalgal biofuel production systems. *Trends Plant Sci.* 2010;15:554–64. doi:10.1016/J.TPLANTS.2010.06.003.
6. Jones CS, Mayfield SP. Algae biofuels: versatility for the future of bioenergy. *Curr Opin Biotechnol.* 2012;23:346–51. doi:10.1016/J.COPBIO.2011.10.013.
7. Schenk PM, Thomas-Hall SR, Stephens E, Marx UC, Mussgnug JH, Posten C, et al. Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Res.* 2008;1:20–43. doi:10.1007/s12155-008-9008-8.
8. Redalje D, Duerr E, Wastewater J de la N-..., 1989 undefined. Algae as ideal waste removers: biochemical. [books.google.com. https://books.google.nl/books?hl=en&lr=&id=63qBH_rzJWEC&oi=fnd&pg=PA91&dq=microalgae+detoxification+waste+water&ots=ik7Zg_GgF3&sig=c5FaFVIsdekuAOT0P5OO1QFcKdE](https://books.google.nl/books?hl=en&lr=&id=63qBH_rzJWEC&oi=fnd&pg=PA91&dq=microalgae+detoxification+waste+water&ots=ik7Zg_GgF3&sig=c5FaFVIsdekuAOT0P5OO1QFcKdE). Accessed 24 Feb 2019.
9. Norsker N-HH, Barbosa MJ, Vermuë MH, Wijffels RH. Microalgal production — A close look at the economics. *Biotechnol Adv.* 2011;29:24–7. doi:10.1016/j.biotechadv.2010.08.005.
10. Specht EA, Karunanithi PS, Gimpel JA, Ansari WS, Mayfield SP. Host Organisms: Algae. In: *Industrial Biotechnology*. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2016. p. 605–41. doi:10.1002/9783527807796.ch16.
11. Yusuf C, Chisti Y, Yusuf C. Biodiesel from microalgae. *Biotechnol Adv.* 2007;25:294–306. doi:10.1016/j.biotechadv.2007.02.001.
12. Benemann JR, Tillett DM, Weissman JC. Microalgae biotechnology. *Trends Biotechnol.* 1987;5:47–53. doi:10.1016/0167-7799(87)90037-0.
13. Breuer G, Lamers PP, Martens DE, Draaisma RB, Wijffels RH. The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresour Technol.* 2012;124:217–26. doi:10.1016/j.biortech.2012.08.003.
14. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, et al. Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. *Plant J.* 2008;54:621–39. doi:10.1111/j.1365-313X.2008.03492.x.
15. Tornabene TG, Holzer G, Lien S, Burris N. Lipid composition of the nitrogen starved green alga *Neochloris oleoabundans*. *Enzyme Microb Technol.* 1983;5:435–40. doi:10.1016/0141-0229(83)90026-1.
16. Ruiz J, Olivieri G, de Vree J, Bosma R, Willems P, Reith JH, et al. Towards industrial products from microalgae. *Energy Environ Sci.* 2016;9:3036–43. doi:10.1039/C6EE01493C.
17. Adarme-Vega T, Lim DKY, Timmins M, Vernen F, Li Y, Schenk PM. Microalgal biofactories: a promising approach towards sustainable omega-3 fatty acid production. *Microb Cell Fact.* 2012;11:96. doi:10.1186/1475-2859-11-96.

-
18. Lohr M, Schwender J, Polle JEW. Isoprenoid biosynthesis in eukaryotic phototrophs: A spotlight on algae. *Plant Sci.* 2012;185–186:9–22. doi:10.1016/J.PLANTSCI.2011.07.018.
 19. Hempel F, Bozarth AS, Lindenkamp N, Klingl A, Zauner S, Linne U, et al. Microalgae as bioreactors for bioplastic production. *Microb Cell Fact.* 2011;10:81. doi:10.1186/1475-2859-10-81.
 20. Hempel F, Lau J, Klingl A, Maier UG. Algae as Protein Factories: Expression of a Human Antibody and the Respective Antigen in the Diatom *Phaeodactylum tricornutum*. *PLoS One.* 2011;6:e28424. doi:10.1371/journal.pone.0028424.
 21. Bosma R, de Vree JH, Slegers PM, Janssen M, Wijffels RH, Barbosa MJ. Design and construction of the microalgal pilot facility AlgaePARC. *Algal Res.* 2014;6:160–9. doi:10.1016/J.ALGAL.2014.10.006.
 22. Slegers PM, Boxtel AJB van, Straten G van, Barbosa MJ, Wijffels RH. AlgaePARC and models for large scale microalgal production. 2012. <https://library.wur.nl/WebQuery/wurpubs/430362>. Accessed 25 Feb 2019.
 23. Ogbonna JC, Tanaka H. Night biomass loss and changes in biochemical composition of cells during light/dark cyclic culture of *Chlorella pyrenoidosa*. *J Ferment Bioeng.* 1996;82:558–64. doi:10.1016/S0922-338X(97)81252-4.
 24. León-Saiki GM, Remmers IM, Martens DE, Lamers PP, Wijffels RH, van der Veen D. The role of starch as transient energy buffer in synchronized microalgal growth in *Acutodesmus obliquus*. *Algal Res.* 2017;25:160–7.
 25. Juergens MT, Disbrow B, Shachar-Hill Y. The Relationship of Triacylglycerol and Starch Accumulation to Carbon and Energy Flows during Nutrient Deprivation in *Chlamydomonas reinhardtii*. *Plant Physiol.* 2016;171:2445–57. doi:10.1104/pp.16.00761.
 26. Benvenuti G, Bosma R, Ji F, Lamers P, Barbosa MJ, Wijffels RH. Batch and semi-continuous microalgal TAG production in lab-scale and outdoor photobioreactors. *J Appl Phycol.* 2016;28:3167–77. doi:10.1007/s10811-016-0897-1.
 27. Klok AJ, Martens DE, Wijffels RH, Lamers PP. Simultaneous growth and neutral lipid accumulation in microalgae. *Bioresour Technol.* 2013;134:233–43.
 28. Remmers IM, Hidalgo-Ulloa A, Brandt BP, Evers WAC, Wijffels RH, Lamers PP. Continuous versus batch production of lipids in the microalgae *Acutodesmus obliquus*. *Bioresour Technol.* 2017;244:1384–92.
 29. Pittman JK, Dean AP, Osundeko O. The potential of sustainable algal biofuel production using wastewater resources. *Bioresour Technol.* 2011;102:17–25. doi:10.1016/j.biortech.2010.06.035.
 30. Christenson L, Sims R. Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts. *Biotechnol Adv.* 2011;29:686–702. doi:10.1016/J.BIOTECHADV.2011.05.015.
 31. Bhattacharjee M, Siemann E. Low algal diversity systems are a promising method for biodiesel production in wastewater fed open reactors. *ALGAE.* 2015;30:67–79. doi:10.4490/algae.2015.30.1.067.
 32. Osundeko O, Davies H, Pittman JK. Oxidative stress-tolerant microalgae strains are highly efficient for biofuel feedstock production on wastewater. *Biomass and Bioenergy.* 2013;56:284–94. doi:10.1016/j.biombioe.2013.05.027.
 33. de Jaeger L, Verbeek RE, Draaisma RB, Martens DE, Springer J, Eggink G, et al. Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*: (I) mutant generation and characterization. *Biotechnol Biofuels.* 2014;7:69. doi:10.1186/1754-6834-7-69.
 34. Breuer G, de Jaeger L, Artus VGPG, Martens DE, Springer J, Draaisma RB, et al. Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*: (II) evaluation of TAG yield and productivity in controlled photobioreactors. *Biotechnol Biofuels.* 2014;7:1–11. doi:10.1186/1754-6834-7-70.
 35. de Jaeger L. Strain improvement of oleaginous microalgae. Wageningen University; 2015.
 36. Winter G, Krömer JO. Fluxomics - connecting 'omics analysis and phenotypes. *Environ Microbiol.* 2013;15:1901–16. doi:10.1111/1462-2920.12064.
-

-
37. Tardif M, Atteia A, Specht M, Cogne G, Rolland N, Brugière S, et al. Predalgo: A new subcellular localization prediction tool dedicated to green algae. *Mol Biol Evol.* 2012;29:3625–39. doi:10.1093/molbev/mss178.
38. J.M.F. Reijnders M, M. Carreres B, J. Schaap P. Algal Omics: The Functional Annotation Challenge. *Curr Biotechnol.* 2015;4:457–63. doi:10.2174/2211550105666151203202455.
39. Koehorst JJ, van Dam JCJ, Saccenti E, Martins dos Santos VAP, Suarez-Diez M, Schaap PJ. SAPP: functional genome annotation and analysis through a semantic framework using FAIR principles. *Bioinformatics.* 2017. doi:10.1093/bioinformatics/btx767.
40. Chisti Y. Constraints to commercialization of algal fuels. *Journal of Biotechnology.* 2013.
41. Bordbar A, Monk JM, King Z a, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014;15:107–20. doi:10.1038/nrg3643.
42. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and Molecular Evolution of the Green Algae. *CRC Crit Rev Plant Sci.* 2012;31:1–46. doi:10.1080/07352689.2011.615705.
43. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421. doi:10.1186/1471-2105-10-421.
44. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40.
45. Harris EH. *Chlamydomonas* as a model organism. *Annu Rev Plant Physiol Plant Mol Biol.* 2001;52:363–406. doi:10.1146/annurev.arplant.52.1.363.
46. Scaife MA, Nguyen GTDT, Rico J, Lambert D, Helliwell KE, Smith AG. Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host. *Plant J.* 2015;82:532–46. doi:10.1111/tpj.12781.
47. Reijnders MJMF, van Heck RGA, Lam CMC, Scaife MA, Santos VAPM dos, Smith AG, et al. Green genes: bioinformatics and systems-biology innovations drive algal biotechnology. *Trends Biotechnol.* 2014;32:617–26. doi:10.1016/J.TIBTECH.2014.10.003.
48. Grossman AR, Harris EE, Hauser C, Lefebvre PA, Martinez D, Rokhsar D, et al. *Chlamydomonas reinhardtii* at the crossroads of genomics. *Eukaryot Cell.* 2003;2:1137–50. doi:10.1128/ec.2.6.1137-1150.2003.
49. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz J, Witman GB, et al. The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science (80-).* 2007;318:245–50. doi:10.1126/science.1143609.The.
50. UniProt Consortium. UniProt: a hub for protein information. *Nucleic Acids Res.* 2015;43:D204–12. doi:10.1093/nar/gku989.
51. Gao C, Wang Y, Shen Y, Yan D, He X, Dai J, et al. Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics.* 2014;15:582. doi:10.1186/1471-2164-15-582.
52. Moreau H, Verhelst B, Couloux A, Derelle E, Rombauts S, Grimsley N, et al. Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biol.* 2012;13:R74. doi:10.1186/gb-2012-13-8-r74.
53. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, et al. The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *Plant Cell.* 2010;22:2943–55. doi:10.1105/tpc.110.076406.
54. Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, et al. The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol.* 2012;13:R39. doi:10.1186/gb-2012-13-5-r39.
-

-
55. Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, et al. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science*. 2009;324:268–72. doi:10.1126/science.1167222.
56. Corteggiani Carpinelli E, Telatin A, Vitulo N, Forcato C, D'Angelo M, Schiavon R, et al. Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion. *Mol Plant*. 2014;7:323–35. doi:10.1093/mp/sst120.
57. Blanc-Mathieu R, Verhelst B, Derelle E, Rombauts S, Bouget F-Y, Carré I, et al. An improved genome of the model marine alga *Ostreococcus tauri* unfolds by assessing Illumina de novo assemblies. *BMC Genomics*. 2014;15:1103. doi:10.1186/1471-2164-15-1103.
58. Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, et al. The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proc Natl Acad Sci U S A*. 2007;104:7705–10. doi:10.1073/pnas.0611046104.
59. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, et al. Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science*. 2010;329:223–6. doi:10.1126/science.1188800.
60. Radivojac P, Clark WT, Oron TR, Schnoes AM, Wittkop T, Sokolov A, et al. A large-scale evaluation of computational protein function prediction. *Nat Methods*. 2013;10:221–7. doi:10.1038/nmeth.2340.
61. Falda M, Toppo S, Pescarolo A, Lavezzo E, Di Camillo B, Facchinetti A, et al. Argot2: A large scale function prediction tool relying on semantic similarity of weighted Gene Ontology terms. *BMC Bioinformatics*. 2012;13 SUPPL4:S14. doi:10.1186/1471-2105-13-S4-S14.
62. Minneci F, Piovesan D, Cozzetto D, Jones DT. FFPred 2.0: improved homology-independent prediction of gene ontology terms for eukaryotic protein sequences. *PLoS One*. 2013;8:e63754. doi:10.1371/journal.pone.0063754.
63. Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH. UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*. 2007;23:1282–8. doi:10.1093/bioinformatics/btm098.
64. Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, et al. Pfam: the protein families database. *Nucleic Acids Res*. 2014;42:D222–30. doi:10.1093/nar/gkt1223.
65. Eddy SR. Profile hidden Markov models. *Bioinformatics*. 1998;14:755–63. doi:10.1093/bioinformatics/14.9.755.
66. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res*. 2014;42:W252–8. doi:10.1093/nar/gku340.
67. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*. 2010;5:93–121. doi:10.1038/nprot.2009.203.
68. Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA, Henry C. Automated Genome Annotation and Metabolic Model Reconstruction in the SEED and Model SEED. 2013. p. 17–45. doi:10.1007/978-1-62703-299-5_2.
69. Karp PD, Paley SM, Krummenacker M, Latendresse M, Dale JM, Lee TJ, et al. Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform*. 2010;11:40–79. doi:10.1093/bib/bbp043.
70. Millar AH, Taylor NL. Subcellular proteomics-where cell biology meets protein chemistry. *Front Plant Sci*. 2014;5:55. doi:10.3389/fpls.2014.00055.
71. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting Subcellular Localization of Proteins Based on their N-terminal Amino Acid Sequence. *J Mol Biol*. 2000;300:1005–16. doi:10.1006/jmbi.2000.3903.
72. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol*. 2004;340:783–95. doi:10.1016/j.jmb.2004.05.028.
-

-
73. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. 2010;26:1608–15. doi:10.1093/bioinformatics/btq249.
74. Barsanti L, Gualtieri P. *Algae: anatomy, biochemistry, and biotechnology*. <https://www.crcpress.com/Algae-Anatomy-Biochemistry-and-Biotechnology-Second-Edition/Barsanti-Gualtieri/p/book/9781439867327>. Accessed 5 Aug 2019.
75. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2015;43:D30–5. doi:10.1093/nar/gku1216.
76. Huntley RP, Sawford T, Mutowo-Meullenet P, Shypitsyna A, Bonilla C, Martin MJ, et al. The GOA database: gene Ontology annotation updates for 2015. *Nucleic Acids Res*. 2015;43 Database issue:D1057–63. doi:10.1093/nar/gku1113.
77. Chibucos MC, Mungall CJ, Balakrishnan R, Christie KR, Huntley RP, White O, et al. Standardized description of scientific evidence using the Evidence Ontology (ECO). *Database (Oxford)*. 2014;2014:bau075–bau075. doi:10.1093/database/bau075.
78. Schaid DJ, Sinnwell JP, Jenkins GD, McDonnell SK, Ingle JN, Kubo M, et al. Using the gene ontology to scan multilevel gene sets for associations in genome wide association studies. *Genet Epidemiol*. 2012;36:3–16. doi:10.1002/gepi.20632.
79. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21:3674–6. doi:10.1093/bioinformatics/bti610.
80. Worm P, Koehorst JJ, Visser M, Sedano-Núñez VT, Schaap PJ, Plugge CM, et al. A genomic view on syntrophic versus non-syntrophic lifestyle in anaerobic fatty acid degrading communities. *Biochim Biophys Acta*. 2014;1837:2004–16. doi:10.1016/j.bbabi.2014.06.005.
81. Elliott LG, Feehan C, Laurens LML, Pienkos PT, Darzins A, Posewitz MC. Establishment of a bioenergy-focused microalgal culture collection. *Algal Res*. 2012;1:102–13. doi:10.1016/J.ALGAL.2012.05.002.
82. Lamers PP, Janssen M, De Vos RCH, Bino RJ, Wijffels RH. Carotenoid and fatty acid metabolism in nitrogen-starved *Dunaliella salina*, a unicellular green microalga. *J Biotechnol*. 2012;162:21–7.
83. Ledford HK, Niyogi KK. Singlet oxygen and photo-oxidative stress management in plants and algae. *Plant, Cell Environ*. 2005;28:1037–45.
84. Li Y, Han D, Yoon K, Zhu S, Sommerfeld M, Hu Q. Molecular and Cellular Mechanisms for Lipid Synthesis and Accumulation in Microalgae: Biotechnological Implications. *Handb Microalgal Cult Appl Phycol Biotechnol*. 2013;:545–65.
85. Frada MJ, Burrows EH, Wyman KD, Falkowski PG. Quantum requirements for growth and fatty acid biosynthesis in the marine diatom *Phaeodactylum tricornutum* (Bacillariophyceae) in nitrogen replete and limited conditions. *J Phycol*. 2013;49:381–8.
86. Müller P, Li XP, Niyogi KK. Non-photochemical quenching. A response to excess light energy. *Plant Physiol*. 2001;125:1558–66.
87. Kolber Z, Zehr J, Falkowski P. Effects of Growth Irradiance and Nitrogen Limitation on Photosynthetic Energy Conversion in Photosystem II. *Plant Physiol*. 1988;88:923–9.
88. Klok AJ, Verbaanderd J a., Lamers PP, Martens DE, Rinzema A, Wijffels RH. A model for customising biomass composition in continuous microalgae production. *Bioresour Technol*. 2013;146:89–100.
89. Dellomonaco C, Fava F, Gonzalez R. The path to next generation biofuels: successes and challenges in the era of synthetic biology. *Microb Cell Fact*. 2010;9:3. doi:10.1186/1475-2859-9-3.
-

-
90. Kliphuis AMJ, Klok AJ, Martens DE, Lamers PP, Janssen M, Wijffels RH. Metabolic modeling of *Chlamydomonas reinhardtii*: energy requirements for photoautotrophic growth and maintenance. *J Appl Phycol.* 2012;24:253–66. doi:10.1007/s10811-011-9674-3.
 91. Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain enhanced lookup time accelerated BLAST. *Biol Direct.* 2012;7:12. doi:10.1186/1745-6150-7-12.
 92. Conesa A, Götz S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008;2008.
 93. Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F, Ton JKH, et al. Colorimetric Method for Determination of Sugars and Related Substances. *Anal Chem.* 1956;28:350–6. doi:10.1021/ac60111a017.
 94. Herbert D, Phipps PJ, Strange RE. Chemical analysis of microbial cells In: *Methods in Microbiology*, JR Norris, DW Ribbons,(eds.) Vol. 5B. 1971.
 95. Weesepeel Y, Vincken JP, Pop RM, Liu K, Gruppen H. Sodiation as a tool for enhancing the diagnostic value of MALDI-TOF/TOF-MS spectra of complex astaxanthin ester mixtures from *Haematococcus pluvialis*. *J Mass Spectrom.* 2013;48:862–74.
 96. Dortch Q, Roberts T, Clayton J, Ahmed S. RNA/DNA ratios and DNA concentrations as indicators of growth rate and biomass in planktonic marine organisms . *Mar Ecol Prog Ser.* 1983;13:61–71.
 97. Stephanopoulos G, Aristidou AA, Nielsen J. *Metabolic engineering: principles and methodologies.* Academic press; 1998.
 98. Van Der Heijden RTJM, Heijnen JJ, Hellinga C, Romein B, Luyben KC a M. Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnol Bioeng.* 1994;43:3–10.
 99. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUASt: Quality assessment tool for genome assemblies. *Bioinformatics.* 2013;29:1072–5.
 100. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc.* 2007;2:953–71.
 101. Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* 2007;35 SUPPL.2:W585–7.
 102. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson D a, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52. doi:10.1038/nbt.1883.
 103. Henschel R, Nista PM, Lieber M, Haas BJ, Wu L-S, LeDuc RD. Trinity RNA-Seq assembler performance optimization. In: *Proceedings of the 1st Conference of the Extreme Science and Engineering Discovery Environment on Bridging from the eXtreme to the campus and beyond - XSEDE '12.* ACM; 2012. p. 8. doi:10.1145/2335755.2335842.
 104. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2009;26:139–40.
 105. Maere S, Heymans K, Kuiper M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics.* 2005;21:3448–9.
 106. Klok AJ, Lamers PP, Martens DE, Draaisma RB, Wijffels RH. Edible oils from microalgae: Insights in TAG accumulation. *Trends in Biotechnology.* 2014.
 107. Caspeta L, Nielsen J. Economic and environmental impacts of microbial biodiesel. *Nat Biotechnol.* 2013;31:789–93. doi:10.1038/nbt.2683.

108. Haider S, Pal R. Integrated analysis of transcriptomic and proteomic data. *Curr Genomics*. 2013;14:91–110. doi:10.2174/1389202911314020003.
109. Yoon K, Han D, Li Y, Sommerfeld M, Hu Q. Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and degradation while synthesizing triacylglycerol in the unicellular green microalga *Chlamydomonas reinhardtii*. *Plant Cell*. 2012;24:3708–24. doi:10.1105/tpc.112.100701.
110. Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, et al. Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J Biol Chem*. 2012;287:15811–25. doi:10.1074/jbc.M111.334052.
111. Gargouri M, Park J-J, Holguin FO, Kim M-J, Wang H, Deshpande RR, et al. Identification of regulatory network hubs that control lipid metabolism in *Chlamydomonas reinhardtii*. *J Exp Bot*. 2015;66:4551–66. doi:10.1093/jxb/erv217.
112. Kajikawa M, Sawaragi Y, Shinkawa H, Yamano T, Ando A, Kato M, et al. Algal Dual-Specificity Tyrosine Phosphorylation-Regulated Kinase, Triacylglycerol Accumulation Regulator1, Regulates Accumulation of Triacylglycerol in Nitrogen or Sulfur Deficiency. *Plant Physiol*. 2015;168:752–64. doi:10.1104/pp.15.00319.
113. Maruyama S, Tokutsu R, Minagawa J. Transcriptional Regulation of the Stress-Responsive Light Harvesting Complex Genes in *Chlamydomonas Reinhardtii*. *Plant Cell Physiol*. 2014;55:1304–10. doi:10.1093/pcp/pcu068.
114. Wobbe L, Schwarz C, Nickelsen J, Kruse O. Translational control of photosynthetic gene expression in phototrophic eukaryotes. *Physiol Plant*. 2008;133:507–15. doi:10.1111/j.1399-3054.2008.01091.x.
115. Durnford DG, Price JA, McKim SM, Sarchfield ML. Light-harvesting complex gene expression is controlled by both transcriptional and post-transcriptional mechanisms during photoacclimation in *Chlamydomonas reinhardtii*. *Physiol Plant*. 2003;118:193–205. doi:10.1034/j.1399-3054.2003.00078.x.
116. Matile P, Hörtensteiner S, Thomas H. Chlorophyll Degradation. *Annu Rev Plant Physiol Plant Mol Biol*. 1999;50:67–95.
117. Neidhardt J, Benemann JR, Zhang L, Melis A. Photosystem-II repair and chloroplast recovery from irradiance stress: Relationship between chronic photoinhibition, light-harvesting chlorophyll antenna size and photosynthetic productivity in *Dunaliella salina* (green algae). *Photosynth Res*. 1998;56:175–84.
118. Ballottari M, Girardon J, Dall L, Bassi R. Biochimica et Biophysica Acta Evolution and functional properties of Photosystem II light harvesting complexes in eukaryotes ☆. *BBA - Bioenerg*. 2012;1817:143–57. doi:10.1016/j.bbabi.2011.06.005.
119. Vavilin D, Vermaas W. Continuous chlorophyll degradation accompanied by chlorophyllide and phytol reutilization for chlorophyll synthesis in *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta - Bioenerg*. 2007;1767:920–9. doi:10.1016/J.BBABI.2007.03.010.
120. Schoeder H-U, Lockau W. Phylloquinone copurifies with the large subunit of photosystem I. *FEBS Lett*. 1986;199:23–7. doi:10.1016/0014-5793(86)81216-9.
121. Li Z, Keasling JD, Niyogi KK. Overlapping Photoprotective Function of Vitamin E and Carotenoids in *Chlamydomonas*. *Plant Physiol*. 2012;158:313–23.
122. Havaux M, Lütz C, Grimm B. Chloroplast membrane photostability in chlP transgenic tobacco plants deficient in tocopherols. *Plant Physiol*. 2003;132:300–10.
123. Niyogi KK, Niyogi KK. PHOTOPROTECTION REVISITED: Genetic and Molecular Approaches. *Annu Rev Plant Physiol Plant Mol Biol*. 1999;50:333–59. doi:10.1146/annurev.arplant.50.1.333.
124. Munné-Bosch S, Alegre L. The Function of Tocopherols and Tocotrienols in Plants. *CRC Crit Rev Plant Sci*. 2002;21:31–57. doi:10.1080/0735-260291044179.

-
125. Tanaka R, Oster U, Kruse E, Rüdiger W, Grimm B. Reduced Activity of Geranylgeranyl Reductase Leads to Loss of Chlorophyll and Tocopherol and to Partially Geranylgeranylated Chlorophyll in Transgenic Tobacco Plants Expressing Antisense RNA for Geranylgeranyl Reductase1. *Plant Physiol.* 1999;120:695–704.
126. Shpilyov A V., Zinchenko V V., Shestakov S V., Grimm B, Lokstein H. Inactivation of the geranylgeranyl reductase (ChlP) gene in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta - Bioenerg.* 2005;1706:195–203. doi:10.1016/j.bbabi.2004.11.001.
127. Rüdiger W. Chlorophyll metabolism: From outer space down to the molecular level. *Phytochemistry.* 1997;46:1151–67. doi:10.1016/S0031-9422(97)80003-9.
128. Rathmann R, Szklo A, energy RS-R, 2010 undefined. Land use competition for production of food and liquid biofuels: An analysis of the arguments in the current debate. Elsevier. <https://www.sciencedirect.com/science/article/pii/S0960148109000974>. Accessed 5 Aug 2019.
129. Fraiture C, Giordano M, policy YL-W, 2008 undefined. Biofuels and implications for agricultural water use: blue impacts of green energy. *iwaponline.com*. <https://iwaponline.com/wp/article-abstract/10/S1/67/19607>. Accessed 5 Aug 2019.
130. Tilman D, Socolow R, Foley J, Hill J, ... EL-, 2009 undefined. Beneficial biofuels—the food, energy, and environment trilemma. *science.sciencemag.org*. <https://science.sciencemag.org/content/325/5938/270.short>. Accessed 5 Aug 2019.
131. Fargione J, Hill J, Tilman D, Science SP-, 2008 undefined. Land clearing and the biofuel carbon debt. *science.sciencemag.org*. <https://science.sciencemag.org/content/319/5867/1235>. Accessed 5 Aug 2019.
132. Greenway H, Munns R. Mechanisms of Salt Tolerance in Nonhalophytes. *Annu Rev Plant Physiol.* 1980;31:149–90. doi:10.1146/annurev.pp.31.060180.001053.
133. Al-Karaki GN. Growth, water use efficiency, and sodium and potassium acquisition by tomato cultivars grown under salt stress. *J Plant Nutr.* 2000;23:1–8. doi:10.1080/01904160009381992.
134. Moya J, ... AG-J of, 2003 undefined. Chloride absorption in salt-sensitive Carrizo citrange and salt-tolerant Cleopatra mandarin citrus rootstocks is linked to water use. *academic.oup.com*. <https://academic.oup.com/jxb/article-abstract/54/383/825/545764>. Accessed 5 Aug 2019.
135. Hauser F, Plant TH-, environment cell &, 2010 undefined. A conserved primary salt tolerance mechanism mediated by HKT transporters: a mechanism for sodium exclusion and maintenance of high K⁺/Na⁺ ratio in leaves. *Wiley Online Libr.* <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3040.2009.02056.x>. Accessed 5 Aug 2019.
136. Ashraf M, Foolad M. Roles of glycinebetaine and proline in improving plant abiotic stress resistance. *Environ Exp Bot.* 2007;59:206–11. doi:http://dx.doi.org/10.1016/j.envexpbot.2005.12.006.
137. Streeter JG, Lohnes DG, Fioritto RJ. Patterns of pinitol accumulation in soybean plants and relationships to drought tolerance. *Plant, Cell Environ.* 2001;24:429–38. doi:10.1046/j.1365-3040.2001.00690.x.
138. Hoekstra FA, Golovina EA, Buitink J. Mechanisms of plant desiccation tolerance. *Trends Plant Sci.* 2001;6:431–8. doi:10.1016/S1360-1385(01)02052-0.
139. Desplats P, Folco E, Biochemistry GS-PP and, 2005 undefined. Sucrose may play an additional role to that of an osmolyte in *Synechocystis* sp. PCC 6803 salt-shocked cells. Elsevier. <https://www.sciencedirect.com/science/article/pii/S0981942805000173>. Accessed 5 Aug 2019.
140. Mackay M, Norton R, Biology LB-M, 1983 undefined. Marine blue-green algae have a unique osmoregulatory system. Springer. <https://link.springer.com/article/10.1007/BF00392256>. Accessed 5 Aug 2019.
141. Reviews KW-F microbiology, 1986 undefined. Osmoregulation in eukaryotic algae. *academic.oup.com*. <https://academic.oup.com/femsre/article-abstract/2/1-2/37/535409>. Accessed 5 Aug 2019.
-

142. Flowers TJ, Troke PF, Yeo AR. The Mechanism of Salt Tolerance in Halophytes. *Annu Rev Plant Physiol.* 1977;28:89–121. doi:10.1146/annurev.pp.28.060177.000513.
143. Hellebusi JA. Osmoregulation. *Annu Rev Plant Physiol.* 1976;27:485–505. doi:10.1146/annurev.pp.27.060176.002413.
144. Blumwald E, Microbiology ET-O-A of, 1982 undefined. Osmoregulation and cell composition in salt-adaptation of *Nostoc muscorum*. Springer. <https://link.springer.com/article/10.1007/BF00508725>. Accessed 5 Aug 2019.
145. Ahmad I, physiology JH-P, 1984 undefined. Osmoregulation in the extremely euryhaline marine micro-alga *Chlorella autotrophica*. *Am Soc Plant Biol.* <http://www.plantphysiol.org/content/74/4/1010.short>. Accessed 5 Aug 2019.
146. HELLEBUST JA, LIN Y-HY ???H. Regulation of glycerol and starch metabolism in *Chlamydomonas pulsatilla* in response to changes in salinity. *Plant Cell Environ.* 1989;12:621–7. doi:10.1111/j.1365-3040.1989.tb01230.x.
147. Band, C.J., Arredondo-Vega, B.O., Vazquez-Duhalt, R., Greppin H, BAND CJ, ARREDONDO-VEGA BO, VAZQUEZ-DUHALT R, GREPPIN H. Effect of a salt-osmotic upshock on the edaphic microalga *Neochloris oleoabundans*. *Plant Cell Environ.* 1992;15:129–33. doi:10.1111/j.1365-3040.1992.tb01466.x.
148. Mastrobuoni G, Irgang S, Pietzke M, Aßmus HE, Wenzel M, Schulze WX, et al. Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*. *BMC Genomics.* 2012;13:215. doi:10.1186/1471-2164-13-215.
149. Santos A, Janssen M, Lamers P, ... WE-B, 2012 undefined. Growth of oil accumulating microalga *Neochloris oleoabundans* under alkaline–saline conditions. Elsevier. <https://www.sciencedirect.com/science/article/pii/S0960852411015720>. Accessed 5 Aug 2019.
150. Gouveia L, Marques AE, da Silva TL, Reis A. *Neochloris oleabundans* UTEX #1185: A suitable renewable lipid source for biofuel production. *J Ind Microbiol Biotechnol.* 2009;36:821–6. doi:10.1007/s10295-009-0559-2.
151. Griffiths MJ, Harrison STL. Lipid productivity as a key characteristic for choosing algal species for biodiesel production. *J Appl Phycol.* 2009;21:493–507. doi:10.1007/s10811-008-9392-7.
152. Chantanachat S, Bold HC. Some algae from arid soils. University of Texas Publications (6218); 1962.
153. Breuer G, Lamers PPP, Martens DE DE, ... RD-B, 2013 undefined, Draaisma RB, et al. Effect of light intensity, pH, and temperature on triacylglycerol (TAG) accumulation induced by nitrogen starvation in *Scenedesmus obliquus*. *Bioresour Technol.* 2013;143:1–9. doi:10.1016/j.biortech.2013.05.105.
154. Breuer G, Evers WAC, de Vree JH, Kleinegriss DMM, Martens DE, Wijffels RH, et al. Analysis of Fatty Acid Content and Composition in Microalgae. *J Vis Exp.* 2013;80 September:e50628. doi:10.3791/50628.
155. Kim, H. K.; Choi, Y. H.; Verpoorte R. NMR-based metabolomic analysis of plants. *Nat Protoc.* 2010;5 MARCH 2010:536–49. doi:10.1038/nprot.2009.237.
156. Kim HK, Saifullah, Khan S, Wilson EG, Kricun SDP, Meissner A, et al. Metabolic classification of South American *Ilex* species by NMR-based metabolomics. *Phytochemistry.* 2010;71:773–84. doi:10.1016/j.phytochem.2010.02.001.
157. Peng Y, Leung HCM, Yiu SM, Chin FYL. IDBA-UD: A de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics.* 2012;28:1420–8.
158. Azachi M. Salt Induction of Fatty Acid Elongase and Membrane Lipid Modifications in the Extreme Halotolerant Alga *Dunaliella salina*. *Plant Physiol.* 2002;129:1320–9. doi:10.1104/pp.001909.

-
159. Rismani-Yazdi H, Hsin C, Peccia J, Article I, Central P, Central B, et al. Transcriptomic analysis of the oleaginous microalga *Neochloris oleoabundans* reveals metabolic insights into triacylglyceride accumulation. *Biotechnol Biofuels*. 2012;5:74. doi:10.1186/1754-6834-5-74.
160. Morales-Sánchez D, Kyndt J, Ogden K, research AM-A, 2016 undefined. Toward an understanding of lipid and starch accumulation in microalgae: A proteomic study of *Neochloris oleoabundans* cultivated under N-limited heterotrophic. Elsevier. <https://www.sciencedirect.com/science/article/pii/S2211926416303332>. Accessed 5 Aug 2019.
161. Bonhert MJ JR. Strategies for engineering water stress tolerance in plants. *Trends Biotechnol*. 1996;14:89–97. <http://www.scopus.com/inward/record.url?eid=2-s2.0-0029996302&partnerID=40&md5=42e440171df55a02999ff9fc5ef1bfff>.
162. Lunn JE. Evolution of sucrose synthesis. *Plant Physiol*. 2002;128:1490–500. doi:10.1104/pp.010898.
163. Curatti L, Porchia AC, Herrera-Estrella L, Salerno GL. A prokaryotic sucrose synthase gene (*susA*) isolated from a filamentous nitrogen-fixing cyanobacterium encodes a protein similar to those of plants. *Planta*. 2000;211:729–35. doi:10.1007/s004250000343.
164. Singer MA, Lindquist S. Thermotolerance in *Saccharomyces cerevisiae*: The Yin and Yang of trehalose. *Trends Biotechnol*. 1998;16:460–8. doi:10.1016/S0167-7799(98)01251-7.
165. Page-Sharp M, Behm CA, Smith GD. Involvement of the compatible solutes trehalose and sucrose in the response to salt stress of a cyanobacterial *Scytonema* species isolated from desert soils. *Biochim Biophys Acta - Gen Subj*. 1999;1472:519–28. doi:10.1016/S0304-4165(99)00155-5.
166. Bremauntz M del P, Torres-Bustillos LG, Cañizares-Villanueva R-O, Duran-Paramo E, Fernández-Linares L. Trehalose and Sucrose Osmolytes Accumulated by Algae as Potential Raw Material for Bioethanol. *Nat Resour*. 2011;02:173–9. doi:10.4236/nr.2011.23023.
167. Csonka LN, Gelvin SBS, Goodner BW, Orser CS, Siemieniak D, Slightom JL, et al. Nucleotide sequence of a mutation in the *proB* gene of *Escherichia coli* that confers proline overproduction and enhanced tolerance to osmotic stress. *Gene*. 1988;64:199–205. doi:10.1016/0378-1119(88)90335-6.
168. Mallick N. Copper-induced oxidative stress in the chlorophycean microalga *Chlorella vulgaris*: response of the antioxidant system. *J Plant Physiol*. 2004;161:591–7. doi:10.1078/0176-1617-01230.
169. Zhang LP, Mehta SK, Physiology ZL-P and C, 2008 undefined, Liu ZP, Yang ZM. Copper-Induced Proline Synthesis is Associated with Nitric Oxide Generation.pdf. *Plant Cell Physiol*. 2008;49:411–9. <http://www.scopus.com/inward/record.url?eid=2-s2.0-40849133965&partnerID=40&md5=3e240dbd25ed34a9f60f900c411d5443>. Accessed 5 Aug 2019.
170. Mehta SKS, Gaur JPJ. heavy-metal-induced proline accumulation and its role in ameliorating metal toxicity in *Chlorella vulgaris*. *New Phytol*. 1999;143:253–259. doi:10.1046/j.1469-8137.1999.00447.x.
171. Sharma P, Shanker Dubey R. Modulation of nitrate reductase activity in rice seedlings under aluminium toxicity and water stress: Role of osmolytes as enzyme protectant. *J Plant Physiol*. 2005;162:854–64. doi:10.1016/j.jplph.2004.09.011.
172. Krell A, Funck D, Plettner I, John U, Dieckmann G. Regulation of proline metabolism under salt stress in the psychrophilic diatom *Fragilariopsis cylindrus* (Bacillariophyceae). *J Phycol*. 2007;43:753–62. doi:10.1111/j.1529-8817.2007.00366.x.
173. Smirnoff N. The role of active oxygen in the response of plants to water deficit and desiccation. *New Phytol*. 1993;125:27–58. doi:10.1111/j.1469-8137.1993.tb03863.x.
174. Zhang Y-MM, Chen H, He C-LL, Wang Q. Nitrogen Starvation Induced Oxidative Stress in an Oil-Producing Green Alga *Chlorella sorokiniana* C3. *PLoS One*. 2013;8:e69225. doi:10.1371/journal.pone.0069225.
-

175. biology NS-C opinion in plant, 2000 undefined. Ascorbic acid: metabolism and functions of a multi-faceted molecule. Elsevier. <https://www.sciencedirect.com/science/article/pii/S1369526600800709>. Accessed 5 Aug 2019.
176. Marrs KA. the Functions and Regulation of Glutathione S-Transferases in Plants. *Annu Rev Plant Physiol Plant Mol Biol.* 1996;47:127–58. doi:10.1146/annurev.arplant.47.1.127.
177. Sappl PG, Carroll AJ, Clifton R, Lister R, Whelan J, Harvey Millar A, et al. The Arabidopsis glutathione transferase gene family displays complex stress regulation and co-silencing multiple genes results in altered metabolic sensitivity to oxidative stress. *Plant J.* 2009;58:53–68. doi:10.1111/j.1365-313X.2008.03761.x.
178. Howe G, Merchant S. Heavy Metal-Activated Synthesis of Peptides in *Chlamydomonas reinhardtii*. *Plant Physiol.* 1992;98:127–36. doi:10.1104/pp.98.1.127.
179. Hu S, Lau KWK, Wu M. Cadmium sequestration in *Chlamydomonas reinhardtii*. *Plant Sci.* 2001;161:987–96. doi:10.1016/S0168-9452(01)00501-5.
180. Chaurasia N, Mishra Y, Rai LC, Research LR-B and B, 2008 undefined. Cloning expression and analysis of phytochelatin synthase (pcs) gene from *Anabaena* sp. PCC 7120 offering multiple stress tolerance in *Escherichia coli*. *Biochem Biophys Res Commun.* 2008;376:225–30. doi:10.1016/j.bbrc.2008.08.129.
181. Zhang H, Xu W, Guo J, He Z, Science MM-P, 2005 undefined, et al. Coordinated responses of phytochelatin and metallothioneins to heavy metals in garlic seedlings. *Plant Sci.* 2005;169:1059–65. doi:10.1016/j.plantsci.2005.07.010.
182. Seki M, Ishida J, Narusaka M, Fujita M, Nanjo T, Umezawa T, et al. Monitoring the expression pattern of around 7 , 000 Arabidopsis genes under ABA treatments using a full-length cDNA microarray. *Funct Integr Genomics.* 2002;2 I : 6:282–91. doi:10.1007/s10142-002-0070-6.
183. de Winter L, Klok AJ, Cuaresma Franco M, Barbosa MJ, Wijffels RH, Winter L De, et al. The synchronized cell cycle of *Neochloris oleoabundans* and its influence on biomass composition under constant light conditions. *Algal Res.* 2013;2:313–20. doi:10.1016/j.algal.2013.09.001.
184. Jia J, Han D, Gerken HG, Li Y, Sommerfeld M, Hu Q, et al. Molecular mechanisms for photosynthetic carbon partitioning into storage neutral lipids in *Nannochloropsis oceanica* under nitrogen-depletion conditions. *Algal Res.* 2015;7:66–77. doi:10.1016/j.algal.2014.11.005.
185. Rangasamy D, Ratledge C. Compartmentation of ATP:citrate lyase in plants. *Plant Physiol.* 2000;122:1225–30. doi:10.1104/pp.122.4.1225.
186. Roesler K, Shintani D, Savage L, ... SB-P, 1997 undefined, Boddupalli S, et al. Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. *Plant Physiol.* 1997;113:75–81. doi:10.1104/pp.113.1.75.
187. Dunahay TG, Jarvis EE, Dais SS, Roessler PG. Manipulation of Microalgal Lipid Production Using Genetic Engineering. *Appl Biochem Biotechnol.* 1996;57:223–31. doi:10.1007/BF02941703.
188. Nikolau B, Ohlrogge J, and EW-A of biochemistry, 2003 undefined. Plant biotin-containing carboxylases. Elsevier. <https://www.sciencedirect.com/science/article/pii/S0003986103001565>. Accessed 5 Aug 2019.
189. Ohlrogge JB, Jaworski JG. Regulation of fatty acid synthesis. *Annu Rev Plant Physiol Plant Mol Biol.* 1997;48:109–36. doi:10.1146/annurev.arplant.48.1.109.
190. Vigeolas H, Waldeck P, Zank T, Geigenberger P. Increasing seed oil content in oil-seed rape (*Brassica napus* L.) by over-expression of a yeast glycerol-3-phosphate dehydrogenase under the control of a seed-specific promoter. *Plant Biotechnol J.* 2007;5:431–41. doi:10.1111/j.1467-7652.2007.00252.x.
191. Chen M, Thelen JJ. The Plastid Isoform of Triose Phosphate Isomerase Is Required for the Postgerminative Transition from Heterotrophic to Autotrophic Growth in Arabidopsis. *Plant Cell.* 2010;22:77–90. doi:10.1105/tpc.109.071837.

-
192. Deng XD, Gu B, Li YJ, Hu XW, Guo JC, Fei XW. The roles of acyl-CoA: Diacylglycerol acyltransferase 2 genes in the biosynthesis of triacylglycerols by the green algae *Chlamydomonas reinhardtii*. *Mol Plant*. 2012;5:945–7. doi:10.1093/mp/sss040.
193. Hsieh HJ, Su CH, Chien LJ. Accumulation of lipid production in *Chlorella minutissima* by triacylglycerol biosynthesis-related genes cloned from *Saccharomyces cerevisiae* and *Yarrowia lipolytica*. *J Microbiol*. 2012;50:526–34. doi:10.1007/s12275-012-2041-5.
194. La Russa M, Bogen C, Uhmeyer A, Doebbe A, Filippone E, Kruse O, et al. Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. *J Biotechnol*. 2012;162:13–20. doi:10.1016/j.jbiotec.2012.04.006.
195. Niu YF, Zhang MH, Li DW, Yang WD, Liu JS, Bai W Bin, et al. Improvement of neutral lipid and polyunsaturated fatty acid biosynthesis by overexpressing a type 2 diacylglycerol acyltransferase in marine diatom *Phaeodactylum tricornutum*. *Mar Drugs*. 2013;11:4558–69. doi:10.3390/md11114558.
196. Li J, Han D, Wang D, Ning K, Jia J, Wei L, et al. Choreography of Transcriptomes and Lipidomes of *Nannochloropsis* Reveals the Mechanisms of Oil Synthesis in Microalgae. *Plant Cell*. 2014;26:1645–65. doi:10.1105/tpc.113.121418.
197. Wang D, Ning K, Li J, Hu J, Han D, Wang H, et al. *Nannochloropsis* Genomes Reveal Evolution of Microalgal Oleaginous Traits. *PLoS Genet*. 2014;10:e1004094. doi:10.1371/journal.pgen.1004094.
198. Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, et al. Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc Natl Acad Sci U S A*. 2013;110:1–6. doi:10.1073/pnas.1309299110.
199. Krienitz L, Bock C. Present state of the systematics of planktonic coccoid green algae of inland waters. *Hydrobiologia*. 2012;698:295–326.
200. Dzhambazov B, Belkinova D, Mladenov R. Sind *Scenedesmus obtusiusculus* CHOD. und *Scenedesmus obliquus* (TURP.) KÜTZ. (Chlorophyta, Chlorococcales) verschiedene Arten? *Arch für Hydrobiol Suppl Algal Stud*. 144:141–50. <http://cat.inist.fr/?aModele=afficheN&cpsidt=13869751>. Accessed 27 Sep 2016.
201. Breuer G, Martens DE, Draaisma RB, Wijffels RH, Lamers PP. Photosynthetic efficiency and carbon partitioning in nitrogen-starved *Scenedesmus obliquus*. *Algal Res*. 2015;9:254–62.
202. Breuer G, Lamers P, Janssen M, ... RW-B, 2015 undefined. Opportunities to improve the areal oil productivity of microalgae. Elsevier. <https://www.sciencedirect.com/science/article/pii/S0960852415004174>. Accessed 20 Jan 2019.
203. Dayarian A, Michael TP, Sengupta AM. SOPRA: Scaffolding algorithm for paired reads via statistical optimization. *BMC Bioinformatics*. 2010;11:345.
204. Hunt M, Newbold C, Berriman M, Otto TD. A comprehensive evaluation of assembly scaffolding tools. *Genome Biol*. 2014;15:R42.
205. Boetzer M, Henkel C V., Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics*. 2011;27:578–9.
206. Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. Contiguous and accurate *de novo* assembly of metazoan genomes with modest long read coverage. 2015.
207. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. *Genome Biol*. 2012;13:R56. doi:10.1186/gb-2012-13-6-r56.
208. de Cambiaire J-C, Otis C, Lemieux C, Turmel M. The complete chloroplast genome sequence of the chlorophycean green alga *Scenedesmus obliquus* reveals a compact gene organization and a biased distribution of genes on the two DNA strands. *BMC Evol Biol*. 2006;6:37. doi:10.1186/1471-2148-6-37.
-

-
209. Guarnieri MT, Pienkos PT. Algal omics: unlocking bioproduct diversity in algae cell factories. *Photosynthesis research*. 2015;123:255–63.
210. Wijffels RH, Barbosa MJ, Eppink MHM. Microalgae for the production of bulk chemicals and biofuels. *Biofuels, Bioproducts and Biorefining*. 2010;4:287–95.
211. Blanken W, Cuaresma M, Wijffels RH, Janssen M. Cultivation of microalgae on artificial light comes at a cost. *Algal Research*. 2013;2:333–40.
212. Farré EM. The regulation of plant growth by the circadian clock. *Plant Biology*. 2012;14:401–10.
213. McClung CR. Plant Circadian Rhythms. *PLANT CELL ONLINE*. 2006;18:792–803. doi:10.1105/tpc.106.040980.
214. León-Saiki GM, Martini TC, van der Veen D, Wijffels RH, Martens DE. The impact of day length on cell division and efficiency of light use in a starchless mutant of *Tetrademus obliquus*. *Algal Res*. 2018;31:387–94.
215. Nikaido SS, Johnson CH. Daily and circadian variation in survival from ultraviolet radiation in *Chlamydomonas reinhardtii*. *Photochem Photobiol*. 2000;71:758–65.
216. Suzuki L, Johnson CH. Algae know the time of day: Circadian and photoperiodic programs. *Journal of Phycology*. 2001;37:933–42.
217. Smith SM, Fulton DC, Chia T, Thorneycroft D, Chapple A, Dunstan H, et al. Diurnal Changes in the Transcriptome Encoding Enzymes of Starch Metabolism Provide Evidence for Both Transcriptional and Posttranscriptional Regulation of Starch Metabolism in *Arabidopsis* Leaves 1. *Plant Physiol*. 2004.
218. Smith SR, Gillard JTF, Kustka AB, McCrow JP, Badger JH, Zheng H, et al. Transcriptional Orchestration of the Global Cellular Response of a Model Pennate Diatom to Diel Light Cycling under Iron Limitation. *PLoS Genet*. 2016.
219. Chauton MS, Winge P, Brembu T, Vadstein O, Bones AM. Gene Regulation of Carbon Fixation, Storage, and Utilization in the Diatom *Phaeodactylum tricornutum* Acclimated to Light/Dark Cycles1[C][W][OA]. *Plant Physiol*. 2013;161:1034–48.
220. Poliner E, Panchy N, Newton L, Wu G, Lapinsky A, Bullard B, et al. Transcriptional coordination of physiological responses in *Nannochloropsis oceanica* CCMP1779 under light/dark cycles. *Plant J*. 2015;83:1097–113.
221. Zones JM, Blaby IK, Merchant SS, Umen JG. High-Resolution Profiling of a Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous Cell and Metabolic Differentiation. *American Society of Plant Biologists*; 2015. doi:10.1105/tpc.15.00498.
222. Suzuki S, Ishida KI, Hirakawa Y. Diurnal transcriptional regulation of endosymbiotically derived genes in the chlorarachniophyte *bigeloviella natans*. *Genome Biol Evol*. 2016;8:2672–82.
223. Fujiwara T, Misumi O, Tashiro K, Yoshida Y, Nishida K, Yagisawa F, et al. Periodic gene expression patterns during the highly synchronized cell nucleus and organelle division cycles in the unicellular red alga *cyanidioschyzon merolae*. *DNA Res*. 2009;16:59–72.
224. Wynne MJ, Hallan JK. Reinstatement of *Tetrademus* G. M. Smith (Sphaeropleales, Chlorophyta). *Feddes Repert*. 2016.
225. Carreres BM, de Jaeger L, Springer J, Barbosa MJ, Breuer G, van den End EJ, et al. Draft Genome Sequence of the Oleaginous Green Alga *Tetrademus obliquus* UTEX 393. *Genome Announc*. 2017;5:e01449-16. doi:10.1128/genomeA.01449-16.
226. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2015;32:767–9.
-

-
227. Dam JCJ Van, Koehorst JJ, Vik JO, Schaap PJ, Suarez-diez M. Interoperable genome annotation with GBOL , an extendable infrastructure for functional data mining. *bioRxiv*. 2017;:1–9.
228. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;3:160018. doi:10.1038/sdata.2016.18.
229. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
230. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol*. 2010;28:511–5.
231. Nguyen N-N, Srihari S, Leong HW, Chong K-F. EnzDP: Improved enzyme annotation for metabolic network reconstruction based on domain composition profiles. *J Bioinform Comput Biol*. 2015;13:1543003. doi:10.1142/S0219720015430039.
232. Nueda MJ, Tarazona S, Conesa A. Next maSigPro: Updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics*. 2014;30:2598–602.
233. Robinson M, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol*. 2010;11:R25. doi:10.1186/gb-2010-11-3-r25.
234. Äijö T, Butty V, Chen Z, Salo V, Tripathi S, Burge CB, et al. Methods for time series analysis of RNA-seq data with application to human Th17 cell differentiation. *Bioinformatics*. 2014;30.
235. Hennig C. fpc: Flexible Procedures for Clustering. 2018. <https://cran.r-project.org/package=fpc>.
236. R Core Team. R: A Language and Environment for Statistical Computing. 2017. <https://www.r-project.org/>.
237. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28:27–30. <http://www.ncbi.nlm.nih.gov/pubmed/10592173>. Accessed 7 Mar 2018.
238. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016;44:D457–62. doi:10.1093/nar/gkv1070.
239. Carlson M, Falcon S, Pages H, Li N. A set of annotation maps describing the entire Gene Ontology. 2007. <ftp://ctan.uib.no/pub/bioconductor/2.7/data/annotation/html/GO.db.html>. Accessed 7 Mar 2018.
240. Christ B, Süßenbacher I, Moser S, Bichsel N, Egert A, Müller T, et al. Cytochrome P450 CYP89A9 is involved in the formation of major chlorophyll catabolites during leaf senescence in Arabidopsis. *Plant Cell*. 2013;25:1868–80. doi:10.1105/tpc.113.112151.
241. Oda-Yamamizo C, Mitsuda N, Sakamoto S, Ogawa D, Ohme-Takagi M, Ohmiya A. The NAC transcription factor ANAC046 is a positive regulator of chlorophyll degradation and senescence in Arabidopsis leaves. *Sci Rep*. 2016;6:23609. doi:10.1038/srep23609.
242. Christ B, Hörtensteiner S. Mechanism and Significance of Chlorophyll Breakdown. *J Plant Growth Regul*. 2014;33:4–20. doi:10.1007/s00344-013-9392-y.
243. Mittag M. Circadian rhythms in microalgae. Academic Press; 2001. doi:10.1016/S0074-7696(01)06023-5.
244. Causton HC, Feeney KA, Ziegler CA, O’Neill JS. Metabolic cycles in yeast share features conserved among circadian rhythms. *Curr Biol*. 2015;25:1056–62.
245. Stitt M, Zeeman SC. Starch turnover: Pathways, regulation and role in growth. *Current Opinion in Plant Biology*. 2012;15:282–92.
246. Bišová K, Zachleder V. Cell-cycle regulation in green algae dividing by multiple fission. *J Exp Bot*. 2014;65:2585–602. doi:10.1093/jxb/ert466.
-

247. Zachleder V, Schläfli O, Boschetti A. Growth-controlled oscillation in activity of histone H1 kinase during the cell cycle of *Chlamydomonas reinhardtii* (chlorophyta). *J Phycol.* 1997;33:673–81. doi:10.1111/j.0022-3646.1997.00673.x.
248. Izumo A, Fujiwara S, Sakurai T, Ball SG, Ishii Y, Ono H, et al. Effects of granule-bound starch synthase I-defective mutation on the morphology and structure of pyrenoidal starch in *Chlamydomonas*. *Plant Sci.* 2011.
249. Jin S-H, Jiang D-A, Li X-Q, Sun J-W. Characteristics of photosynthesis in rice plants transformed with an antisense Rubisco activase gene. *J Zhejiang Univ Sci.* 2004;5:897–9. <http://www.ncbi.nlm.nih.gov/pubmed/15236471>. Accessed 25 Jun 2018.
250. Portis AR, Jr. Rubisco activase – Rubisco’s catalytic chaperone. *Photosynth Res.* 2003;75:11–27. doi:10.1023/A:1022458108678.
251. Lopez-Ruiz a, Verbelen JP, Roldan JM, Diez J. Nitrate reductase of green algae is located in the pyrenoid. *Plant Physiol.* 1985;79:1006–10. doi:10.1104/pp.79.4.1006.
252. Afify AEMMR, El Baroty GS, El Baz FK, Abd El Baky HH, Murad SA. *Scenedesmus obliquus*: Antioxidant and antiviral activity of proteins hydrolyzed by three enzymes. *Journal of Genetic Engineering and Biotechnology.* 2018.
253. Villarejo A, Martinez F, Pino Plumed M, Ramazanov Z. The induction of the CO₂ concentrating mechanism in a starch-less mutant of *Chlamydomonas reinhardtii*. *Physiol Plant.* 1996;98:798–802. doi:10.1111/j.1399-3054.1996.tb06687.x.
254. Hellebust JA, Ahmad I. Regulation of Nitrogen Assimilation in Green Microalgae. *Biol Oceanogr.* 1989.
255. Mallick N, Rai LC, Mohn FH, Soeder CJ. Studies on nitric oxide (NO) formation by the green alga *Scenedesmus obliquus* and the diazotrophic cyanobacterium *Anabaena Doliolum*. *Chemosphere.* 1999;39:1601–10. doi:10.1016/S0045-6535(99)00058-2.
256. Cullimore J V., Sims AP. Pathway of ammonia assimilation in illuminated and darkened *Chlamydomonas reinhardtii*. *Phytochemistry.* 1981.
257. de Winter L, Klok AJ, Cuaresma Franco M, Barbosa MJ, Wijffels RH. The synchronized cell cycle of *Neochloris oleoabundans* and its influence on biomass composition under constant light conditions. *Algal Res.* 2013;2:313–20.
258. Breuer G, Lamers PP, Martens DE, Draaisma RB, Wijffels RH. Effect of light intensity, pH, and temperature on triacylglycerol (TAG) accumulation induced by nitrogen starvation in *Scenedesmus obliquus*. *Bioresour Technol.* 2013;143:1–9.
259. Ho SH, Chen CY, Chang JS. Effect of light intensity and nitrogen starvation on CO₂ fixation and lipid/carbohydrate production of an indigenous microalga *Scenedesmus obliquus* CNW-N. *Bioresour Technol.* 2012.
260. Mandal S, Mallick N. Microalga *Scenedesmus obliquus* as a potential source for biodiesel production. *Appl Microbiol Biotechnol.* 2009.
261. De Jaeger L, Verbeek REM, Draaisma RB, Martens DE, Springer J, Eggink G, et al. Superior triacylglycerol (TAG) accumulation in starchless mutants of *Scenedesmus obliquus*: (I) mutant generation and characterization. *Biotechnol Biofuels.* 2014;7:69. doi:10.1186/1754-6834-7-69.
262. Carreres BM, León-Saiki GM, Schaap PJ, Remmers IM, van der Veen D, Martins dos Santos VAP, et al. The diurnal transcriptional landscape of the microalga *Tetrademus obliquus*. *Algal Res.* 2019;40:101477. doi:10.1016/J.ALGAL.2019.101477.
263. Mulders KJM, Lamers PP, Wijffels RH, Martens DE. Dynamics of biomass composition and growth during recovery of nitrogen-starved *Chromochloris zofingiensis*. *Appl Microbiol Biotechnol.* 2014.

-
264. DuBois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric Method for Determination of Sugars and Related Substances. *Anal Chem.* 1956;28:350–6. doi:10.1021/ac60111a017.
265. Postma P, Miron T, Olivieri G, ... MB-B, 2015 undefined. Mild disintegration of the green microalgae *Chlorella vulgaris* using bead milling. Elsevier. <https://www.sciencedirect.com/science/article/pii/S096085241401284X>. Accessed 20 Jan 2019.
266. Hennig C. fpc: Flexible Procedures for Clustering. 2018. <https://cran.r-project.org/package=fpc>.
267. Breuer G, Lamers PP, Martens DE, Draaisma RB, Wijffels RH. The impact of nitrogen starvation on the dynamics of triacylglycerol accumulation in nine microalgae strains. *Bioresour Technol.* 2012;124:217–26.
268. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, et al. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J.* 2008;54:621–39. doi:10.1111/j.1365-313X.2008.03492.x.
269. Fan J, Yan C, Andre C, ... JS-P and C, 2012 undefined. Oil accumulation is controlled by carbon precursor supply for fatty acid synthesis in *Chlamydomonas reinhardtii*. *academic.oup.com*. <https://academic.oup.com/pcp/article-abstract/53/8/1380/1827959>. Accessed 20 Jan 2019.
270. Berg JM, Tymoczko JL SL. Triacylglycerols are highly concentrated energy stores. In: *Biochemistry.* 2002.
271. Lancelot C, Mathot S. Biochemical fractionation of primary production by phytoplankton in Belgian coastal waters during short- and long-term incubations with ¹⁴C-bicarbonate - I. Mixed diatom population. *Mar Biol.* 1985.
272. Johnson X, cell JA-E, 2013 undefined. Central carbon metabolism and electron transport in *Chlamydomonas reinhardtii*, metabolic constraints for carbon partitioning between oil and starch. *Am Soc Microbiol.* <http://ec.asm.org/content/early/2013/03/26/EC.00318-12.short>. Accessed 20 Jan 2019.
273. Li Y, Han D, Hu G, Sommerfeld M, Hu Q. Inhibition of starch synthesis results in overproduction of lipids in *Chlamydomonas reinhardtii*. *Biotechnol Bioeng.* 2010;107:258–68. doi:10.1002/bit.22807.
274. Baker KE, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. *Curr Opin Cell Biol.* 2004;16:293–9. doi:10.1016/J.CEB.2004.03.003.
275. Maquat LE. Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol.* 2004;5:89–99. doi:10.1038/nrm1310.
276. Bagnato C, Prados MB, Franchini GR, Scaglia N, Miranda SE, Beligni M V. Analysis of triglyceride synthesis unveils a green algal soluble diacylglycerol acyltransferase and provides clues to potential enzymatic components of the chloroplast pathway. *BMC Genomics.* 2017.
277. Yoon K, Han D, Li Y, Sommerfeld M, Hu Q. Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and degradation while synthesizing triacylglycerol in the unicellular green microalga *Chlamydomonas reinhardtii*. *Plant Cell.* 2012;24:3708–24. doi:10.1105/tpc.112.100701.
278. Kong F, Romero IT, Warakanont J, Li-Beisson Y. Lipid catabolism in microalgae. *New Phytol.* 2018;218:1340–8. doi:10.1111/nph.15047.
279. Rylott EL, Hooks MA, Graham IA. Co-ordinate regulation of genes involved in storage lipid mobilization in *Arabidopsis thaliana*. *Biochem Soc Trans.* 2001;29 Pt 2:283–7. <http://www.ncbi.nlm.nih.gov/pubmed/11356168>. Accessed 3 Jan 2019.
280. Richter CD, Allen JWA, Higham CW, Koppenhofer A, Zajicek RS, Watmough NJ, et al. Cytochrome cd1, reductive activation and kinetic analysis of a multifunctional respiratory enzyme. *J Biol Chem.* 2002;277:3093–100. doi:10.1074/jbc.M108944200.
281. Cooke RJ. Protein degradation in plants. *Science Progress (1933-).* 1981;67:461–80. doi:10.2307/43420541.
-

-
282. Geider RJ, Osborne B a. Respiration and microalgal growth: a review of the quantitative relationship between dark respiration and growth. *New Phytol.* 1989.
283. Visser R DE, SPITTERS It C, BOUMA T, Spitters C, Molecular TB-, And B, et al. ENERGY COST OF PROTEIN TURNOVER: THEORETICAL CALCULATION AND EXPERIMENTAL ESTIMATION FROM REGRESSION OF RESPIRATION ON PROTEIN CONCENTRATION OF FULL-GROWN LEAVES. 1992;:493–508. <https://library.wur.nl/WebQuery/wurpubs/fulltext/216567>. Accessed 23 Aug 2018.
284. Penning De Vries FWT. The cost of maintenance processes in plant cells. *Ann Bot.* 1975.
285. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet.* 2012;13:227–32. doi:10.1038/nrg3185.
286. Coin L, Bateman A, Durbin R. Enhanced protein domain discovery using taxonomy. *BMC Bioinformatics.* 2004;5:56. doi:10.1186/1471-2105-5-56.
287. Dutilh BE, Jurgelenaite R, Szklarczyk R, van Hijum SAFT, Harhangi HR, Schmid M, et al. FACIL: Fast and Accurate Genetic Code Inference and Logo. *Bioinformatics.* 2011;27:1929–33. doi:10.1093/bioinformatics/btr316.
288. Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G. The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. *Genome Res.* 2000;10:819–31. doi:10.1101/gr.10.6.819.
289. Noutahi E, Calderon V, Blanchette M, El-Mabrouk N, Lang BF. Rapid Genetic Code Evolution in Green Algal Mitochondrial Genomes. *Mol Biol Evol.* 2019;36:766–83. doi:10.1093/molbev/msz016.
290. Tsimring LS. Noise in Biology. *Rep Prog Phys.* 2014;77:026601. doi:10.1088/0034-4885/77/2/026601.
291. Lüring M. The smell of water: grazer-induced colony formation in *Scenedesmus*. 1999. <https://library.wur.nl/WebQuery/wurpubs/62264>. Accessed 25 May 2019.
292. Pickett-Heaps JD, Staehelin LA. The ultrastructure of *scenedesmus* (chlorophyceae). II. cell division and colony formation. *J Phycol.* 1975;11:186–202. doi:10.1111/j.1529-8817.1975.tb02766.x.
293. Trainor FP, Burg CA. *Scenedesmus obliquus* Sexuality. *Science.* 1965;148:1094–5. doi:10.1126/science.148.3673.1094.
294. Trainor FR. Reproduction in *Scenedesmus*. *ALGAE.* 1996;11:183–201. doi:10.1086/331338.
295. Edmundson SJ, Huesemann MH. The dark side of algae cultivation: Characterizing night biomass loss in three photosynthetic algae, *Chlorella sorokiniana*, *Nannochloropsis salina* and *Picochlorum* sp. *Algal Res.* 2015;12:470–6. doi:10.1016/J.ALGAL.2015.10.012.
296. Lüttge U, Beck F. Endogenous rhythms and chaos in crassulacean acid metabolism. *Planta.* 1992;188:28–38. doi:10.1007/BF00198936.
297. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, et al. A survey of best practices for RNA-seq data analysis. *Genome Biol.* 2016;17:13. doi:10.1186/s13059-016-0881-8.
298. Troein C, Corellou F, Dixon LE, van Ooijen G, O’Neill JS, Bouget F-Y, et al. Multiple light inputs to a simple clock circuit allow complex biological rhythms. *Plant J.* 2011;66:375–85. doi:10.1111/j.1365-313X.2011.04489.x.
299. Matsuo T, Okamoto K, Onai K, Niwa Y, Shimogawara K, Ishiura M. A systematic forward genetic analysis identified components of the *Chlamydomonas* circadian system. *Genes Dev.* 2008;22:918–30. doi:10.1101/gad.1650408.
300. Corellou F, Schwartz C, Motta J-P, Djouani-Tahri EB, Sanchez F, Bouget F-Y. Clocks in the green lineage: comparative functional analysis of the circadian architecture of the picoeukaryote *ostreococcus*. *Plant Cell.* 2009;21:3436–49. doi:10.1105/tpc.109.068825.
-

-
301. Linde A-M, Eklund DM, Kubota A, Pederson ERA, Holm K, Gyllenstrand N, et al. Early evolution of the land plant circadian clock. *New Phytol.* 2017;216:576–90. doi:10.1111/nph.14487.
302. Yang Y, Gao K. Effects of CO₂ concentrations on the freshwater microalgae, *Chlamydomonas reinhardtii*, *Chlorella pyrenoidosa* and *Scenedesmus obliquus* (Chlorophyta). *J Appl Phycol.* 2003;15:379–89. doi:10.1023/A:1026021021774.
303. Roenneberg T, Foster RG. Twilight times: light and the circadian system. *Photochem Photobiol.* 1997;66:549–61. <http://www.ncbi.nlm.nih.gov/pubmed/9383985>. Accessed 10 Jun 2019.
304. Roenneberg T, Hastings JW. Are the effects of light on phase and period of the *Gonyaulax* clock mediated by different pathways? *Photochem Photobiol.* 1991;53:525–33. doi:10.1111/j.1751-1097.1991.tb03665.x.
305. Roenneberg T, Hastings JW. Two photoreceptors control the circadian clock of a unicellular alga. *Naturwissenschaften.* 1988;75:206–7. doi:10.1007/BF00735584.
306. Winkler U, Saftel W, Stabenau H. beta-Oxidation of fatty acids in algae: Localization of thiolase and acyl-CoA oxidizing enzymes in three different organisms. *Planta.* 1988;175:91–8. doi:10.1007/BF00402885.

Acknowledgments

After more than 6 years, a lot of efforts, struggle, procrastination, and dedication, this thesis is finally completed. During this time, many things happened and it was a great adventure within a unique friendly environment. It is finally time to step back and proudly look at what was achieved and thank all the people that took part in it, even remotely.

It all started on the Friday 15h of March 2013, at the very same day my beautiful niece Mila was born. If I was superstitious, I could have thought that this is a sign of great achievements for the both of us.

First of all, Peter, thank you for giving me this opportunity, for supervising me and providing me the material and support to work on all these projects. You gave me the independence to work on the projects according to my own judgment, all the while kicking me back into the right path if needed.

Dirk, thank you for all the collaborations and many discussions. I particularly enjoyed our regular meetings as it would help keeping a certain efficiency. We both learned from each other, you helped me with the biological and bioprocessing aspects of these microalgae, and I taught you about the computational aspects. It was valuable that that you always try to understand every details so you could input your thoughts, and of course, it served as a form of quality control. Thank you for your dedicated help and attention on the manuscripts that became articles.

Maria, many thanks to you! While you were not my supervisor, you are by far the person I bothered the most. I slowly started to request help from you about R programming and statistics, until you said (approximately) “I do not mind helping you, but my contribution should be rewarded by authorship”. This was quickly settled and that gave me a “all you can ask” free pass. I enjoyed collaborating with you, and that’s probably thanks to your patience, cheerful words and great wisdom. Many times you and Dirk were advising me to refocus to the main objectives, I often couldn’t help fulfilling my need to look for too much details. You showed me how simple writing a manuscript can be (I still had to practice a lot:), and you also taught me how to find the quickest way to deal with problems in the manuscript. Again, thank you for all the help that you provided me out of simple altruism.

I would like to thank both of my promoters Vitor and René. I am glad your collaboration resulted in the creation of the position that was given to me. While you were more distantly taking part in my thesis, I am very thankful for your scientific input and more importantly for your confidence to fund the large sequencing project that resulted in the chapters 6 and 7.

Thanks to my collaborators, without whom this thesis would not have been possible, namely Anne, Lenny, Maarten, Mitsue and Ilse.

While I enjoyed to overall atmosphere of SSB and MIB, the warmest place (literally and figuratively) was my office, thanks to my office mates. Javi (aka bastardo, cabron, pu*illa, etc) and Kal (aka the longest name in the world), we spent so much time together that I don't think I have to tell you how much I enjoyed your company and how much that made the office life unique. I have one regret, though we had few trio trips, I missed all the ones you guys did before I joined the office. Special thanks to Kal for bartending the office! Kal, you are a smart guy and I admired your combativity to fight against what would seem to be fate. Never give up! You tend to be proud of your weirdnesses and like to joke about it (quite funny though), I think you have entered some sort of a vicious loop or inception of weirdness, find the exit! Javi, you are the king of bash and that's probably for a good reason. It's like your English speaking, people think they can understand it until you show off. Took me a while but I progressively but the decoder in my brain.

Carolyn, I am sorry that we might have been too noisy at times. While you could appear shy, you were actually very opened for conversation and you were always ready to help. It was funny that sometimes you tried to argue some logic in some senseless joke chats we had with the guys. It was great having you in the office! Then Carolyn left, and Niels joined us in the adventure. Niels, you fitted well in the office, though you worked a lot from home in the beginning and we had to share you with Jasper and Jesse. We did not see each other much outside of the office, but you chose the hard path of building a family all the while working on your thesis, and I admire you for that (same goes to Jesse!). Then we moved to the Helix building and my worst nightmare came to life... I shared the office with Maarten! Actually it was not that bad and he was very quite in the office. That is unless Peter came and they had an improvised meeting literally behind my back. It was actually much worse to work (and shout) with him on the chapter 2. We had more fun at lunch times and in the outside world. For some reason, we always end up drunk WWE fighting, even in the snow and shirtless... I moved you to Lausanne to make sure you were far from Wageningen, but coincidence made that I ended up coming to the same place. Now that I will work on the campus next to yours, we'll have to hang out more often, and not just three times in a year. You may have matured but you are more boring now :). Take good care of Baillou's offspring Zaza. I think I've seen Javi and Kal more than that. Last office, Tjerko. You finished your thesis too fast (congrats for that) and you were there too rarely. It was fun having you and your vicious smile in the office, and thanks for the help in bothering Maarten. Then there was all the many temporary students from Rob, too many weirdos but also nice people.

Because I was too often alone in the Helix office, I was often heading to the office next door for random brakes with Bart (aka Bartardo, el barto, etc), Nikolas, Niels, Kal (for a short while) and Bastie. These poor students that did not dare to tell us we were laughing too loud and they could not concentrate ^^ . Nikolas, ex-brother in PC, it was great having a casual gamer in this world of scientist nerds that never play. Even if it hurts me to say that, I have to admit you are definitely not casual gamer unlike me. Seriously, who else is crazy enough to redo the same game levels over and over again to end up loosing everything because it's only fun playing in hardcore mode... Many great memories with you, but it's too bad you were too busy playing games instead of joining parties :P. Barto, it was great having you around, especially for the computer nerdy stuff, but unlike Nikolas, more hardware oriented. Together countless number of coffees were dunk, countless number of times the server were either fixed or setup (also with Jasper, thanks for dragging me into that). Bastie, you are irreplaceable, and together with Maarten, you were a funny pair of weirdos. You changed a lot over the years. On my first day, you seemed so depressed, and later you transformed into a party animal (never missing one), gym animal, dance animal, etc. You even started adding whiskey to you usual plain cola drinks! Such a drastic change in the right direction. Tough your German redneck rudeness might scare people, you are a too kind person and it's no surprise everyone appreciated you around.

Jasper and Jesse, it was great helping you working on the ontologies and on the transcriptomics part of SAPP. We should write the application note for that... It's a pity I had to keep this work as extra. Not only it's a cool technology, it was also fun working with you. Even though you were both more distant friends, particularly I enjoyed your company, especially when it was surrounding computer nerd stuff. Special thanks to Jesse, it is partly because of you that I am writing these words now. When we met at the SWAT4LS, Pierre-Yves "shouted" that I was looking for a PhD position, you mentioned that one position was available in SSB and you gave my CV to Peter.

Milad (aka milady, milado), we had so much fun time together! Just like many people in SSB, you were one special kind of weirdo. We were like two stupid children, always playing, fighting, joking, and talking some random and unusual stuff. At lunch, while I always had great food made by my back then other half, you were sneaking like a raccoon to steal my food! I would still invite you all the time for diner. I moved you so many times, I even collected (you gave me of course) few things from you each time. You knew how to praise the good-stuff to get some more. You were gone for a while now, and these moments are already missed.

Angela, you were here for a rather short period of time, but it was definitely during the best of my PhD. That's probably because of all the good times we had with Milad, Teresita, Alicia and the rest of the Spanish community (and others of course).

Teresita, thank you for all these moments we share together. You were one of the most important person of SSB/MIB that helped in creating these great years. That's also thanks to the Spanish community that was revolving around (Javi, Juanan, Irene, Alicia, and probably many more I am forgetting, sorry). Thanks for letting us all have BBQs at your place. I hope we get another chance to travel somewhere again like we did in Andalusia with Javi, Gerben, and my back-then other half.

Ruben and Dorett, it was great having you around, especially for chats and cheese based diners. Dorett, thank you so so so much for all the mini cheese fondue you brought back from Switzerland especially for me. I am glad I could bring it to work for lunch and share the smell with everyone, most certainly my biggest fan Rob. Ruben, SSB would not have been a great place without you either, especially during the coffee breaks. It was so much fun watching you tricking Maarten with your wordplay. I am soon done with this, maybe we can resume our A team of Rocket League, together with Nikolas and maybe Barto.

Talking about Rob, the post-doc that we often forgot he was. You arrived shortly after me and it was great having you around. Thank you for all the scientific nerdy chats, beers, rugby, and bbq here and there. You might not have realized, but at times, you were of a great help for this thesis. You could easily follow me when I discussed details, even though sometimes I was not being clear at all. I also enjoyed hearing about your research. Those discussion were probably the most scientific nerdy talks.

Catalina (aka, batman, you are such a fun and crazy person, keep up! We have been close to each other only starting from the later years, thanks to A common friend. It's funny to say, but you are the person to go when someone complain about whatever. Probably that's because of your great empathy.

Niru, I never understood why you had such a bad opinion of me. I hope I have proved you otherwise over the time! Still it was great chatting with you, whether it was random jokes or serious talks. Linde, thanks for your never ending smile that helps bringing a lot of good vibes around you. Thanks again for the stress balls, that ended up being thrown all over the office(s) with Maarten... Some were lost and may hopefully be found by the next generations of students sitting in my office (they are stuck above the hanging ceiling). If only you did not betrayed us (the dry SSBs) and moved downstairs to the wet SSBs. Erika, too bad you were not

so often around (at least at the same time as me) for the social activities, you are a fun girl and it's fun talking with you. By the way, Italian wine is too sour!

Sven, the legendary ex-PhD of MIB, friend of everyone, and now, the should-be-official photographer of PhD defenses of Wageningen University. Together with Sara, we had awesome moments for many different excuses that somewhat always involved drinking and often eating. You were probably at every parties in Wageningen, of course that was for "strictly" professional reasons. It was great meeting for the awesome diners with very special dishes made by you. You are the definition of German efficiency and punctuality. So much that even before it was time, you would start complaining that we were late. I am sorry for all these frequent 5-10 minute delays. I hope we keep contact over the coming years, especially for men's day. With Sara, take care of the little cutie Charlotte and don't forget the cat (sven, don't be mean with this poor animal).

Aleks, it was great time with you, you were a important part of the social mood. I am glad you found your other half. Johanna, you are my favorite hippie, it's was chilling with you. Jasper S, great time during the PhD trip, but we rarely hang out in wag. Monika, thank you for the Yoga lesson, it was awesome and I have a much better opinion of it now (Yuan now pushes me to practice sometimes). I hope we can find some time, maybe in spain with Juanan and Sven, so you can teach us kite surf. Irene, what can I say besides that you are a crazy bastarda. You were so much fun during these years, keep that up. Stamatios, it was great chatting with you. It's too bad you did not join more often the social activities, especially when you and Kal were living in the huge house.

There are many more people I would like to thank for these great years, and you all contributed to that. I honestly would like to spend more lines writing every single one of you, but I am running out of time and space. I am also certain I am shamelessly forgetting someone. If that's the case, come to me and I'll tell you personally :). Thanks to Mark Davids, Rienk, Peer, Lara, Emma, Rik, Edoardo, Nong, Nhung Marta, Tom, Carolien, Melanie, Jueeli, Cristina, Giulio (crazy dude).

I would also like to thank my friends from the almost religious Thursday drinking group, amateurs of fine Belgian beers. After I started this thesis, and probably because I wasn't heavily incited, I progressively gave up on leaving work at normal time to join you for a weekly beer, but time flies and my heart remained with all of you. Hopefully, memories get better over time, and these nasty little games of Heleen to torture me are forgiven :). Erik, the other Erik and his wife Petra (congrats guys!), Paul, Heleen, thank you guys for the good times at the Vlaamsche Reus. By the way, sorry for I have sinned, I fixed so many Null Pointer

Exception errors on Thursdays! No regret though, that was my way to revenge from Pierre-Yves not telling me the rules before.

Emilie Doutre, when we met in Wageningen, you were following my steps, doing the same bachelor, you even became my neighbor and did the same master as me. Just like me, you loved Wageningen and you came back multiple times, even living in bornsesteeg for a while. I kept great memories over these many years. I regret that after I started the PhD we meet less and less often, also because I was not taking the time to come to the Thursdays.

Rico, you were and will remain my b*tch. You are the only student are really supervised, and I had fun doing so. It's too bad it was one of these dead-on-arrival projects. If only they told us before! Also at that time I was a bit lost in what could be done and what was worth trying. That was a good training the hard way.

Special thanks to Pierre-Yves Chibon, thanks for everything you taught me about linux and programming. While I already knew a lot, you taught much more and more importantly you taught me the best practices. You are a real passionate and I am super happy and not so surprised that RedHat linux hired you.

Thanks to Jerome Collemare and Pauline Kerbirou for being my last french friends in Wageningen and to have teach me the taste for good wine. We had great times and we ate great food, specially thanks to Jerome's dexterity. Jerome, I always enjoy these long science nerd talks, especially about some of your most passionate topics. If we had the occasion, I believe we could have worked well as a team. Pauline, you had to leave far and now even further. At least I saw you regularly to catch up, and I hope to come some day in NZ.

Tim and Xiaoqian, I have always know you together and I cannot think of one without another. I've known you shortly before I started the PhD, and we have seen each other on a rather regular basis since then, even though you guys were a little more outside of my other circle of friends. I enjoyed talking about tech and game stuff with Tim, all the while talking about science and computer programming with Xiaoqian. That's quite a unique case. It's been a while now and I hope this friendship will last. Anyway, you kind of have to give me news, since that I gave you Baillou's offspring Sir Inti Mc Purface.

I want to particularly thank a special person with whom I shared more than seven years of my life and without whom I would probably never have settled in Wageningen. I thank you for all this time we shared and for your help in guiding me towards a more serious lifestyle. We basically grew together into adulthood life and I will keep great memories of these years.

Without you, I would certainly not be writing this today, and I am very thankful to you for that. I hope I also contributed positively to your future.

I would also like to thank my friends and classmates during my studies in bachelor and master. Though you did not directly take part in my thesis, you somehow helped me reaching it. Among them, I would like to particularly thank Julien Faramont, Ludivine, Clémentine, PierrO et Le Padré, kader, Aymeric, Marine Chasseray, Benoît Colrat, Quantin, Fabien, Jérôme Funel avec toute la clique, Sébastien Guizard, Emilie Amblard, Omar, Florian and Maude Jacquot.

Yuan, thank you for many many things during this second stage of my thesis. Thank you for this short but precious time were shared together. Thank you for pushing and helping me to be a better man. Thank you for all your support during this painful year. Even though there were difficult times, I believe the hardest part is now over and that things can only be better from now on. Of course, you still have to finish your PhD thesis. It will be tough with the work aside, I can tell, but I know you can make it. You should accept my help when you feel like it, we just have to make sure it does not end up like me and Maarten.

Also special thanks to my lovely and little gray ball of fur. Baillou, thanks for all the happiness your brought me. Though you don't understand emotions, you were of good comfort and you showed unconditional love to me. At least it always works with food. You were always next to me during the long writing nights, sleeping next to me. You would every now and then, come hugging me while trying to walk on the keyboard and releasing as much hair as possible. Thank you for your morale support, your comfort, and for the extra ctrl-z work.

Thanks for all these good moments I had with the people until this very moment. For all the people cited here, and probably many other I did not had time to write something special for. For many of my friends, it's a pity that we are now living so far apart. I have moved places so many times, that's kind of the story of my life.

Papa, Maman, merci de m'avoir donné la vie, de m'avoir donné une éducation, et de d'avoir tout fait pour que puisse en arriver là aujourd'hui. Tout ça c'est grace a vous, je vous aime et je vous en serrais éternellement reconnaissant. Merci à mon frère Cyril et ma sœur Ingrid pour cette enfance passée ensemble. Cyril, tu as toujours été un modèle pour moi et tu as été un grand frère génial.

Curriculum vitae



Benoît Manuel Carrères was born on Monday April 29th 1985 in Draguignan, France. In 2004 he started to study biotechnology at the University of Toulouse III, switched his studies to bioinformatics, and obtained his BSc in 2009 from the University of Clermont-Ferrand I. In 2011, continuing in the same field, he obtained his MSc from the University of Clermont-Ferrand II. In 2009, he performed his BSc thesis at the department of Plant Breeding at the University of Wageningen, during which he analyzed *Solanum Tuberosum* transformants microarray data, selected genes and verified their expressions qPCR. In 2010, he performed his MSc internship at the department of Plant Breeding at the University of Wageningen, during which he updated the “*Solanaceae* Phenotype Ontology” with trait terms used in BreeDB and developed a tool to automatically fill ontologies with diverse information such as term definition and cross-ontology references. In 2011, he returned to the same group to perform his MSc thesis, during which he developed a method to identify gene candidates responsible for a trait within a QTL. This work relied on integrating data from different sources using semantic web technologies. In 2012, he worked for one year as a scientific programmer within the same group of Plant Breeding Wageningen University where he did his MSc internship and thesis. During this time, he worked on the Virtual Lab Plant Breeding (VLPB-I) project that aims to automate data treatment and visualization for plant breeders. In 2013, Benoît started his PhD research project at the University of Wageningen at the group of Systems and Synthetic Biology and in tight collaboration with the group of Bioprocess Engineering. During this time, he was supervised by Dr. Peter Schaap, Dr. Dirk Martens, Prof. René Wijffels and Prof. Vitor Martin dos Santos. The work done in the following five years focused on analyzing RNA sequencing data of green microalgae cultured in diverse conditions that could help understanding the underlying physiological mechanisms that limit the conversion efficiency from photons to Triacylglycerides.

Publication list

Reijnders MJMF, Carreres BM, Schaap PJ. Algal omics: The functional annotation challenge. *Curr Biotechnol.* 2015;4:457–63. doi:10.2174/2211550105666151203202455.

de Jaeger L, Carreres BM, Springer J, Schaap PJ, Eggink G, Martins Dos Santos VAP, et al. *Neochloris oleoabundans* is worth its salt: Transcriptomic analysis under salt and nitrogen stress. *PLoS One.* 2018;13:e0194834. doi:10.1371/journal.pone.0194834.

Carreres BM, de Jaeger L, Springer J, Barbosa MJ, Breuer G, van den End EJ, et al. Draft Genome Sequence of the Oleaginous Green Alga *Tetradismus obliquus* UTEX 393. *Genome Announc.* 2017;5:e01449-16. doi:10.1128/genomeA.01449-16.

Carreres BM, León-Saiki GM, Schaap PJ, Remmers IM, van der Veen D, Martins dos Santos VAP, et al. The diurnal transcriptional landscape of the microalga *Tetradismus obliquus*. *Algal Res.* 2019;40:101477. doi:10.1016/J.ALGAL.2019.101477.

Overview of completed training activities

Discipline specific activities

Linear programming, University of Colorado Boulder	2013
Workshop interdisciplinary lifesciences : modeling, Lorentz Center, Leiden,	2014
WCSB Symposium, **	2014
WCSB Symposium, **	2015
SB@NL2013 Conference, SBNL, *	2013
SB@NL2014 Conference, SBNL,	2014
NBIC 2014 Conference NBIC, *	2014
European conference in computational biology (ECCB), *	2016
Symposium Algeapark/BPE,	2016

General courses

Scientific Writing, WGS,	2014
VLAG PhD Week, VLAG,	2013
Teaching and supervising theses students,WGS,	2015
Mobilising your scientific network, WUR,	2016
Reviewing a Scientific Paper, WUR,	2016
Scientific publishing, WUR,	2016
Brain training, WUR,	2016
Workshop SWAT4LS,	2013

Optionals

Weekly meetings SSB, WUR,	2013-17
B-Wise seminars,	2015-17
Preparation of research proposal,	2013
PhD Study trip,**	2015

* Poster presentation, ** Oral presentation

The research described in this thesis was financially supported by the IPOP programs Systems Biology and Biorefinery at the Wageningen University, and by EU projects IBISBA (nr 730976) and Empowerputida (No 635536).

Cover design by Benoît M. Carrères

Thesis layout by Benoît M. Carrères

Printed by Proefschriftmaken on FSC-certified paper.