GENERATION OF LIPID-RICH MUTANTS AND CHARACTERIZATIONS IN GREEN ALGA MICROALGA *DUNALIELLA TERTIOLECTA*

LINA YAO

(*B.Sc. (Hons.)*, Zhejiang University)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MICROBIOLOGY AND IMMUNOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2016

Declaration

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Lina Yao

(18 July 2016)

Acknowledgements

My utmost and sincere thanks to:

My supervisor, A/P Lee Yuan Kun, for his superb guidance throughout my PhD years. Thank you for providing me with this opportunity to work in this green lab. You supervised me in my research direction patiently, and encouraged me by generously sharing your own experience. Your grace, vitality, and passion for work inspired me a lot. As a newbie who just stepped into the research gate and independent life, I am extremely grateful to you for all you did to enhance my experience here. Thank you for sparing a lot of your busy time helping me through the emails or my unexpected visits.

A/P Tan Tin Wee, for co-supervising me in the bioinformatics. Very special thanks to you for igniting the spark of the powerful programming languages in my head three years ago. It was from you, I realized how important to apply this technology into conventional molecular biological research. And it was from you, I found out enthusiasms and creativities could really motivate a scientist in his/ her research. Your sincere attitude towards work inspired me a lot for my future career. I would also like to express my thanks to Assistant Professor Ban Hon Kim Kenneth, who also guided me in the programming skills. The great people I met and wonderful experiences will help me to continue understanding the multidisciplinary science in the future.

Our lab officer, Mr Low Chin Seng, for his impressive management to make the lab run in order so that we could focus on our own research. And also for his good humor and generousness to us juniors, which brought many joys

and fun during these years. From him, I learnt many skills to solve small technical problems in lab creatively. To our previous research assistant Ms Lin Hui Xin (current PhD student) for helping me to begin with my research from small tiny lab techniques and supporting me all along with the Nile red quantitative work. Without her, I would have a hard time struggling with the trials and errors by myself. It is Mr Low and Hui Xin who introduced the colorful life of Singapore to me when I came here for the first time. I would also like to express my thanks to our research assistant Mr Ezekiel Ho Hee Zhern (previous), Mr Chong Si Cong (previous), laboratory technologist Ms Voong Miao Lian for their great help in supporting the lab work and consumables. Without those people, all things would descend into a mass.

Senior post-doctoral research fellow, Dr Shen Hui, who spent a lot of time guiding me in molecular biology in details by her expertise, and always encouraged me when I felt down by her own experience. Dr Shen also did a lot to create a lovely lab environment for us to feel easy and warm. I also want to express my thanks to Dr Ng Yi Kai, post-doctoral research fellow, who not only guided us in the techniques and supported us by creating the general platforms of next-generation sequencing, total fatty acid measurement and metabolite profiling, but also played with us and brought us a lot of laughs and entertainments brotherly. Thanks to Dr Ng, for his kind help to read and check this thesis and provided me with great suggestions. And to all my lab colleagues, post-graduate students, Dr Zhao Ran (graduated), Dr Ng Hui Ping Daphne (graduated), Koh Ting Wei Kelvin, Tan Wei Min Kenneth, Siti Radiah Binte Safie, post-doctoral research fellow Dr Junhui Huang, visiting Assistant Professor Jin Liu and exchange students

Xiaonian Ma, Zhao Zhang and Dongzhe Sun from PKU, for all their help, time spent together, and good cheer.

To our collaborator, Professor Liang Li's group from University of Alberta, for providing us with great support in our metabolomic study.

As always, first and foremost, my extraordinary parents and family members, for their endless support and encouragement. My boyfriend, Yuan, for his accompany during these years and technical support in my research using his expertise in computer science. We shared our happiness and frustration in our PhD life together, and brainstormed using our multidisciplinary knowledge.

I would also like to express my gratitude to Department of Microbiology and Immunology, National Research Foundation (NRF) under its Campus for Research Excellence and Technological Enterprise (CREATE) Programme, and the NUS Research Scholarship to support me for this work.

Lastly, to all the specialists, seniors, friends, administration staffs who I met in my PhD life that I did not mention in this brief list, I am eternally grateful for all you did to enhance my PhD life physically and mentally.

TABLE OF CONTENTS

Summary

As an alternative sustainable resource for fossil fuels, microalgal lipids have received significant attention in recent years. However, few of the current algal species achieved lipid productivity that is economically competitive with fossil fuels and commercially viable. Metabolic engineering to increase biofuel-relevant lipid (neutral lipid, mainly in the form of triacylglycerol (TAG)) yields in these microalgae is a common approach in advancing economic feasibility.

In this project, *Dunaliella tertiolecta*, a non-model oleaginous microalgal species was used to generate lipid-rich mutants via insertional mutagenesis. D9 mutant was generated with a lipid yield that is 2 to 4-fold higher than that of wild type. Nile red method for rapid quantification of TAG was optimized for mutant library screening. The oleagenesis of mutants through different growth phases was traced.

However, the lack of genomic and transcriptomic information has limited genetic manipulation and global transcriptome profiling for characterizations of microalgal mutants and microalgae cultured under different culture conditions, particularly for those with potential as biofuel sources. To address this problem, in-house transcriptome database was constructed based on high quality high-throughput sequencing data from Illumina MISEQ and HISEQ that were explored using high performance computing in a petascale data center. After subjecting to *de novo* assembly using high memory (1TB RAM) nodes and parallelized mpiBLASTX search in NCBI non-redundant protein

database, an enlarged transcriptome database of 17,845 was constructed, covering ~95% of the total genes. In addition, the pipeline for analyzing RNA-Seq data for comparative transcriptome analysis using this draft transcriptome database was developed and optimized.

Microalgal neutral lipids are typically accumulated during the stationary growth phase. To trigger *D. tertiolecta* TAG accumulation from the exponential growth phase, target screening using fluorescence-activated cell sorting followed by random insertional mutagenesis, *D. tertiolecta* mutant with TAG accumulation from the exponential growth phase (TAEP) was generated. This is crucial for advancing economic feasibility in *D. tertiolecta*. From the characterization of the mutant, we hypothesized that elevation of acetyl-CoA (AcCoA) level would result in TAEP in microalga *D. tertiolecta*. It was discovered that refilling of AcCoA pool through branched-chain amino acid catabolism contributed to six-fold TAEP with marginal compromise (4%) on growth in a TAG-rich *D. tertiolecta* mutant from integration of transcriptomics and metabolomics.

This study provides novel approaches to increase lipid productivity and compress lipid production phase, which could serve for optimization of renewable microalgal biofuel to fulfill the demanding fuel market.

List of Tables

List of Figures

List of Abbreviation

CO_2	carbon dioxide
FA	fatty acid
TAG	triacylglycerol
LD	lipid droplet
ER	endoplasmic reticulum
DAG	diacylglycerol
DGAT	diacylglycerol acyltransferase
PDAT	phospholipid:diacylglycerol acyltransferase
ME	malic enzyme
PEPC	phosphoenolpyruvate carboxylase
ND	nitrogen deprivation
NGS	next-generation sequencing
AcCoA	acetyl-CoA
TAEP	TAG accumulation in the exponential growth phase
WT	wild-type
GC-MS	gas chromatography mass spectrometry
FAME	fatty acid methyl ester
BLAST	Basic Local Alignment Search Tool

NCBI	National Center for Biotechnology Information
KO	KEGG orthology
GO	gene ontology
FACS	fluorescence-activated cell sorting
CIL LC-MS	chemical isotope labeling liquid chromatography mass spectrometry
Q-TOF	quadrupole time-of-flight
PDK	pyruvate dehydrogenase kinase
PDHC	pyruvate dehydrogenase complex
DCA	sodium dichloroacetate
CuAO/ AMX1	copper amine oxidase family
IVD	isovaleryl-CoA-dehydrogenase
MCCB	3-methylcrotonyl-CoA carboxylase
ACCA	acetyl-CoA C-acetyltransferase
FabD	malonyl-CoA:acyl-carrier-protein transacylase
petE	photosynthetic electron transport
petC	cytochrome b6-f complex
LHCA1	light-harvesting complex I chlorophyll a binding protein 1
LHCB1	light-harvesting complex I chlorophyll b binding protein 1
pfkA	6-phosphofructokinase 1

PPC	phosphoenolpyruvate carboxykinase (ATP)
FBP	fructose-1,6-bisphosphatase I
BCKDH	alanine-glyoxylate transaminase / (R)-3-amino-2-methylpropionate-pyruvate transaminase 2-oxoisovalerate dehydrogenase E1 component alpha subunit
BCA2	branched-chain amino acid aminotransferase
α-KG	α-ketoglutarate
Glu	glutamate
Pro	proline
SETD2	histone-lysine N-methyltransferase
petF	ferredoxin
psbS	photosystem II 22kDa protein

Chapter 1 Introduction and literature review

1.1 Fossil fuel crises and prospect of microalgae as biofuel sources

Conventional fossil fuel crises started in the 1970s caused an increase in a global energy demand [1]. The demand for fossil fuels leads to political strife, climate changes, and exhaustion of finite energy reserves [2]. To replace fossil fuels and to develop sustainable energy production, identifying and reassessing sources of biologically derived energy, such as biodiesels, become increasingly urgent. Biodiesel converted from plant or animal neutral lipids have received increasing attention in recent years. One of the most important advantages of biodiesel is carbon neutrality since it can ultimately generate biofuels from solar energy through photosynthetic carbon dioxide (CO_2) sequestration, which also addresses the issue of continued increased atmospheric CO_2 as greenhouse gas [3].

As one of the renewable sources of biofuel feedstock, microalgae have been intensively explored in recent years [4]. Chief advantages over other biodiesel sources are that they do not compete for land and irrigation water with crops, since many of them could be cultured on any land surface, in marine water, and even in wastewater. Microalgal hydrocarbons and TAGs can be used directly or indirectly for chemical conversions to biodiesels [5, 6].

1.2 Lipid accumulation in microalgae

Lipids comprise fatty acids (FAs) and their derivatives, and substances related biosynthetically or functionally to these compounds [7]. They encompass hydrophobic and amphipathic compounds, which are often soluble in organic solvents rather than water [8]. Lipids play many crucial biological functions in algal cells, including forming biological membranes (phosphoglycerolipids, galactoglycerolipids, sterols), storage compounds (TAG is the most abundant), and signaling molecules (phosphoinositides, oxylipins) [9].

Under normal growth conditions, algae synthesize principally glycerol-based polar lipids, which mainly reside within the plasma membrane and endoplasmic membrane systems, as well as the membranes of chloroplasts and mitochondria [10-13]. Under unfavorable growth conditions, however, many algae switched the metabolic pathways towards the biosynthesis of storage lipids in the form of neutral lipids (mainly TAG) [14]. In response to environmental stress, TAGs are accumulated to provide energy deposits that can be readily catabolized to allow rapid growth of the cells [15]. TAGs are stored in densely packed lipid droplet (LD) after being synthesized [14]. Besides TAGs, the neutral lipid core, LDs are surrounded by a monolayer of polar lipids decorated by proteins; some minor free FAs, carotenes, etc, are also present depending on species, to form a monolayer membrane interface to delineate TAGs from an outside aqueous subcellular environment [16].

TAG consists of a glycerol moiety, each hydroxyl group of which is esterified to a FA [7]. The identity of the FA is due to its origin [3]. TAG is one of the

most energy-rich reduced carbon and the most abundant form of simple lipid class that is commercially important as a biodiesel precursor in nature [7, 17]. TAG can be readily converted to mixture of FA esters via transesterification reaction [5, 6].

1.3 Biosynthesis of lipid droplets (LDs)

FAs are biosynthesized in the plastid, LDs are largely present in the cytosol, which are formed via extension of specific domains of endoplasmic reticulum (ER) [18]. The plastid and ER share some reactions during TAG biosynthesis[16, 19, 20].

TAG assembly shares common pathways with synthesis of membrane lipids until the formation of diacylglycerols (DAGs) [16], from which TAG is synthesized by the acylation reaction via two routes: acyl-CoA-dependent route, catalyzed by diacylglycerol acyltransferase (DGAT), as the well-known Kennedy pathway [21] in Figure 1.1 [22, 23]; and acyl-CoA-independent route, catalyzed by phospholipid:diacylglycerol acyltransferase (PDAT) [16, 19, 20].

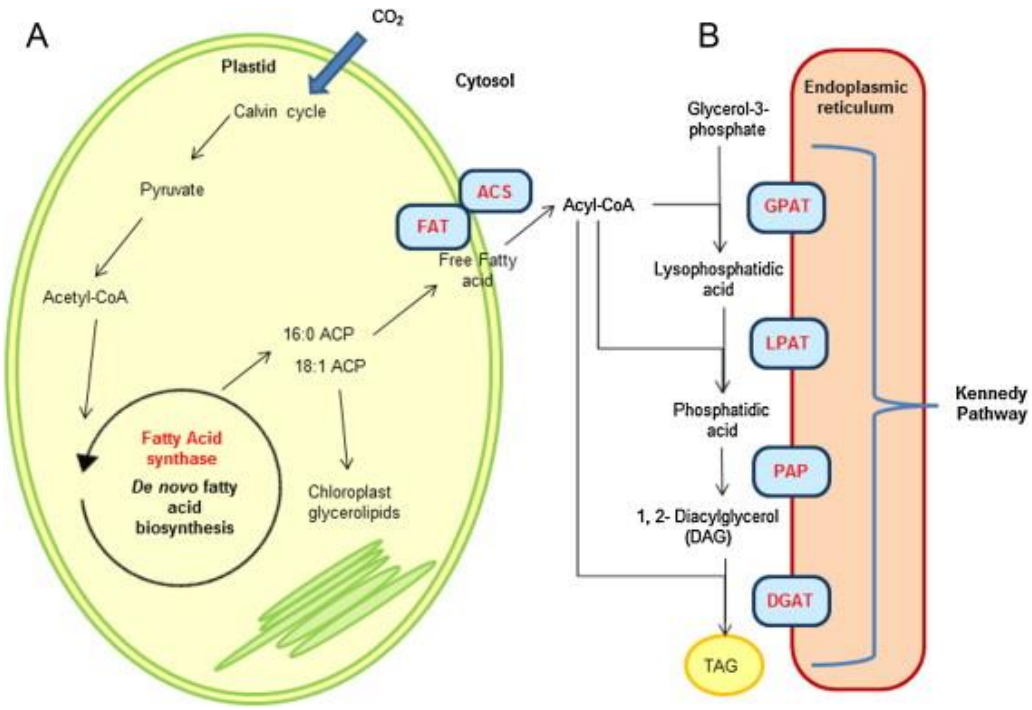


Figure 1.1 Simplified scheme of TAG biosynthesis in higher plant cells (adapted from [22, 23]).

1.4 Numerous approaches to increase triacylglycerol (TAG) productivity

Despite the ability for microalgae to produce TAG as a precursor for biofuels, microalgal derived biofuels are still not economically viable compared to fossil fuels or other alternative fuels [2].

Significant improvements of microalgal lipid productivity and optimization of extraction methods are crucial. In recent years, many attempts have been made to enhance lipid overproduction in microalgae, which mainly focuses on genetic engineering and biochemical engineering approaches to promote and enhance the lipid biosynthesis pathways [23-25].

1.4.1 Genetic engineering approach

The genetic engineering approach is widely used to enhance lipid production [24]. It focuses on increasing activities of lipid biosynthesis pathways or blocking other competing pathways [23], to channel metabolite fluxes to lipid biosynthesis [24]. Numerous trials have been made in different microalgal species to investigate the feasibility of manipulating genes that regulate lipid synthesis or related pathways to enhance lipid production, such as DGAT [26, 27], malic enzyme (ME) [28] [29], β-oxidation pathway enzymes [30], phospholipid biosynthesis pathway enzymes [24], phosphoenolpyruvate carboxylase (PEPC) [31]. In some cases, a "secondary bottleneck" might also be manipulated to achieve significant enhancement of lipid biosynthesis [32]. However, all these approaches only led to lipid accumulation in the stationary growth phase with lengthy lipid production phase, thus low overall lipid productivity [33-35]. The ability to trigger lipid accumulation while maintaining high growth rates is imperative for microalgal biofuel production in large economic scales [36]. To achieve this, finding ways within the cells to channel precursors of FA biosynthetic pathways appears to be a reasonable strategy using genetic engineering.

1.4.2 Biochemical approach

Biochemical engineering strategy triggers lipid accumulation by imposing environmental stress conditions such as nutrient stresses (nitrogen/ phosphorus deprivation, etc), pH, temperature, heavy metals, osmotic stress, radiation, and other chemical stresses to channel carbon fluxes to lipid accumulation [25].

The most commonly adopted way is to use nitrogen deprivation (ND) to trigger neutral lipid accumulation in microalgae. Nevertheless, deficiency of nutrients may impede photosynthesis capability that responsible for generating the initial biomass carbon and energy, thus resulting in a decrease of overall lipid productivity. A two-stage culture system was proposed by Li and Rodolfi et al. [37, 38]: for the first stage, cells were cultured in nutrient sufficient medium, and for the second stage, cells were transferred to nutrient deprivation medium for lipid accumulation. Though it has achieved better lipid performance than the single stage nutrient deprivation culture system, there is apparently a dilemma in the separation of the two stages, which takes longer for the lipid production time.

1.5 Organism of interest: *Dunaliella tertiolecta*

Over the past few decades, many oleaginous algal and cyanobacteria species have been isolated and characterized for high lipid content [4]. Among them, *Dunaliella tertiolecta* is considered one of the most promising species. It is a flagellated unicellular microalga, under the Chlorophyta phylum. It has a high tolerance to salt, which reduces contaminations of most environmental microalgae, protozoa, fungi and bacteria. *D. tertiolecta* can also utilize inorganic nutrients that present in saltwater, wastewater or brackish water, and it also lacks a rigid cell wall, hence a reduction in complexity to harvest commercially and pharmaceutically significant products and genetic manipulation [39-44]. The chief rational of selecting it as the research organism in this study lies in its ability to produce large quantities of lipids

(up to 67% of organism dry weight) [44].

1.6 Research gap

Because of limited knowledge of underlying biochemistry, cell biology and genetic characteristics of microalgae [3], and high cost of growing microalgae for biofuel production, algal biofuel production is currently economically not competitive with fossil fuel [45-52], the potential of microalgae as precursor of biodiesel has yet to be realized [14].

1.6.1 Rapid TAG measurement method

Traditionally, the analysis of lipid content is largely based on solvent extraction of the lipids from the samples and then quantified by gravimetric [53] or spectrophotometric methods [54]. However, the techniques are time-consuming, which do not allow high-throughput screening. Moreover, the traditional protocols may strongly affect yields of lipid [55]. Therefore, an easy and accurate quantitative measurement method is essential for screening *D. tertiolecta* and mutants for lipid accumulation in large scale. Nile red, a lipid-soluble fluorescent probe that allows *in situ* staining of neutral lipids in microalgal samples [56, 57], will be investigated and optimized in this study.

1.6.2 *D. tertiolecta* transcriptome database

Genetic manipulation is a common strategy for enhancement of lipid overproduction in microalgae by manipulating one or more genes of key enzymes in microalgal strains [24]. To perform genetic manipulation in

microalgal strains, their genome information is necessary. As of now, only few comprehensively annotated model microalgae genomes are available, such as *Coccomyxa subellipsoidea* C-169 v2 [58], *Chlorella variabilis* NC64A v1 [59], *Chlamydomonas reinhardtii* v4 [60], *Micromonas pusilla* RCC299 v3 [61], *Ostreococcus lucimarinus* v2 [62], *Volvox carteri* [63] and *Thalassiosira pseudonana* CCMP 1335 v3 [64].

However, a large number of microalgal species, like *D. tertiolecta*, with unique traits and considerable advantages over model microalgae as feedstock of biofuel production and many other purposes are not sequenced yet [65]. The insufficient genome sequences in non-model microalgae with potential for biofuel production limits the development of genetic engineering of these microalgae. Although Hamid et al. [43] provided a good approach to investigating into the transcriptome and annotating partial transcripts, the nuclear and chloroplast genome sequences of *D. tertiolecta* are still incomplete and limited for global transcriptome studies. Herein, in this study, a more complete transcriptome database in this non-model microalgae, *D. tertiolecta*, is urgent to be developed, which will pave the way for a more thorough characterization and development of elite algal strains for biofuels production.

1.6.3 Understanding of regulatory mechanisms of lipid metabolism by next-generation sequencing (NGS) technology

Lipid metabolism in microalga is complex, and distinct from the well studied lipid metabolic pathways in vascular plants [8], and even within algal groups [66]. Current genetic engineering approach can barely depend on single

regulatory factors. The whole picture of lipid regulatory mechanism network in microalgae needs to be depicted.

Enlarged data analytical capability and improved downstream processing in the next-generation sequencing (NGS) technology have been developed [43, 67, 68]. RNA-Seq is a recently developed approach for global transcriptome profiling. It uses deep-sequencing technologies to elucidate the complexity of eukaryotic transcriptomes [69]. To date, numerous studies have been reported to use RNA-Seq to quantify transcriptional expression under different culture conditions or strains to elucidate metabolic regulatory mechanisms of microalgae, such as nitrogen deprivation-included TAG accumulation [43, 65, 70-75].

1.6.4 Cell growth and lipid accumulation

Most of approaches to increase TAG accumulation take place in the stationary growth phase or at the expense of biomass accumulation [34, 76] and overall lipid productivity. Vigorous growth and TAG accumulation appear to be mutually exclusive as TAG is a secondary (storage) metabolite and the pyruvate to AcCoA pathway is tightly regulated by the growth-dependent pyruvate dehydrogenase complex activity [72, 77].

The oleaginous diatom *Fistulifera solaris* JPCC DA0580 was the first to be reported to have a temporal overlap of TAG accumulation and cell growth during the exponential growth phase [78]. Such a feature that triggers TAG accumulation while maintaining high growth rate is a critical advantage in the microalgal large-scale culture for TAG production. To further exploit this

potential in the green oleaginous microalgae *D. tertiolecta*, molecular mechanisms and novel pathways governing related processes will be investigated.

1.7 Project objectives and hypotheses

In this chapter, we outlined the advantages of algae as a biofuel producer, discussed the lipid accumulation mechanism, introduced approaches to increase lipid production and the research organism, and proposed potential solutions for research gaps that remained to be solved to make microalgae efficient and economically competitive.

The overall objective of this project is to optimize *Dunaliella* to be a potential candidate of neutral lipid accumulation for biofuel production. Firstly, our aim is to select and optimize a rapid approach to quantify TAG and lipid content in *Dunaliella*, as it has not been established in our experimental condition. This is to allow rapid screening of high TAG and lipid mutants following genetic manipulation. Secondly, *Dunaliella* genome has not been sequenced, so that it is difficult to do any bioinformatics related work or manipulation, hence one of the aim is to construct a draft transcriptome database for the purpose of genetic engineering and NGS analysis. Thirdly, this is also to allow the study of transcriptomic changes within the cells under different growth conditions as well as comparative transcriptomic studies between the mutants and wild type cells. Lastly, transcriptome and metabolomic studies will be performed to uncover

mechanisms of the carbon flux in mutant that could trigger TAEP in *D. tertiolecta.*

1.8 Significance and scope of this study

The work will provide substantial insights into the below aspects:

1. Nile red quantification method for neutral lipid will be established for mutant library screening and characterization.
2. A user-friendly approach to construct non-model microalgae *D. tertiolecta* transcriptome database will be provided and optimized for characterization of gene expression and analysis of metabolic profiles.
3. *D. tertiolecta* lipid-rich mutants will be generated with overall enhancement of TAG production and early TAG productivity in the growth phase. The mutants will be characterized by the in-house developed RNA-Seq analysis pipeline and metabolomics, to elucidate the regulation of lipid synthetic pathways and the related network.

In this regard, the discoveries and conclusions in this project will contribute to the improvement of microalgal-derived biodiesel. The availability of draft transcriptome database paves way for genetic and metabolic engineering of high lipid producing microalgae and their characterization. The regulation of microalgal cell growth and lipid accumulation will ultimately allow the development of novel genetic engineering targets.

Chapter 2 Materials and methods

2.1 Culture conditions

The algal culture *D. tertiolecta* strain UTEX LB-999 was obtained from the UTEX Culture Collection of Algae (University of Texas at Austin, TX). The microalgal cells were cultivated using ATCC-1174 DA liquid medium (American Type Culture Collection at Manassas, Virginia) containing 0.5M NaCl in flasks with shaking at 100 rpm at 25°C, under 12 h light/12 h dark with light intensity of 30 μmol photons·m^{-2}·s^{-1} [79]. Culture was supplied with 2% CO_2 every two days.

Under high light condition, cells were cultured at light intensity of 320 μmol photons·m^{-2}·s^{-1}. Under ND condition, cells were grown under 0.5 M NaCl ATCC medium with 10% KNO_3 and 90% KCl to substitute KNO_3 (N^-). A comparison group of normal growth condition group was set (N^+).

The growth of microalgal cells was monitored by counting cell number and measuring absorbance at OD680nm by GENESYS 10S UV-Vis spectrophotometer (Thermo Scientific™). The specific growth rate (μ) was calculated from the equation [80]:

$$\mu = \frac{\ln(N2 - N1)}{t2 - t1}$$

where N2 and N1 are the cell number concentration measured at times t2 and t1, respectively.

For the purpose of clarity, work conducted throughout the paper is based on biological and technical triplicates unless otherwise stated.

2.2 RNA extraction

D. tertiolecta was collected for total RNA extraction using an RNeasy plant mini kit (Qiagen, Valencia, CA, USA), according to the manufacturer's instructions. DNase was added to eliminate genomic DNA contamination. Total RNA was used to synthesize Oligo-dT-primed cDNA using the SuperScripte III first-strand synthesis system (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. The synthesized cDNA was used for cloning target gene sequences. Total RNA was also used to synthesize Random-hexamer-primed cDNA according to the manufacturer's instructions for quantitative real-time PCR.

2.3 DNA extraction

1-5 mL microalgal cells were collected after centrifugation at 3,000 × g, 5min. 0.5mL DNA extraction buffer (2% SDS, 400 mM NaCl, 40 mM EDTA, 100 mM Tris-HCl, pH 8.0) was added and the pellets was re-suspended prior to 65 °C for 30 min-1h. 500 uL phenol/ chloroform/ isoamyl alcohol (25:24:1) was added and mixed well. The aqueous phase was extracted after centrifugation at 12,000 × g for 30min at 4 °C. 1:1 volume isopropanol was added, and inverted to mix well. After incubating at -20 °C for 1h, DNA was precipitate by centrifugation at 12,000 × g for 20min at 4 °C, and wash with 75% ethanol twice (7,500 × g for 5min, 4 °C). 30 uL of

elution buffer was added to the DNA after air dry.

2.4 Determination of photosynthetic rate and photosynthetic efficiency

D. tertiolecta cells were collected by centrifugation at 12,000 × g for 15 min at 4 °C. The high centrifugal force served to weaken the cell structures to facilitate extraction of chlorophyll. Chlorophyll content of individual sample was estimated spectrophotometrically [81]. Subsequently, photosynthetic performance was carried out using an oxygen electrode according to the operating manual (Rank Brothers, Bottisham, Cambridge, UK).

2.5 Determination of glycerol and starch content

One milliliter of *D. tertiolecta* liquid culture were harvested for the determination of intracellular and extracellular glycerol contents [82]. The level of glycerol was determined using Free Glycerol Determination Kit (Sigma FG0100), according to the protocol provided by the manufacturer. Ten milliliters of the *D. tertiolecta* cells were collected and the starch content was determined using Starch Assay kit (Sigma STA20), according to the instructions provided by the manufacturer.

2.6 Cloning of *D. tertiolecta* gene

Most gene sequences of the *D. tertiolecta* are not available partially or

completely. Herein, RACE PCR combined with a draft *D. tertiolecta* transcriptome database that mentioned below was used in this study to clone genes. RACE primers of the corresponding genes were designed and worked out based on the *D. tertiolecta* database. The 5' and 3' ends of the *D. tertiolecta* genes were cloned using a SMART^{TM} RACE cDNA amplification kit (Clontech, Mountain View, CA, USA). The RACE-PCR products were ligated into pGEM-T Easy vector (Promega, Madison, WI, USA) and sequenced directly. RACE 5' and 3' overlapping sequences were assembled manually to obtain the full-length sequences. The coding sequences were cloned and confirmed by PCR amplification using high-fidelity Taq polymerase (Invitrogen) as reported previously by our group [79]. In cases that a gene sequence derived from the *D. tertiolecta* database is complete, the high-fidelity Taq polymerase was directly used for PCR amplification and preceded to following steps.

2.7 Analysis of *D. tertiolecta* gene and amino acid sequences

Phylogenetic tree of protein based on various species was constructed by software MEGA 5 (neighbor-joining (NJ) method) [20]. Subcellular localization of the corresponding genes was predicted by SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/), ChloroP (http://www.cbs.dtu.dk/services/ChloroP/), MITOPROT (https://ihg.gsf.de/ihg/mitoprot.html), Hectar (http://webtools.sb-roscoff.fr) online. Default parameters were set in the above software and websites.

2.8 *D. tertiolecta* transformation

2.8.1 Plasmid construction

The pGreenII0000 plasmid incorporated a 431 bps *CrRBCS2* promoter and a 526 bps bleomycin resistance gene (*ble*) to confer zeocin-resistance was designed and constructed in our laboratory [83].

2.8.2 *D. tertiolecta* transformation

Via glass-bead method

D. tertiolecta cells at exponential growth phase were transformed using the glass-bead method [84] with minor modifications. Cultures grown in the exponential growth phase in 0.08 M NaCl ATCC medium were harvested by centrifugation at 3,000 × g, 10 min, and concentrated to a cell density of 1-2 × 10^8 cell·mL^{-1}. A 300-500 μl aliquot of the cell suspension was added to a 12 mL round-bottom tube containing 0.3 g acid-washed glass beads (425-600 μm diameter; Sigma), 100 μl 20% polyethylene glycol (PEG-8000; Sigma), 1 μg plasmid (with a cassette conferring resistance to zeocin) after linearization and 5 μl fish sperm DNA. After 20 s of vortex at maximum speed, cells were plated immediately onto 0.08 M NaCl ATCC medium agar plates containing 10 μg·mL^{-1} zeocin (Invitrogen). Colonies that appeared within 3 weeks were picked and inoculated into 0.5M NaCl ATCC medium. Secondary selection in 0.08M NaCl ATCC medium with 5 μg·mL^{-1} zeocin was performed to reduce the false positive transformants [83].

Via electroporation method

D. tertiolecta transformation and random mutagenesis could also be conducted via electroporation [85] (double consecutive pulses were delivered at intervals of 10 to 15 s in a Bio-Rad Gene Pulser Xcell™ Electroporation Systems, with capacitance = 500 μFD, resistance = 400 Ω, voltage = 400 V), 500 μl cells were incubated with 50 μl of linearized plasmid (1 μg/ 10 μl, total amount 1-5 μg) and 5 μl of carrier DNA (fish sperm DNA) in 4-mm gap cuvette. The plasmid with a cassette conferring resistance to zeocin served to generate random insertional mutagenesis [86]. The cells were plated on the selection agar plates and screened as described in the previous section 2.8.2.1.

2.8.3 Mutant screening and selection

After inoculated in 6-well plates, genomic DNA was extracted for genotyping PCR to confirm the existence of the transgene using the primers ble_F and ble_R designed from zeocin cassette (Appendix 1). The PCR reaction was set in a volume of 20 μl using 1 μg template, hot-started at 95 °C and continued for 35 cycles of denaturation at 95 °C for 30 s, annealing at 58 °C for 30 s, and extension at 72 °C for 1 min, ended in an addition extension at 72 °C for 10 min as described by Lin et al. [83].

Chapter 3 Optimization of quantitative measurement of neutral lipids using Nile red staining

3.1 Introduction

For the purpose of selecting lipid-rich strains using high-throughput screening for neutral lipid content in microalgae, the fluorescent dye staining method is commonly used to quantify neutral lipids, like Nile red, offering a rapid, inexpensive and time-saving analysis approach [87]. Two methods of Nile red quantitative measurement using Tecan Infinite M200 PRO were developed and applied to a nitrogen deprivation study. Flow cytometry was carried out as a comparison on the quantitative measurement of TAG with the Nile red method. Additionally, fluorescent microscopy was used to visualize the correct staining of the neutral lipids. This method of TAG measurement was also compared with total FA quantification using gas chromatography mass spectrometry (GC-MS).

3.2 Materials and Methods

3.2.1 Culture and reagent preparation for Nile red staining

A stock solution of Nile red (Sigma, 72485) was prepared by adding 5 mg of Nile red to 50 mL of acetone in an amber colored bottle and stored in the dark at 4 ^{o}C. 25 μg/mL of Nile red was used as the working solution. A 5 mg/mL isopropanol solution of triolein (Triolein T7140 Sigma (Glyceryl Trioleate) 99%) was prepared as the lipid standard stock. ATCC liquid medium containing 0.5 M NaCl was prepared as the carrier for standard

curve plotting and cell sample measurement. The detail protocol for the two methods described below could be found in Appendix 2.

Direct staining method

Cell density was checked by spectrophotometer and hemocytometer, and harvested by centrifugation 3000 × g for 5 min and re-suspended in fresh ATCC medium with the same salinity, in order to eliminate impurities potentially present in the medium used for cell growth and to ensure the same salinity level for all the samples. After re-suspension, the cell density measured at OD_{750} was adjusted to 0.1-0.5. 200 µl of cell samples, culture medium blank and triolein standards were transferred to the same 96-well black, clear bottom plate. The plate was read at the selected parameters before and after additional 2 µl Nile red working solution, 5 minutes incubation in the dark.

Standard addition method

For the analysis of neutral lipids content with the standard addition method, cells were re-suspended in the fresh ATCC medium with the same salinity. From each 1.98 mL re-suspended sample, isopropanol and lipid standard solution were added in order to obtain a final volume of 2 mL. Since the final concentration of triolein changed for each time point of cultivation, the added volumes of triolein standard and isopropanol varied for each sample (e.g. for the "0" point of the curve, 20 µl of isopropanol were added; for the 2.5 µg/mL point of the curve, 1 µl triolein standard stock solution and 19 µl of isopropanol were added; etc.). After triolein addition, the fluorescence of each sample was detected at the selected parameters before and after

additional 2 µl Nile red working solution, 5 minutes incubation in the dark.

3.2.2 Optimization of instrumental parameters

The fluorescence signal of samples was measured before and after Nile red addition, in order to subtract the intrinsic fluorescence value of the sample as mentioned by Bertozzini et al. [55]. The fluorescence intensity of the medium containing Nile red was also measured, which served as the blank control and subtracted.

Excitation and emission wavelengths were identified by scanning the standard triolein in the medium. Different concentrations of triolein solution and different Nile red staining time were measured. A range of microalgal cell densities was also tested. From these studies, an optimal condition was determined for routine measurement of TAG content.

3.2.3 Application on nitrogen deprivation (ND) study

As aforementioned, microalgae produce more TAGs stored in oil bodies in response to the ND condition. *D. tertiolecta* cells were cultured under normal and ND conditions according to section 2.1 for TAG quantification using Nile red staining.

3.2.4 Gas chromatography mass spectrometry (GC-MS) for total FA measurement

Fatty acid methyl esters (FAMEs) were prepared by direct transmethylation with sulfuric acid in methanol [7]. The FAMEs were analyzed by gas chromatograph (model 7890B, Agilent Technologies) equipped with a mass-

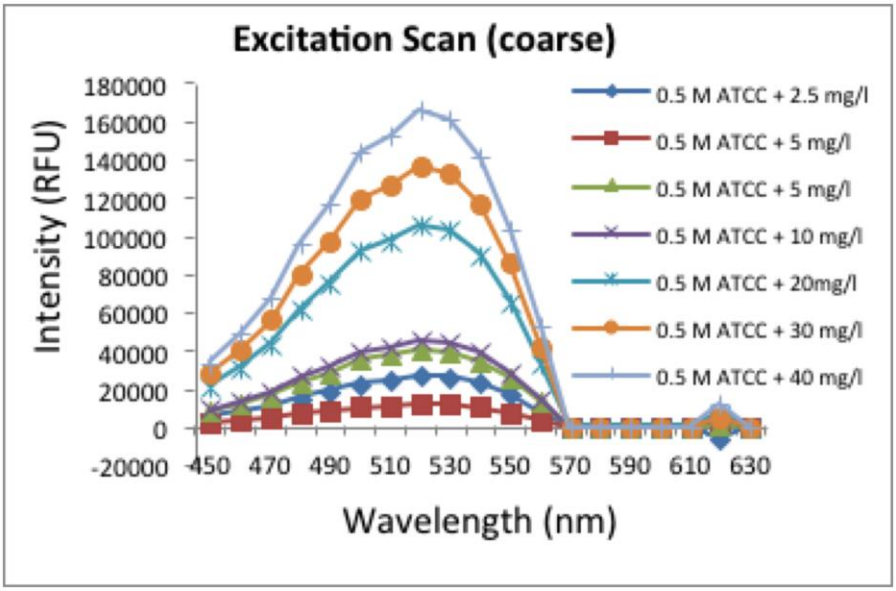
selective detector (model 5977A, Agilent Technologies) [88, 89].

3.3 Results and Discussion

3.3.1 Optimizing instrument parameters and staining conditions in Nile red method

Excitation and emission wavelength were selected at 524 nm and 586 nm respectively after scanning (Figure 3.1), to eliminate the interference of the characteristic autofluorescence of chlorophyll, whose emission peak was at ~668 nm. 5 minutes was selected as the optimal incubation time, since a longer incubation time resulted in signal decay (Figure 3.2). The fluorescence intensity of Nile red stained triolein standard solution was linear up to 40 mg/L (Figure 3.3). A range of cell concentrations that resulted in a linear relationship between the cell concentration and fluorescence intensity was between OD_{750} at 0.1 (equivalent to 8×10^5 cells/mL) and OD_{750} at 0.5 (equivalent to 3×10^6 cells/mL) for *D. tertiolecta*. In our current study, the concentration of 2×10^6 cells/mL (equal to $OD_{750} = 0.3$) was chosen as the working cell density.

a



b

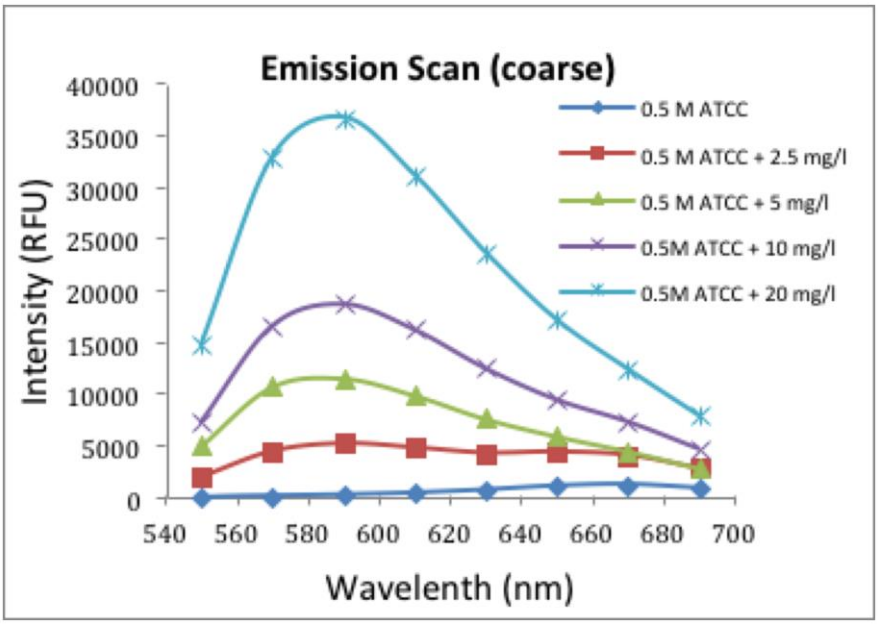


Figure 3.1 Excitation and emission wavelength scanning.

(a) Excitation wavelength scanning. (b) Emission wavelength scanning.

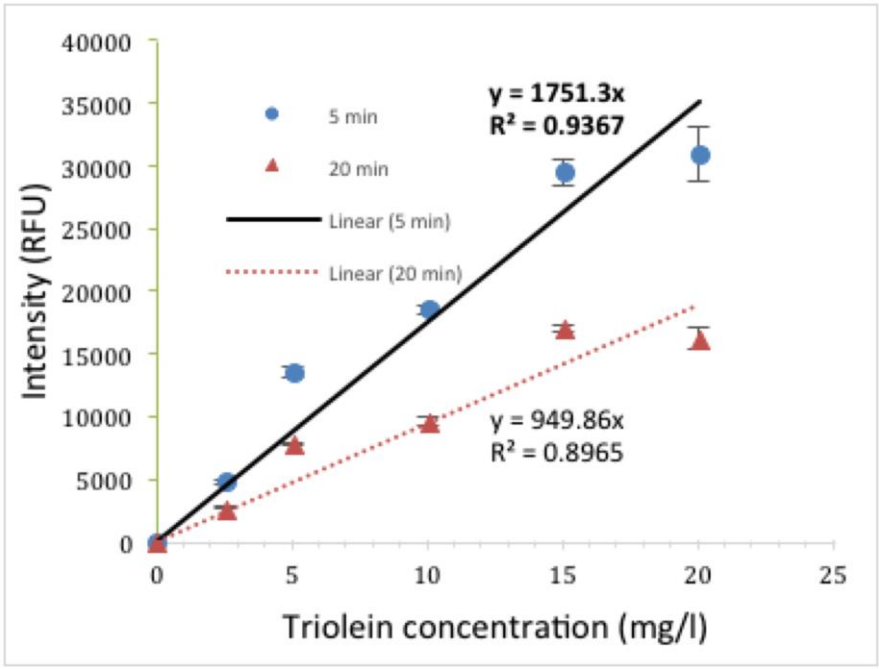


Figure 3.2 Comparison of Nile red staining time based on the standard curve.

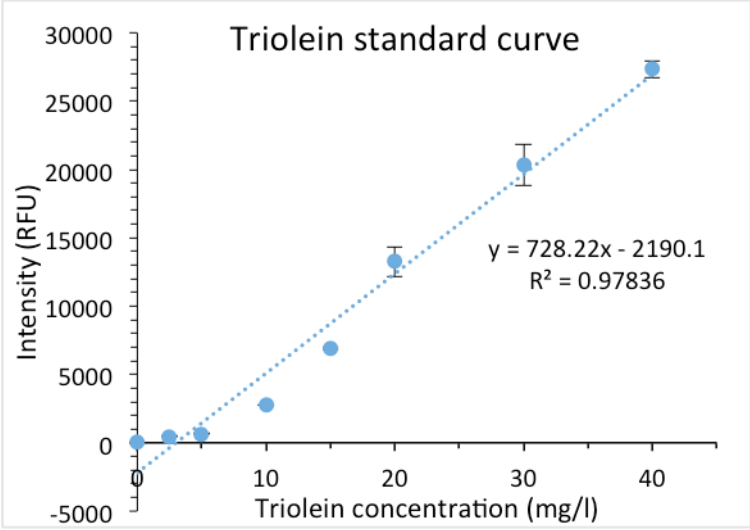
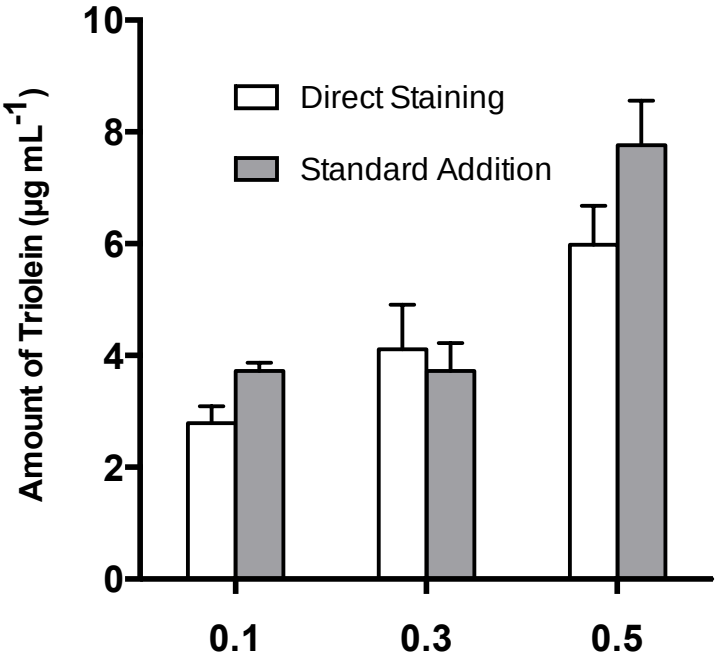


Figure 3.3 Linearity range of the concentration of triolein standard.

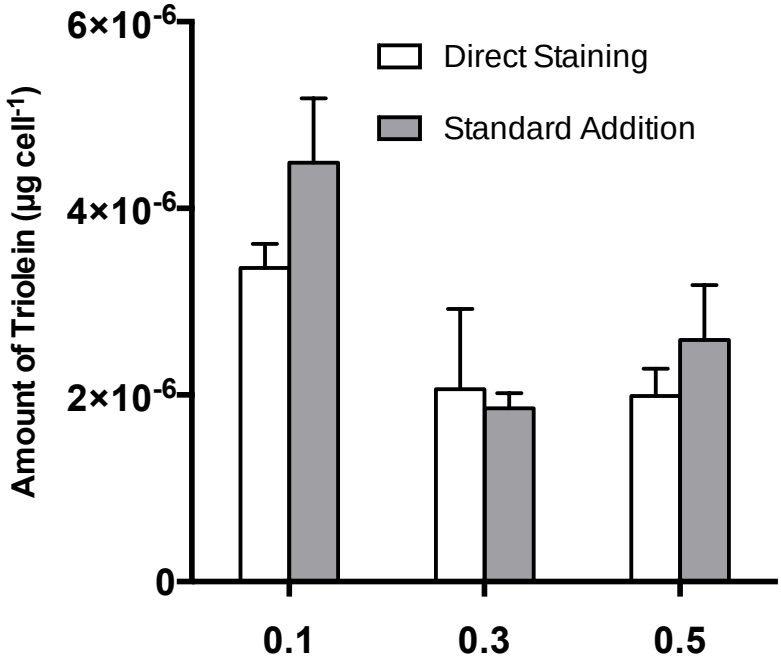
3.3.2 Comparing the direct staining method and standard addition method

Both the direct staining and standard addition method are rapid as the

incubation time only takes five minutes compared to other methods that may take up to several hours. The advantage of the direct staining method is that it only requires a small sample volume (200 µl) compared to the larger sample volume (10 mL) in the standard addition method. However, the standard addition method was found to be useful in determining lipid content for cells containing fewer lipids. In addition, the standard curve is not required to be plotted in the standard addition method, while it is required for every plate in the direct staining method. As the sample measurement goes together with the reference triolein solutions in the standard addition method, fewer samples (maximum 8 samples) could be assayed in one micro plate compared to the direct staining method (maximum 40 samples). The quantitative comparison results of the two methods are presented in Figure 3.4, which shows a good similarity.



a



b

Figure 3.4 Comparison of two Nile red quantitative measurement methods. (a) Neutral lipid content per cell basis. (b) Neutral lipid content per mL basis.

3.3.3 Neutral lipid is triggered by ND stress

In the late exponential growth phase, a huge quantity of neutral lipids were triggered by ND stress which was determined by direct Nile red quantitative measurement (Figure 3.5) and flow cytometry (Figure 3.6).

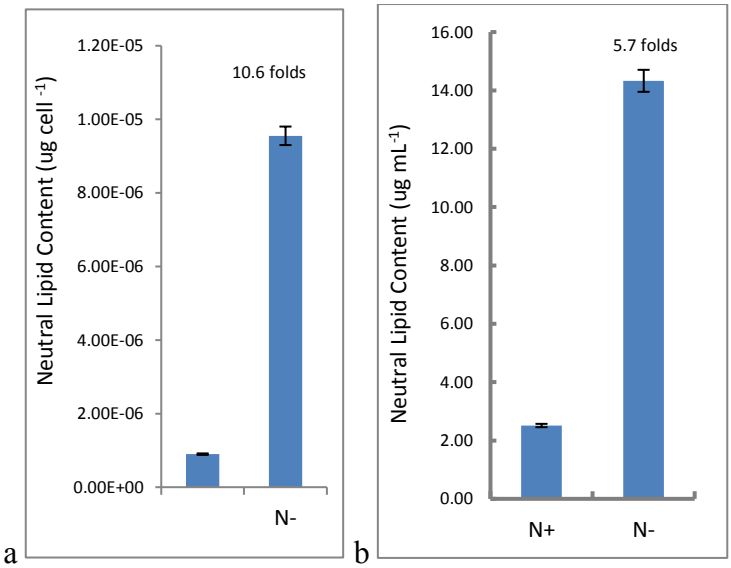


Figure 3.5 Neutral lipid measurement of cells cultured under normal condition (N+) and nitrogen deprivation condition (N-) by direct Nile red method.

(a) Neutral lipid content per cell basis. (b) Neutral lipid content per mL basis.

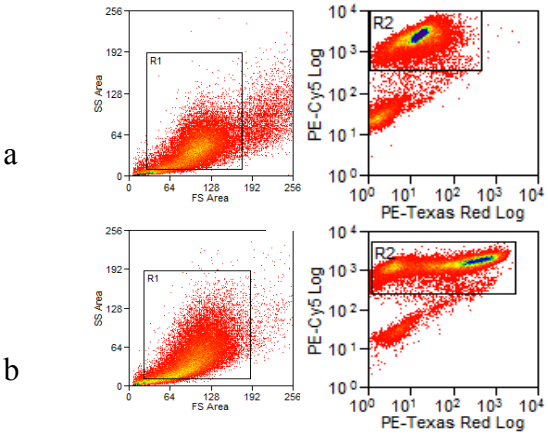


Figure 3.6 Neutral lipid measurement of cells cultured under normal condition and nitrogen deprivation condition by flow cytometry.

(a) Normal condition, mean = 18.28. (b) Nitrogen deprivation condition, mean = 276.3.

3.3.4 Visualization of neutral lipid droplet through fluorescence microscopy

To visualize the staining of Nile red on the neutral lipids, fluorescence microscopy was used to observe stained neutral lipids (Figure 4.4d). There appears to be a larger amount of neutral lipids visualized under the higher lipid strain compared to its control. This is similar to the observation of higher neutral lipid content determined quantitatively by the Nile red staining as mentioned earlier.

3.3.5 Comparison of Nile red assay with GC-MS measurement

We used this rapid TAG detection protocol of Nile red assays [55, 86, 90, 91] to track neutral lipid accumulation. As an attempt to use it as a standard method for routine rapid TAG detection, we compared quantification results obtained from the GC-MS analysis with those obtained from the Nile red assay. The Nile red data showed a good correlation with the GC-MS data (R^2=0.86 for D9 strain and WT strain, Figure 3.7; R^2=0.97 for G11_7 strain, R^2=0.92 for WT strain, Figure 3.8. D9 and G11_7 are two mutants generated in Chapter 4), indicating such an assay could potentially be used as a high-throughput screen for identifying the next generation of FA-overproducing mutant strains, which was also tested and suggested by Peng Xu [92].

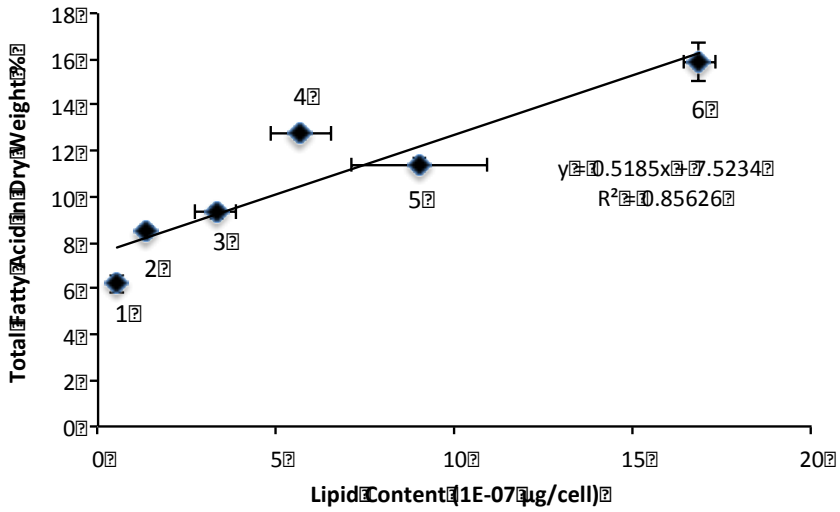
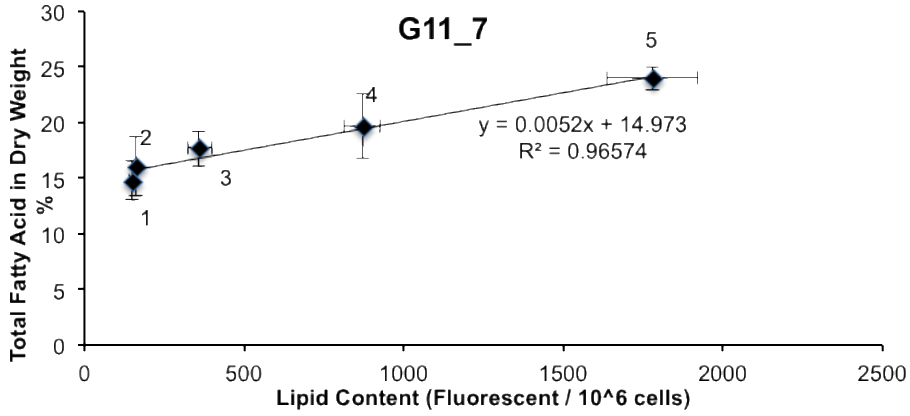
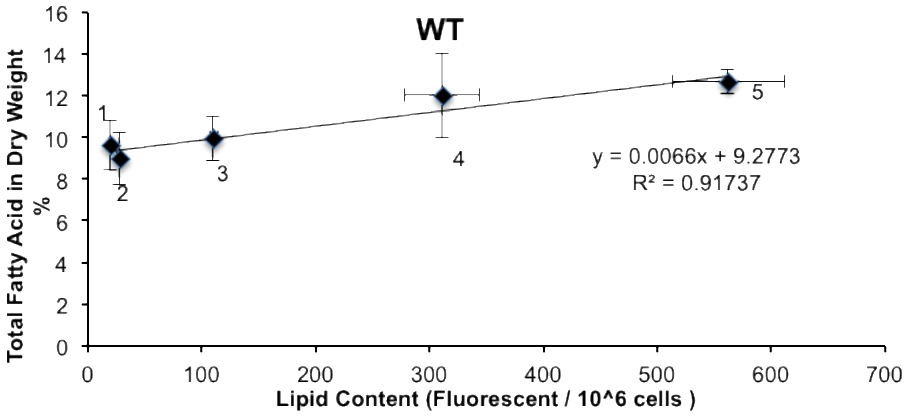


Figure 3.7 Correlation between Nile red assays and GC-MS measurements using D9 mutant and WT.

Horizontal error bars show standard deviation for Nile red assay; Vertical error bars show standard deviation for GC-MS assay. The linear line shows the correlation between these two methods. Strain genotypes: 1, 3, 5, represent culture day 8, 13, 16 for WT strain, respectively; 2, 4, 6, represent culture day 8, 13, 16 for D9 mutant strain. All experiments were performed in triplicate.

a





b

Figure 3.8 Linear regression of Nile red assay and GC-MS measurement.

(a) and (b), Horizontal error bars show standard deviation for Nile red assay. Vertical error bars show standard deviation for total FA from GC-MS measurement. The line shows linear regression between the two methods. Strain genotypes: 1, 2, 3, 4, 5 represent culture day 6, 8, 16, 28, 33 for G11_7 strain in a), respectively; 1, 2, 3, 4, 5 represent culture day 6, 8, 16, 28, 33 for WT strain in b). All experiments were performed in triplicate.

3.4 Conclusion

To facilitate rapid lipid screening in *D. tertiolecta*, the Nile red method has been developed which could benefit selection of high lipid producers from transformations to treatments. Moreover, the Nile red method also could be applied to other microalgae where additional treatments may needed for species with rigid cell walls. Currently, several other techniques, such as TAG kit, BODIPY staining and solvent methods could also be investigated into in the future.

Chapter 4 Generation of lipid-rich mutants

4.1 Introduction

The lack of genomic and transcriptomic information largely limited our investigating into specific genes for target genetic engineering. Herein, we used random insertional mutagenesis to generate TAG-rich mutants directly first. *D. tertiolecta* mutants D9 and G11_7 with enhanced neutral lipid production on different phases were generated in this study. In this chapter, detailed strategies to generate such biofuel-dependent oleaginous microalgal strains and screen for their lipogenesis at different growth stages will be discussed.

4.2 Materials and Methods

4.2.1 Culture strains and physiological study

D. tertiolecta cells were cultured under the low light condition and cell growth was traced as mentioned in Chapter 2.1. Photosynthetic rate measurement and photosynthetic efficiency evaluation, glycerol and starch content detection were performed following Chapter 2.4, 2.5, respectively.

4.2.2 Generation of *D. tertiolecta* mutants

Random insertional mutagenesis was performed for generating lipid-rich mutants in *D. tertiolecta*. Nile red quantification method was used for rapid screening according to TAG content. For different purpose, a D9 mutant with overall increased TAG production, and a G11 (0.G11) mutant with early TAG

accumulation (TAEP) were generated on different aspects as below. And subsequent studies showed that Nile red measurement in the tested range was parallel that of the total fatty acid measurement by GC-MS (Chapter 3.3.5).

Mutant with overall increased TAG production

Transformation using random insertional mutagenesis system was performed according to Chapter 2.8.2.1 via glass-bead method carrying a *ble* cassette. Genotyping PCR using ble_F/ bel_R was used for mutant library screening for positive insertions in 96-well plates. Positive mutants were subsequently transferred to 6-well plates for Nile red staining quantification. Mutants with overall higher TAG content were selected. Among all those mutants, a high-lipid rich mutant D9 with remarkably increase of TAG content was selected and applied in the development of the RNA-Seq pipeline in the following Chapter 5.

Mutant with TAG accumulation in exponential growth phase (TAEP)

D. tertiolecta transformation and random mutagenesis was conducted via electroporation [85] as previously described in Chapter 2.8.2.2. Lipid-rich mutant 0.G11 that had an early neutral lipid accumulation was selected.

Further optimization using flow cytometric analyses of *Dunaliella* mutant 0.G11 and its WT were performed on Nile red-stained cells on Beckman-Coulter CyAN. High-TAG sorting strategy was applied to enable enrichment of high-TAG mutant (0.G11) (5^{th} day of cultivation), by two rounds of FACS on pooled mutants stained with Nile Red using Beckman-Coulter Mo-Flo Legacy Cell Sorter [93]. The same strain treated with acetone was used as the

background. The distribution of an identically treated culture of WT was used as the control to determine the TAG gates for mutant pool sorting. A high-TAG gate that captured about 1.91% of WT with maximal Nile red fluorescence signal for a given chlorophyll fluorescence signal were performed (Figure 4.3a). The same gate captured about 18.99% of the cells in the 0.G11 mutant cells (Figure 4.3b). All cells that fell into this gate were collected into vials containing 0.5 mL 0.5 M NaCl ATCC medium. The cells were spun down at 750 × g for 5min, re-suspended in the fresh 0.5M ATCC medium, and plated onto 0.5 M ATCC 1.5% agar plates and incubated for two weeks. The colonies were subsequently transferred to a flask for sub-culturing. Mutant pool (1.G11) and WT pool on day 5 were collected and used for second round of FACS. The second cell sorting on 1.G11 mutant and WT pools was carried out using the same amount of cells at a higher-TAG gate (Figure 4.3c), and captured 11.07% of the cells in the mutant pool (Figure 4.3d). A wider Nile red and chlorophyll signal distribution in the pool of mutants indicates the presence of mutants that accumulate higher and lower amounts of TAG and chlorophyll. All cells that fell into this gate were collected into 96 well plates containing 0.5M NaCl ATCC media and incubated under the same culture condition. The cells from each well were then transferred to flasks for further TAG analysis. Kinetic studies of TAG accumulation in mutants and WT *D. tertiolecta* were carried out using Nile red quantitative assay together with GC-MS technique as described previously [86]. On culture day 6, G11_7 mutant and WT were harvested to visualize the lipid droplets using fluorescent microscopy (Olympus BX63, Tokyo, Japan), and measure the photosynthetic rate using an oxygen

electrode according to the operating manual (Rank Brothers, Bottisham, Cambridge, UK).

4.3 Results and Discussion

4.3.1 Screening and selection of lipid-rich mutants

Mutants of *D. tertiolecta* subjected to random insertional mutagenesis were generated by transformation using pGreenII0000 plasmid with a bleomycin selection cassette. Zeocin-resistant transformants were screened on 0.08M ATCC medium agar plates with zeocin. About 30 mutants resistant to zeocin were selected. One transformant with constantly enhanced lipid production from rapid TAG screening was selected and named D9 for further characterization. The bleomycin transgene was detected through genotyping PCR (Figure 4.1).

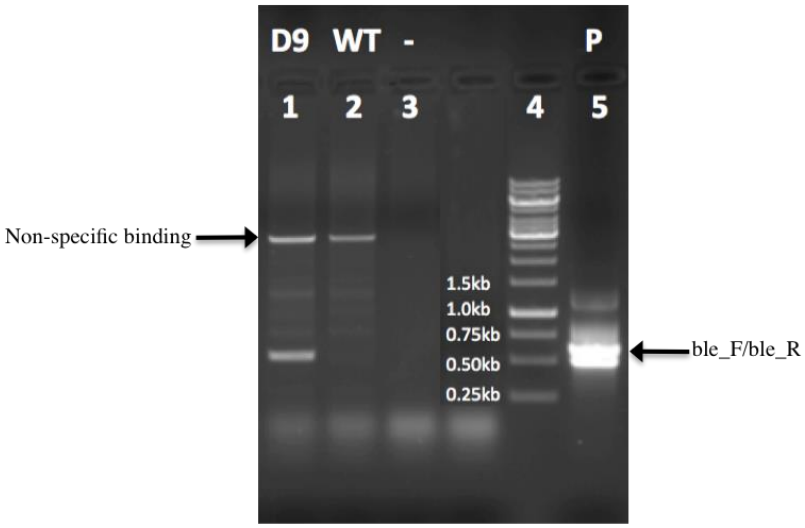


Figure 4.1 Genotyping PCR results of D9 mutant and WT *D. tertiolecta.*

The template for lane 1, 2, 3, 4, 5: D9 mutant, WT, negative control (without template), 1kb ladder, positive control (using the reconstructed plasmid as the template), respectively. The primer pair for lanes 1, 2, 3, and 5: ble_F + ble_R (the predicted size is 503bp). The two arrows indicate a non-specific

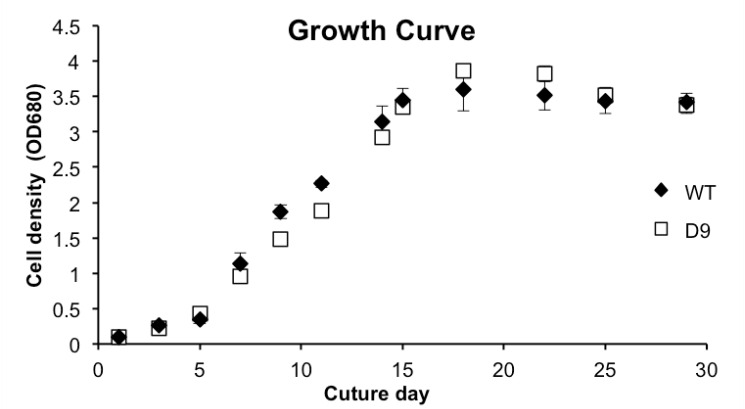
binding band and the target bleomycin band.

4.3.2 Physiological studies for the D9 mutant

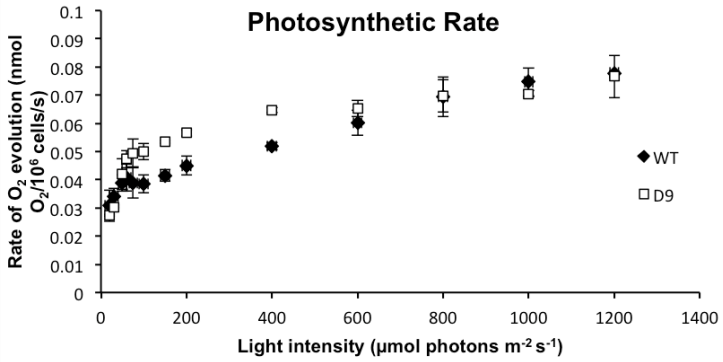
Growth kinetics of the D9 and WT *D. tertiolecta* cells grown in the low light condition were examined and shown in Figure 4.2a. In comparison with the WT, D9 shared the similar growth kinetics. Neutral lipids in D9 was examined and compared with that in the WT. As shown in Figure 4.2d, the D9 mutant produced neutral lipids at the late exponential and early stationary phases at about 2 to 4-fold of that in the WT, indicating that some carbon flow channeling may be occurring.

Photosynthetic activities (including rates of photosynthesis, photosynthetic efficiency and the maximum photochemical capacity) of the D9 and WT *D. tertiolecta* cells were investigated. D9 mutant in general showed increased photosynthetic performance (Figure 4.2b, c), with around 2-fold increase in photosynthetic efficiency compared to that of WT *D. tertiolecta.*

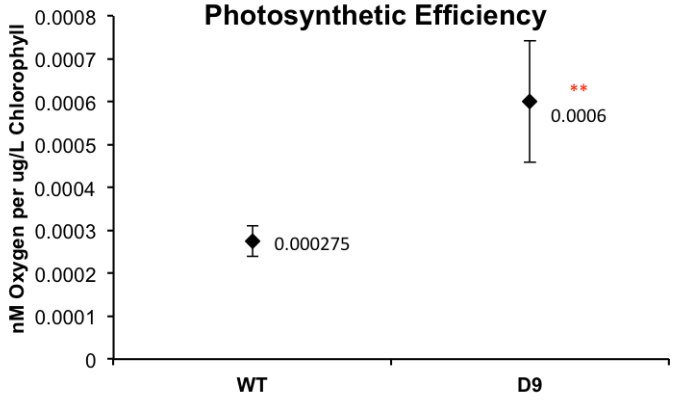
The glycerol production and starch content of D9 and WT cells at stationary phase were investigated. Regardless of the neutral lipids production and photosynthetic performance, glycerol production and starch contents in the two genotypes were comparable, as shown in Figure 4.2e-f. Lipids and starch are known as endogenous carbon storage compounds [23], and glycerol is the osmo-regulator for *Dunaliella* [94]. The enhanced photosynthetic efficiency resulted in a carbon flux to lipid synthesis and accumulation, rather than starch or glycerol production.


Growth Curve
Cell density (OD680)
Cuture day
WT
D9

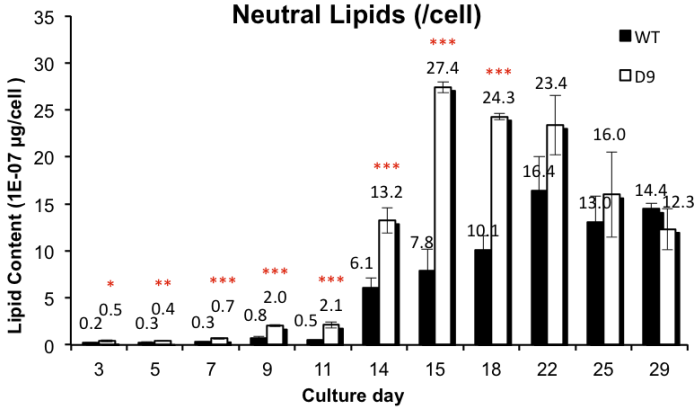
a


Photosynthetic Rate
Rate of O2 evolution (nmol O2/10^6 cells/s)
Light intensity (μmol photons m-2 s-1)
WT
D9

b


Photosynthetic Efficiency
nM Oxygen per ug/L Chlorophyll
0.000275
0.0006
**
WT
D9

c


Neutral Lipids (/cell)
Lipid Content (1E-07 μg/cell)
Culture day
WT
D9

d

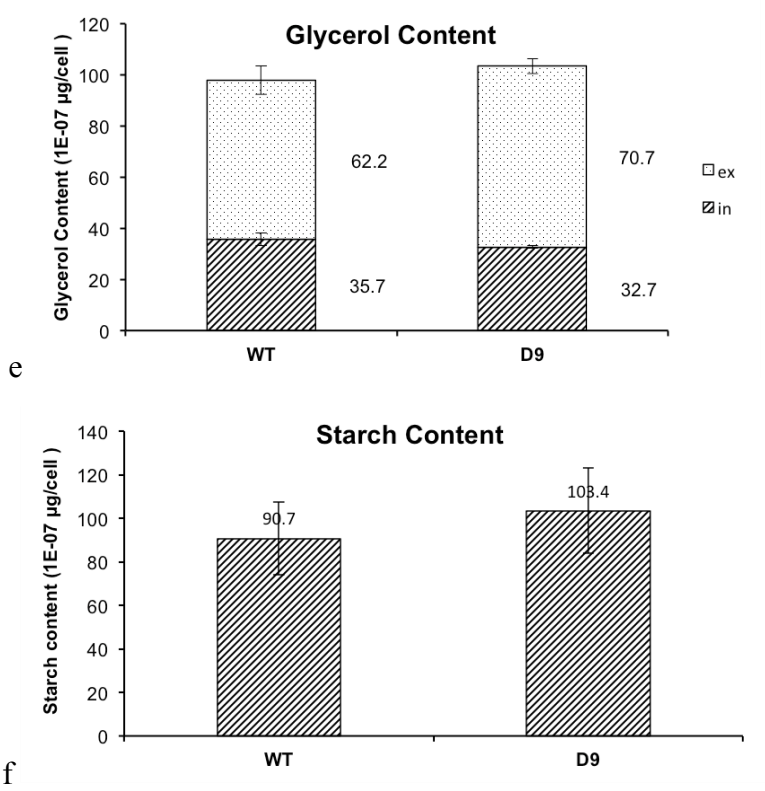


Figure 4.2 Physiological performance of D9 mutant and WT *D. tertiolecta.*

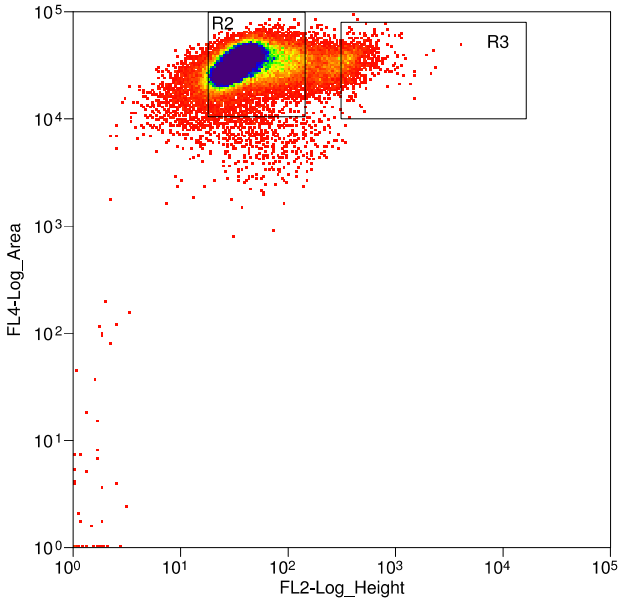
(a) Growth curve of D9 mutant and WT *D. tertiolecta*. (b) Photosynthetic rate and maximum photochemical capacity of D9 mutant and WT *D. tertiolecta*. (c) Photosynthetic efficiency of D9 mutant and WT *D. tertiolecta*. (d) to (f) Detection of carbon fixation parameters, namely, neutral lipid content, glycerol content, and starch content.

* indicating the statistically significant difference between D9 mutant and WT *D. tertiolecta* after two-tailed T test (* 0.01 < p value <= 0.05; ** 0.001 < p value <= 0.01; *** p value <= 0.001).

4.3.3 FACS enriched a pool of mutant strains and generation of G11_7

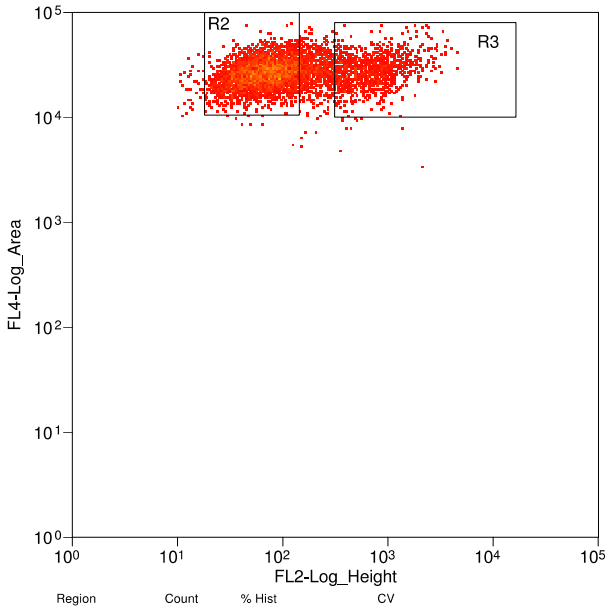
We generated TAG-rich mutant library of G11_7 via genetic engineering and two rounds of fluorescence-activated cell sorting (FACS) [93]. All the 27 isolated strains showed reproducible increase in Nile red signal (Figure 4.3e). Further observation on top 6 mutants showed consistent higher TAG content with a statistical significant P value < 0.01 (t test) comparing to the WT

(Figure 4.3f).



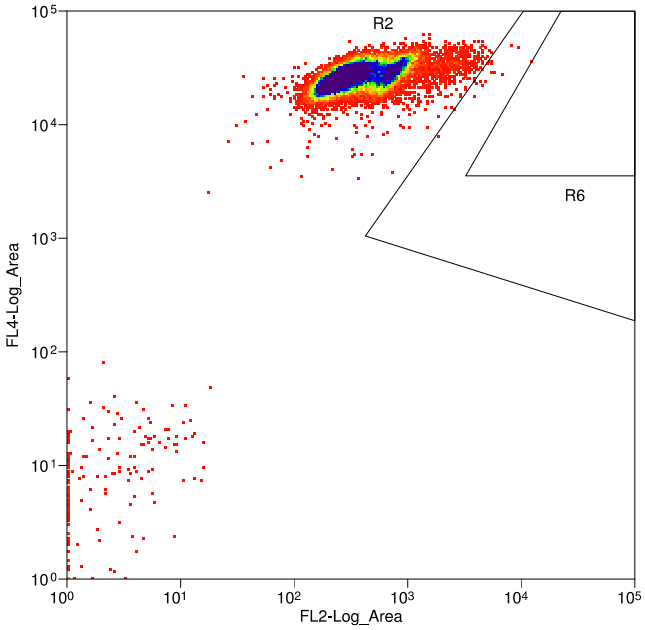
Region	Count	% Hist	CV
Total	66491	100.00	133.28, 27.04
R2	60469	90.94	51.91, 24.90
R3	1271	1.91	47.00, 24.24

a



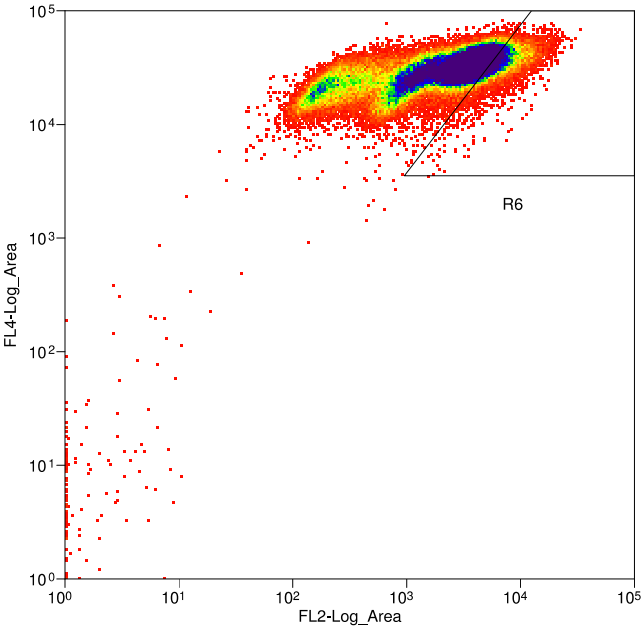
Region	Count	% Hist	CV
Total	5715	100.00	165.96, 31.69
R2	3826	66.95	45.75, 29.55
R3	1085	18.99	69.08, 34.15

b



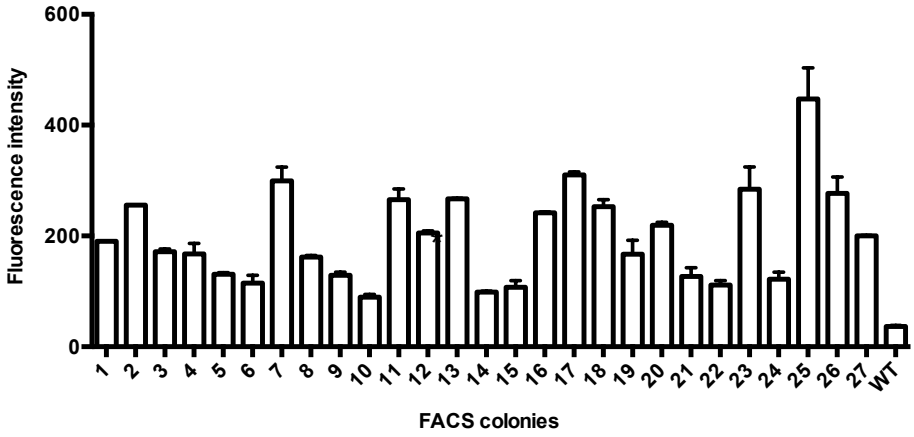
Region	Count	% Hist	CV
Total	40363	100.00	89.18, 22.54
R2	21	0.05	41.62, 33.60
R6	1	0.00	0.00, 0.00

c

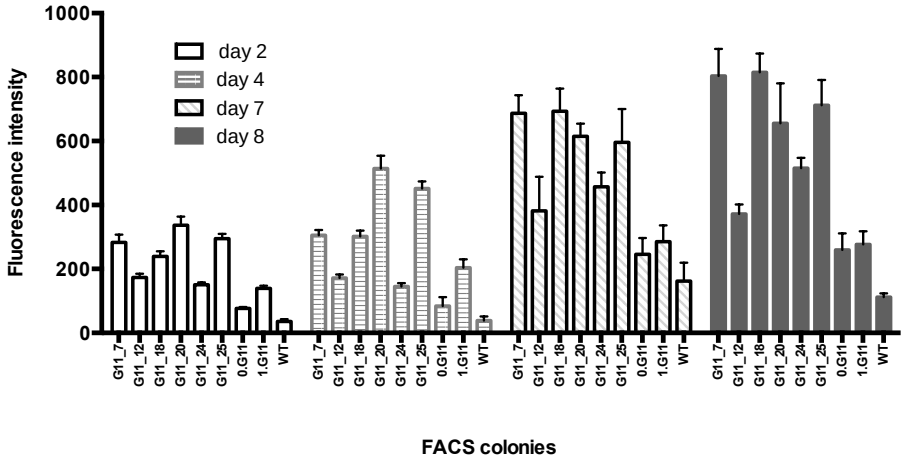


Region	Count	% Hist	CV
Total	100493	100.00	77.54, 29.95
R6	11122	11.07	37.94, 25.90

d



e



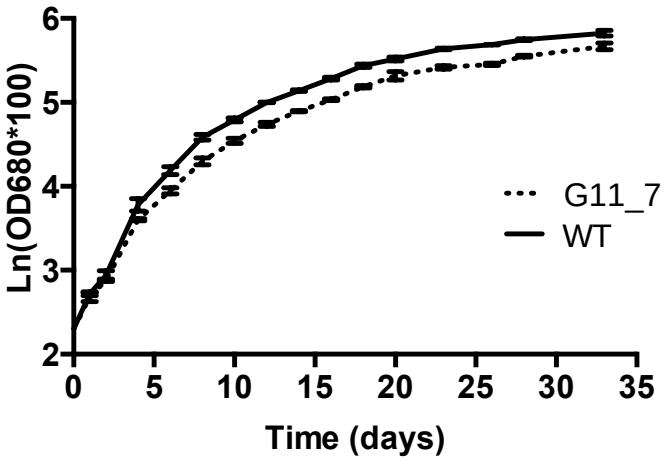
f

Figure 4.3 Fluorescence-activated cell sorting for mutant and WT *D. tertiolecta*.

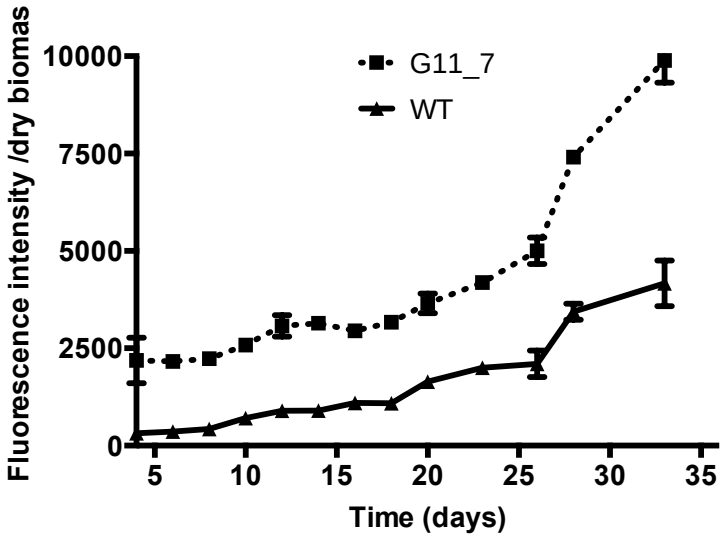
(a) Fluorescent cell sorting image from first sorted WT cells (R3). (b) Fluorescent cell sorting image from first sorted G11 cells (R3). (c) Fluorescent cell sorting image from second sorted WT cells (R6). (d) Fluorescent cell sorting image from second sorted G11 cells (R6). (e) Fast screening of *D. tertiolecta* colonies from FACS using Nile red staining assay. (f) Detailed screening of *D. tertiolecta* colonies with top TAG accumulation ability in biological triplicate using Nile red staining assay.

Among all the mutants, we selected the top mutant (G11_7) with highest TAG increase at its exponential growth phase for characterization. G11_7 mutant accumulated TAG at its exponential growth phase (Figure 4.4b) with marginal difference on growth (Figure 4.4a). It consistently showed enhanced TAG production (about 2-7 folds) compared to WT *D. tertiolecta* (WT). In

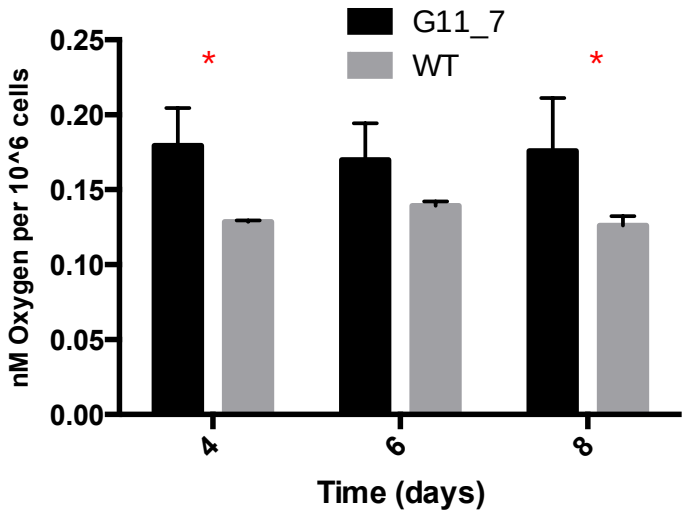
addition, the mutant had a significant better photosynthetic performance (Figure 4.4c, 33% higher than WT). It is evidence that the mutant has enhanced energy/carbon capture capacity. LDs were visualized by fluorescent microscopy (Figure 4.4d) in G11_7 mutant, showing golden-yellow fluorescence. The TAG accumulation in cultures grown under high light condition showed a similar trend as low light condition (Figure 4.4e, and Figure 4.4f).



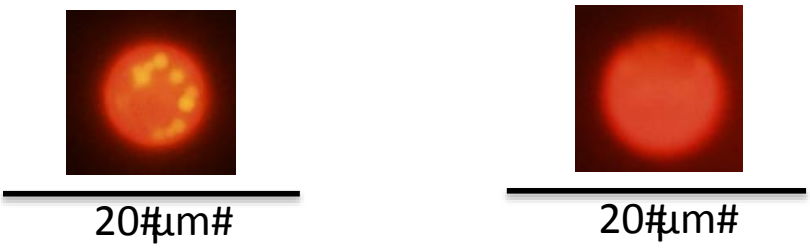
a

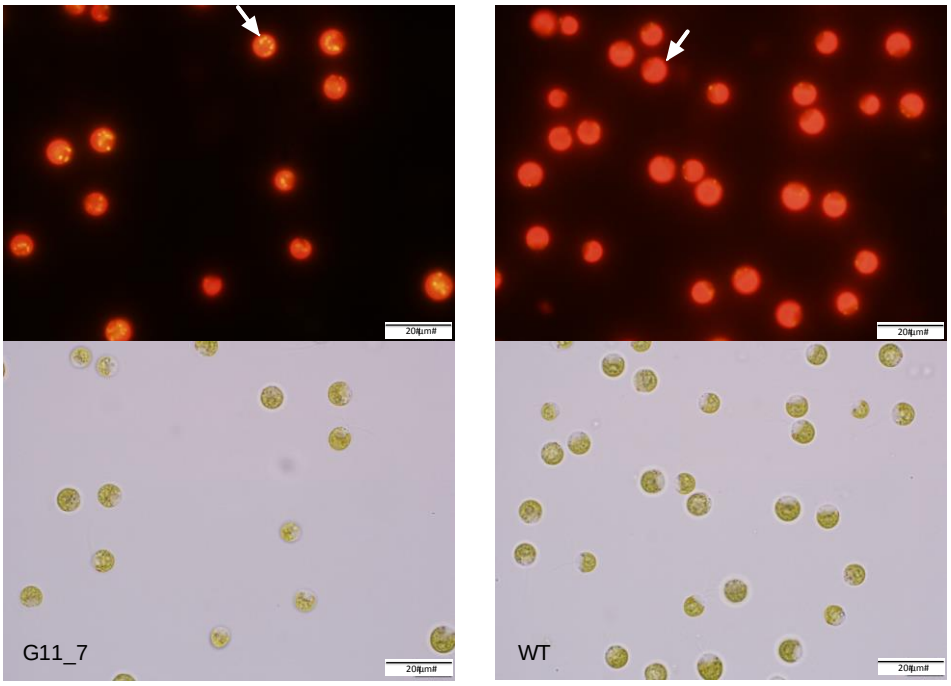


b

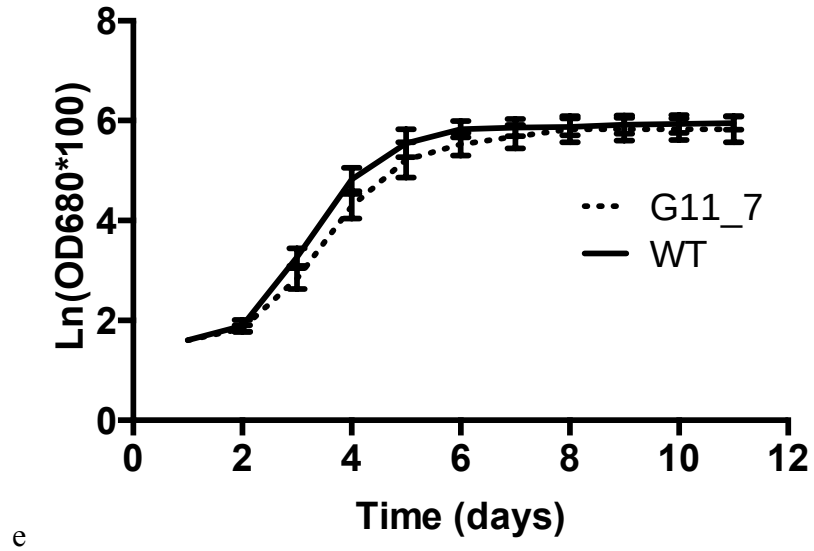

G11_7
WT
*
*
nM Oxygen per 10^6 cells
0.25
0.20
0.15
0.10
0.05
0.00
4
6
8
Time (days)

c

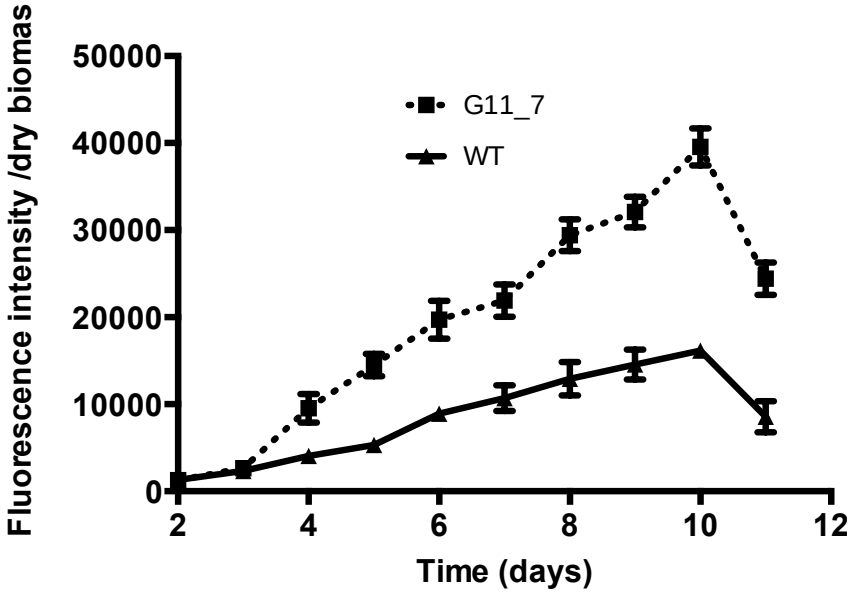

20#µm#
20#µm#


G11_7
WT

d



e



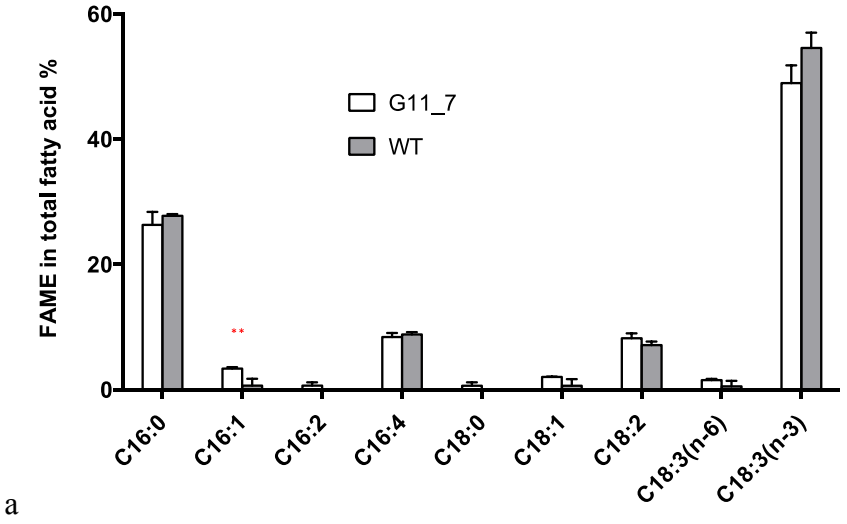
f

Figure 4.4 Physiological performance of G11_7 mutant versus WT *Dunaliella tertiolecta*.

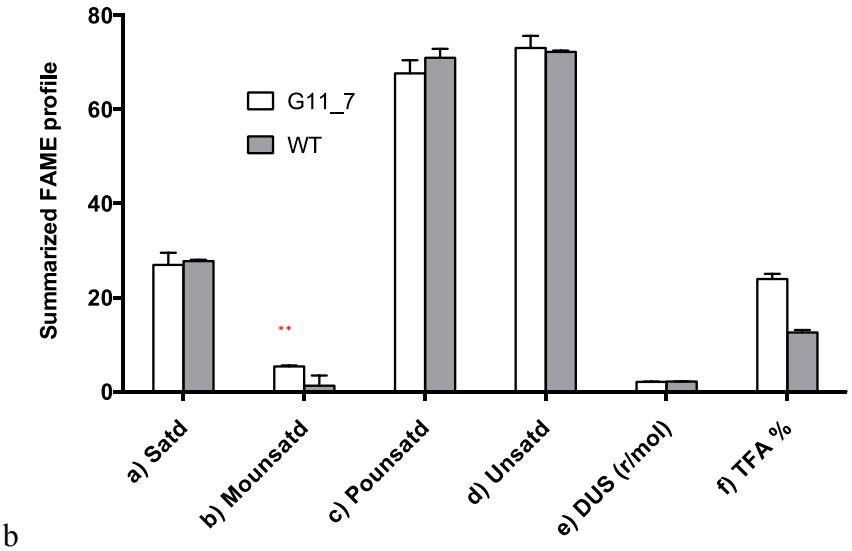
(a) Growth curve monitored by spectrophotometry under low light. (b) TAG quantitative assay by Nile red staining method under low light. (c) Photosynthetic rate of G11_7 mutant and WT *D. tertiolecta*. (d) Microscopy images (above, Nile red fluorescence; below, bright field) of G11_7 mutant and WT under low light (under 100 × objective). (e) Growth curve monitored by spectrophotometry under high light. (f) TAG quantitative assay by Nile red staining method under high light.

The FA composition of the mutant and WT are presented in Figure 4.5a, with a typical profile of unsaturated FAs 16:1 and 18:3(n-3) being the predominant

FAs. There is a significant increase in the monounsaturated FA (MUFA) with most others remained similarly (Figure 4.5b).



a



b

Figure 4.5 Fold change of G11_7 and WT FA profile.

(a) FA content is expressed as percentage of total FAs of G11_7 and WT (n = 3). (b) *Satd* saturated FAs, *Mounsatd* monounsaturated FAs, *Pounsatd* polyunsaturated FAs, *Unsatd* unsaturated FAs, *DUS* the degree of FA unsaturation = [1.0 × (% monoenes) + 2.0 × (% dienes) + 3.0 × (% trienes) + 4.0 × (% tetraenes)]/100.

Error bars, SEM. Statistical analyses were performed using Student t test, *P < 0.05, **P < 0.01, ***P < 0.001.

4.4 Conclusion

In this project, instead of using conventional target genetic engineering approach on one or multiple genes, random insertional mutagenesis and FACS techniques were applied to generate *D. tertiolecta* lipid-rich mutants. It is a feasible strategy to quickly and purposely screen for microalgal mutants of desire characteristics. Their lipid regulatory mechanisms and carbon flow channeling were investigated and reported below.

Chapter 5 Construction of Bag2D package and its application in characterizing D9 mutant

5.1 Introduction

To characterize the mutants generated, NGS (RNA-Seq) is a useful tool. However, the analysis of NGS data is highly dependent on the genome/ transcriptome database. Therefore, we develop an in-house program using Basic Local Alignment Search Tool, BLAST suite, by comparing with related green algal lineages and plant species to construct the draft transcriptomic database of *D. tertiolecta* from RNA-Seq analysis. The information revealed regulation of lipid synthetic pathways in the D9 mutant.

5.2 Materials and Methods

5.2.1 Construction of cDNA libraries for NGS analysis

Total RNA of D9, WT was extracted as mentioned in section 2.2. Approximately 2 μg of the resulting total RNA was used for synthesis of cDNA using the TruSeq Stranded Total RNA LT Sample Prep Kit (with Ribo-Zero Plant) (Illumina) according to manufacturer's instruction including synthesis of first and second strands cDNA, end repair, 3'-end adenylaton, adapter ligation, fragment enrichment (e.g., ~350 bps, in length), library validation and quantification. The libraries were sequenced using Illumina MISEQ Sequencer (Illumina, Inc., San Diego, CA, USA) according to the manufacturer's instructions. To check the variance from the effect of RNA-Seq technology on the mutant and WT samples, we extracted another set of

duplicate samples on Day 12, and the same process was carried out as the previous experiment group.

5.2.2 *De novo* assembly of Illumina short reads and RNA-Seq data processing

The FASTQ datasets of *D. tertiolecta* generated from Illumina MISEQ were imported into Partek® Flow® software (version 3.0 Copyright ©; 2014 Partek Inc., St. Louis, MO, USA) for quality assessment. The raw data were then trimmed from both ends based on the following parameter setting: Min read length = 25; Quality encoding = Auto detect; End min quality level (Phred) = 20. *D. tertiolecta* database was *de novo* assembled via Velvet and subsequently Oases assembler. The parameter used for Velvet assembly were: Hash length = 21; Expected coverage = 10; Max coverage = 500; Min coverage = 1; Min contig length = 50; Min long cutoff = 2; Max branch length = 100; Max divergence rate = 0.2; Max gap count = 3; Min read-pair validation =10. The parameter used for Oases assembly were: Coverage cutoff = 3; Min paired cutoff = 4; Min observed to estimate ratio = 0.01; Edge sensitivity = 0.01; Contig uniqueness = 3. A user-friendly script count_geneLength.py was used to determine the length of the assembled contigs.

The paired-end reads were mapped back to the assembled contigs. Data aligned to the transcriptome in RNA-Seq analysis were selected to estimate transcript abundance. Gene specific analysis was subsequently used to compare the samples from mutant D9 and WT at transcript level using the default setting (Poisson model was used). Read hits per contig were

normalized to RPKM (Reads Per Kilobase per Million mapped reads) used for estimation at transcription level. One-Way ANOVA analysis was used for the differentially expressed transcripts in at least one comparison for day 12 RNA-Seq data.

5.2.3 Comparisons among other green algae and high plant species

The FASTA files of transcript and protein sequences of *C. reinhardtii*, *V. carteri*, *C. subellipsoidea* C-169, *C. variabilis* NC64A were downloaded from JGI website (www.jgi.doe.gov), and *Arabidopsis thaliana* from the Arabidopsis Information Resource (www.arabidopsis.org) was used as reference sequences for alignment with *D. tertiolecta* transcripts as query sequences. The *D. tertiolecta* hit profiles among different species were compared.

5.2.4 Optimization of the *D. tertiolecta* database and Bag2D program

To make the *D. tertiolecta* database more complete and accurate, more MISEQ RNA-Seq data were included together with the NCBI published sequence information. Since the Bag2D program is in-house constructed, we validate it using a set of RNA-Seq data from *Chlamydomonas* as a benchmark.

5.2.5 Functional annotation of the *D. tertiolecta* contigs and biological interpretation

D. tertioleta is a species that has not been sequenced or annotated. Using the reference genome of related species such as *C. reinhardtii*, *V. carteri*, *C.*

subellipsoidea C-169, and *C. variabilis* NC64A, the percentages of hit for *D. tertiolecta* in Partek Flow STAR aligner were less than 5% (data not shown). To improve the hits, a novel Bag2D-workflow program using python language based scripts was developed. A typical run of Bag2D consists of two steps. From the precursor *de novo* assembly that was obtained from the Partek Flow using Velvet + Oases, a set of blast output was generated by BLASTX using the query information of *D. tertiolecta* transcript list and the reference protein sequences of related species with a customer e value of E-value < =1E-6. Subsequently, *D. tertiolecta* contigs were selected from the top hits, and genes with reference to the model organisms were constructed. The redundant transcripts were filtered using a "delete" program. With the remaining *D. tertiolecta* contigs, BLASTX was conducted for a second time. The corresponding gene ID, transcript ID, and protein name were sorted and the resulting files were used as a new 'annotation' file (on the author's Github page) to perform further gene analysis and biological interpretation by Partek® Genomics Suite® software (version 6.6 Copyright ©; 2014 Partek Inc., St. Louis, MO, USA). The two steps were carried out by the user-friendly scripts extract_blast.py, and get_geneID/ transcriptID/ proteinname.py (including count_geneLength.py). The steps of the Bag2D-workflow appear as red box in Figure 5.1. The gene specific analysis was subsequently imported into PGS with the annotation file. They were merged according to the name of *D. tertiolecta* contigs. The annotated data were used to perform GO enrichment with a modified *C. reinhardtii* GO annotation file downloaded from JGI website (www.jgi.doe.gov), using Fisher's Exact test with a restrict analysis of more than 2 genes. The representative pathways

were generated using pathway analysis tool, by Fisher's Exact test and the *C. reinhardtii* KEGG database (http://www.genome.jp/). The novel Bag2D program package and computationally processed *Dunaliella tertiolecta* draft transcriptome database and its annotation files are hosted at the author's GitHub page https://github.com/SPURc-Lab/NGS-D9 with the step-by-step user manual for public access. The raw data from RNA-Seq and processed files were deposited into GEO with accession number of GSE70876. We have made the simulation data for this experiment available on the website.

5.3 Results and Discussion

5.3.1 Pre-analysis and *de novo* assembly of sequenced data

The RNA-Seq data analysis workflow for the non-model species used in this study is illustrated in Figure 5.1, which includes upstream cell culturing, harvest and downstream data interpretation steps.

The paired-end reads (130 bps in length/read) from each library were generated using Illumina MISEQ Sequencer. The short paired-end reads from D9, WT were pooled and examined with pre-alignment QA/QC and a secondary pre-alignment QA/QC after trimming was subsequently used. QA/QC reports mentioned throughout the paper are presented in Table 5.1. The trimmed *D. tertiolecta* short paired-reads were merged and subjected to the assembly programs, Velvet and Oases.

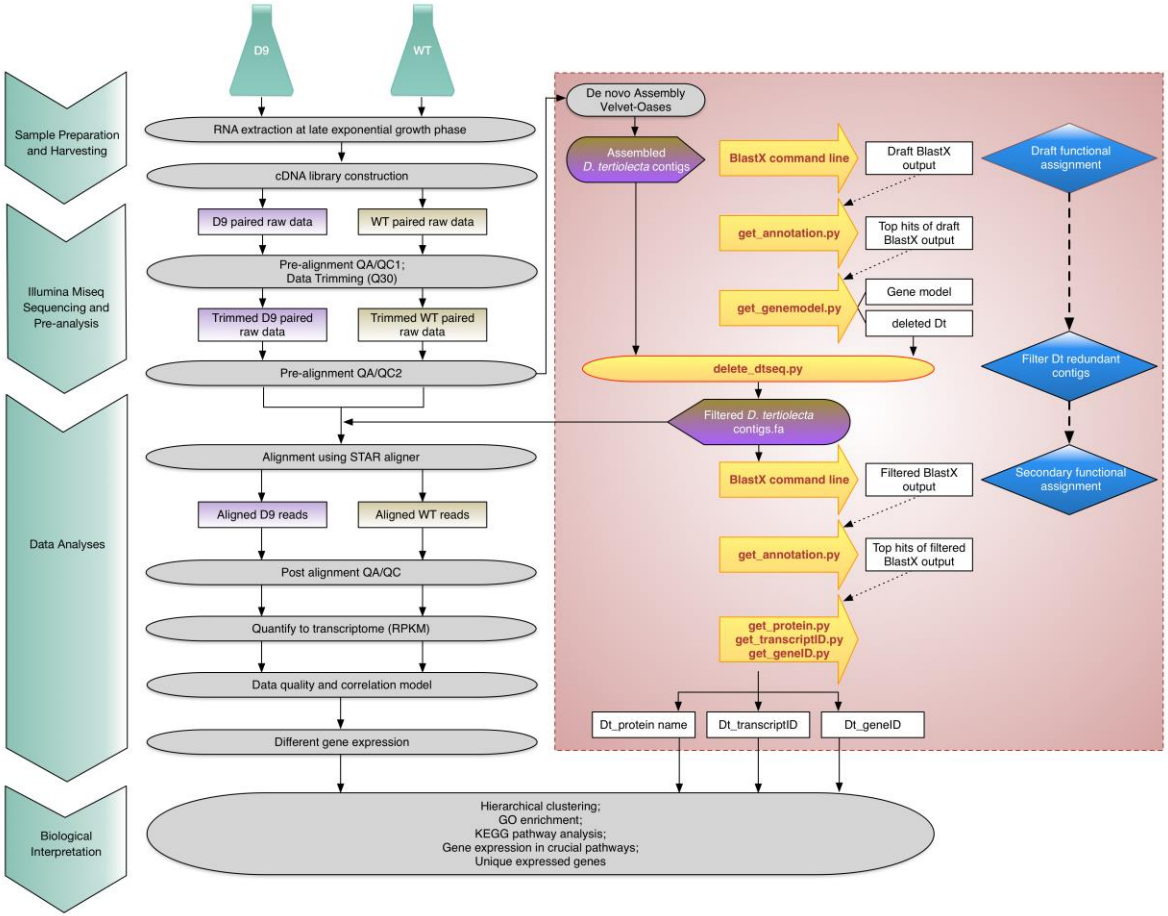


Figure 5.1 RNA-Seq data analysis flowchart.

Table 5.1 Run summary on the Illumina MISEQ platform.

Sample name	Sequencing stats and pre-alignment QA/QC of raw data /pre-alignment QA/QC of post-trimming data					Post-alignment QA/QC after alignment	
	Total reads	Avg. read length	Avg. read quality	% N	% GC	Total reads	Aligned
D9_S7_L001_R1	954,340/ **954,300**	137.28/ **137.2**	38/ **38.01**	9.62E-04 /**0**	47.74/ **47.73**	954,300	56.72%
D9_S7_L001_R2	954,340/ **954,300**	137.07/ **136.98**	37.69/ **37.7**	1.04E-03 /**0**	47.67/ **47.66**		
WT_S8_L001_R1	1,084,599/ **1,084,546**	136.67/ **136.59**	38/ **38.01**	1.11E-03 /**0**	47.41/ **47.4**	1,084,546	57.24%
WT_S8_L001_R2	1,084,599/ **1,084,546**	136.36/ **136.28**	37.76/ **37.78**	1.25E-03 /**0**	47.34/ **47.33**		

QA/QC: Quality assurance and quality control.

Avg. read quality > 30 indicating 99.9% accuracy or greater.

5.3.2 Functional annotation of the genes

Various strategies in optimizing growth conditions and phases have been suggested to stimulate production and accumulation of lipids or starch; these include expression of genes involved in lipids or starch biosynthesis, and the maximization of the diversity of expressed genes [41, 43, 76, 95-97]. To verify this and maximize the assembled outputs, sequencing results using different datasets for *de novo* assembly were compared: sequencing data from (i) WT only, (ii) WT and D9 mutant, (iii) enlarged Dt database with additional *D. tertiolecta* sample groups (Dt_v1, obtained from a WT *D. tertiolecta* strain cultured in a different growth phase, another random insertional mutant from *D. tertiolecta*) were then separately used to generate reference contigs and gene construction. The contigs, hit percentage and resulting transcriptomes were compared in Table 5.2, among which the (iii) enlarged Dt database Dt_v1 stood out with more complete transcriptome information, and was used for the following sections. The command line of BLASTX (from NCBI) was applied to the FASTA data generated from the *de novo* assembly, to obtain their functional assignments with reference to *C. reinhardtii* protein (Chlre4_best_proteins.fasta, Creinhardtii_169_peptide.fa, Creinhardtii_281_v5.5.protein.fa). By extracting the top hits with a BLASTX E-value <= 1E-6, *D. tertiolecta* contigs were annotated. However, there were redundant *D. tertiolecta* contigs assembled with different lengths at distinct positions and junctions. To avoid redundant gene modeling, a filtering step was included with the E-value, confidence of assembled contigs and length of the contigs as the thresholds to select the top hit contig for the unique

annotation name. The complete program used here was named as Bag2D-workflow (Blast1-annotation1-gene model1-Delete redundant genes-Blast2-annotation2-gene model2, namely protein name, transcriptID, geneID), using python language based scripts.

Table 5.2 Determine the reference transcripts through de novo assemblies using different sets of data.

No.	Samples used for assembly	Size of the assembled file (MB)	Gene model	Number of non-redundant ranscripts
(i)	WT	9.4	3,818	14,903
(ii)	WT+D9	16.9	5,360	23,564
(iii)	Enlarged Dt-1	45.8	7,023	47,276

Although the contigs with E-value <= 1E-6 may not be real transcripts, we decided to keep them to avoid complications with false negatives, accepting that the incidence of false positive could increase. In addition, different cutoff expectation values were set and compared afterwards.

5.3.3 Comparison of global transcriptome of D9 mutant with WT

The gene specific analysis generated from Partek Flow was imported into Partek Genomics Suite for statistical analyses. The data quality of log2-transformed RPKM values for these two RNA-Seq datasets was checked using the sample histogram. The peak and trend of the curves indicated that the normalization method was suitable for this dataset and the RNA-Seq data were of high quality and similarly distributed. Gene list was created with threshold of false discovery rate (FDR) <=0.05, multiple fold change (FC) >= 1.5 or FC <= -1.5 (to detect as many genes as possible that is different between the mutant and WT), and subsequently merged with the above

generated annotation file with the common column of *D. tertiolecta* contig names (results were available in GEO with accession number of GSE70876). The annotated gene list created was used for further biological interpretation. Gene ontology (GO) enrichment categories also provided insights into differentially distributed protein functional families and categories. The up-regulated GOs were highly represented in the photosynthesis and ATP synthesis families, as shown in Figure 5.2. FA synthesis requires stoichiometric amount of ATP, acetyl CoA and NADPH for each two carbon added to the growing acyl chain [98]. Light-driven electron transport is coupled to ATP synthesis in chloroplasts [99]. Photosynthetic reactions are thus essential not only in providing a carbon source but also in generating reducing power (NADH and NADPH) and energy (ATP) for FA synthesis [98]. According to this, mechanisms for enhanced lipid production could be elucidated, while comparable growth rate was observed between D9 mutant and WT cultures.

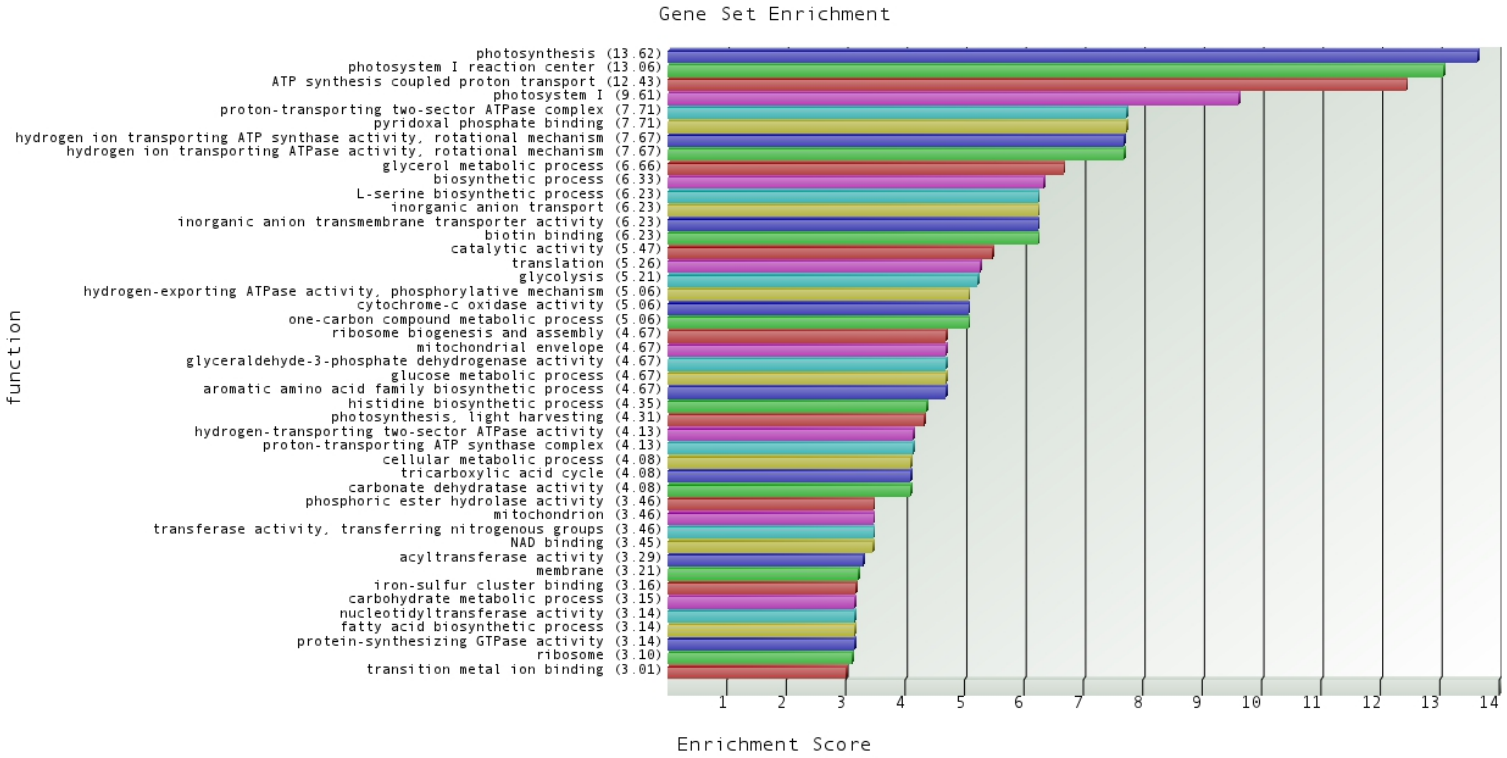


Figure 5.2 GO enrichment result of up-regulated genes in D9 mutant compared to WT *D. tertiolecta.*

The KEGG pathway analysis showed that inositol phosphate metabolism, FA biosynthesis, biosynthesis of secondary metabolites, FA metabolism, purine metabolism were significantly correlated with higher lipid yield based on the cutoff of enrichment p-value less than or equal to 0.05. From the pathway analysis, the top two pathways with low-abundance contigs were the inositol phosphate metabolism and FA biosynthesis pathways.

From the inositol phosphate metabolism, the gene that encoding myo-inositol oxygenase (MIOX) [EC: 1.13.99.1] showed significant up-regulation, which led to enhanced generation of D-Glucuronate as the substrate for both ascorbate acid synthesis pathway and generation of D-Glucarate. FA biosynthesis was the second top hit pathway. The up-regulated gene coding for ACCase [EC: 6.4.1.2] was detected in spite of a down-regulation of the gene (*FabG*) coding for 3-oxoacyl-[acyl-carrier protein] reductase [EC: 1.1.1.100], which is involved in the FA chain elongation reactions.

In Figure 5.3, we attempted to correlate the transcriptomic data with metabolite pathways. Changes within the metabolic intermediates with the metabolic pathways between D9 mutant and WT *D. tertiolecta*, namely the inositol phosphate metabolism, ascorbate acid synthesis, citrate cycle, pyruvate metabolism, and FA biosynthesis pathways were reconstructed based on the information from *C. reinhardtii* ChlamyCyc database [100] and the transcript abundance of enzymes in *D. tertiolecta*.

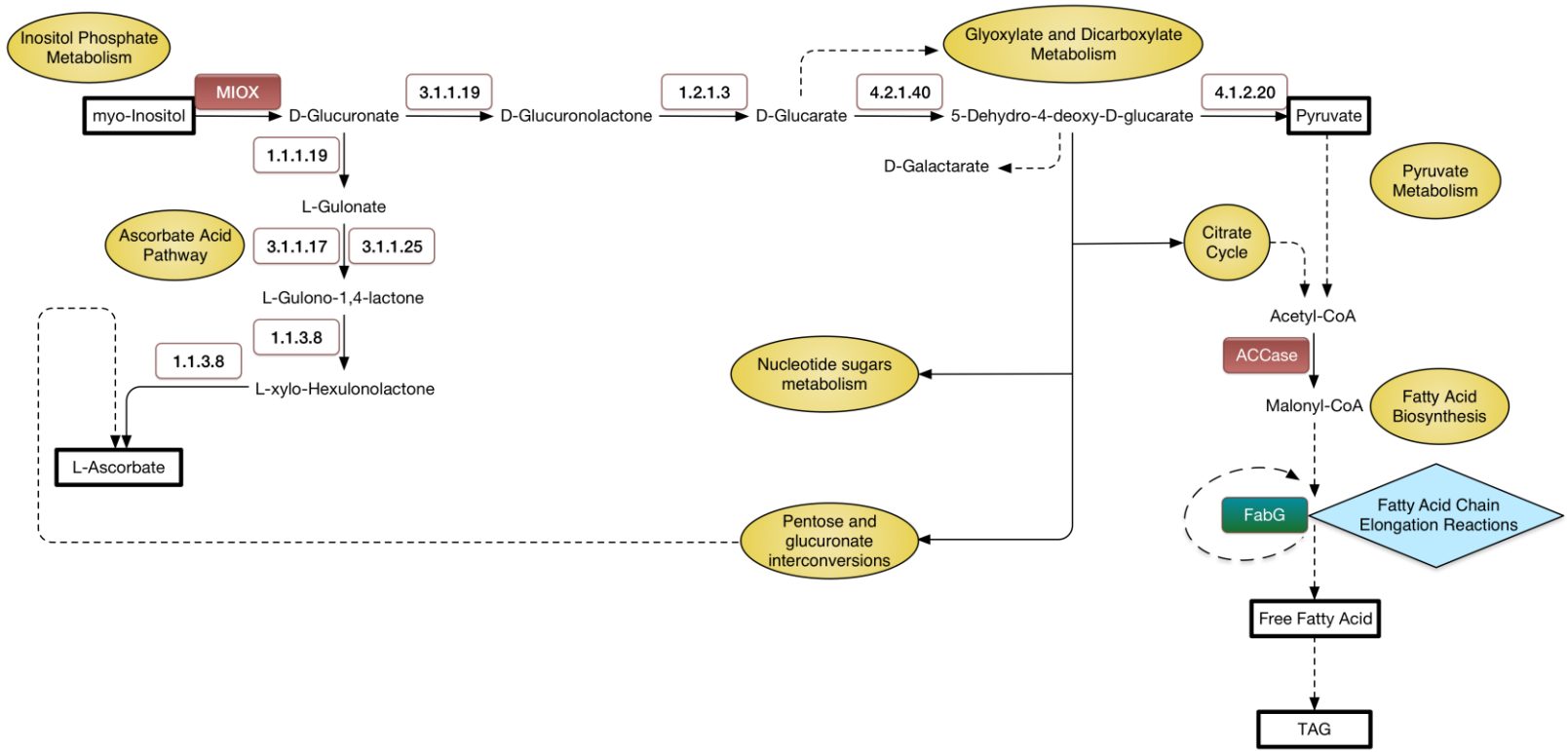


Figure 5.3 Pathway analyses for mutant D9 and WT *D. tertiolecta.*

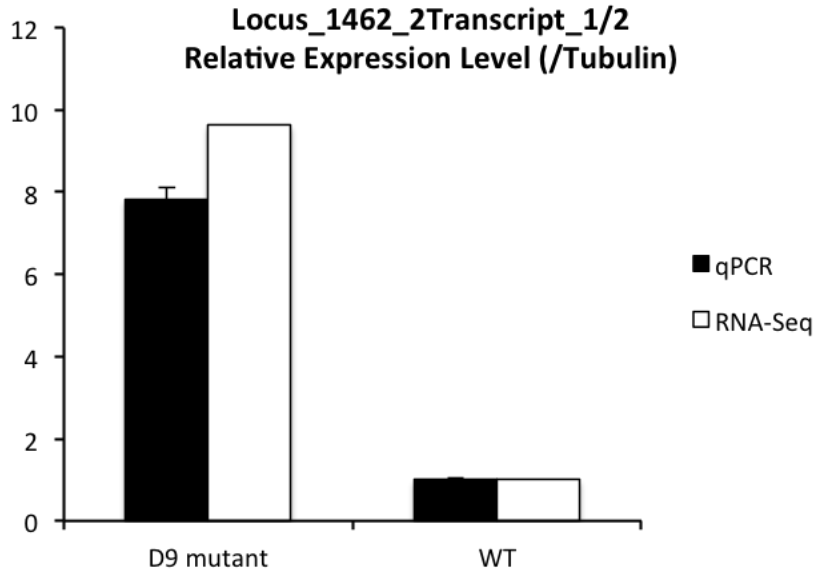
It is noted that the horizontal branch in the pathways in Figure 5.3 goes into the production of pyruvate. No significant differentially expressed genes in glyoxylate and dicarboxylate metabolism were identified in D9 mutant as compared to WT.

Up-regulation of the gene coding for ACCase was observed. As ACCase catalyzes the first reaction of the FA biosynthesis pathway, namely the formation of malonyl CoA from acetyl CoA and CO_2, its up-regulation may channel carbon flux to FA synthesis. This suggests that concurrent over-expression of genes particularly in ACCase as the first committed step along the Inositol Phosphate Pathway and the FA Biosynthesis Pathway will enhance neutral lipid production in algae.

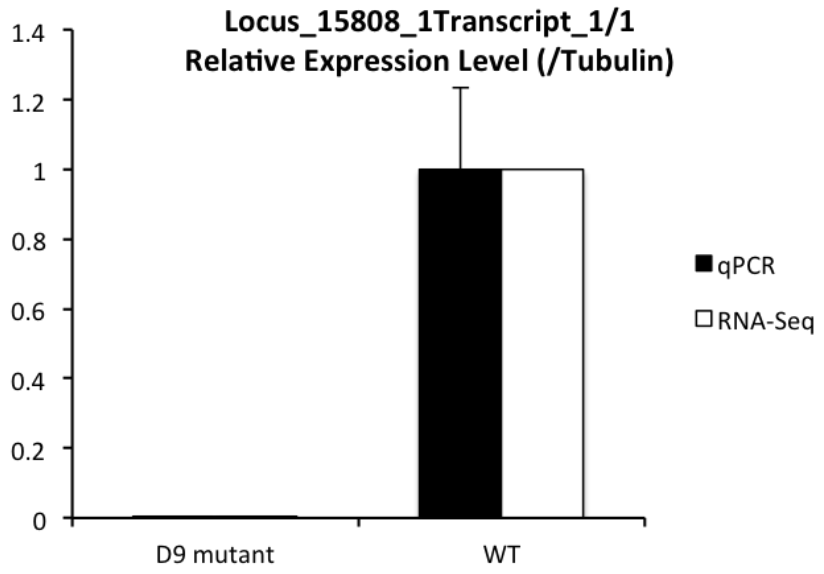
To form a saturated FA, the 3-ketoacyl ACP product is reduced by the enzyme FabG [101]. Herein, we predict that the down-regulation of the gene coding for FabG in the D9 mutant may change the length and saturation level of the FA chain, leading to accumulation of more short chain FAs or unsaturated FAs. The short chain and unsaturated FAs were indeed found to increase among the FAs in the mutant. The 16- or 18-carbon FAs are formed by a series of two-carbon chain elongating reactions catalyzed by a multi-subunit enzyme in most plants and algae [17]. Viscosity increases with FA chain length [102], so that most plant TAGs have a viscosity range that is much higher than that of conventional diesel [103]. The higher viscosity results in poor fuel atomization in modern diesel engines, leading to problems derived from incomplete combustion such as carbon deposition and coking [104]. Medium chain FAs (C8-C14) are favorable for production of

biofuels, for they have properties that mimic current diesel fuels and they improve fluid characteristic of the liquid fuel [17, 105].

Pre-analysis for an additional D9 biological duplicate RNA-Seq data for validation was presented in Appendix 3. With more stringent criteria of FC >= 2 or FC <= -2, and FDR <= 0.05, we found an overall similar pattern from gene expression profiles (available in GEO with accession number of GSE70876) and KEGG pathway analysis results in Appendix 4. The expression levels of many genes in the metabolic pathways, photosynthesis pathways, carbon fixation pathways, and FA pathways were increased. Two randomly selected genes showed similar expression patterns in both real-time PCR and RNA-Seq (Figure 5.4). This batch of experiment gave us a good hint that our results from the previous RNA-Seq experiment are reproducible.



a



b

Figure 5.4 Comparison of gene expression profiles from real-time PCR and RNA-Seq.

(a) The ratio of Locus_1462_2Transcript_1/2 expression level in D9 compared to WT *D. tertiolecta* is 7.8 from qPCR and 9.6 from RNA-Seq. (b) The ratio of Locus_15808_1Transcript_1/1 expression level in D9 compared to WT *D. tertiolecta* is 6.2E-04 from qPCR and 1.1E-06 from RNA-Seq.

5.3.4 Comparison of *D. tertiolecta* transcriptome database with other microalgae

Contigs from aforementioned Dt-1 database were compared to the transcripts/ proteins from five other chlorophyta species with known genome sequences (*C. reinhardtii, C. subellipsoidea* C-169*, O. lucimarinus, V. carteri* and *C. variabilis* NC64A) by BlastN/ BLASTX. Compared with BLASTX results, BlastN results retrieved fewer hits (Table 5.3). For example, 7,023 *D. tertiolecta* contigs matched *C. reinhardtii* proteins using BLASTX with a cut-off E-value of 1E-6 indicating 35.97% of the *C. reinhardtii* proteins could be assigned in *D. tertiolecta.* While only 611 (3.13%) *D. tertiolecta* contigs matched *C. reinhardtii* using BlastN with cut-off E-value of 1E-6.

Similar patterns were also observed in the Blast results from the other three green algae at cut-off E-values of 0, 1E-10, 1E-6, 1E-3, respectively. This result indicates that the sequences of microalgae are more conserved at the protein level than that at the nucleotide level; therefore, BLASTX results were subsequently used in this study for gene annotation.

Table 5.3 Comparisons of homologues between *D. tertiolecta* and other five green algae with reference genome sequences.

Algal reference	***Cre***	***Csu***	***Olu***	***Vca***	***ChlN***
Reference Transcripts or proteins #	19,526	9,629	7,796	15,285	9,791
BLASTN					
Transcript hit (E-value = 0)	82 (0.42%)	18 (0.19%)	0 (0.00%)	24 (0.16%)	20 (0.20%)
Transcript hit (E-value <= 1E-10)	574 (2.94%)	229 (2.38%)	15 (0.19%)	246 (1.61%)	281 (2.87%)
Transcript hit (E-value <= 1E-6)	611 (3.13%)	249 (2.59%)	16 (0.21%)	269 (1.76%)	307 (3.14%)
Transcript hit (E-value <= 1E-3)	654 (3.35%)	252 (2.62%)	16 (0.21%)	278 (1.82%)	315 (3.22%)
BLASTX					
Transcript hit (E-value = 0)	583 (2.99%)	356 (3.70%)	205 (2.63%)	523 (3.42%)	239 (2.44%)
Transcript hit (E-value <= 1E-10)	6610 (33.85%)	5025 (52.19%)	4002 (51.33%)	6249 (40.88%)	5106 (52.15%)
Transcript hit (E-value <= 1E-6)	7023 (35.97%)	5643 (58.60%)	4592 (58.90%)	6986 (45.70%)	5773 (58.96%)
Transcript hit (E-value <= 1E-3)	8561 (43.84%)	6393 (66.39%)	5267 (67.56%)	8162 (53.40%)	6604 (67.45%)

***** The number before and within brackets mean: the number of *D. tertiolecta* transcripts with Blast hits from reference alga and the hitting percentage in the corresponding reference transcripts/ proteins.

Cre *C. reinhardtii,* ***Csu*** *C. subellipsoidea,* ***Olu*** *O. lucimarinus,* ***Vca*** *V. carteri,* and ***ChlN*** *C. variabillis* N64A.

With the decrease of E-value stringency from 0, 1E-10, 1E-6, 1E-3, the BLASTX hit number for each transcript increased and the percentage of total contigs matching reference sequences increased (Table 5.3). To exclude possible false positives, only those contigs with E-value of 0 from BLASTX results were used for identification of *D. tertiolecta* true genes. The results indicate that these six green algae have 115 contigs in common, and a total of 640 genes in *D. tertiolecta* could be annotated and identified (the overlaps and unique areas in the Venn diagram in Figure 5.5) in the five references, while 583 contigs were in common with *C. reinhardtii*. From the Venn diagram, some *Dunaliella* genes could not match that of the *Chlamydomonas* genes, but they do overlap with genes from the other four model algae. Through a comparison study on E-value from 0 to 0.1, the number of corresponding transcripts was plotted out in Figure 5.6, and 1E-6 was used as the threshold from the distribution of the hitting contigs. Therefore, a modification process was applied with a combination of identified *D. tertiolecta* transcripts from the BLASTX results with the above five green algae, by increasing E-value number to 1E-6 BLASTX for the construction of *D. tertiolecta* database.

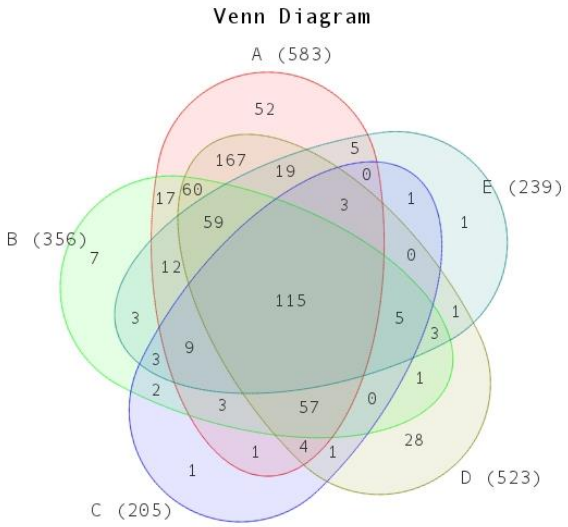


Figure 5.5 Venn diagram of the numbers of *D. tertiolecta* transcripts with BLASTX hits from five model algae.

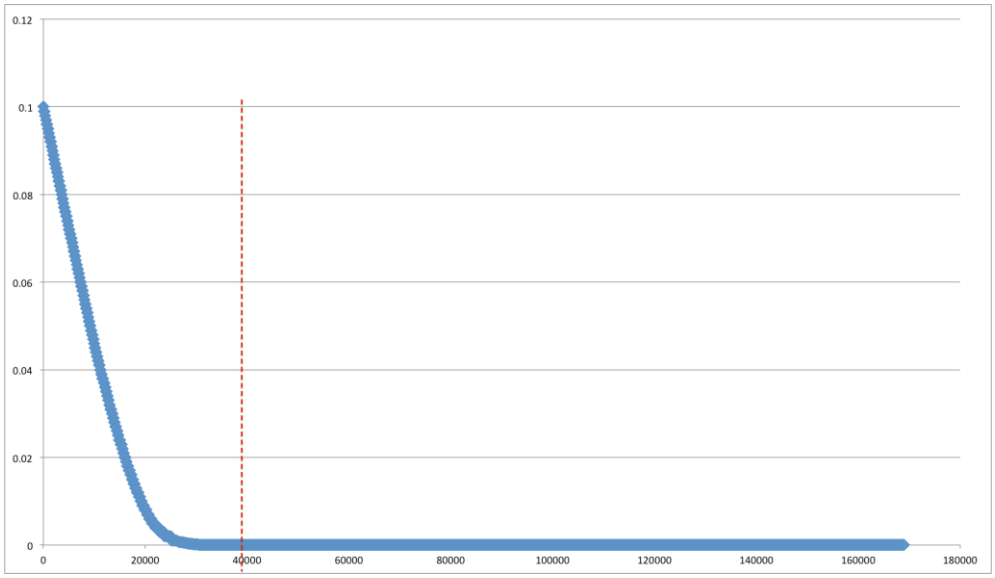


Figure 5.6 Length distributions of the contigs from Dt_v10.

5.3.5 Validation and optimization of the *D. tertiolecta* transcriptomic database

To validate the accuracy of the *D. tertiolecta* transcriptomic database from

Bag2D, reported *D. tertiolecta* nucleotides on National Center for Biotechnology Information (NCBI) were compared. The transcripts with the same *Chlamydomonas* protein annotation were compared in Table 5.4. The one with a lower E-value compared to the reference species will be used in the final *D. tertiolecta* database. From such comparison, we found that most contig sequences from Bag2D were identical to the NCBI published one, concluded from the high alignment scores generated from multiple alignment tools. Though we found some low alignment score pairs, which might be caused by the existence of isoforms or errors of the online available sequences. Besides, we found there are some redundant nucleotides, like gi|46981381|gb|AY575952.1| *Dunaliella tertiolecta* assimilatory nitrate reductase (nar) gene (partial cds), gi|311818483|emb|HH768845.1| Sequence 5032 from Patent EP2221382, gi|18913154|gb|AY078279.1| *Dunaliella tertiolecta* assimilatory nitrate reductase (nar) mRNA (complete cds), with different NCBI ID numbers and names, which might cause misunderstanding. Some new genes have been found and the sequences are available, e.g., acetyl-coa carboxylase beta-carboxyltransferase subunit of plastidic multimeric ACCase, with the name of >Locus_2576_1Transcript_1/1_Confidence_1.000_Length_1955. The *D. tertiolecta* database was optimized with a better sequence or the original sequence elongated.

Table 5.4 Comparison studies of transcripts from Bag2D program and NCBI.

D. tertiolecta **transcripts from Bag2D**	**E-value**	**Length (bps)**	***D. tertiolecta*** **nucleotide from NCBI**	**E-value**	**Length (bps)**	**Aligned score**	**Elongated**
>Locus_1123_4Transcript_1/2_Confidence_0.800_Length_1500	0	1500	>gi|46981381|gb|AY575952.1| Dunaliella tertiolecta assimilatory nitrate reductase (nar) gene, partial cds	2e-18 0	1313	-	
			>gi|311818483|emb|HH768845.1| Sequence 5032 from Patent EP2221382	0	2531	90.6	
			>gi|18913154|gb|AY078279.1| Dunaliella tertiolecta assimilatory nitrate reductase (nar) mRNA, complete cds		3447	90.6	+
>Locus_96_4Transcript_1/1_Confidence_1.000_Length_1881	0	1881	>gi|3869303|gb|AF065142.1| Dunaliella tertiolecta glutamine synthetase mRNA, partial cds	2e-109	601	92.8453	
>Locus_797_6Transcript_1/1_Confidence_1.000_Length_1463	1e-144	1463	>gi|2645974|gb|AF034201.1| Dunaliella tertiolecta proliferating cell nuclear antigen (PCNA)	7e-122	616	97.8896	

			mRNA, partial cds				
>Locus_6018_7Transcript_2/2_Confidence_0.667_Length_959	2e-115	959	>gi|12232559|gb|AF036312.2| Dunaliella tertiolecta mitotic cyclin mRNA, partial cds	6e-47	490	96.1224	+
>Locus_1375_7Transcript_7/9_Confidence_0.632_Length_1220	1e-132	1220	>gi|167984|gb|M60049.1|DUNCAB D.tertiolecta 28.5-kDa LHCII apoprotein mRNA, complete cds	2e-132	1041	91.0663	
>Locus_169_2Transcript_3/5_Confidence_0.632_Length_1062	3e-122	1062	>gi|225322931|gb|FJ769282.1| Dunaliella tertiolecta ascorbate peroxidase mRNA, partial cds	1e-72	546	87.5458	
>Locus_1725_6Transcript_1/1_Confidence_1.000_Length_1811	0	1811	>gi|585087513|gb|KF193066.1| Dunaliella tertiolecta sedoheptulose-1,7-bisphosphatase (SBP) mRNA, complete cds	2e-180	1631	97.9767	
>Locus_613_6Transcript_1/1_Confidence_1.000_Length_1405	0	1405	**>gi|682124627|gb|KJ930518.1| Dunaliella tertiolecta mitogen-activated protein kinase (MAPK) mRNA, complete cds**	0	1687	90.3915	+

>Locus_284_4Transcript_1/1_Confidence_1.000_Length_1355	2e-16	1355	**>gi|371532816|gb|JQ039042.1| Dunaliella tertiolecta ATP-dependent Clp protease proteolytic subunit (clpP) mRNA, complete cds; plastid**	2e-16	1560	98.2288	
>Locus_17606_9Transcript_1/1_Confidence_1.000_Length_2042	0	2042	**>gi|682124625|gb|KJ930517.1| Dunaliella tertiolecta phosphofructokinase (PFK) mRNA, complete cds**	0	2049	99.3144	
>Locus_979_2Transcript_1/2_Confidence_0.833_Length_2662	4e-139	2662	>gi|144601644|gb|EF471039.1| Dunaliella tertiolecta isolate 1 NRT2 (Nrt2) gene, partial cds	5e-50	644	94.4099	
>Locus_1062_4Transcript_1/1_Confidence_1.000_Length_630	2e-82	630	**>gi|19879329|gb|AY032598.1| Dunaliella tertiolecta nucleoside diphosphate kinase mRNA, complete cds**	7e-57	840	42.5397	*
>Locus_2576_1Transcript_1/1_Confidence_1.000_Length_1955	0	1955	>gi|459938225|gb|KC572136.1| UNVERIFIED: Dunaliella tertiolecta isolate PL1 acetyl-CoA carboxylase beta subunit-like (accD) gene, partial sequence;	1e-42	925	44.973	

			chloroplast				
>Locus_468_7Transcript_1/1_Confidence_1.000_Length_1448	0	1448	>gi|4165328|gb|AF038570.1| Dunaliella tertiolecta cyclin-dependent kinase 1 (DUNCDC2) mRNA, complete cds	0	1061	99.7172	
>Locus_652_4Transcript_1/1_Confidence_1.000_Length_2435	0	2435	**>gi|371532796|gb|JQ039032.1| Dunaliella tertiolecta ATP synthase CF1 alpha subunit (atpA) mRNA, complete cds; plastid**	0	1515	42.8383	*
>Locus_7608_9Transcript_1/1_Confidence_1.000_Length_2170	0	2170	>gi|371532798|gb|JQ039033.1| Dunaliella tertiolecta ATP synthase CF1 beta subunit (atpB) mRNA, partial cds; plastid	5e-177	1167	46.8723	*
>Locus_3958_1Transcript_1/1_Confidence_1.000_Length_2042	0	2042	>gi|371532912|gb|JQ039091.1| Dunaliella tertiolecta elongation factor Tu (tufA) mRNA, complete cds; plastid	8e-154	1257	47.494	

>Locus_3579_4Transcript_1/1_Confidence_1.000_Length_943	1e-64	943	>gi|371532800|gb|JQ039034.1| Dunaliella tertiolecta ATP synthase CF1 epsilon subunit (atpE) mRNA, complete cds; plastid	3e-07	408	26.4706	*
>Locus_6870_1Transcript_3/3_Confidence_0.778_Length_1870	1e-117	1870	>gi|761262915|gb|KJ930371.1| Dunaliella tertiolectaglucose-6-phosphate dehydrogenase (G6PDH) mRNA, complete cds;chloroplast	3e-72	3246	34.6524	*

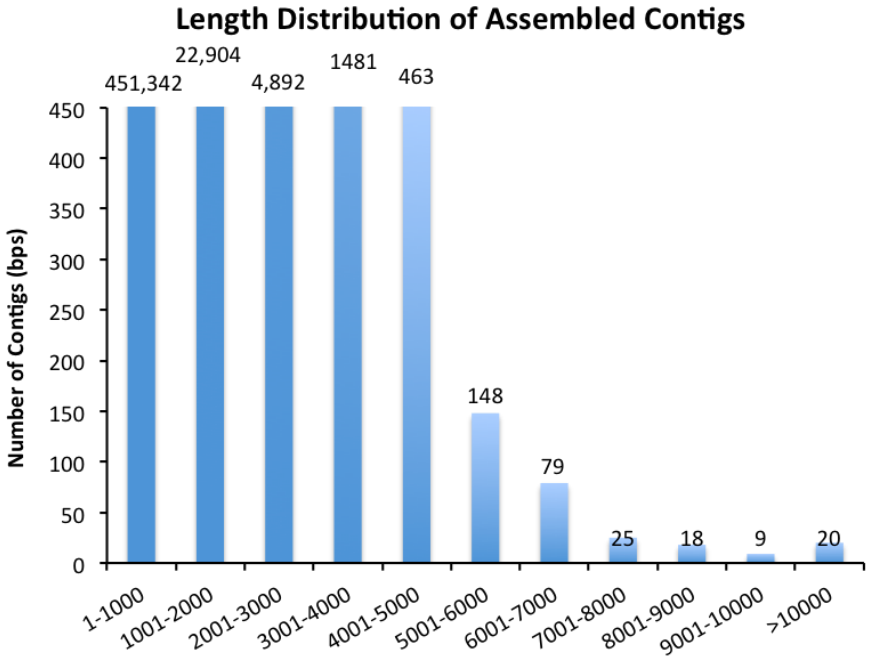
BLASTX with *C. reinhardtii* using cut-off value at E-value <= 1E-6 for this comparison.

The one in bold font was chosen in the database optimization.

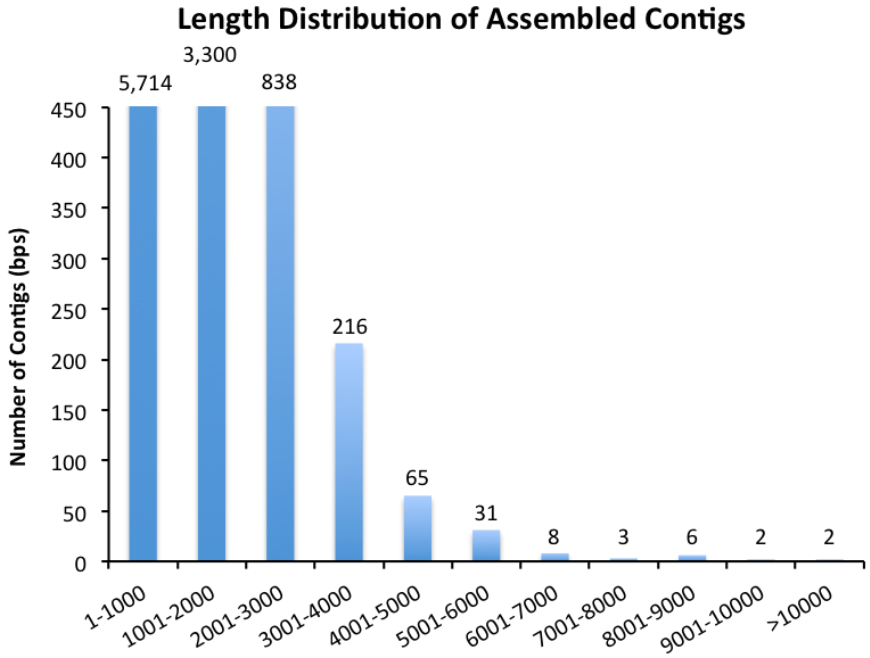
+ - the NCBI nucleotide was chosen for the corresponding gene, but it was elongated by part of the sequence from Bag2D.

* - the two nucleotides did not quite match from the two methods, which may be caused by the exists of isoforms or errors of the one from NCBI (determined by E-value and analyses of the sequences).

We optimized the *D. tertiolecta* transcriptome database mainly by two strategies: 1) Enlarge the *D. tertiolecta* transcriptome sequence in the database. MISEQ output data after each run were included in the latest version, and spiked with all the reported *D. tertiolecta* nucleotide sequences from NCBI to generate Dt_v10 after running through Bag2D pipeline. 2) Annotate the *D. tertiolecta* sequence with an addition of model high plant species. Considering the functional annotation of some microalgae may be problematic because annotation is largely based on plants and there is large phylogenetic distance between them. We also compared *D tertiolecta* with a high quality annotation of plant species *A. thaliana*, using the latest large Dt merged database (Dt_v10). From RNA-Seq alignment, a total of 181 Mega-base nucleotides (481,381 contigs) ranging from 74 bps to 17,995 bps in length were obtained from Dt_v10 database. The length distribution of the assembled transcripts was described in Figure 5.7. Dt_v10-hit database (only including the contigs that can be annotated in Dt_v10 by 6 reference species) has a total 11 Mega-base nucleotides (10,185 contigs) with minimum length of 101 bps, maximum length of 15,975 bps, and average length of 1,106 bps (7,052 was generated from *C. reinhardtii* and for RNA-Seq analysis). This allows us to identify a good quantity of *D. tertiolecta* gene sequences.



a



b

Figure 5.7 Length distribution of the contigs in Dt_v10 and Dt_v10-hit.

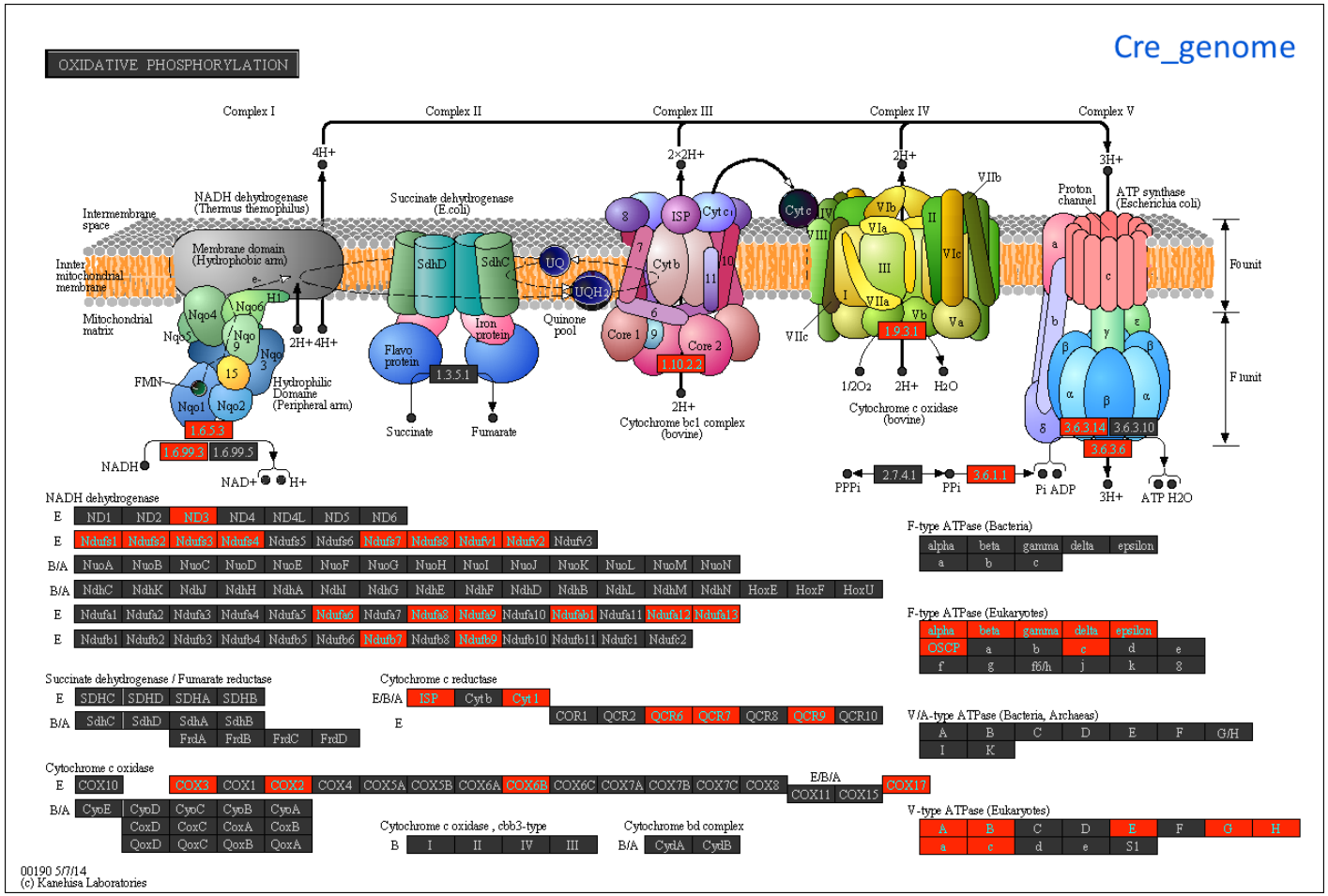
5.3.6 *Chlamydomonas* RNA-Seq data were used as the benchmark

We used a set of *Chlamydomonas reinhardtii* RNA-Seq data as the

benchmark to check the performance of Bag2D program and Partek software. Since *Chlamydomonas* has the genome sequence annotated, it can be performed in the standard way in the Partek Flow software for the alignment and annotation. It can also be assembled and annotated using in-house Bag2D software. From Table 5.5, we found 88% of the pathways were hit in both methods. A comparison of the top hit KEGG pathway generated from both methods was presented in Figure 5.8, where a similar hit pattern of the over-representative genes in both methods were found.

Table 5.5 Comparison two different methods to analyze *Chlamydomonas* RNA-Seq data.

Cre_genome			**Assembly**	
Pathway Name		**Enrichment p-value**	**Pathway Name**	**Enrichment p-value**
Oxidative phosphorylation	✔	6.33E-13	Oxidative phosphorylation	7.78E-12
Photosynthesis - antenna proteins	✔	3.81E-10	Phagosome	2.22E-08
Phagosome	✔	8.66E-09	Metabolic pathways	1.15E-05
Biosynthesis of secondary metabolites	✔	1.22E-05	Energy Metabolism	1.65E-05
Glyoxylate and dicarboxylate metabolism	✔	0.000152945	Alanine, aspartate and glutamate metabolism	2.15E-05
Metabolic pathways	✔	0.000501193	Biosynthesis of secondary metabolites	0.000183858
Starch and sucrose metabolism	✔	0.00273003	Photosynthesis - antenna proteins	0.000486186
Ribosome		0.00288034	Starch and sucrose metabolism	0.00124077
Alanine, aspartate and glutamate metabolism	✔	0.00328567	Pyruvate metabolism	0.00158613
Energy Metabolism	✔	0.00901442	Citrate cycle (TCA cycle)	0.00195073
Carbon metabolism	✔	0.00949291	Carbon metabolism	0.00276037
Citrate cycle (TCA cycle)		0.0106282	Plant hormone signal transduction	0.00351455
Porphyrin and chlorophyll metabolism	✔	0.0133897	Glyoxylate and dicarboxylate metabolism	0.00918645
Protein processing in endoplasmic reticulum		0.013603	Protein processing in endoplasmic reticulum	0.012208
2-Oxocarboxylic acid metabolism		0.0238692	Glycine, serine and threonine metabolism	0.0378914
Amino sugar and nucleotide sugar metabolism	✔	0.0305854	Amino sugar and nucleotide sugar metabolism	0.0454971
Nitrogen metabolism		0.0343084	N-Glycan biosynthesis	0.0516308
Pyruvate metabolism	✔	0.0390328	Cyanoamino acid metabolism	0.0558074
Steroid biosynthesis		0.0401102	Carbon fixation in photosynthetic organisms	0.0652938


Cre_genome
OXIDATIVE PHOSPHORYLATION
Complex I
Complex II
Complex III
Complex IV
Complex V
4H+
2×2H+
2H+
3H+
NADH dehydrogenase (Thermus thermophilus)
Succinate dehydrogenase (E.coli)
Intermembrane space
Inner mitochondrial membrane
Mitochondrial matrix
Membrane domain (Hydrophobic arm)
Hydrophilic Domaine (Peripheral arm)
Flavo protein
Iron protein
Quinone pool
Cytochrome bc1 complex (bovine)
Cytochrome c oxidase (bovine)
Proton channel
ATP synthase (Escherichia coli)
Fo unit
F unit
NADH
NAD+
Succinate
Fumarate
PPPi
PPi
Pi ADP
ATP H2O
1/2O2
H2O
NADH dehydrogenase
F-type ATPase (Bacteria)
F-type ATPase (Eukaryotes)
Succinate dehydrogenase / Fumarate reductase
Cytochrome c reductase
V/A-type ATPase (Bacteria, Archaeas)
Cytochrome c oxidase
Cytochrome c oxidase , cbb3-type
Cytochrome bd complex
V-type ATPase (Eukaryotes)
00190 5/7/14
(c) Kanehisa Laboratories

a

b

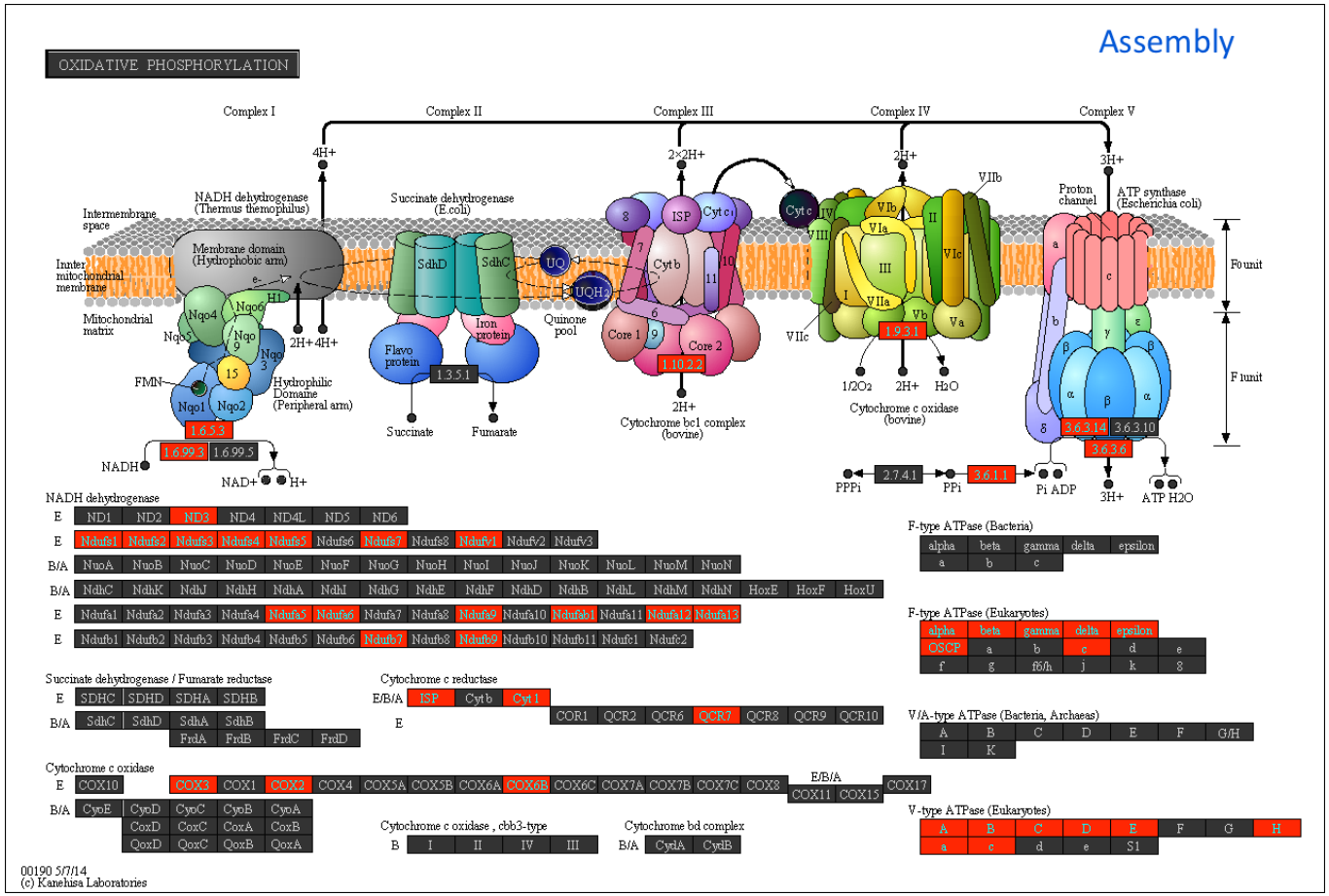


Figure 5.8 Comparison of the top hit pathway - Oxidative Phosphorylation Pathway from the two methods.

5.4 Conclusions

This study provided a workflow (Bag2D) to construct a transcriptome database for the analysis of RNA-Seq data on the lipid-rich mutant. We have generated a lipid-rich microalgal mutant, named D9, via random mutagenesis. To characterize it on the transcriptomic level, RNA-Seq technology was applied, and Bag2D was developed. After validated in a model algal species *Chlamydomonas* as the benchmark, Bag2D is currently free for public access. It could serve as a model tool to construct transcriptome databases for those organisms without complete transcriptome information. Similarly, Hamid et al. [9] presented the first next-generation sequencing effort and transcriptome annotation of a non-model marine microalgae that is relevant to biofuel production by extracting cells from various growth conditions and different phases of growth cycle, using the 454 Genome Sequencer FLX with Titanium Chemistry. As a result, they identified 33,307 assembled isotigs, 409,789 unique isotigs and singletons, 8,466 isotigs within E-value threshold of 1E-6. Despite their achievement, using Bag2D-workflow to analyze the RNA-Seq data from Illumina MISEQ, we could obtain and identify more assembled contigs (181.2 MB) and unique transcripts (10,185 transcripts, with E-value threshold of 1E-6) with functional assigned.

From the RNA-Seq result, it can been seen that elimination of the gene (*FabG*) coding for 3-oxoacyl-[acyl-carrier protein] reductase can relieve the feedback inhibition of β-ketoacyl-ACP synthase (*fabB* or *fabH*) caused by the accumulation of fatty acyl-ACPs

[106] and terminate the chain elongation cycle [92], along with the boosting of rate-limiting precursor (malonyl-CoA) flux by the expression of the acetyl-CoA carboxylase (ACCase), which leads to the D9 mutant producing about 2 to 4-fold more FAs. The upstream enzyme myo-inositol oxygenase was overexpressed to convert more precursors for the biosynthesis of FA. In addition, from KEGG pathway analysis and GO enrichment analysis, many photosynthesis related genes were found to be overexpressed. Photosynthetic rate measurement also proved that more energy was generated from photosynthesis in the mutant. Based on the results, it appears that this workflow is able to provide biological insights into the expression patterns of transcripts associated with energy metabolism and carbon flow pathways. This multiple pathway engineering approach that focuses on optimizing the expression of several related pathways is potentially useful for gene manipulation that can be directly applied to engineering microalgal FA production.

Chapter 6 Exploring the transcriptome of *Dunaliella tertiolecta* through high-throughput sequencing and high performance computing

6.1 Introduction

With the advances in sequencing technologies that most widely using Illumina MISEQ/ HISEQ today, and as sequencing depth becomes higher, the assembly of raw data now requires high capacity processing, which could not be fulfilled by off-the-shelf PCs [107]. Herein, the high performance computer (HPC) in a petascale data center was introduced in our study [108]. The use of HPC opened up great opportunities for applications in many areas, including NGS data analysis.

Recently, the whole genome sequence of the *D. salina* v1.0 was released (genome. jgi-psf.org), which aids in our comparison study as the most related species. Equipped with more input sequencing data (from Illumina HISEQ 4000), advanced *de novo* assembler, a wider reference species annotation database (all plants and bacterial proteins), and HPC in high performance data center, a much more complete *D. tertiolecta* transcriptome database (~95% of the total gene numbers) was constructed herein and applied in a case study of RNA-Seq data analysis from nitrogen-deprived cells. From the nitrogen deprivation (ND) study, potential regulatory mechanisms of cell growth and triacylglycerol (TAG) accumulation were proposed. Further, alternative-splicing variants in *D. tertiolecta* was predicted and compared with related species for the first time. This approach could be applied to other non-model

microalgae for further applications.

6.2 Materials and Methods

6.2.1 Microalgal sample preparation for sequencing

The algal culture *D. tertiolecta* strain UTEX LB-999 was cultured under different growth conditions according to Chapter 2.1. Cells were harvested for total RNA extraction and cDNA library construction as mentioned in Chapter 5.2.1. The validation and quality assessment of each library was performed from gel electrophoresis and a bioanalyzer (Agilent Technologies; Santa Clara, CA, USA). The concentration of each library was quantified via KAPA Library Quantification Kit (Illumina® platforms). The resulting libraries were sequenced by Illumina MISEQ sequencer (KR, for ND and highlight cultures) and Illumina HISEQ sequencer (G, for mutant cultures).

6.2.2 Measurement of dry cell weight, TAG and fatty acid content

Dry cell weight (DCW) measurement was performed by harvesting 10 mL of cells and collected by filtration on pre-weighed Advantec GB-140 filter paper (0.4 μm pore size; diameter 47 mm). The filter paper was then washed with isotonic 0.5 M ammonium formate (40 mL) to remove salts without causing the cells to burst. Cells captured on filter paper discs were dried in oven at 95^{o}C, and measured for DCW.

A modified Nile red staining method [86] was used to quantify intracellular TAGs.

Briefly, cells were harvested by centrifugation (3000 *g* for 10 min at 4ºC), supernatant was removed and the pellet resuspended in fresh 0.5M ATCC-1174 DA media to an OD_{680} of 0.3. Two hundred microliters of triolein standards (40, 20, 10, 5, 2.5, 0 μg/mL) and cell suspensions were loaded as technical triplicates onto a 96-well black, clear bottom plate (CLS3603; Sigma-Aldrich). Prior to staining, Nile red stock is diluted in acetone to obtain a working solution (25 μg/mL), and 2 μL of the Nile red working solution is added to each well of sample and standard, followed by a 5 min incubation in the dark. Fluorescence of each sample was detected using a microplate reader (Infinite M200 PRO, Tecan) at excitation and emission wavelengths of 524 nm and 586 nm. Fluorescence imaging of Nile Red-stained cells was performed with an automated fluorescence microscope (Olympus BX63). Acquisition and processing of data was done using the cellSens software.

To analyze the accumulation of total lipids, cells were harvested, snap-frozen in liquid nitrogen and stored at -80ºC until analysis. Frozen culture samples were lyophilized by freeze-drying and lipids were extracted by hexane using direct transesterification [109] as it was reported to be a convenient and accurate method for analyzing total fatty acids [110]. Biomass quantities of between 5 and 10 mg of biomass were weighed into glass 55-mL PYREX culture tubes with polytetrafluoroethylene (PTFE)-lined phenolic caps (25 mm diameter × 150 mm height, PYREX #9826-25, Corning). To each sample, 0.2 mL of chloroform-methanol (2:1, *v*/*v*) was added and mixed by vortexing, followed by simultaneous transesterification of lipids with 0.3 mL of 1.25M methanolic HCl and vortexed to mix. An internal standard (100 μg Methyl

tridecanoate, C13-Fatty Acid Methyl Ester, C13-FAME; Cat. no. 91558, Sigma-Aldrich) was included to correct for the loss of FAME during the reaction, and to correct for subsequent incomplete extraction of hexane [111]. The culture tube was then incubated in a 50^{o}C waterbath overnight. After 24 hours, 1 mL of hexane was added and mixed by vortex, and incubated at room temperature for 1 hour. The upper organic phase containing FAMEs was removed using a glass pipette, filtered through a 0.22-µm PTFE syringe filter (Agilent Technologies), and collected in a 250-µL glass vial insert (Part no. 5181-1270, Agilent Technologies). FAME extracts were injected into a GC system (Model 7890B, Agilent Technologies) equipped with an Agilent Agilent HP-5ms Ultra Inert column (30m x 250µm x 0.25µm) (Cat. no. 19091S-433UI, Agilent Technologies) interfaced with a mass spectrometric detector (Model 5977A, Agilent Technologies). Injection volume was set at 1 µL with a 5:1 split ratio at a GC inlet temperature of 250^{o}C. Helium was used as the carrier gas in a fixed flow of 1 mL/min throughout. Temperature program is as follows: initial oven temperature of 70^{o}C held for 3 mins, ramp to 130^{o}C at 20^{o}C/min, 178^{o}C at 4^{o}C/min, 190^{o}C at 1^{o}C/min, and 290^{o}C at 10^{o}C/min. The total run time was 40 minutes. Shifting of retention times (RTs) were eliminated by comparing the RTs of each FA compound to the C13-FAME internal standard. Analysis was performed using the MassHunter WorkStation Qualitative Analysis B.07.00 software (Agilent Technologies) and compounds were identified with the NIST mass spectral library (National Institute of Standards and Technology, Data Version: NIST 14).

6.2.3 Pre-analysis and *de novo* transcriptome assembly

After the sequencing reads were trimmed by QA/QC, Trinity assembler v2.2.0 [112, 113] was used to obtain strand-specific paired-end short reads data, with the default setting. To have a more complete transcriptome database, a draft database constructed by Shin et al. [74] (http://cholab.or.kr/data/) was adopted in the following pipeline.

6.2.4 Annotation of the transcriptome

To have the functional annotation of the assembled transcripts, Basic Local Alignment Search Tool (BLAST suite), was used to compare against the 'best' proteins in the comprehensively annotated plant and bacterial Nr database from NCBI website (http://www.ncbi.nlm.nih.gov/refseq/). Protein IDs and their hypothetical function names were obtained for corresponding transcripts. To avoid multiple counting of contigs, only the best alignment ('top hit') from BLASTX was kept.

Generally, we filtered our BLASTX results using a three-step criterion: 1. The best alignment was chosen with E-value <= 1E-10; 2. Length percentage of the query sequence >= 80% of the subject protein sequence alignment; 3. Redundant contigs with the same ncbi_proteinID were deleted. The resulting transcripts were served as the protein coding sequences (Dtertiolecta_v11.transcript_primaryTranscriptOnly.fa, short as 'Dt_v11' below). Transcripts associated with a Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathway or a Gene Ontology (GO) biological process were predicted to represent a certain expression pattern [71]. To identify it, we

applied the online KEGG database for conversion of BLASTX results (ncbi_proteinID) into KEGG gene, KEGG Orthology (KO), GO and GO definition through KEGG (http://www.kegg.jp) and GenomeNet (http://www.genome.jp) websites for pathway mapping and GO analysis based on the KEGG/GO enrichment scores.

The resulted transcriptome annotation information was compared to available transcriptome information from *C. reinhardtii* v5.5 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Creinhardtii), *V. carteri* v2.1 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Vcarteri), *D. salina* v1.0 (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Dsalina_er).

6.2.5 Differential expression analysis

We used Nr transcript dataset constructed as the reference for mapping the sequencing reads using RSEM version 1.2.29 with default settings [114], and subsequently imported and normalized in EBSeq for gene differential expression analysis [115]. Deferentially transcribed contigs upon ND is obtained using a cutoff of fold change (post) >= 2 or <= -2, and PPEE (FDR) value <= 0.05.

6.2.6 Retrieval essential redundant contigs

Normally, Nr transcripts after BLASTX best-hit search were selected for transcriptome database construction. There are five basic types of non-top-hit events:

1) false *de novo* assembly; 2) isoforms with alternative transcription starts or ends; 3) breakage of an integrated gene; 4) alternative splicing (exon skipping, intron retention, alternative 5' or 3' splices site, mutually exclusive exons) [116-118]. Python scripts using ClustalW algorithm [119] were proposed to check the similarities of all the redundant contigs, which hit the same ncbi_proteinID. Alternative-splicing variants from *A. thaliana*, *C. reinhardtii*, *V. carteri*, *D. salina* were extracted and their homogeneities were compared.

6.3 Results and Discussion

6.3.1 System environment for experimental software

In general, the complete workflow follows Figure 6.1. Experiments using HPC were completed on the petascale National Supercomputing Centre (NSCC), which comprises of 1,288 nodes (dual socket, 12 cores/CPU E5-2690v3), 128 GB DDR4/ node. Additionally, 9 nodes are equipped with more than 1 TB memory RAM for enabling large memory applications. All the software settings used for construction and analyses of the transcriptome in this study are described in Appendix 5.

It is reported that the conventional BLASTX is computationally intensive and embarrassingly parallel [120]. As the input high-throughput data size continuously increases, time cost becomes the major issue. An open-source parallelization of BLASTX, mpiBLASTX version 1.6.0), that segments and distributes a BLASTX database among cluster nodes such that each node searches a unique portion of the

database was a great advantage for speedup than the conventional single-core BLASTX [120]. Thus, instead of using normal single-core BLASTX, we used mpiBLASTX. In mpiBLASTX, database needs to be segmented into 24 fragments prior to do BLAST by using mpiformatdb (http://www.mpiblast.org). The database segmentation can save time from producing heavy intercommunication between nodes to realize the elimination of high overhead of disk I/O [120]. Figure 6.2 shows the scalability test based on the subsampling from the four sources of datasets, where we increased the number of cores in the system for mpiBLASTX application from 24 (1 node) to 1680 (70 nodes) cores and measured the speedup achieved. It was concluded that using 960 (40 nodes) cores was optimal regarding the time cost in this study. The performance flourishes when increasing core count from 24 to 960, as a result of the abundant parallelism. For configurations with more than 960 cores, however, the performance begins to diminish because the communication cost becomes the predominant factor, rendering the computing cores underutilized. Therefore, prompted by the subsampling results, the optimal configuration was used in our real study.

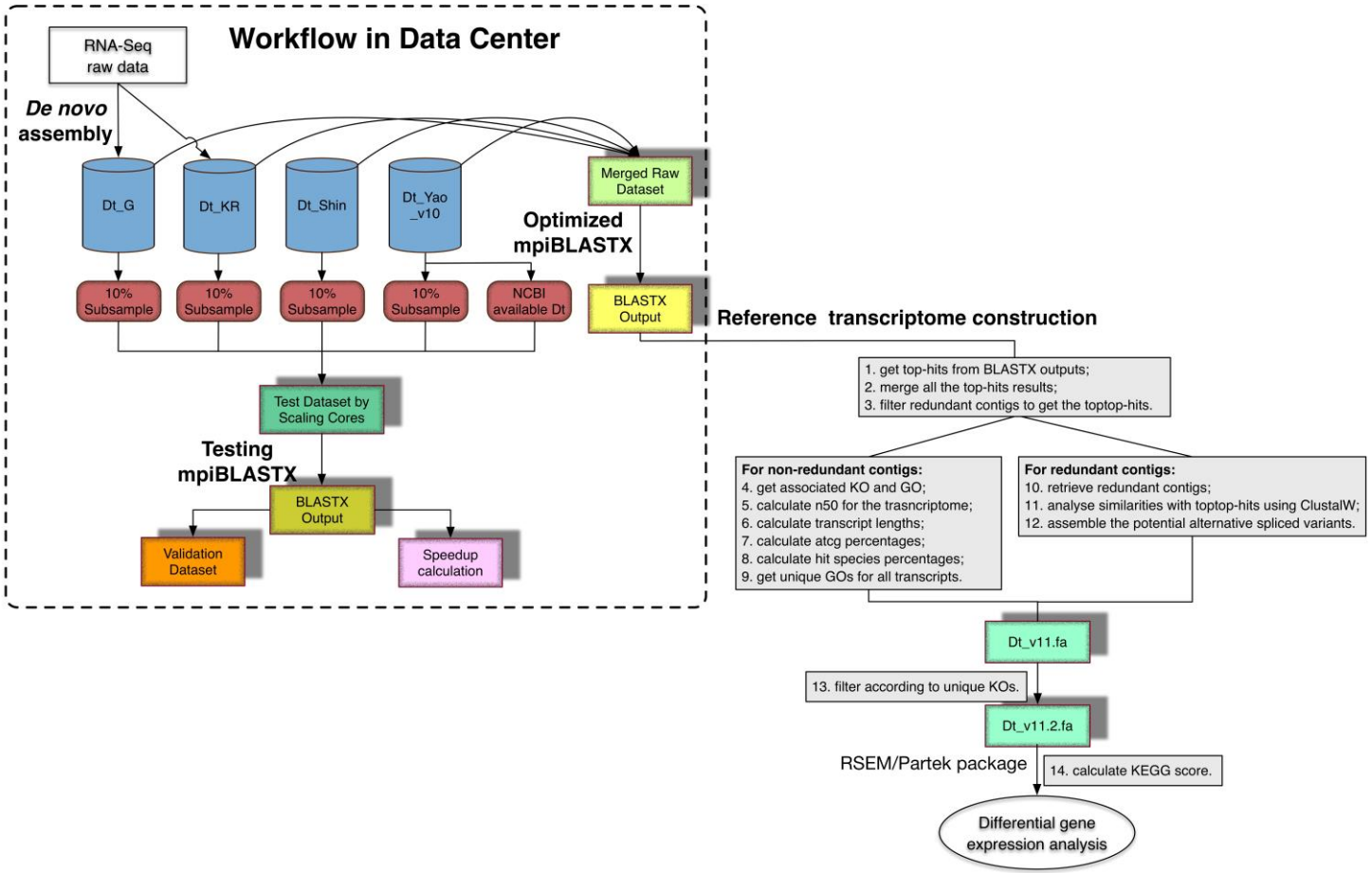


Figure 6.1 Pipeline of RNA-Seq data analysis workflow from short sequence raw data.

The general pipeline includes workflow for transcriptome database construction, annotation, and differential gene expression analysis. Workflow in data center mainly consists of *de novo* assembly, mpiBLASTX. Dt_G and Dt_KR are in-house constructed samples. Dt_Shin and Dt_Yao_v10 are two published datasets. 10% of each datasets are randomly picked for mpiBLASTX test to get the best parameters for mpiBLASTX with NCBI database for the merged *D. tertiolecta* datasets. The mpiBLASTX output was further extracted and filtered for annotation using python script.

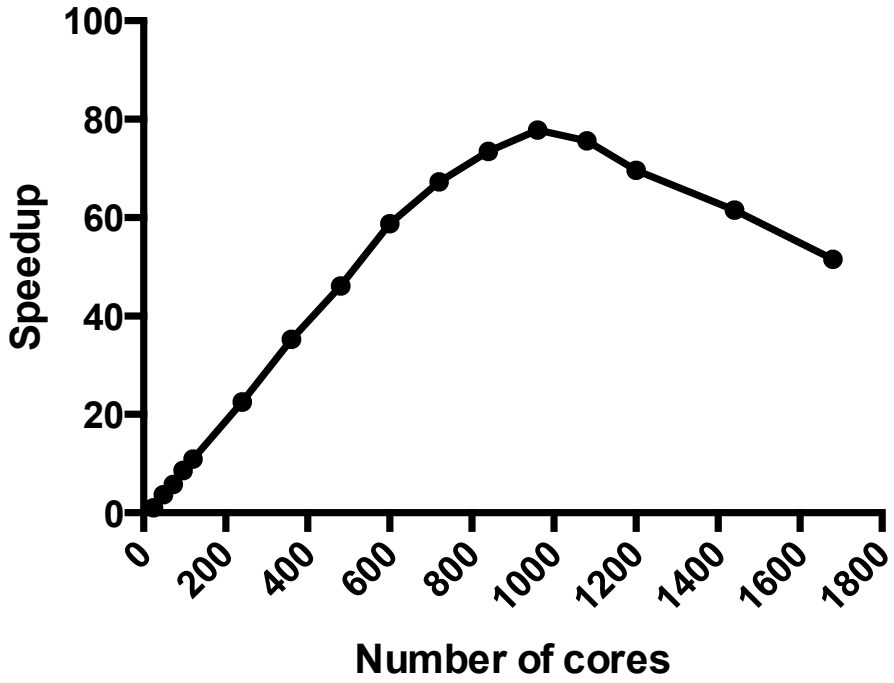


Figure 6.2 Speedup achieved by mpiBLAST calculated over run of 24 cores.

It shows the scalability test based on the subsampling from the four sources of datasets, where we increased the number of cores in the system for mpiBLASTX application from 24 (1 node) to 1680 (70 nodes) cores and measured the speedup achieved. It was concluded that using 960 (40 nodes) cores was optimal regarding the time cost in this study.

6.3.2 *De novo* assembly of *D. tertiolecta* transcriptome

Strand-specific RNA sequencing data from each condition were pooled together for *de novo* transcriptome assembly of *D. tertiolecta* transcriptome database and subsequent gene expression analysis. Due to the unavailability

of complete genome or transcriptome information of *D. tertiolecta* LB 999, the *de-novo* assembled RNA-seq data can be used to identify genes. In particular, bigger input pooled data could enlarge the output transcriptome. To evaluate the transcripts created from different input datasets, assembly statistics between different raw data were compared and pooled together (Table 6.1). As a result, 87,197 contigs were merged for annotation.

Table 6.1 Input raw data and post-analyzed data from MISEQ and HISEQ.

Data source	Number of protein-coding transcripts
Dt_G (HISEQ 4000)	27,797
Dt_Shin (MISEQ) [74]	13,861
Dt_KR (MISEQ)	25,475
Dt_Yao_v10 (MISEQ) [86]	20,229
Merged contigs	87,197
Non-redundant (Nr) contigs	17,845

6.3.3 Annotation of *de novo*-assembled transcriptome

Using the *de novo*-assembled contigs, annotation based on BLASTX with plant and bacterial Nr protein database was performed, to obtain protein identification from the taxonomy of plant and bacteria. *De novo* assembly methods are known to produce false positive contigs proportional to the sequencing depth [121]. Among the total 87,197 contigs obtained from the pooled library, only 17,845 transcripts were matched to proteins falling into our criteria to remove any false positives. The 17,845 annotated *D. tertiolecta* transcripts were subjected to functional analysis. Transcript length was ranged from 114 bps to 16,518 bps.

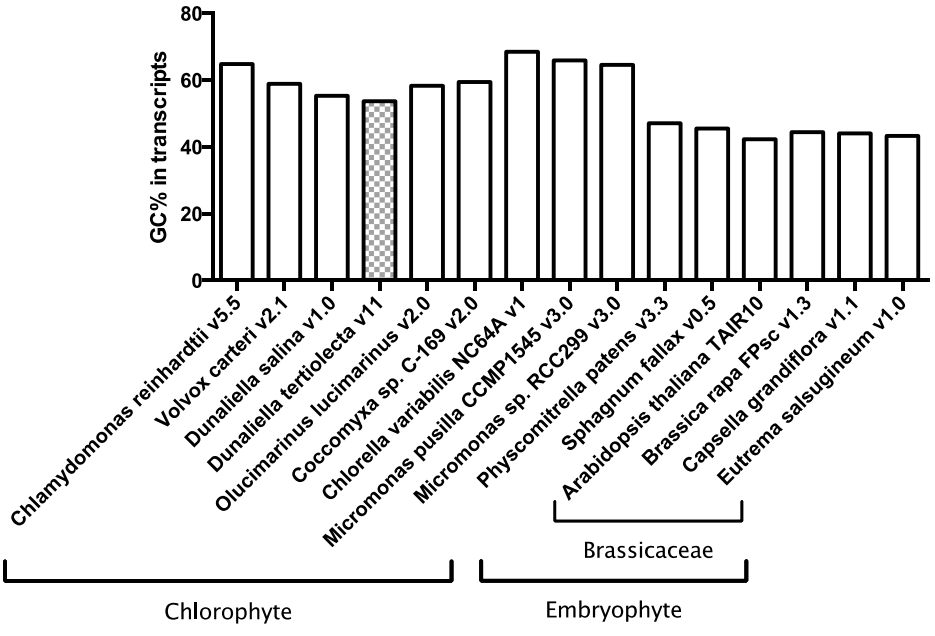
Of the 17,845 transcripts, 2,525 are associated with at least one GO function, and 10,790 were found to have KEGG gene name and 5,227 associated KO. However, some transcripts with the different KEGG gene name would end into the same KO. We further filtered the transcripts that have the same KO using python scripts, resulting in 15,336 transcripts (regarded as Dt_v11.2). Subsequently, 2,718 out of the 15,336 transcripts have unique KO and associated with at least one metabolic pathway.

6.3.4 Analysis of the *D. tertiolecta* transcriptome information

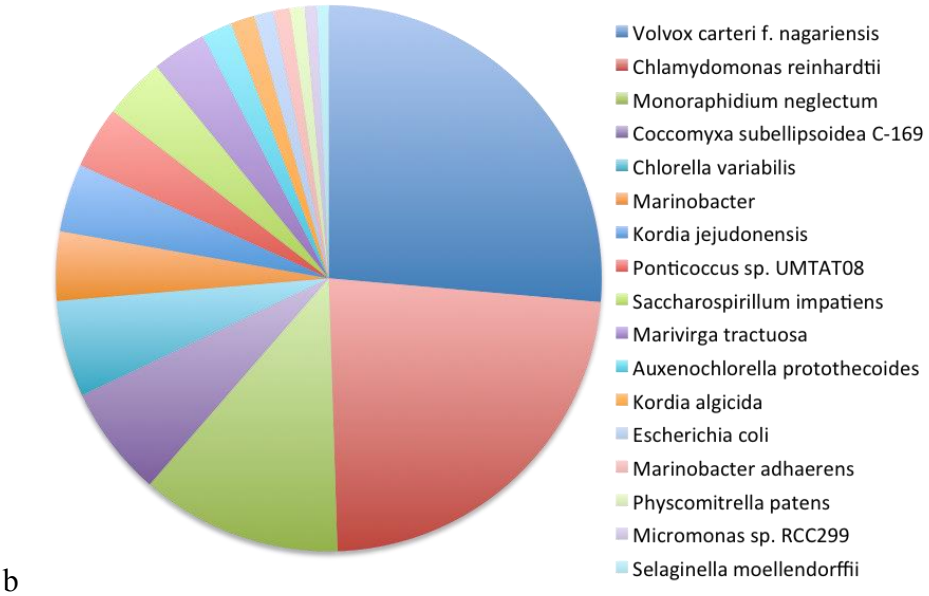
GC content is an indicator for many features of an organism, and it is correlated with various genomic features, including repeat element distribution, methylation pattern, and gene density [122-125]. GC contents of transcriptome database in *D. tertiolecta* and other related species were calculated and compared. The transcriptome of *D. tertiolecta* (54%) and *D. salina* (55%) showed the lowest GC content among the microalgal species studied, in comparison to *C. reinhardtii* (64%), and *C. variabilis* NC64A v1 (69%) (Figure 6.3a). Species in *chlorophyte* also had lower GC content than moss species, such as *Physcomitrella patens* v3.3 (47.1%) and *Sphagnum fallax* v0.5 (45.5%). Further, higher plants even had lower GC content; such as the ones in *Brassicaceae*, *A. thaliana* TAIR10 (42.3%), *Brassica rapa* FPsc v1.3 (44.4%), *Capsella grandiflora* v1.1 (44%), and *Eutrema salsugineum* v1.0 (43.3%). This phenomenon may reveal new insights into the gene regulatory mechanisms required for the evolution among *Viridiplantae*, or green plants [126]. It was observed that GC content varied considerably within *Chlorophyte* and among lineages. Although there was

extensive heterogeneity in GC content across all species, there were relatively little variations among the same taxa. This figure provides a brief view of the major relationship within *Viridiplantae* according to the ancestor nodes from Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html#!search).

We also looked into the best-hit annotation models in the BLASTX search. Majority (73.7 %) of the best-hit annotations in *D. tertiolecta* were found to derive from green algal species (Figure 6.3b). *D. tertiolecta* transcriptome appears to mostly resemble that of *V. carteri* (26.4%), followed by *C. reinhardtii* (23.1%), *Monoraphidium neglectum* (11.9%), *C. subellipsoidea* C-169 (6.6%), and *C. variabilis* (5.7%).



a



b

Figure 6.3 *D. tertiolecta* transcriptome information.

(a) GC content in D. tertiolecta transcriptome compared with other species; (b) Identification and verification of the protein-coding transcripts in *D. tertiolecta*. This bar chart is the result from BLASTX output. The sum of top hit transcripts from each individual species. The right side species names: descending numbers of hits.

Furthermore, approximately 97.2% of the core eukaryotic genes (CEGs) were found in the 17,845 transcripts, suggesting a rather high coverage of transcripts that has been obtained to represent the *D. tertiolecta* LB 999 transcriptome [127]. A detailed comparison of transcriptome information of *D. tertiolecta* with other species is presented in Table 6.2. This has shown that the newly constructed database (Dt_v11.2) has a relatively good coverage (~95% that of *D. salina*) of transcripts and high assembly and annotation quality. N50, maximum contig length, total size of contigs, number of protein-coding transcripts, and average contig length all increased based on the current available *D. tertiolecta* databases. To this end, this enhanced transcriptome database comprising core genes in *D. tertiolecta* was used as a reference for following studies.

Table 6.2 Transcriptome assembly and annotation descriptions of different species compared with *D. tertiolecta* after cutoff by E-value and length percentage.

	***C. reinhardtii* v5.5**	***V. carteri* v2.1**	***D. salina* v1.0**	***D. tertiolecta* Shin et al.**	**Dt_v11/** *Dt_v11.2*
Genome description	111.1 Mb arranged on 17 chromosomes and 37 minor scaffolds	131.2 Mb arranged in 434 scaffolds	343.7Mb arranged in 5,512 scaffolds	-	-
N50 (bp)	3,938	4,188	2,291	1,540	**1,797**
Maximum contig length (bp)	72,700	24,197	17,353	15,234	**16,518**
Total size of contigs (bp)	63,797,006	51,775,597	33,246,103	16,600,538	**24,538,468**
Protein-coding transcripts	19,526	16,075	18,801	13,861	-
transcript_primaryTranscriptOnly	17,741	14,247	16,697	9,839	**17,845/** *15,336*
Average length (bp)	3,267	3,220	1,768	1,197	**1,375**
Alternatively spliced transcripts	1,785	1,828	2,104	-	-

Microalgae are a highly diverse group with largely unexplored genetic information, and there was the enormous amount (67.7 % distinct) of diversity among microalgae at genetic level, which indicated that the functional genetic information is very diverse and species-dependent in microalgae, though they could be morphologically similar [66]. Therefore, the enrichment of *D. tertiolecta* transcriptome database is a necessity for accurate genetic engineering and RNA-Seq analysis, using larger input data, and multiple annotation species.

Construction of transcriptome coverage can vary due to expression differences and input data depth [128]. Theoretically, when addition input reads does not provide new output information, a sequencing saturation depth was hit. Several studies suggested that saturation depths at 95% gene coverage [129-132]. However, in this study, the use of the increasing number of high-throughput sequencing data enlarged the *de novo* transcriptome assembly to ~95% of *Dunaliella* genes. The enhancement and exploration of the database gave us essential and additional information for comparative analysis of the transcriptome data.

Alternative splicing, an essential mechanism for increasing transcriptome and proteome diversity in eukaryotes, are quite common [126]. It is however less clear, and has few reports in microalgae. And the alternative-splicing cases among different species usually have little homogeneity. The retrieved *D. tertiolecta* potential splicing variants appear to be diverse and do not resemble much of those in the close related species. Further, different GC content (Figure 6.3a) might also cause differences between the species in

alternative splicing as reported elsewhere [133, 134]. To further verify the predicted alternative-splicing variants, genome sequencing or third generation sequencing (single-molecule long-read sequencing) is necessary.

6.3.5 Case study of RNA-Seq data from nitrogen-deprived cells

D. tertiolecta ND cells were chosen as a case study for comparing results using RSEM-EbSeq pipeline and Partek software, as the transcriptomic and physiological responses are well documented in microalgae to promote TAG accumulation [135]. We found that nitrogen-deprived *D. tertiolecta* cells on culture day 5 had comparable DCW but remarkably increase in TAG. It was reported carbohydrate accumulation during the early stages of ND conditions existed, which could account for its little increase in DCW ([74, 136]).

Through Illumina MISEQ sequencing, over 27 million qualified raw reads with 150bps in length were used. Besides contributing to our large database construction, these data were analyzed for differential gene expression. Raw data were deposited in SRA database (SRR4011625, SRR4011626, SRR4011627, SRR4011628). Using Partek pipelines based on *C. reinhardtii* annotation [86], the top-hit representative KEGG pathways, differential expressed genes, significant GO output were presented in Appendix 6-7 and Figure 6.4, respectively. From the perspective of significant differential expressed genes, the number increased from 482 to 582 after updated to Dt_v11 analysis.

Table 6.3 Comparison of dry cell weight and TAG content in *D. tertiolecta* ND culture on day 5.

	Dry cell weight (g/L)	TAG content (pg/cell)	Fatty acid content (% DCW)
N-replete	0.31 ± 0.08	0.15 ± 0.02	6.2 ± 0.27
N-deplete	0.34 ± 0.04	1.29 ± 0.12	5.14 ± 0.45

The values are presented as the mean ± the standard deviation.

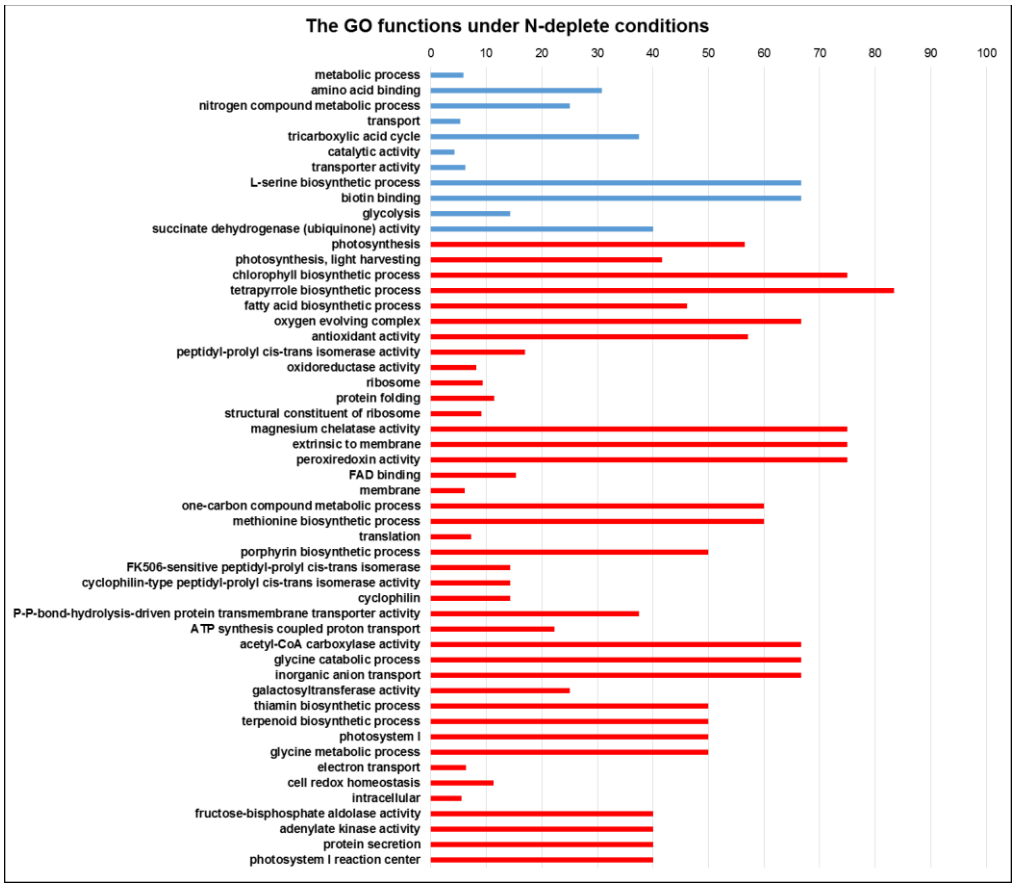


Figure 6.4 GO functional enrichment of upregulated (blue) and downregulated (red) genes under nitrogen depleted conditions.

Categories were filtered by Fisher's exact test with an FDR-corrected p-value ≤ 0.05.

Equipped with more genes annotated from different plant and bacterial species of the newly constructed database, we reported results of differential

gene expression with top fold changes from RSEM and Partek pipelines. Based on the Dt_v11 GO annotation, a .gmt file requested by biological analysis with unique GO reflecting to different transcripts.

Regarding KO, given one culture condition a, and one KEGG term $KEGG_b$, the KEGG enrichment score in the network x can be computed by the below equation (Eq.1) [137-140]:

$$S_{GO/KEGG}(a, KEGG_b) = -log_{10}(p - value)$$

$$= -log_{10}(\sum_{k=m}^{n} \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}})$$

where N, is the total number of genes with KEGG in a certain number of species (x), M, is the number of proteins that are annotated to the KEGG term $KEGG_b$, n, is the number of proteins in K(a), and m is the number of proteins both in K(a) and annotated to the KEGG term $KEGG_b$, respectively. Larger the enrichment score of one KEGG term is, more overrepresented such a term is under the culture condition a [141, 142].

In this study, the annotation was performed based on all plants and bacterial protein database; this x is not easily adopted, therefore we could still compare the enrichment score according to the relative values. Through in-house constructed scripts to calculate KEGG enrichment score, the output of significant KEGG pathways was presented in Additional file 3b. Combined with Dt_v10 (Dt_Yao_v10 (MISEQ) [86]) as the baseline, we found additional KEGG pathways that were significantly representative in the ND treatment with change of gene expression levels on the basis of a larger annotation. These included nitrogen metabolism, fatty acid biosynthesis, fatty

acid metabolism, biotin metabolism, terpenoid backbone biosynthesis, propanoate metabolism, pentose phosphate pathway, oxidative phosphorylation, 2-oxocarboxylic acid metabolism, DNA replication, insulin resistance, starch and sucrose metabolism. Genes participating in fatty acid biosynthesis/ metabolism and nitrogen metabolism pathway, which were detected in Dt_v11 analysis but not in Dt_v10 analysis, were summarized in Table 6.4. Using the ND study, pathway maps showing the well annotated and enriched transcriptome database help shed light on metabolic pathways and regulation of *D. tertiolecta* and guide rational synthetic biology approaches.

Table 6.4 Genes participating in important pathways that are exclusively found in Dt_v11.

ko00061 Fatty acid biosynthesis & ko01212 Fatty acid metabolism			
KO	**Name**	**Definition**	**Fold Change (ND/ N replete)**
ko:K00059	fabG	3-oxoacyl-[acyl-carrier protein] reductase [EC:1.1.1.100]	-3.100679049
ko:K00208	fabI	enoyl-[acyl-carrier protein] reductase I [EC:1.3.1.9 1.3.1.10]	-3.38232061
ko:K00645	fabD	[acyl-carrier-protein] S-malonyltransferase [EC:2.3.1.39]	-3.52911698
ko:K01962	accA	acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2]	-2.126940035
ko:K01963	accD	acetyl-CoA carboxylase carboxyl transferase subunit beta [EC:6.4.1.2]	-3.075257707
ko:K02160	accB	acetyl-CoA carboxylase biotin carboxyl carrier protein	-7.80359902

ko:K02372	fabZ	3-hydroxyacyl-[acyl-carrier-protein] dehydratase [EC:4.2.1.59]	-2.198560822
ko:K09458	fabF	3-oxoacyl-[acyl-carrier-protein] synthase II [EC:2.3.1.179]	-3.332311424
ko00910 Nitrogen metabolism			
ko:K00264	GLT1	glutamate synthase (NADPH/NADH) [EC:1.4.1.13 1.4.1.14]	23.8846
ko:K00366	nirA	ferredoxin-nitrite reductase [EC:1.7.7.1]	-13.89942929
ko:K01915	glnA	glutamine synthetase [EC:6.3.1.2]	2.84411
ko:K02575	NRT	MFS transporter, NNP family, nitrate/nitrite transporter	-4.415225463
ko:K10534	NR	nitrate reductase (NAD(P)H) [EC:1.7.1.1 1.7.1.2 1.7.1.3]	-9.464319515

6.3.6 Global gene expression level change under ND conditions

In this ND study, we observed all the photosynthetic genes were repressed (Appendix 6). Consistent with this, Yang et al. observed the similar trend in *Phaeodactylum tricurnutum*, and extensive degradation of chloroplast membranes under an electron microscope, with photosynthetic efficiency declined [143]. It suggested that nitrogen is an essential factor for photosynthetic activity, which reduced the *de novo* absorption of carbon and energy and eventually would compromise the cell growth and affect the total TAG production (reported in our subsequent study).

Genes in fatty acid biosynthesis and metabolism were also found repressed, while TAG levels significantly increased. Under nitrogen depleted conditions, the cells gradually stop dividing and then lipid accumulation

within the cell can be substantial over the time even though the genes in fatty acid biosynthesis and metabolism are repressed to certain extent. Interestingly, from the nitrogen metabolism, we found the glutamate synthase (NADPH/NADH) was greatly activated under ND conditions. Similarly, Shin et al. also observed similar results, proving that intracellular glutamate levels were the main factor for the regulation of cell growth and carbon accumulation [74]. Further studies on metabolite profiling of glutamate and related amino acid levels could be potential targets to uncover the regulatory mechanism more specifically. As a signal of the growth delimitation under ND, genetic engineering to activate or block certain enzyme coding genes would promote the biofuel-relevant productions in microalgae.

6.4 Conclusion

In this study, fuelled by a high performance data center (NSCC), high quality high-throughput RNA-Seq data were *de novo* assembled and annotated, which resulted in 17,845 protein-coding transcripts in *D. tertiolecta*, accounting for ~95% of the total *Dunaliella* genes. The enhanced transcriptome database subjected to the analysis of RNA-Seq data from ND study gave us new insights of regulation mechanisms of cell growth and lipid biosynthesis. Overall, these results laid the foundation for demonstrating integration of supercomputing with large input datasets to improve microalgal transcriptome database and elucidate the regulatory response of ND conditions for promoting biofuel production.

Chapter 7 Elevated acetyl-CoA by amino acid recycling fuels G11_7 mutant TAG accumulation in exponential growth

7.1 Introduction

In microalgae, AcCoA, malonyl-CoA, and NADPH are the major substrates supporting FA synthesis. Malonyl-CoA is also generated from carboxylation of AcCoA. Thus AcCoA is the primary precursor for FA synthesis [144]. The AcCoA balance in an algal cell could be described by the following equation:

$$[\mathrm{AcCoA_T}] - [\mathrm{AcCoA_B}] = [\mathrm{AcCoA_{NL}}]$$

Total AcCoA ($[AcCoA_T]$) and reduced NADH are produced via glycolysis. In the exponential growth phase, NADH is mainly oxidized through respiration to yield ATP, and AcCoA is used predominantly for biomass growth ($[AcCoA_B]$), including that for structural lipid (glycerophospholipids) synthesis, while a minor fraction of AcCoA ($[AcCoA_{NL}]$) and reduced NAD(P)H is used for FA synthesis to accumulate neutral lipids (TAG). When microalgal cells enter the stationary growth phase, carbon metabolism for biomass growth diminished, which leads to accumulation of AcCoA and reduced NADH. Thus in stationary phases or growth hindering stress conditions, a conspicuous fraction of AcCoA and reduced NAD(P)H is channeled to FAs biosynthesis, resulting in TAG accumulation in cells [145]. Recent studies also suggested that intracellular membrane remodeling contributed to TAG accumulation during stationary phase or nitrogen deprivation [146-148]. To accelerate TAG accumulation in exponential

growth phase (TAEP) while maintaining cell growth, AcCoA ([$AcCoA_T$]) level should be elevated over a certain set point that is needed for biomass growth ([$AcCoA_B$]).

There are three principal sources of AcCoA during growth phase, namely FA oxidation, glycolysis pathway, and amino acid degradation [144]. FA oxidation is the reversal of FA synthesis, and does not generate *de novo* AcCoA. Instead, it is thought that AcCoA is largely derived from the glycolytic pathway via pyruvate. Pyruvate is converted to AcCoA by PDHC in mitochondria and chloroplasts, and this step has been suggested as the key rate-limiting step [77, 144, 149]. One approach to increase AcCoA production is to relieve pyruvate dehydrogenase kinase (PDK) control of pyruvate dehydrogenase complex (PDHC) resulting in the activation of PDHC. This would facilitate the bioconversion of pyruvate to AcCoA, and enhance the metabolic flux toward both cell growth via the TCA cycle, and FA biosynthesis in the growth phase. The third source of AcCoA, which derived from amino acid degradation, has largely been ignored as a relevant pathway for bioengineering. Despite the fact that it bypasses pyruvate and the highly controlled PDHC/PDK regulatory process, it was considered insufficient for FA biosynthesis [144].

We hypothesized that increase of AcCoA pool by multiple routes could trigger TAEP. In our study, from the activation of pyruvate to AcCoA reaction by addition of sodium dichloroacetate (DCA) to release the PDHC/PDK regulatory process, we achieved TAEP in the WT *D. tertiolecta*. Besides this conventional *de novo* synthetic pathway, we questioned the contribution of

amino acid degradation on TAEP, although it has largely been ignored. Through performing genetic engineering, we generated mutants, which exhibited pronounced TAEP with little compromise on growth rate. By employing transcriptomics and metabolomics, key phenotypic regulatory characteristics of lipogenesis in this microalga were uncovered, implying that a secondary contributor of AcCoA derived from amino acid catabolism, in particular branch-chain amino acid catabolism, contributed to TAEP. These two major approaches were proposed in our 3-step α loop model. The results highlight the complex interplay between microalgal cellular proliferation and carbon flux in lipogenesis, and suggested that genetic and metabolic manipulations targeted at amino acid catabolism could be used to increase accumulation of fuel-relevant molecules in microalgae in the exponential growth phase.

7.2 Materials and Methods

7.2.1 Strains and culture conditions

D. tertiolecta WT strain was cultured under low light condition as previously described (section 2.1). Mutant G11 was generated according to section 4.2.2 by random insertional mutagenesis. Followed by two rounds of FACS, G11_7 mutant was purposely obtained.

In the DCA treated experiment, WT *D. tertiolecta* with spike of 750 μM DCA (Sodium dichloroacetate, 98%, ACROS Organics™) and its net control were cultured under low light condition as aforementioned. TAG content was

measured using quick Nile red assay. RNA was extracted for quantitatively measurement of *PDK* mRNA levels using quantitative real-time PCR. AcCoA levels were quantified using Acetyl-Coenzyme A Assay Kit (Sigma, MAK039). We did the same of DCA treatment on G11_7 mutant both under low light (30 μmol photons•m^{-2}•s^{-1}) and high light conditions (320 μmol photons•m^{-2}•s^{-1}), respectively, with a net G11_7 strain as the control.

In the amino acid spike experiment, different concentrations of leucine (20μM, 50μM, 500μM, 1000μM) were spiked into the microalgal culture medium of WT and G11_7 mutant *D. tertiolecta*, with its medium as blank control.

7.2.2 Next-generation sequencing using HISEQ 4000

RNA was extracted from G11_7 and WT on culture day 6. RNA quality was verified using Agilent RNA 6000 Nano Kit in Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, Calif.) and gel electrophoresis. The samples were linearized with 0.1N NaOH into single-stranded forms, they were then neutralized and diluted into 200pM loading concentration with Examp master mix (EPX1 to 3) and loaded into 1 lane of Paired End flowcell using the Illumina cBOT machine. The DNA were attached and amplified simultaneously inside each oligo well on the flowcell surfaces, as a proprietary clustering method known as exclusion amplification. This method ensures that only a single DNA template binds and forms a cluster within a single well, reducing the occurrence of polyclonal wells thus increasing the usable reads. The sequencing primer was then attached to the reads, preparing for sequencing run. The flowcell was then loaded into the Illumina

HISEQ4000 High Sequencers with the sequencing reagents and run at 2x151 cycles, the second read turnaround was done using the Sequencers after the first read was completed. The images were captured by the HiSeq Control Software (HCS), and the Real Time Analysis (RTA) software converted the images into Cycle Intensity Files (CIF) and later Basecall (bcl) files. All the bcl files were then transferred to the server for storage and primary analysis. In the primary analysis, the bcl files were converted into fastq files using the bcl2fastq pipeline. After the conversion, the fastq reads were filtered to remove all the reads that did not pass filtering, leaving only useable Passed Filtered (PF) reads. The usable reads were then analyzed and bin to each of the barcode file known as demultiplexing, those that did not pass the filtering are not demultiplexed. The primary analysis result was then generated as the Demultiplexed report and reviewed. The paired-end raw data (150 bps in length/read) of G11_7 and WT were pre-analyzed as section 5.2.2 (Table 7.1) in Partek® Flow® software (version 4.0, Partek Inc., St. Louis, MO, USA) with Dt_v10 *D. tertiolecta* transcriptome database (available on author's website: https://github.com/SPURc-Lab) as the reference using Reads Per Kilobase of transcript per Million mapped reads (RPKM). Subsequently the different expressed transcripts were imported into Partek® Genomics Suite® software (version 6.6, Partek Inc., St. Louis, MO, USA) for gene annotation and KEGG pathway analysis [86]. We also used an in-house pipeline to analyze the RNA-Seq data. The transcriptome database was enlarged via Trinity assembler [112] and BLASTX against reference protein sequences from all plants and bacterial from National Center for Biotechnology Information (NCBI)[74]. Transcript level comparison was

performed using RSEM [114] with default settings. The count data from RSEM was imported for normalization in the Ebseq pipeline [115], which was used for differential expression analysis, with the design matrix formulated to fit the experimental conditions. The National Supercomputing Centre (NSCC) was used for running the aforementioned software. The sequencing data were deposited into GEO with accession number of GSE82121.

Table 7.1 Run summary of G11_7 and WT *D. tertiolecta* on the Illumina HISEQ4000 platform.

Sample name	**Sequencing stats and pre-alignment QA/QC of raw data /pre-alignment QA/QC of post-trimming data**					**Post-alignment QA/QC after alignment**	
	Total reads	Avg. read length	Avg. read quality	% N	% GC	Total reads	% Aligned
G11_7-1	131,836,866/ **131,836,590**	150.00/ **149.80**	38.78/ **38.81**	.03/ **.02**	55.40/ **55.40**	65,918,295	65.13
G11_7-2	175,867,666/ **175,867,448**	150.00/ **149.80**	38.82/ **38.85**	.03/ **.02**	55.16/ **55.16**	87,933,724	63.65
G11_7-3	129,743,218/ **129,742,886**	150.00/ **149.79**	38.73/ **38.76**	.03/ **.02**	55.00/ **54.99**	64,871,443	65.75
WT-1	135,893,898/ **135,893,720**	150.00/ **149.82**	38.92/ **38.95**	.03/ **.02**	55.35/ **55.34**	67,946,860	63.47
WT-2	110,407,944/ **110,407,800**	150.00/ **149.81**	38.89/ **38.92**	.03/ **.02**	54.83/ **54.83**	55,203,900	63.79
WT-3	123,035,916/ **123,035,586**	150.00/ **149.81**	38.87/ **38.90**	.03/ **.02**	54.54/ **54.54**	61,517,793	66.31

7.2.3 Cloning and analysis of important genes

The full-lengths of the coding region of putative genes of key enzymes, *DtCuAO* (copper amine oxidase family), *DtIVD* (isovaleryl-CoA-dehydrogenase), *DtMCCB* (3-methylcrotonyl-CoA carboxylase) were amplified from *D. tertiolecta* cDNA by RACE PCR via the predicted region from the *D. tertiolecta* database. All the primers used in this study were listed in Appendix 1. Phylogenetic tree of protein clusters and subcellular localization of the related genes was predicted according to section 2.7.

The complete cDNA sequence of *D. tertiolecta* copper amine oxidase gene (1524-bp encoding for 507 amino acids), isovaleryl-CoA dehydrogenase gene (1053-bp encoding for 350 amino acids), and 3-methylcrotonyl-CoA carboxylase beta subunit gene (1725-bp encoding for 574 amino acids) were obtained using RACE PCR. The candidate *D. tertiolecta* copper amine oxidase gene contained a copper amine oxidase enzyme domain and showed the highest (60%) amino acid homology compared to copper amine oxidase of *Volvox carteri* f. nagariensis, and was designated *DtCuAO* (*AMX1*). The candidate *D. tertiolecta* isovaleryl-CoA dehydrogenase gene contained an isovaleryl-CoA dehydrogenase domain and showed the highest (76%) homology in amino acid sequence compared to that of *Chlamydomonas reinhardtii*, and was designated *DtIVD* (*ACAD*). The candidate *D. tertiolecta* 3-methylcrotonyl-CoA carboxylase beta subunit gene contained a 3-methylcrotonyl-CoA carboxylase beta chain domain and showed the highest (67%) homology in amino acid sequence compared to that of *C. reinhardtii*, and was designated *DtMCCB*. Amino acid sequences of *CuAO*, *IVD*, and

MCCB from other species were obtained by BLAST search in NCBI database with the putative DtCuAO, DtIVD, DtMCCB. The phylogenetic tree constructed by MEGA5 demonstrated that the putative DtCuAO, DtIVD and DtMCCB showed high homology with CuAO, IVD and MCCB, respectively from other species (Appendix 8). Using the predicted sequence of these three genes from the *D. tertiolecta* database, we designed primers and confirmed by high-fidelity DNA polymerases sequencing. The cDNA sequences of the three genes were also attached as Appendix 9a-c. The *DtPDK* gene that subsequently measured in the DCA treatment experiment was discovered by the prediction from the *D. tertiolecta* in-house database (known as Locus_5000_7Transcript_1/1_Confidence_1.000_Length_2194) and confirmed by experimental sequencing using high-fidelity DNA polymerases and shown in Appendix 9d.

7.2.4 Metabolomics analysis

200 mL algal cells of G11_7 and WT were collected on day 6 respectively, and freeze-dried for metabolomics analysis. Metabolite extraction method was first optimized using a test sample. 4 solvent mixtures were tested and the optimal condition was chosen to ensure the maximum number of peak pairs that could be detected.

To compare the metabolome profiles of the two genotypes of microalgae with biological triplicates were analyzed using chemical isotope labeling liquid chromatography mass spectrometry (CIL LC-MS) method. The CIL LC-MS analysis focused on amine/ phenol compounds and carboxylic acid submetabolomes extracted from the samples. 3 aliquots of each sample were

weighed out and each aliquot was labeled according to protocols reported previously [150-152]. Samples were then mixed appropriately and analyzed on the Bruker Impact HD quadrupole time-of-flight (Q-TOF) mass spectrometer (Bruker, Billerica, MA) connected to a Dionex Ultimate 3000 (Dionex, Sunnyvale, CA, USA). The samples were injected onto an Agilent reversed-phase Eclipse Plus C18 column (2.1 mm × 10 cm, 1.8 μm, 95 Å) for separation. Solvent A was 0.1% (v/v) formic acid / 5% (v/v) ACN /H_2O, and solvent B was 0.1% (v/v) formic acid / ACN. The chromatographic conditions for dansyl labeling were: t = 0 min, 20% B; t = 3.5 min, 35% B; t = 18 min, 65% B; t = 21 min, 95% B; t = 26 min, 95% B. The gradient for DmPA labeling was t = 0 min, 20% B; t = 9 min, 50% B; t = 22 min, 65% B; t = 26 min, 80% B; t = 29 min, 98% B; t = 40 min, 98% B. Column temperature was set at 30□, and a flow rate of 180 μL/min was applied. All MS spectra were obtained in the positive ion mode. The mass range was set as 220-1000 m/z for danzyl labeling and 110-1000 m/z for acid labeling. The MS spectral rate was 1.0 Hz.

For targeted acyl-CoA analysis, LC-MS with standard addition method was used to quantify the CoAs in the algae samples [153]. A mixture of fifteen CoA standards was used. 30 uL of each biological triplicate sample was injected onto a Kinetex Coreshell HILIC column (100 mm x 2.1 mm, 1.7 μm) for separation. LC-MS analysis was done using a Dionex Ultimate 3000 (Dionex, Sunnyvale, CA, USA) connected to a Bruker maXis II Q-TOF instrument (Bruker, Billerica, MA, USA). Column temperature was set at 40℃ and a solvent gradient of 5 minutes at 500 μL/min was used with an equilibration of 8 minutes at 600 μL/min. Solvent A was 10 mM ammonium

acetate (pH 5.6) and solvent B was 95% ACN /5% 10 mM ammonium acetate (pH 5.6). The gradient was as follows: 0-3.0 mins. (90% B-5% B); 3.0-5.5 mins. (5% B-5% B). For the Q-TOF instrument, the mass range was set as: 750-1100 m/z and the spectral acquisition rate was 3 Hz. All MS spectra were obtained in the positive ion mode. Peak areas of all the CoAs were extracted using Bruker Target Analysis software for quantification.

Profiling the amine- and phenol-containing metabolites (i.e., the amine/phenol submetabolome) was carried out by using differential CIL LC-MS (Figure 7.1).

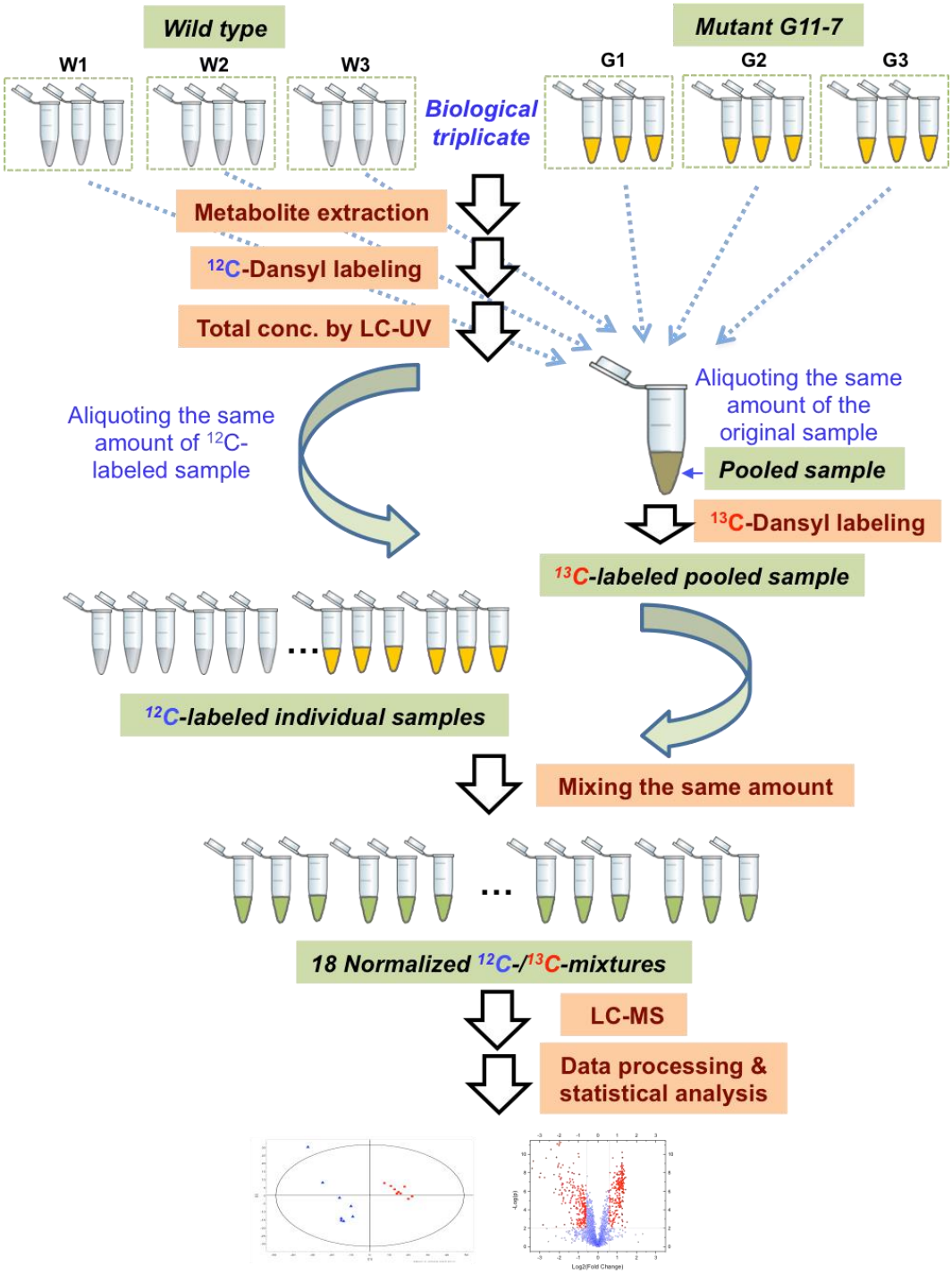


Figure 7.1 Experimental workflow of isotopic labeling LC-MS for quantifying the changes of metabolites in the WT and G11-7 mutant *D. tertiolecta*.

7.3 Results and Discussion

7.3.1 DCA treatment elevated AcCoA pool

After addition of DCA to the WT *D. tertiolecta*, TAG was found to accumulate in the exponential growth phase (data not present here).

AcCoA is *de novo* converted from pyruvate, which is tightly regulated by PDHC and catalyzes the oxidative decarboxylation of pyruvate. PDKC could be deactivated by PDK through reversible ATP dependent phosphorylation mainly in mitochondria [154]. To deactivate PDK, DCA was added to the algal culture medium. DCA is a byproduct of chlorine disinfection process, which inhibit PDK, through formation of DCA helix bundle in the N-terminal domain of PDK [154, 155]. Bound DCA promotes local conformational changes that are communicated to both nucleotide-binding and lipoyl-binding pockets of PDK, leading to the inactivation of kinase activity [154]. Thus, when DCA was included in the culture medium, PDHC became active since PDK was blocked resulting in an elevation of AcCoA.

7.3.2 FACS enriched a pool of mutant strains with higher-TAG content

G11_7 mutant that can accumulate TAG from the exponential growth phase was generated by random insertional mutagenesis and two rounds of FACS, according to section 4.3.3. Data were shown in Figure 4.3.

The growth and specific growth rate (d^{-1}) was traced according to Chapter 2.1, and presented in Figure 7.2.

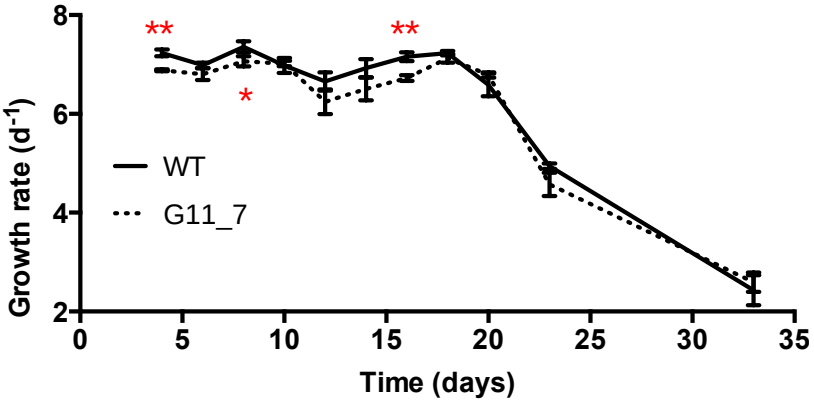


Figure 7.2 Specific growth rate of G11_7 mutant versus WT *Dunaliella tertiolecta* under low light.

7.3.3 Altered expression level of genes in amino acid catabolism

Differential expressed genes in the mutant as compared to the WT are available on GEO with accession number of GSE82121. KEGG enrichment scores were calculated and shown in Figure 7.3 [141, 156]. There are in total 105 KEGG enrichment scores featured in the three predicted analyses, among which 19 KEGG pathways were found significant in at least two enrichment analyses according to scores. Valine, leucine and isoleucine (branched-chain amino acids, BCAA) degradation pathway is the most significantly affected pathway detected in all three analyses. *CuAO* (or *AMX1*, K00276), *IVD* (K00253), and *MCCB* (K01969) were the top hits at expression level in both Partek analysis and in-house workflow. Upregulation of *ACCA* (K00626) (acetyl-CoA C-acetyltransferase) was also detected, which resulted in the accumulation of AcCoA. Upregulation of the downstream *FabD* (K00645) (malonyl-CoA:acyl-carrier-protein transacylase) contributed to accumulation of FAs in chloroplasts. The fast FA accumulation might cause a drawn-down in AcCoA, thus initiating a pull down from upstream photosynthetic

pathways. The pull down may have caused an enhancement of photosynthetic rate (Figure 4.4c) to support AcCoA *de novo* synthesis, which was supported by the upregulation of upstream photosynthesis and glycolysis genes, including *petE* (K02638) (photosynthetic electron transport), *petC* (K02636) (cytochrome b6-f complex), *LHCA1* (K08907) and *LHCB1* (K08912) (light-harvesting complex I chlorophyll a/b binding protein 1), *pfkA* (K00850) (6-phosphofructokinase 1), *PPC* (K01610) (phosphoenolpyruvate carboxykinase (ATP)), and *FBP* (K03841) (fructose-1,6-bisphosphatase I).

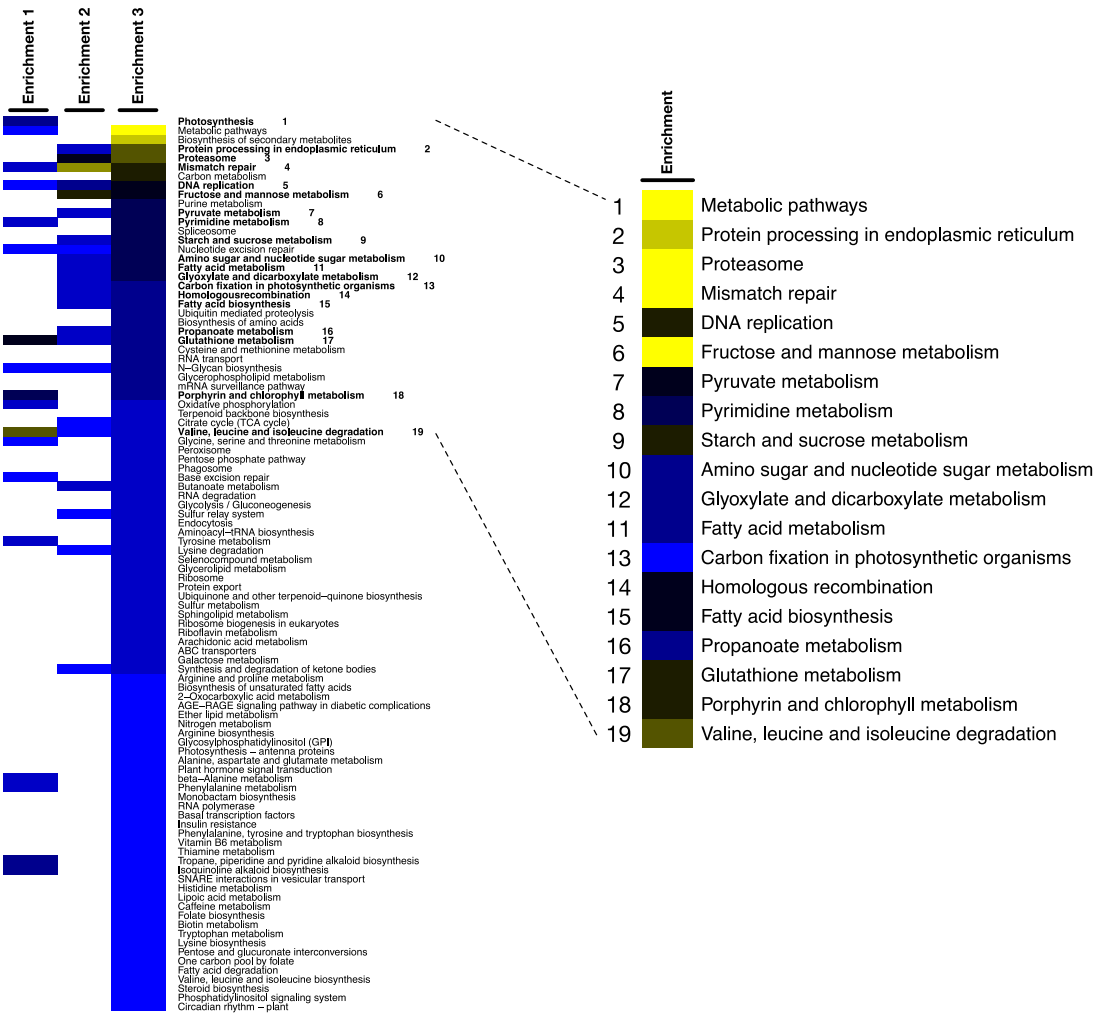


Figure 7.3 Heat map of the KEGG profiles.

Colors represent abundance of KEGG enrichment score (blue to yellow, ascending enrichment score). Rows of the heat map represent pathways from

three analyses (Enrichment 1: Partek workflow; Enrichement 2: in-house workflow, background KEGG gene that derived from *C. reinhardtii* as the reference was set at 3,321; Enrichment 3: in-house workflow, background KEGG gene derived from *D. tertilecta* transcriptome non-redundant database was set at 10,579. Images were made with R (http://www.r-project.org/). Significant KEGG pathways are marked by bold font with number on the *left* and expansion on the *right*.

We arbitrarily tested two genes for their expression levels by quantitative real time PCR at different growth phases (Figure 7.4) using the *D. tertiolecta* beta-tubulin gene (DtTUB) as the internal standard for normalization [79]. It showed a similar tendency with that of RNA-Seq data. Interestingly, the fold changes in the gene expression level and lipid level are correlated, suggesting that they are the key genes regulating the lipid accumulation process.

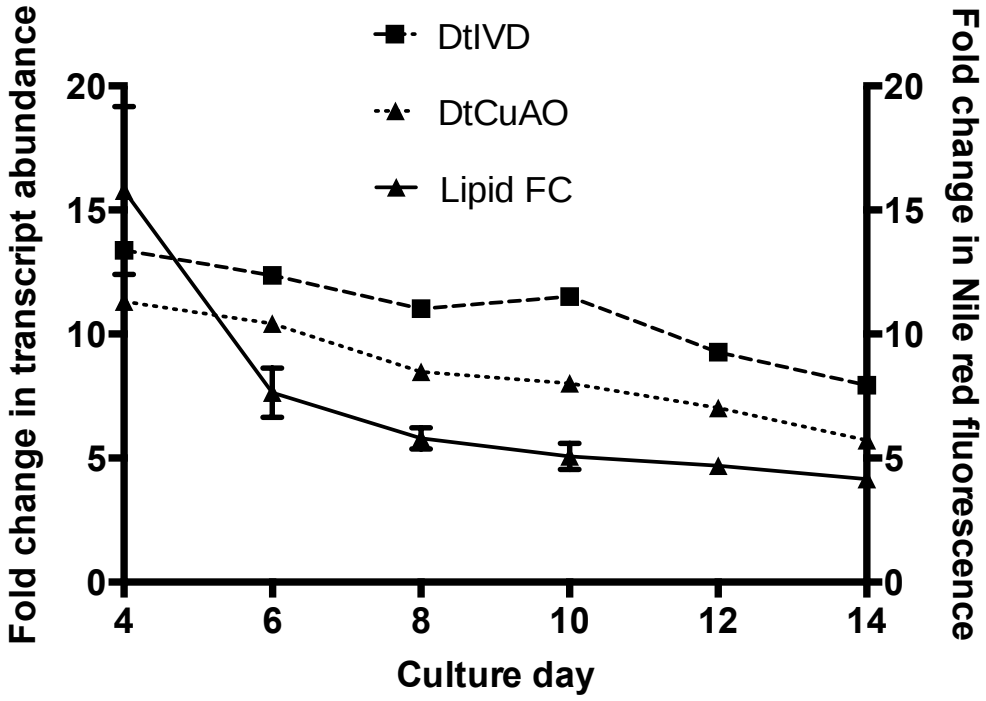


Figure 7.4 Temporal expression of predicted genes and lipid accumulation fold changes.

Fold change in abundance of *DtIVD* and *DtCuAO* transcript (primary axis) during lipid accumulation in G11_7/WT *D. tertiolecta* (secondary axis).

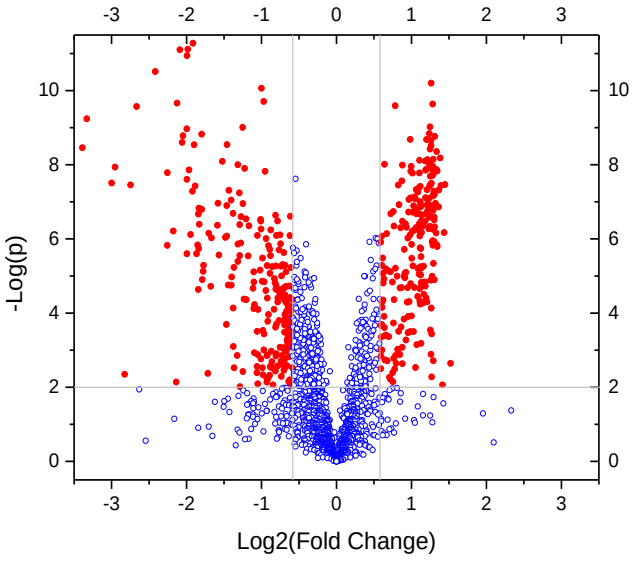
7.3.4 Pilot study using PCA and PLS-DA for the analysis of metabolic profiles of G11_7 and WT

A total of 2246 metabolites (isotope peak pairs) were detected in the 18 LC-MS runs. Among those metabolites, 276 were considered significant contributors to the differentiation of the two genotypes from Volcano Plot, and the two groups of samples can be well separated through either Principal Component Analysis (PCA) or Partial least squares Discriminant Analysis (PLS-DA) (Figure 7.5a-c). Data from G11_7 (G) and WT (W) composed of biological and technical triplicates were quickly identified by Volcano Plot (Figure 7.5a) to plot Fold Change versus significance on the x and y axes, respectively. The scores plot (PCA, unsupervised method) obtained from all observations of the two strains of *D. tertiolecta* is displayed (Figure 7.5b). Further analysis using a discriminant analysis (supervised method) was applied. PLS-DA was performed on the data to classify the observations and supervise their separation. Similar to PCA scores plot represented by G and W (Figure 7.5b), PLS-DA analysis shows a clear separation among the strains (Figure 7.5c). The statistical validation was assessed and presented in the caption (Figure 7.5), indicating excellent prediction ability of the model. Among those 276 metabolites, 18 were positively identified against dansyl standards library based on both mass and retention time matches; 46 could be putatively identified using MyCompoundID search engine against HMDB based on 0 reaction and Metlin.

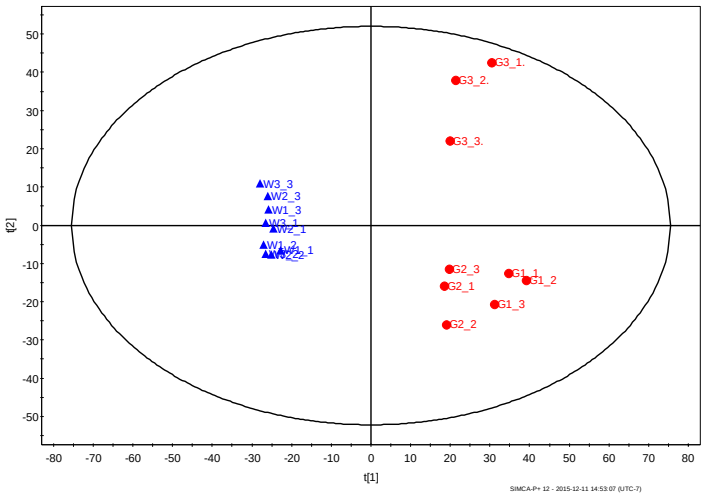
In the carboxylic acid submetabolome profiling, 2246 metabolites (isotope peak pairs) were detected in 18 LC-MS runs. Among those metabolites, 142

were considered significant contributors to the differentiation of the two genotypes (Figure 7.5d-f). Similarly, Fold Change versus significance was plotted in Figure 7.5d. Separation of the two groups of samples can be very well observed in PCA and PLS-DA plots (Figure 7.5e-f).

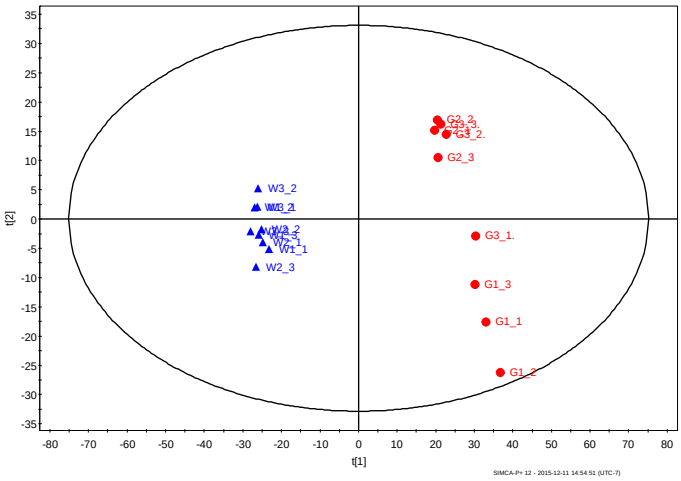
Importantly, fifteen targeted acyl-CoAs with different acyl groups were also performed using LC-MS and their fold changes between the two groups, along with CoA concentrations. Our emphases were mainly in the FA biosynthesis pathway, protein metabolic pathway, citrate acid cycle (TCA cycle) and their precursors.



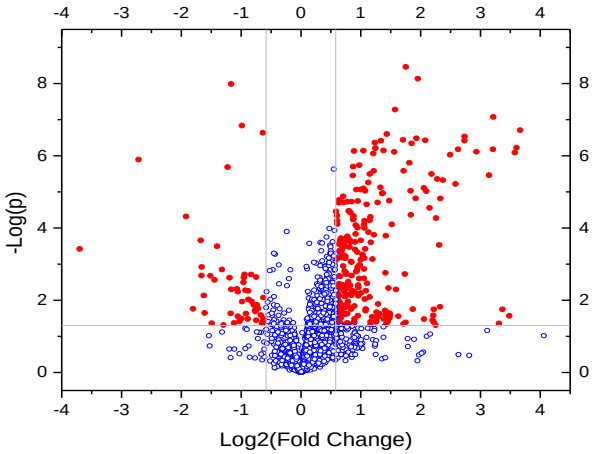
a


G3_1.
G3_2.
G3_3.
W3_3
W2_3
W1_3
G2_3
G1_1
G1_2
G2_1
G1_3
G2_2
t[1]
t[2]

b


G2_3
G3_2.
W3_2
W2_3
W1_1
G3_1.
G1_3
G1_1
G1_2
t[1]
t[2]

c


-Log(p)
Log2(Fold Change)

d

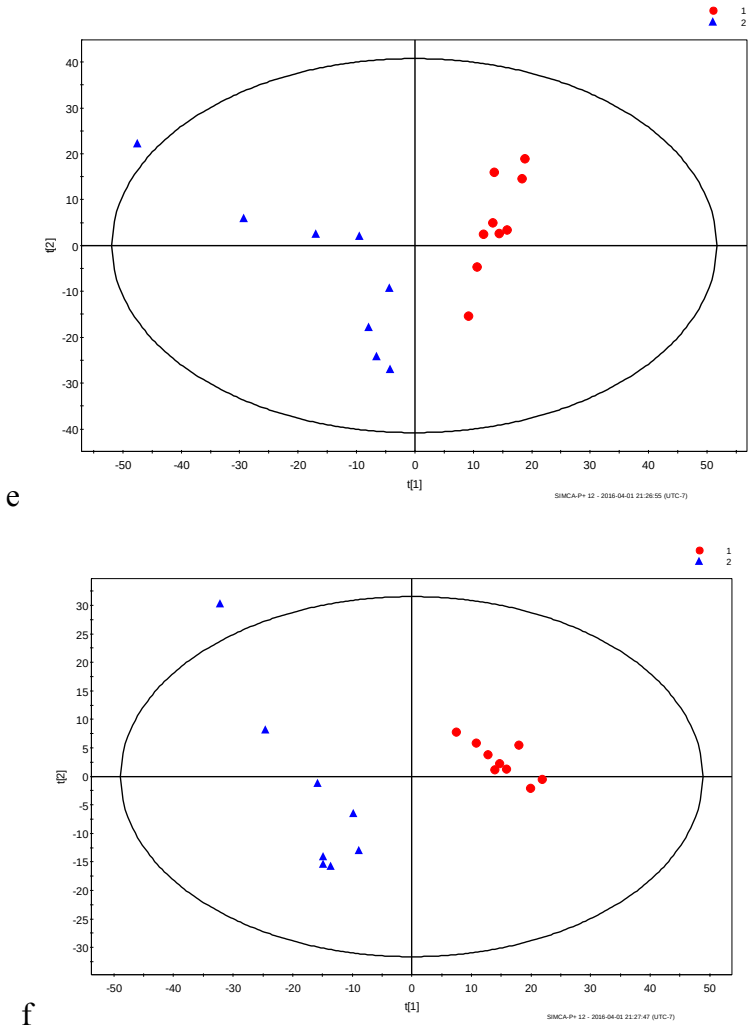


Figure 7.5 Volcano Plot, and PCA and PLS-DA score plot.

(a) Volcano plot from amine/phenol analysis. 276 metabolites has Fold Change (mutated/WT)> 1.5, p< 0.01, with 100 upregulated and 176 downregulated. (b) PCA plot from amine/phenol analysis. (c) PLS-DA score plot from amine/phenol analysis, R^2= 0.997, Q^2= 0.96. (d) Volcano Plot from global carboxylic acids profiling. 303 metabolites has Fold Change (mutated/WT)> 1.5, p< 0.05; 248 up-regulated and 55 down-regulated. (e) PCA plot from global carboxylic acids profiling. (f) PLS-DA score plot from global carboxylic acids profiling. R^2= 0.995, Q^2= 0.949.

7.3.5 AcCoA pool was maintained at high level in the mutant

According to the congruent transcriptomic and metabolomic analyses (Table 7.2), we hypothesized the tentative pathways for channeling of metabolites

towards FA syntheses as depicted in Figure 7.6. AcCoA was found 1.3-fold in G11_7 mutant.

Table 7.2 Important genes and metabolites affected in the mutant strain G11_7.

Metabolite Names	Fold Change (G11_7/ WT)	p-value	Participating pathways	Gene Names	Fold Change (G11_7/ WT)	FDR
L-Glutamic Acid	1.59	7.18E-07	Valine, leucine and isoleucine degradation & Transamination	isovaleryl-CoA-dehydrogenase (IVD)	4.93 */9.37*	3.3E-02 */0*
n-Propionyl-CoA	1.7	2.3E-03		3-methylcrotonyl-CoA carboxylase (MCCB)	4.87 */6.96*	2.5E-02 */-*
				acetyl-CoA C-acetyltransferase (ACCA)	*111.609*	*0*
				alanine-glyoxylate transaminase / (R)-3-amino-2-methylpropionate-pyruvate transaminase	*0.0115924*	*7.38E-08*
				2-oxoisovalerate dehydrogenase E1 component alpha subunit (BCKDHA)	*0.03*	*-*
Glutaryl-CoA	1.8	5.6E-03	Lysine degradation	*acetyl-CoA C-acetyltransferase (ACCA)*	*111.609*	*0*
Butyryl-CoA	1.3	1.4E-02		*histone-lysine N-methyltransferase (SETD2)*	*0.0560417*	*2.83E-02*

Pyruvate	1.55	1.75E-02	*			
Acetyl-CoA	1.3	3.1E-05	*			
Citrate	2.12	2.6E-04	TCA cycle			
Succinyl-CoA	5.9	4.3E-04				
Malonyl-CoA	2.1	4.9E-03	FA biosynthesis & elongation	*malonyl-CoA:acyl-carrier-protein transacylase (FabD)*	*5.67*	*1.0E-04*
Myristoyl-CoA (14:0-CoA)	0.1	1.2E-04				
Palmitoleoyl-CoA (16:1-CoA)	0.3	1.0E-03				
Oleoyl-CoA (18:1-CoA)	0.5	1.3E-02				
Arachidonyl CoA (20:4-CoA)	0.3	4.6E-03				
L-Proline	1.52	7.05E-05	Arginine and proline metabolism & β-	copper amine oxidase family protein	7.59	1.3E-02

1-Pyrroline-3-hydroxy-5-carboxylate	1.59	7.18E-07	Alanine & Transamination	(AMX1)	*/10.95*	/-
Ornithine	-2.26	2.88E-06				
Spermidine	-1.70	1.99E-06				
			Photosynthesis	plastocyanin (petE)	*22.4279*	*2.39E-07*
				cytochrome b6-f complex iron-sulfur subunit (petC)	*2.35716*	*1.02E-03*
				ferredoxin (petF)	*1.548*	-
				photosystem II 22kDa protein (psbS)	*0.0143979*	*1.75E-09*
				light-harvesting complex I chlorophyll a/b binding protein 1 (*LHCA1*)	*40.9366*	*1.05E-02*
				light-harvesting complex II chlorophyll a/b binding protein 1 (*LHCB1*)	*4.4559*	-
			Glycolysis / Gluconeogenesis	6-phosphofructokinase 1 (*pfkA*)	*8.3757*	-
				phosphoenolpyruvate carboxykinase (ATP) (PPC)	*21.3294*	*1.1E-09*
				fructose-1,6-bisphosphatase I (*FBP*)	*43.363*	-

The changes in gene expression level found from in-house pipeline are shown in italic. Changes found from Partek pipeline are shown in normal font.

* Pyruvate and acetyl-CoA are core intermediates participating in many pathways.

- No FDR is provided for specific genes since multiple contigs are detected.

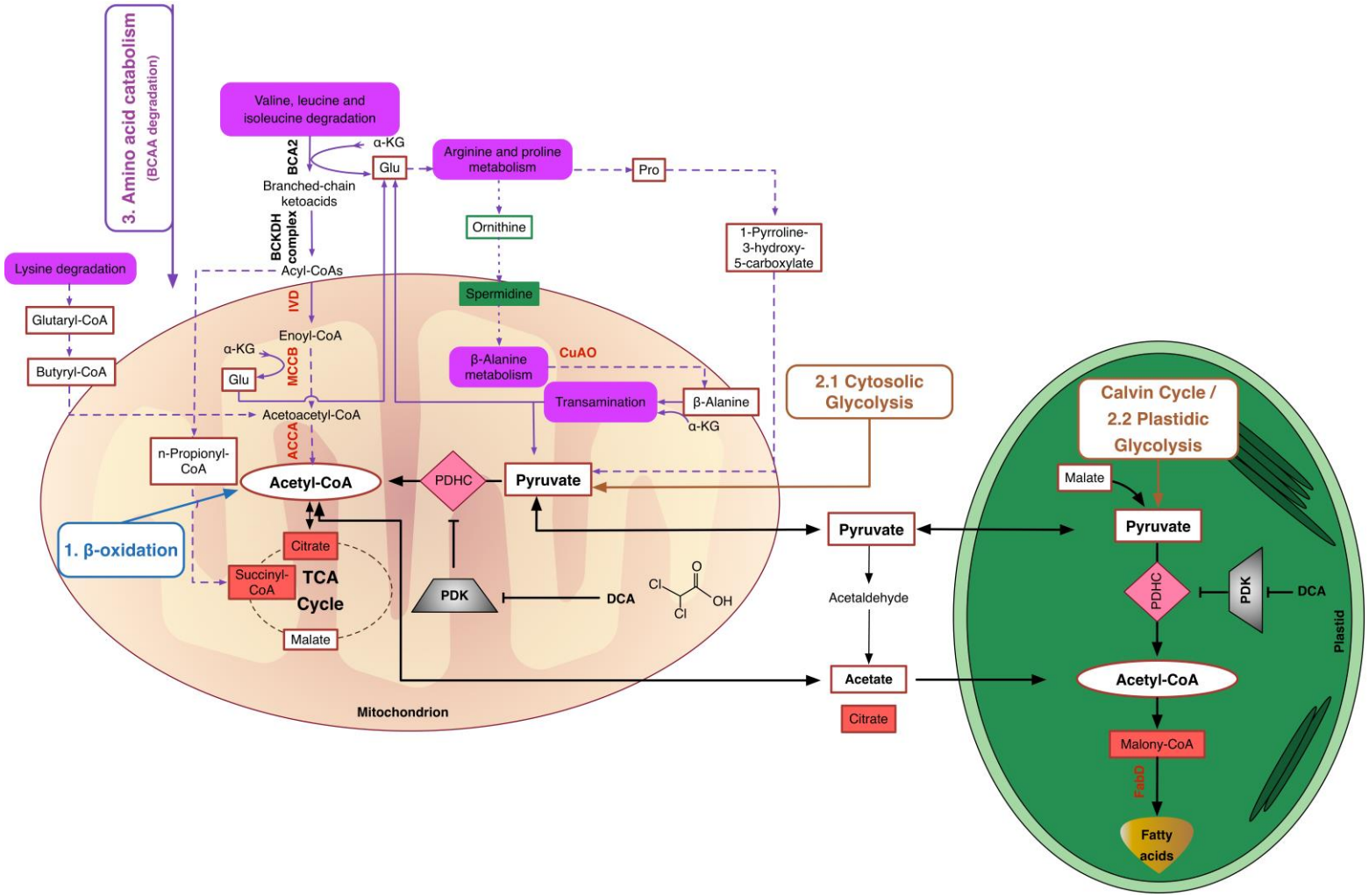
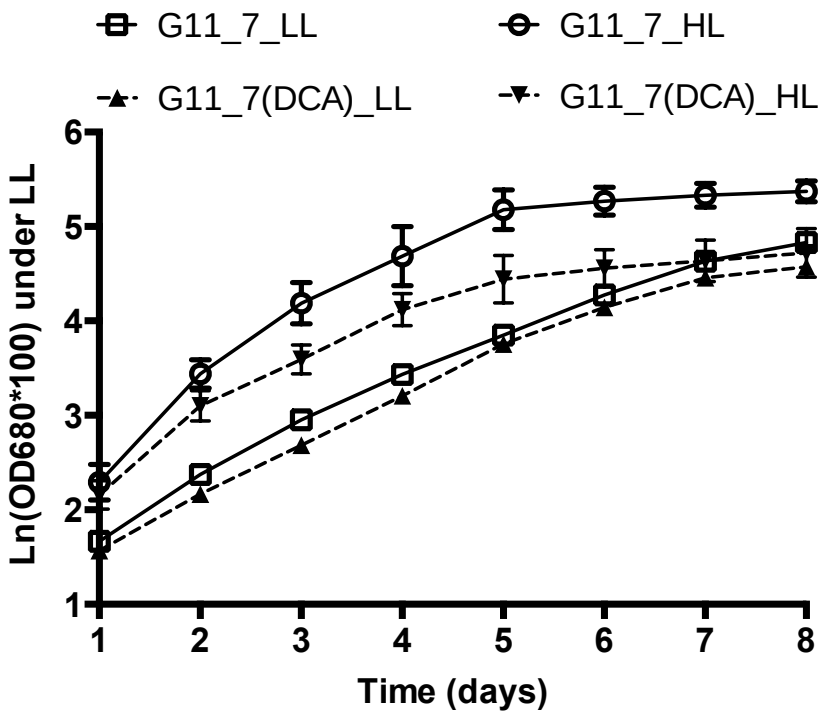

3. Amino acid catabolism
(BCAA degradation)
Valine, leucine and isoleucine degradation
BCA2
α-KG
Glu
Branched-chain ketoacids
BCKDH complex
Acyl-CoAs
Arginine and proline metabolism
Pro
Ornithine
1-Pyrroline-3-hydroxy-5-carboxylate
Lysine degradation
Glutaryl-CoA
Butyryl-CoA
IVD
Enoyl-CoA
α-KG
Glu
MCCB
Acetoacetyl-CoA
Spermidine
β-Alanine metabolism
CuAO
Transamination
β-Alanine
α-KG
2.1 Cytosolic Glycolysis
n-Propionyl-CoA
ACCA
Acetyl-CoA
PDHC
Pyruvate
1. β-oxidation
Citrate
Succinyl-CoA
TCA Cycle
Malate
PDK
DCA
Mitochondrion
Pyruvate
Acetaldehyde
Acetate
Citrate
Calvin Cycle / 2.2 Plastidic Glycolysis
Malate
Pyruvate
PDHC
PDK
DCA
Plastid
Acetyl-CoA
Malony-CoA
FabD
Fatty acids

Figure 7.6 Hypothesized *D. tertiolecta* G11_7 mutant FA metabolic pathways.

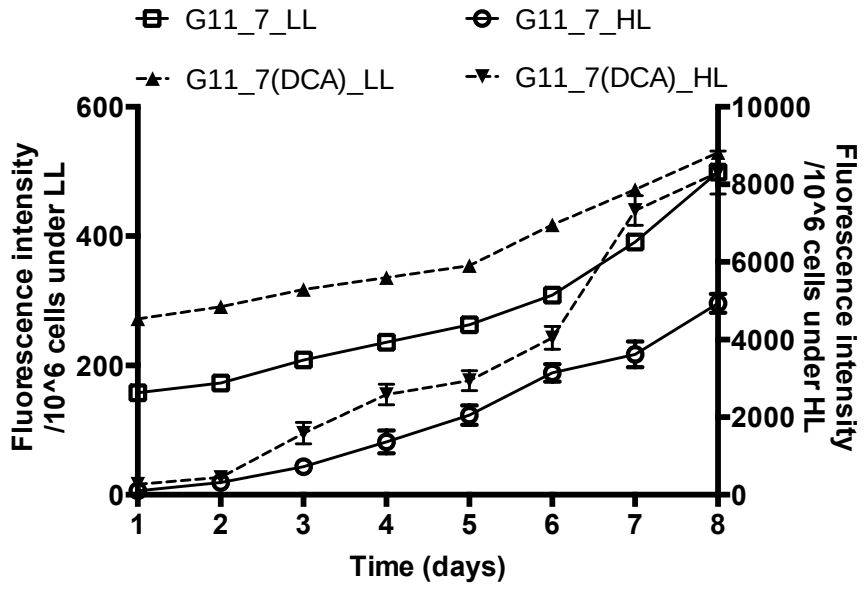
Enzyme names and intermediates are abbreviated in the figure. Validated overexpressed gene-coding enzymes are shown in red color font sitting on the pathway lines. Validated significantly increased intermediates are surrounded by red solid line rectangles, and decreased intermediates are surrounded by green solid line rectangles. The increased/decreased intermediates with a fold-change >= 2 or <= -2 are shown in filled rectangles with red/green color. The common three sources of acetyl-CoA are 1, β-oxidation, 2, glycolysis, and 3: amino acid catabolism. The specific core pathways are shown in solid arrows for one-step reactions and dash arrows for multiple-step reactions. The DCA treatment related pathways have two critical enzymes, PDH and PDK in pink prismatic and grey trapezoid shapes, respectively.

The enzyme and intermediate abbreviations are listed below: BCA2, branched-chain amino acid aminotransferase; BCKDH, branched-chain ketoacid dehydrogenase; IVD, isovaleryl-CoA dehydrogenase; MCCB, 3-methylcrotonyl-CoA carboxylase beta unit; ACCA: acetyl-CoA C-acetyltransferase; AMX1, copper amine oxidase family protein. PDHC, pyruvate dehydrogenase complex; PDK, pyruvate dehydrogenase kinase; FabD: malonyl-CoA:acyl-carrier-protein transacylase; α-KG, α-ketoglutarate; Glu, Glutamate; Pro, Proline.

Pyruvate, malate, proline and 1-Pyrroline-3-hydroxy-5-carboxylate were accumulated in G11_7 mutant, suggesting the relief of PDHC/PDK regulatory process was necessary for efficient bioconversion of the pyruvate-dependent FA precursors. To achieve this, DCA was further added to the mutant culture medium. TAG accumulation and growth curve in the DCA treated G11_7 mutant under both low light and high light conditions are shown in Figure 7.7a-b. There was a significant enhancement of TAG accumulation detected in DCA treated mutant cells from exponential growth phase, with an elevated AcCoA level (Figure 7.7c). Consistently, mRNA expression level of the *DtPDK* was significantly reduced in DCA treated groups (Figure 7.7d)

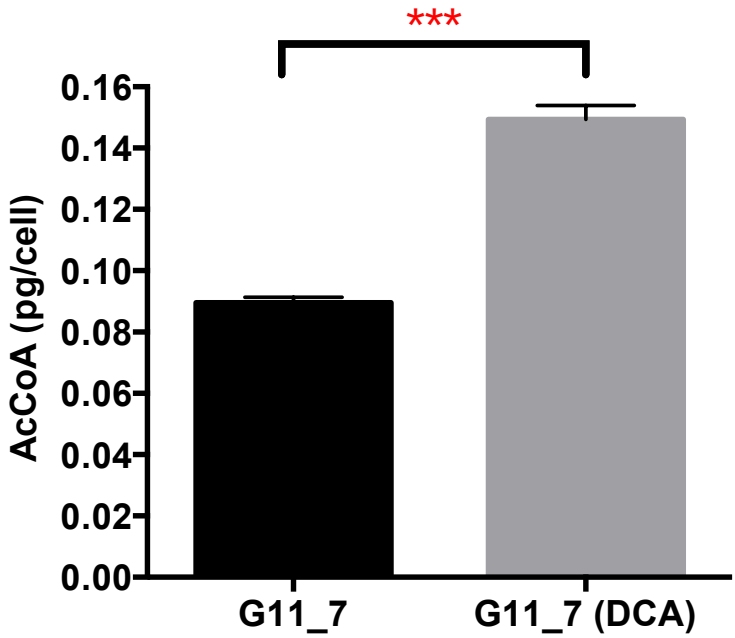

G11_7_LL
G11_7_HL
G11_7(DCA)_LL
G11_7(DCA)_HL
Ln(OD680*100) under LL
Time (days)

a


G11_7_LL
G11_7_HL
G11_7(DCA)_LL
G11_7(DCA)_HL
Fluorescence intensity /10^6 cells under LL
Fluorescence intensity /10^6 cells under HL
Time (days)

b

c



d

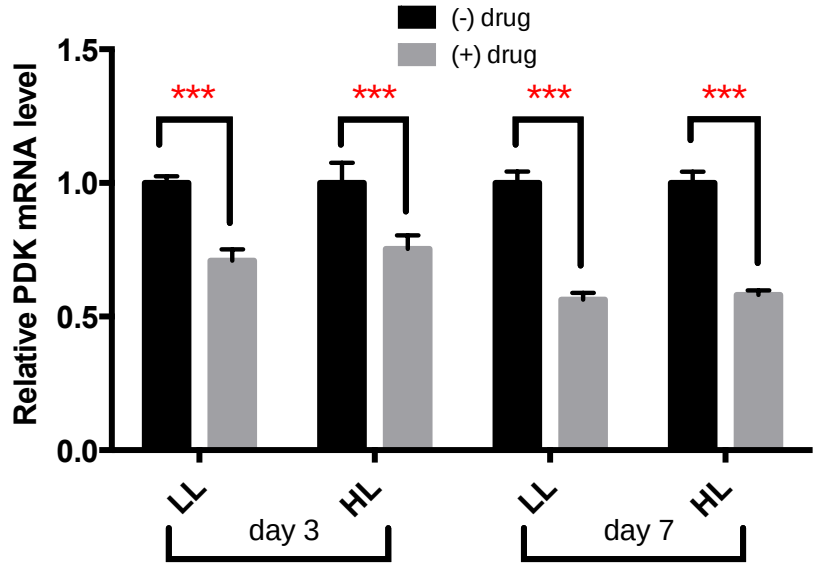


Figure 7.7 Physiological study and DtPDK mRNA expression levels of G11_7 mutant after DCA treatment.

(a) Growth curve monitored by spectrophotometry under low light (LL) and high light (HL) conditions (LL: 30 μmol photons•m^{-2}•s^{-1}, HL: 320 μmol photons•m^{-2}•s^{-1}). (b) TAG quantitative assay by Nile red staining method under low light (LL) and high light (HL) conditions (left y axis shows the LL fluorescent performance, right y axis shows the HL fluorescent performance). (c) Acetyl-CoA level change after addition of DCA to G11_7 under the low light condition. (d) Relative *DtPDK* mRNA expression levels on day 3 and day 7 under different light conditions (LL: 30 μmol photons•m^{-2}•s^{-1}, HL: 320 μmol photons•m^{-2}•s^{-1}). Error bars, SEM. Statistical analyses were performed using Student t test, ***$P < 0.001$.

7.3.6 BCAAs as a potential precursor for acetyl-CoA production in TCA metabolism and lipogenesis

In contrast to the WT, the mutant G11_7 has an activated BCAA catabolism via intensive upregulation of IVD, MCCB and ACCA, resulting in an enhancement of the AcCoA pool. The increase in glutaryl-CoA (1.8 fold) and butyryl-CoA (1.3 fold) participating in the lysine degradation pathway also contributed to AcCoA via the upregulation of ACCA. The acetyl residue of AcCoA enters the TCA cycle by reaction with oxaloacetate, and subsequently incorporated into citrate (2.12 fold). According to the targeted CoA analyses, n-propionyl-CoA, a prerequisite for succinyl-CoA, had a 1.7-fold increase in the mutant. Propionyl-CoA is known to be carboxylated to generate methylamalonyl-CoA, which is racemized and then isomerized to form succinyl-CoA (5.9 fold), a member of the TCA cycle [157]. The flux from both citrates in the TCA cycle and propionyl-CoA accounts for the massive increase of succinyl-CoA in the mutant.

Thus besides priming the TCA cycle for cell growth, the overproduced AcCoA was channelled into the FA reservoir. The enhancement of malonyl-CoA (2.1 fold), which plays a key role in the FA biosynthesis and chain elongation, provides the evidence for this.

BCAA catabolic flux was activated by the aforementioned genes, particularly in the leucine degradation pathway. To confirm the contribution of leucine in TAG synthesis in BCAA catabolism, G11_7 mutant and WT were given a spike of different concentrations of leucine in the normal ATCC medium. In contrast to the WT, which did not metabolize leucine as the precursor for cell

growth and lipid production, G11_7 mutant strain showed increase TAG production in a dose dependent manner (Figure 7.8). Increased in TAG content with addition of leucine in the medium accounted for as much as 39% of TAG pools.

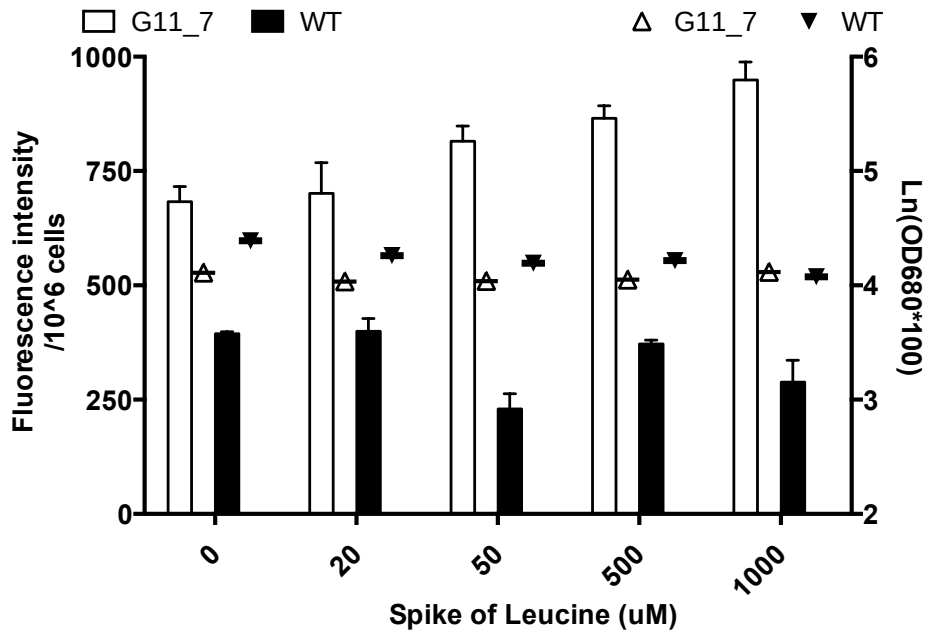


Figure 7.8 Effect of leucine spike to TAG accumulation and culture growth.

TAG contents of G11_7 and WT are presented in bars (primary axis, white bar - G11_7, black bar - WT), and the corresponding culture growth is presented in scatter dot plots (secondary axis, white upward triangle - G11_7, black downward triangle - WT) under different leucine concentration.

These data provide experimental evidence for the functional role of BCAA catabolism, reinforcing its importance in regulating lipogenesis in the exponential growth phase. There might be bottlenecks in the synthesis of structural carbohydrates, proteins and lipids for additional cell growth, since no increase in cell growth rate was observed. Instead, pronounced increase of FA was detected providing a hint that FA biosynthesis related processes might not be rate limiting during the exponential cell growth phase, and

AcCoA generated from the BCAA pathway was transported into chloroplasts for FA accumulation in the mutant.

7.3.7 Contribution of amino acids to TAG accumulation

The primary purpose of amino acids is to serve as the building blocks for protein biosynthesis in living cells, as a consequence, the amount of free amino acids is trivial under most circumstances especially in cultures under high growth rate [144]. Amino acids are derived from the TCA cycle, which provides carbon skeletons via 2-oxoglutarate or oxaloacetate [156, 158, 159]. New amino acids can also be formed from transamination by transferring the amino group to a ketoacid [160]. In our case, a significant proportion of free amino acids were degraded, driven by transcriptional upregulation of *DtIVD*, *DtMCCB*, and *DtACCA* encoding for the key enzymes in the BCAA catabolic pathway, leading to AcCoA synthesis. This strategy shows the feasibility of using the aforementioned third source for AcCoA, which was previously ignored.

The functional role of BCAA catabolic process in lipogenesis has been demonstrated in other various organisms. In the diatom, *P. tricurnutum*, inhibition of *MCCB* expression by RNA interference disturbed the carbon flux, resulting in decreased TAG accumulation and impaired biomass growth [157]. Green et al. [161] highlighted the contribution of BCAAs to adipocyte metabolism in mouse cell line (3T3-L1 cells) and demonstrated that amino acids (BCAAs in particular) from both extracellular sources and protein catabolism were highly utilized by differentiated adipocytes. Inhibition of BCAA catabolism negatively influenced 3T3-L1 adipogenesis. In the study

of Peng et al. [162], BCAA catabolic mutants defective in enzymes both upstream and downstream of IVD displayed enhanced senescence in prolonged darkness, showing that function of BCAA catabolism in providing TCA cycle substrates in energy-limited conditions. It also demonstrated that IVD influences energy homeostasis in multiple ways, providing BCAA catabolic CoA intermediates to the mitochondrial electron transport chain, as well as catabolizing additional substrates such as phytanoyl-CoA and aromatic amino acids [163, 164].

Interestingly, other amino acid catabolic pathways were found in concordant in contributing to the TCA cycle and AcCoA production. Lysine metabolism was previously demonstrated to interact with plant energy metabolism [165]. Vorapreeda et al. [166] also reported that leucine and lysine degradation in oleaginous fungi provided the alternative substrate for AcCoA as the precursor for lipid production, by contrast to that in non-oleaginous fungi. This is confirmed by the free amino acid (leucine) uptake study reported here. Free amino acid uptake (transport, assimilation /accumulation) and excretion had been observed in microalgae [167, 168]. Three exogenous transamination and deamination cycles introduced by Huo et al. [168] also reengineered carbon flux for fuel-convertible amino acids, and enabled protein hydrolysates to be used for fuel production.

7.4 Conclusion

In conclusion, we have studied a TAG-rich mutant strain of *D. tertiolecta*

under controlled laboratory conditions to advance our understanding of lipid metabolic pathways in the growth phase. Our study revealed a "3-step α loop" model to elucidate lipogenesis in the exponential growth phase as shown in Figure 7.9 and summarized below:

The vertical path: Under normal conditions, a number of key enzymes in microalgae supply carbon precursors for *de novo* FA synthesis, which include those involved in PDHC, glycolysis, and suites of specific transporters. They were found substantially upregulated under ND conditions [72]. This pyruvate-dependent glycolysis pathway is tightly regulated by cell growth via PDHC/PDK cascade (Figure 7.9, route 1, 2). AcCoA and NADPH produced from this pathway are readily used by the TCA cycle to produce amino acids, macromolecules, and energy for biomass growth. During this process, a minor fraction of AcCoA and NADPH is used for FAs synthesis [145].

The back loop: While in G11_7 mutant, genes involved in the amino acid catabolism in mitochondria were enhanced. The overflow free amino acids and catabolized proteins were channeled into AcCoA. This provides an additional source of AcCoA for FA synthesis (Figure 7.9, route 5, 6). The recycling of amino acids at a moderate level led to a temporal increase in AcCoA concentration, with little compromise on the biomass growth rate (4%). This elevated AcCoA concentration (30%) exceeds the demand (set point) for biomass growth, which shunts the carbon and energy precursors to the FA synthesis route (Figure 7.9, route 7). Genes responsible for FA synthesis were constantly overexpressed at the mRNA level and this equilibrium leads to ultimate generation of TAG in the growth phase.

The pull down: In respond to the drawn down of [AcCoA] for lipogenesis, a pull down of carbon flux from photosynthetic process takes effect, leading to the increase in photosynthetic rate (33%), and overexpression of genes participating in photosynthesis and glycolysis. The inclusion of DCA removed the regulation of PDHC, led to an increase in AcCoA and accumulation in TAG (42%) (Figure 7.9, route 8). In consistency with our observation, bioengineering manipulation, such as antisense knockdown of PDK in the diatom *P. tricurnutum*, was reported to promote TAG production by 82% [169].

The proposed "3-step α loop" model suggests that the rate of lipogenesis in the growth phase is determined by the balance between the carbon/ energy supply, biomass synthesis (growth) and amino acid catabolism. When the carbon metabolism is at high gears, a rapid catabolic rate would lead to accumulation of AcCoA and FA synthesis, without compromise on biomass growth. However, an overrun amino acid catabolism would result in reduced growth rate, lower biomass concentration and ultimately lower TAG productivity of the culture. Collectively, TAEP in microalgae could be stimulated by elevated AcCoA level through multiple approaches. Besides aforementioned two major approaches, a balance for FA oxidation is also of great interest to be investigated. To our knowledge, few genetic manipulations have been achieved in promoting TAG production in the growth phase. Collectively, the deliberated investigation provides targets for metabolic engineering of eukaryotic microalgae for efficient lipid production, and may inspire novel biofuel production technology based on growth-phase lipid producing oleaginous microalgae.

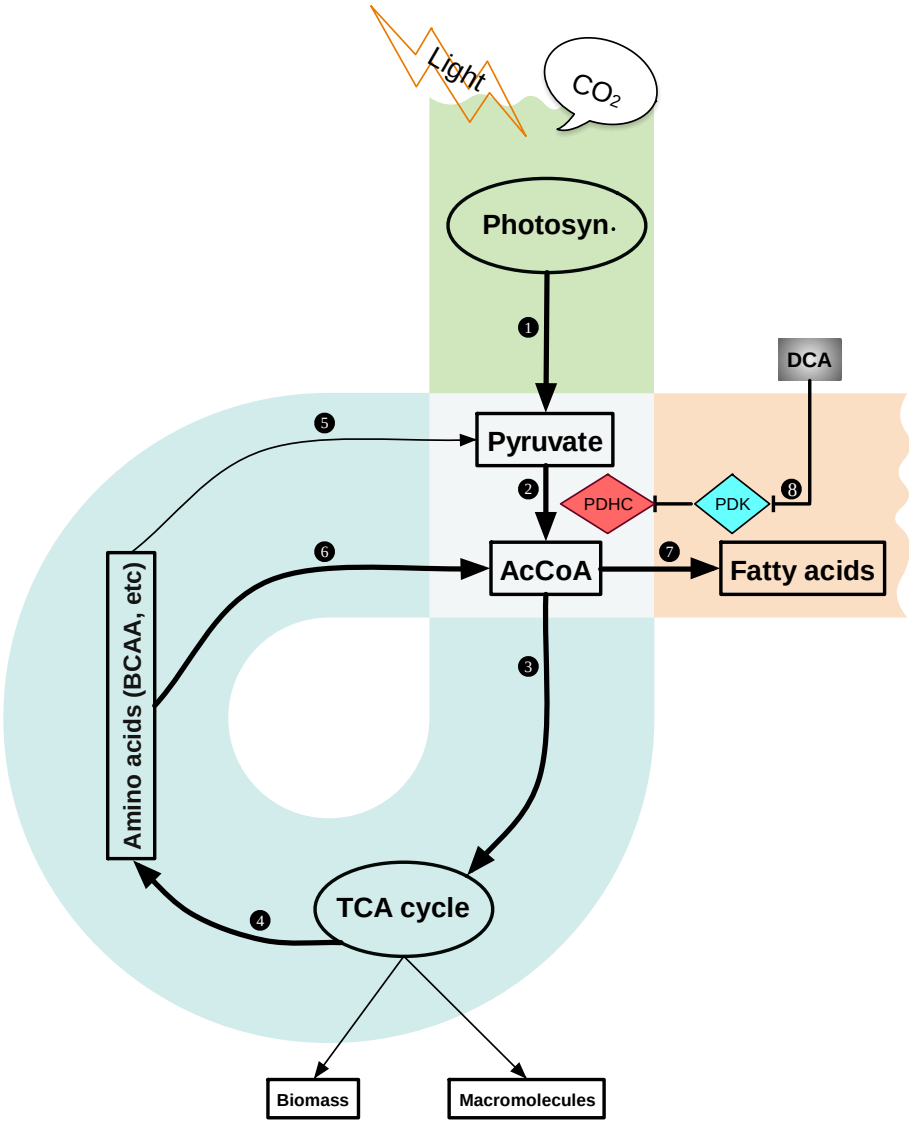


Figure 7.9 Regulation of metabolic pathways related to carbon capture and conversion in the growth phase of *Dunaliella tertiolecta* mutant G11_7.

Chapter 8 Conclusion and future directions

The potential of the *D. tertiolecta* mutants as microalgal candidate for biofuel production were investigated and demonstrated in this project. Via the optimized Nile red quantitative measurement of neutral lipids, high-lipid *D. tertiolecta* strains were screened and generated. *D. tertiolecta* mutant D9 was observed to have a consistent 2-4 folds increase in lipid content compared to the WT strain without compromise on the cell growth, which was applied in the RNA-Seq study. To characterize it at the transcriptome level, Bag2D workflow for transcriptome database construction and annotation was developed. Moreover, in-house pipeline to analyze the RNA-Seq data from this non-model organism was proposed. *D. tertiolecta* mutant G11 was subsequently discovered through FACS followed by random insertional mutagenesis, resulting in enhanced TAG productivity from exponential growth phase, implying that the lipid sink might not be limited to stationary phase or stress conditions.

It was demonstrated in the mutant characterizations and ND study that more energy should be captured from photosynthesis for the purpose of overall lipid production. G11_7 was deeply investigated through the integration of transcriptomics and metabolomics, showing that AcCoA pool regulated neutral lipid synthesis in *D. tertiolecta* in the exponential growth phase.

Generation and characterizations of such mutants with potential commercial feasibility have provided useful insights into perspectives of the development of carbon bioconversion processes and optimization of microalgal strains. It is hoped that the information gained from this project would contribute to the

feasibility of *D. tertiolecta* used as a microalgal candidate for biofuel. Despite these progresses, there remains detailed demonstration in the understanding of carbon channeling of lipid metabolism in different growth phases, and critical engineering breakthroughs related to algal mass culture and downstream processing. For further study, carbon isotope flux tracing, aimed at the corresponding amino acids to TAG metabolic route, could be used to validate the models proposed here.

Reference

1. **US Energy Information Administration**.
2. Georgianna DR, Mayfield SP: **Exploiting diversity and synthetic biology for the production of algal biofuels**. *Nature* 2012, **488**(7411):329-335.
3. Merchant SS, Kropat J, Liu B, Shaw J, Warakanont J: **TAG, You're it! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation**. *Current opinion in biotechnology* 2012, **23**(3):352-363.
4. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A: **Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances**. *Plant J* 2008, **54**(4):621-639.
5. Dellomonaco C, Fava F, Gonzalez R: **The path to next generation biofuels: successes and challenges in the era of synthetic biology**. *Microbial Cell Factories* 2010, **9**(1):1.
6. Hossain AS, Salleh A, Boyce AN, Chowdhury P, Naqiuddin M: **Biodiesel fuel production from algae as renewable energy**. *American journal of biochemistry and biotechnology* 2008, **4**(3):250-254.
7. Christie W: **Lipid analysis: isolation, separation, identification, and structural analysis of lipids. Ayr, Scotland**. In.: Oily Press. ISBN 0-9531949-5-7; 2003.
8. Li‐Beisson Y, Beisson F, Riekhof W: **Metabolism of acyl‐lipids in *Chlamydomonas reinhardtii***. *The Plant Journal* 2015, **82**(3):504-522.
9. Somerville CR, Browse, J., Jaworski, J. and Ohlrogge, J.: **Lipids. In Biochemistry and Molecular Biology of Plants (Buchanan, B.B., Gruis- sem, W. and Jones, R.L. eds)**. *Rockville, MD: American Society of Plant Physiologists, pp Chap 10* 2000.
10. Guckert JB, Cooksey KE: **Triglyceride accumulation and fatty acid profile changes in *Chlorella* (Chlorophyta) during high Ph‐induced cell cycle Inhibition1**. *Journal of Phycology* 1990, **26**(1):72-79.
11. Harwood JL: **Fatty acid metabolism**. *Annual Review of Plant Physiology and Plant Molecular Biology* 1988, **39**(1):101-138.
12. Pohl P, Zurheide F: **Fatty acids and lipids of marine algae and the control of their biosynthesis by environmental factors**. *Marine algae in pharmaceutical science, editors, Heinz A Hoppe, Tore Levring, Yukio Tanaka* 1979.
13. Wada H, Murata N: **Membrane lipids in cyanobacteria**. In: *Lipids in photosynthesis: structure, function and genetics.* Springer; 1998: 65-81.
14. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A: **Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances**. *The Plant Journal* 2008, **54**(4):621-639.
15. Rogers L, Gallon JR: **Biochemistry of the Algae and Cyanobacteria**: Clarendon Press; 1988.
16. Goold H, Beisson F, Peltier G, Li-Beisson Y: **Microalgal lipid droplets: composition, diversity, biogenesis and functions**. *Plant cell reports* 2015, **34**(4):545-555.
17. Durrett TP, Benning C, Ohlrogge J: **Plant triacylglycerols as feedstocks for the production of biofuels**. *The Plant Journal* 2008, **54**(4):593-607.
18. Murphy DJ: **The biogenesis and functions of lipid bodies in animals, plants and microorganisms**. *Progress in lipid research* 2001, **40**(5):325-438.
19. Goodson C, Roth R, Wang ZT, Goodenough U: **Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost**. *Eukaryotic Cell* 2011, **10**(12):1592-1606.
20. Nguyen HM, Baudet M, Cuine S, Adriano JM, Barthe D, Billon E, Bruley C, Beisson F, Peltier G, Ferro M: **Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: with focus on proteins involved in lipid metabolism**. *Proteomics* 2011, **11**(21):4266-4273.
21. Kennedy EP, Weiss SB: **The function of cytidine coenzymes in the biosynthesis**

of phospholipides. *Journal of Biological Chemistry* 1956, **222**(1):193-214.
22. Chen JE, Smith AG: **A look at diacylglycerol acyltransferases (DGATs) in algae**. *Journal of biotechnology* 2012, **162**(1):28-39.
23. Radakovits R, Jinkerson RE, Darzins A, Posewitz MC: **Genetic engineering of algae for enhanced biofuel production**. *Eukaryotic cell* 2010, **9**(4):486-501.
24. Courchesne NMD, Parisien A, Wang B, Lan CQ: **Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches**. *Journal of biotechnology* 2009, **141**(1):31-41.
25. Sharma KK, Schuhmann H, Schenk PM: **High lipid induction in microalgae for biodiesel production**. *Energies* 2012, **5**(5):1532-1553.
26. Bouvier‐Navé P, Benveniste P, Oelkers P, Sturley SL, Schaller H: **Expression in yeast and tobacco of plant cDNAs encoding acyl CoA: diacylglycerol acyltransferase**. *European journal of biochemistry* 2000, **267**(1):85-96.
27. Jako C, Kumar A, Wei Y, Zou J, Barton DL, Giblin EM, Covello PS, Taylor DC: **Seed-specific over-expression of an Arabidopsis cDNA encoding a diacylglycerol acyltransferase enhances seed oil content and seed weight**. *Plant Physiology* 2001, **126**(2):861-874.
28. Wynn JP, bin Abdul Hamid A, Ratledge C: **The role of malic enzyme in the regulation of lipid accumulation in filamentous fungi**. *Microbiology* 1999, **145**(8):1911-1917.
29. Li Z, Sun H, Mo X, Li X, Xu B, Tian P: **Overexpression of malic enzyme (ME) of Mucor circinelloides improved lipid accumulation in engineered *Rhodotorula glutinis***. *Applied microbiology and biotechnology* 2013, **97**(11):4927-4936.
30. Shen Y-Q, Burger G: **Plasticity of a key metabolic pathway in fungi**. *Functional & integrative genomics* 2009, **9**(2):145-151.
31. Song D, Fu J, Shi D: **Exploitation of oil-bearing microalgae for biodiesel**. *Chinese Journal of Biotechnology* 2008, **24**(3):341-348.
32. Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J: **Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds**. *Plant Physiology* 1997, **113**(1):75-81.
33. Bigogno C, Khozin-Goldberg I, Boussiba S, Vonshak A, Cohen Z: **Lipid and fatty acid composition of the green oleaginous alga *Parietochloris incisa*, the richest plant source of arachidonic acid**. *Phytochemistry* 2002, **60**(5):497-503.
34. Chiu S-Y, Kao C-Y, Tsai M-T, Ong S-C, Chen C-H, Lin C-S: **Lipid accumulation and CO_2 utilization of *Nannochloropsis oculata* in response to CO_2 aeration**. *Bioresource technology* 2009, **100**(2):833-838.
35. Widjaja A, Chien C-C, Ju Y-H: **Study of increasing lipid production from fresh water microalgae *Chlorella vulgaris***. *Journal of the Taiwan Institute of Chemical Engineers* 2009, **40**(1):13-20.
36. Trentacoste EM, Shrestha RP, Smith SR, Glé C, Hartmann AC, Hildebrand M, Gerwick WH: **Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth**. *Proceedings of the National Academy of Sciences* 2013, **110**(49):19748-19753.
37. Li Y, Horsman M, Wang B, Wu N, Lan CQ: **Effects of nitrogen sources on cell growth and lipid accumulation of green alga *Neochloris oleoabundans***. *Applied microbiology and biotechnology* 2008, **81**(4):629-636.
38. Rodolfi L, Chini Zittelli G, Bassi N, Padovani G, Biondi N, Bonini G, Tredici MR: **Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low‐cost photobioreactor**. *Biotechnology and bioengineering* 2009, **102**(1):100-112.
39. Barzegari A, Hejazi MA, Hosseinzadeh N, Eslami S, Aghdam EM, Hejazi MS: ***Dunaliella* as an attractive candidate for molecular farming**. *Molecular biology reports* 2010, **37**(7):3427-3430.
40. Ben-Amotz A, Polle JE, Rao DS: **The alga *Dunaliella*: biodiversity, physiology, genomics and biotechnology**: Science Publishers Enfield, NH; 2009.
41. Goyal A: **Osmoregulation in *Dunaliella*, Part II: Photosynthesis and starch contribute carbon for glycerol synthesis during a salt stress in *Dunaliella tertiolecta***. *Plant Physiology and Biochemistry* 2007, **45**(9):705-710.
42. Hosseini Tafreshi A, Shariati M: ***Dunaliella* biotechnology: methods and**

applications. *Journal of Applied Microbiology* 2009, **107**(1):14-35.
43. Rismani-Yazdi H, Haznedaroglu BZ, Bibby K, Peccia J: **Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels**. *BMC genomics* 2011, **12**(1):148.
44. Takagi M, Yoshida T: **Effect of salt concentration on intracellular accumulation of lipids and triacylglyceride in marine microalgae *Dunaliella* cells**. *Journal of bioscience and bioengineering* 2006, **101**(3):223-226.
45. Acién F, Fernández J, Magán J, Molina E: **Production cost of a real microalgae production plant and strategies to reduce it**. *Biotechnology advances* 2012, **30**(6):1344-1353.
46. Daelman MR, Sorokin D, Kruse O, van Loosdrecht MC, Strous M: **Haloalkaline bioconversions for methane production from microalgae grown on sunlight**. *Trends in biotechnology* 2016, **34**(6):450-457.
47. Norsker N-H, Barbosa MJ, Vermuë MH, Wijffels RH: **Microalgal production—a close look at the economics**. *Biotechnology advances* 2011, **29**(1):24-27.
48. Reijnders L: **Do biofuels from microalgae beat biofuels from terrestrial plants?** *Trends in biotechnology* 2008, **26**(7):349-350.
49. Sialve B, Bernet N, Bernard O: **Anaerobic digestion of microalgae as a necessary step to make microalgal biodiesel sustainable**. *Biotechnology advances* 2009, **27**(4):409-416.
50. Stephens E, Ross IL, King Z, Mussgnug JH, Kruse O, Posten C, Borowitzka MA, Hankamer B: **An economic and technical evaluation of microalgal biofuels**. *Nature biotechnology* 2010, **28**(2):126-128.
51. Uduman N, Qi Y, Danquah MK, Forde GM, Hoadley A: **Dewatering of microalgal cultures: a major bottleneck to algae-based fuels**. *Journal of renewable and sustainable energy* 2010, **2**(1):012701.
52. Zamalloa C, Vulsteke E, Albrecht J, Verstraete W: **The techno-economic potential of renewable energy through the anaerobic digestion of microalgae**. *Bioresource technology* 2011, **102**(2):1149-1158.
53. Bligh EG, Dyer WJ: **A rapid method of total lipid extraction and purification**. *Can J Biochem Physiol* 1959, **37**(8):911-917.
54. Wawrik B, Harriman BH: **Rapid, colorimetric quantification of lipid from algal cultures**. *J Microbiol Methods* 2010, **80**(3):262-266.
55. Bertozzini E, Galluzzi L, Penna A, Magnani M: **Application of the standard addition method for the absolute quantification of neutral lipids in microalgae using Nile red**. *Journal of microbiological methods* 2011, **87**(1):17-23.
56. Cooksey KE, Guckert JB, Williams SA, Callis PR: **Fluorometric-determination of the neutral lipid-content of mircroalgal cells using Nile red**. *Journal of Microbiological Methods* 1987, **6**(6):333-345.
57. Elsey D, Jameson D, Raleigh B, Cooney MJ: **Fluorescent measurement of microalgal neutral lipids**. *Journal of Microbiological Methods* 2007, **68**(3):639-642.
58. Blanc G, Agarkova I, Grimwood J, Kuo A, Brueggeman A, Dunigan DD, Gurnon J, Ladunga I, Lindquist E, Lucas S: **The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation**. *Genome Biol* 2012, **13**(5):R39.
59. Blanc G, Duncan G, Agarkova I, Borodovsky M, Gurnon J, Kuo A, Lindquist E, Lucas S, Pangilinan J, Polle J: **The *Chlorella variabilis* NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex**. *The Plant Cell Online* 2010, **22**(9):2943-2955.
60. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Maréchal-Drouard L: **The *Chlamydomonas* genome reveals the evolution of key animal and plant functions**. *Science* 2007, **318**(5848):245-250.
61. Worden AZ, Lee J-H, Mock T, Rouzé P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV: **Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas***. *Science* 2009, **324**(5924):268-272.
62. Palenik B, Grimwood J, Aerts A, Rouzé P, Salamov A, Putnam N, Dupont C,

Jorgensen R, Derelle E, Rombauts S: **The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation**. *Proceedings of the National Academy of Sciences* 2007, **104**(18):7705-7710.

63. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK: **Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri***. *Science* 2010, **329**(5988):223-226.
64. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M: **The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism**. *Science* 2004, **306**(5693):79-86.
65. Shang C, Bi G, Yuan Z, Wang Z, Alam MA, Xie J: **Discovery of genes for production of biofuels through transcriptome sequencing of *Dunaliella parva***. *Algal Research* 2016, **13**:318-326.
66. Gimpel JA, Specht EA, Georgianna DR, Mayfield SP: **Advances in microalgae engineering and synthetic biology applications for biofuel production**. *Current opinion in chemical biology* 2013, **17**(3):489-495.
67. Radakovits R, Jinkerson RE, Fuerstenberg SI, Tae H, Settlage RE, Boore JL, Posewitz MC: **Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropsis gaditana***. *Nature communications* 2012, **3**:686.
68. Yang S, Guarnieri MT, Smolinski S, Ghirardi M, Pienkos PT: ***De novo* transcriptomic analysis of hydrogen production in the green alga *Chlamydomonas moewusii* through RNA-Seq**. *Biotechnol Biofuels* 2013, **6**(1):118.
69. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics**. *Nature Reviews Genetics* 2009, **10**(1):57-63.
70. Boyle NR, Page MD, Liu B, Blaby IK, Casero D, Kropat J, Cokus SJ, Hong-Hermesdorf A, Shaw J, Karpowicz SJ: **Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas***. *Journal of Biological Chemistry* 2012, **287**(19):15811-15825.
71. Fang L, Sun D, Xu Z, He J, Qi S, Chen X, Chew W, Liu J: **Transcriptomic analysis of a moderately growing subisolate *Botryococcus braunii* 779 (Chlorophyta) in response to nitrogen deprivation**. *Biotechnology for biofuels* 2015, **8**(1):1.
72. Li J, Han D, Wang D, Ning K, Jia J, Wei L, Jing X, Huang S, Chen J, Li Y: **Choreography of transcriptomes and lipidomes of *Nannochloropsis* reveals the mechanisms of oil synthesis in microalgae**. *The Plant Cell* 2014, **26**(4):1645-1665.
73. Rismani-Yazdi H, Haznedaroglu BZ, Hsin C, Peccia J: **Transcriptomic analysis of the oleaginous microalga *Neochloris oleoabundans* reveals metabolic insights into triacylglyceride accumulation**. *Biotechnology for Biofuels* 2012, **5**(1):1.
74. Shin H, Hong S-J, Kim H, Yoo C, Lee H, Choi H-K, Lee C-G, Cho B-K: **Elucidation of the growth delimitation of *Dunaliella tertiolecta* under nitrogen stress by integrating transcriptome and peptidome analysis**. *Bioresource technology* 2015, **194**:57-66.
75. Sun D, Zhu J, Fang L, Zhang X, Chow Y, Liu J: ***De novo* transcriptome profiling uncovers a drastic downregulation of photosynthesis upon nitrogen deprivation in the nonmodel green alga *Botryosphaerella sudeticus***. *BMC genomics* 2013, **14**(1):715.
76. Wang ZT, Ullrich N, Joo S, Waffenschmidt S, Goodenough U: **Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii***. *Eukaryotic cell* 2009, **8**(12):1856-1868.
77. Oliver DJ, Nikolau BJ, Wurtele ES: **Acetyl-CoA—life at the metabolic nexus**. *Plant Science* 2009, **176**(5):597-601.
78. Satoh A, Ichii K, Matsumoto M, Kubota C, Nemoto M, Tanaka M, Yoshino T, Matsunaga T, Tanaka T: **A process design and productivity evaluation for oil production by indoor mass cultivation of a marine diatom, *Fistulifera* sp. JPCC DA0580**. *Bioresource technology* 2013, **137**:132-138.
79. Lin H, Fang L, Low CS, Chow Y, Lee YK: **Occurrence of glycerol uptake in**

Dunaliella tertiolecta **under hyperosmotic stress**. *FEBS Journal* 2013, **280**(4):1064-1072.

80. Wahidin S, Idris A, Shaleh SRM: **The influence of light intensity and photoperiod on the growth and lipid content of microalgae *Nannochloropsis* sp**. *Bioresource technology* 2013, **129**:7-11.
81. Shoaf WT, Lium BW: **Improved extraction of chlorophyll a and b from algae using dimethyl sulfoxide**. *Limnology and Oceanography* 1976, **21**(6):926-928.
82. Fang L, Lin HX, Low CS, Wu MH, Chow Y, Lee YK: **Expression of the *Chlamydomonas reinhardtii* Sedoheptulose‐1, 7‐bisphosphatase in *Dunaliella bardawil* leads to enhanced photosynthesis and increased glycerol production**. *Plant biotechnology journal* 2012, **10**(9):1129-1135.
83. Lin H, Fang L, Low CS, Chow Y, Lee YK: **Occurrence of glycerol uptake in *Dunaliella tertiolecta* under hyperosmotic stress**. *FEBS J* 2013, **280**(4):1064-1072.
84. Kindle KL: **High-frequency nuclear transformation of *Chlamydomonas reinharditii***. *Proceedings of the National Academy of Sciences of the United States of America* 1990, **87**(3):1228-1232.
85. Shimogawara K, Fujiwara S, Grossman A, Usuda H: **High-efficiency transformation of *Chlamydomonas reinhardtii* by electroporation**. *Genetics* 1998, **148**(4):1821-1828.
86. Yao L, Tan TW, Ng Y-K, Ban KHK, Shen H, Lin H, Lee YK: **RNA-Seq transcriptomic analysis with Bag2D software identifies key pathways enhancing lipid yield in a high lipid-producing mutant of the non-model green alga *Dunaliella tertiolecta***. *Biotechnology for biofuels* 2015, **8**(1):1.
87. Rumin J, Bonnefond H, Saint-Jean B, Rouxel C, Sciandra A, Bernard O, Cadoret J-P, Bougaran G: **The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae**. *Biotechnology for biofuels* 2015, **8**(1):1.
88. Lee S-Y, Kim S-H, Hyun S-H, Suh HW, Hong S-J, Cho B-K, Lee C-G, Lee H, Choi H-K: **Fatty acids and global metabolites profiling of *Dunaliella tertiolecta* by shifting culture conditions to nitrate deficiency and high light at different growth phases**. *Process Biochemistry* 2014, **49**(6):996-1004.
89. Liu J, Sun Z, Zhong Y, Huang J, Hu Q, Chen F: **Stearoyl-acyl carrier protein desaturase gene from the oleaginous microalga *Chlorella zofingiensis*: cloning, characterization and transcriptional analysis**. *Planta* 2012, **236**(6):1665-1676.
90. Chen M, Tang H, Ma H, Holland TC, Ng KS, Salley SO: **Effect of nutrients on growth and lipid accumulation in the green algae *Dunaliella tertiolecta***. *Bioresource Technology* 2011, **102**(2):1649-1655.
91. Chen W, Zhang C, Song L, Sommerfeld M, Hu Q: **A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae**. *Journal of microbiological methods* 2009, **77**(1):41-47.
92. Xu P, Gu Q, Wang W, Wong L, Bower AG, Collins CH, Koffas MA: **Modular optimization of multi-gene pathways for fatty acids production in *E. coli***. *Nature communications* 2013, **4**:1409.
93. Terashima M, Freeman ES, Jinkerson RE, Jonikas MC: **A fluorescence‐activated cell sorting‐based strategy for rapid isolation of high‐lipid *Chlamydomonas* mutants**. *The Plant Journal* 2015, **81**(1):147-159.
94. Gee R, Goyal A, Byerrum RU, Tolbert NE: **Two isoforms of dihydroxyacetone phosphate reductase from the chloroplasts of *Dunaliella tertiolecta***. *Plant physiology* 1993, **103**(1):243-249.
95. Hirokawa T, Hata M, Takeda H: **Correlation between the starch level and the rate of starch synthesis during the developmental cycle of *Chlorella ellipsoidea***. *Plant and Cell Physiology* 1982, **23**(5):813-820.
96. Ratledge C: **Fatty acid biosynthesis in microorganisms being used for single cell oil production**. *Biochimie* 2004, **86**(11):807-815.
97. Ramazanov A, Ramazanov Z: **Isolation and characterization of a starchless mutant of *Chlorella pyrenoidosa* STL‐PI with a high growth rate, and high protein and polyunsaturated fatty acid content**. *Phycological Research* 2006, **54**(4):255-259.
98. Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S,

Bird D, DeBono A, Durrett TP: **Acyl-lipid metabolism**. *The Arabidopsis book/American Society of Plant Biologists* 2013, **11**.
99. Allen JF: **Photosynthesis of ATP—electrons, proton pumps, rotors, and poise**. *Cell* 2002, **110**(3):273-276.
100. May P, Christian J-O, Kempa S, Walther D: **ChlamyCyc: an integrative systems biology database and web-portal for *Chlamydomonas reinhardtii***. *Bmc Genomics* 2009, **10**(1):209.
101. Sharma SK, Kapoor M, Ramya T, Kumar S, Kumar G, Modak R, Sharma S, Surolia N, Surolia A: **Identification, characterization, and inhibition of *Plasmodium falciparum* β -hydroxyacyl-acyl carrier protein dehydratase (FabZ)**. *Journal of Biological Chemistry* 2003, **278**(46):45661-45671.
102. Chow CK: **Fatty acids in foods and their health implications**: CRC Press; 2007.
103. Knothe G, Steidley KR: **Kinematic viscosity of biodiesel fuel components and related compounds. Influence of compound structure and comparison to petrodiesel fuel components**. *Fuel* 2005, **84**(9):1059-1065.
104. Ryan III T, Dodge L, Callahan T: **The effects of vegetable oil properties on injection and combustion in two different diesel engines**. *Journal of the American Oil Chemists Society* 1984, **61**(10):1610-1619.
105. Blatti JL, Beld J, Behnke CA, Mendez M, Mayfield SP, Burkart MD: **Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions**. *PloS one* 2012, **7**(9):e42949.
106. Heath RJ, Rock CO: **Inhibition of-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in *Escherichia coli***. *Journal of Biological Chemistry* 1996, **271**(18):10996-11000.
107. Corwin J, Köhler S, Zerick J: **A virtual machine program-suite for distributed de novo genome construction and motif finding**.
108. Peréz-Sánchez H, Cecilia JM, Merelli I: **The role of High Performance Computing in Bioinformatics**.
109. Lee AF, Bennett JA, Manayil JC, Wilson K: **Heterogeneous catalysis for sustainable biodiesel production via esterification and transesterification**. *Chemical Society Reviews* 2014, **43**(22):7887-7916.
110. Cavonius LR, Carlsson N-G, Undeland I: **Quantification of total fatty acids in microalgae: comparison of extraction and transesterification methods**. *Analytical and bioanalytical chemistry* 2014, **406**(28):7313-7322.
111. Laurens LM, Quinn M, Van Wychen S, Templeton DW, Wolfrum EJ: **Accurate and reliable quantification of total microalgal fuel potential as fatty acid methyl esters by in situ transesterification**. *Analytical and bioanalytical chemistry* 2012, **403**(1):167-178.
112. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature biotechnology* 2011, **29**(7):644-652.
113. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M: ***De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis**. *Nature protocols* 2013, **8**(8):1494-1512.
114. Li B, Dewey CN: **RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome**. *BMC bioinformatics* 2011, **12**(1):1.
115. Leng N, Dawson JA, Thomson JA, Ruotti V, Rissman AI, Smits BM, Haag JD, Gould MN, Stewart RM, Kendziorski C: **EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments**. *Bioinformatics* 2013:btt087.
116. Black DL: **Mechanisms of alternative pre-messenger RNA splicing**. *Annual review of biochemistry* 2003, **72**(1):291-336.
117. Reddy AS: **Alternative splicing of pre-messenger RNAs in plants in the genomic era**. *Annu Rev Plant Biol* 2007, **58**:267-294.
118. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj T, Soreq H: **Function of alternative splicing**. *Gene* 2005, **344**:1-20.
119. Thompson JD, Gibson T, Higgins DG: **Multiple sequence alignment using ClustalW and ClustalX**. *Current protocols in bioinformatics* 2002:2.3. 1-2.3. 22.
120. Darling A, Carey L, Feng W-c: **The design, implementation, and evaluation of**

mpiBLAST. *Proceedings of ClusterWorld* 2003, **2003**:13-15.
121. Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP: **Sequencing depth and coverage: key considerations in genomic analyses**. *Nature Reviews Genetics* 2014, **15**(2):121-132.
122. Duret L, Mouchiroud D, Gautier C: **Statistical analysis of vertebrate sequences reveals that long genes are scarce in GC-rich isochores**. *Journal of Molecular Evolution* 1995, **40**(3):308-317.
123. Galtier N, Piganeau G, Mouchiroud D, Duret L: **GC-content evolution in mammalian genomes: the biased gene conversion hypothesis**. *Genetics* 2001, **159**(2):907-911.
124. Jabbari K, Bernardi G: **CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families**. *Gene* 1998, **224**(1):123-128.
125. Mouchiroud D, D'Onofrio G, Aissani B, Macaya G, Gautier C, Bernardi G: **The distribution of genes in the human genome**. *Gene* 1991, **100**:181-187.
126. Kianianmomeni A, Ong CS, Rätsch G, Hallmann A: **Genome-wide analysis of alternative splicing in *Volvox carteri***. *BMC genomics* 2014, **15**(1):1117.
127. Parra G, Bradnam K, Ning Z, Keane T, Korf I: **Assessing the gene space in draft genomes**. *Nucleic acids research* 2009, **37**(1):289-297.
128. Bullard JH, Purdom E, Hansen KD, Dudoit S: **Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments**. *BMC bioinformatics* 2010, **11**(1):1.
129. Blencowe BJ, Ahmad S, Lee LJ: **Current-generation high-throughput sequencing: deepening insights into mammalian transcriptomes**. *Genes & development* 2009, **23**(12):1379-1386.
130. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G: **Stem cell transcriptome profiling via massive-scale mRNA sequencing**. *Nature methods* 2008, **5**(7):613-619.
131. Francis WR, Christianson LM, Kiko R, Powers ML, Shaner NC, Haddock SH: **A comparison across non-model animals suggests an optimal sequencing depth for *de novo* transcriptome assembly**. *BMC genomics* 2013, **14**(1):1.
132. Li H, Lovci MT, Kwon Y-S, Rosenfeld MG, Fu X-D, Yeo GW: **Determination of tag density required for digital transcriptome analysis: application to an androgen-sensitive prostate cancer model**. *Proceedings of the National Academy of Sciences* 2008, **105**(51):20179-20184.
133. Goodall GJ, Filipowicz W: **Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants**. *The EMBO journal* 1991, **10**(9):2635.
134. White O, Soderlund C, Shanmugan P, Fields C: **Information contents and dinucleotide compositions of plant intron sequences vary with evolutionary origin**. *Plant molecular biology* 1992, **19**(6):1057-1064.
135. de Lomana ALG, Schäuble S, Valenzuela J, Imam S, Carter W, Bilgin DD, Yohn CB, Turkarslan S, Reiss DJ, Orellana MV: **Transcriptional program for nitrogen starvation-induced lipid accumulation in *Chlamydomonas reinhardtii***. *Biotechnology for biofuels* 2015, **8**(1):1.
136. Siaut M, Cuiné S, Cagnon C, Fessler B, Nguyen M, Carrier P, Beyly A, Beisson F, Triantaphylidès C, Li-Beisson Y: **Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves**. *BMC biotechnology* 2011, **11**(1):7.
137. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A: **GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists**. *Genome biology* 2007, **8**(1):R3.
138. Huang T, Shi X-H, Wang P, He Z, Feng K-Y, Hu L, Kong X, Li Y-X, Cai Y-D, Chou K-C: **Analysis and prediction of the metabolic stability of proteins based on their sequential features, subcellular locations and interaction networks**. *PloS one* 2010, **5**(6):e10972.
139. Huang T, Wan S, Xu Z, Zheng Y, Feng K-Y, Li H-P, Kong X, Cai Y-D: **Analysis and prediction of translation rate based on sequence and functional features of the mRNA**. *PLos one* 2011, **6**(1):e16036.

140. Huang T, Wang P, Ye Z-Q, Xu H, He Z, Feng K-Y, Hu L, Cui W, Wang K, Dong X: **Prediction of deleterious non-synonymous SNPs based on protein interaction network and hybrid properties**. *PLoS One* 2010, **5**(7):e11900.
141. Chen L, Chu C, Lu J, Kong X, Huang T, Cai Y-D: **Gene ontology and KEGG pathway enrichment analysis of a drug target-based classification system**. *PloS one* 2015, **10**(5):e0126492.
142. Huang T, Zhang J, Xu Z-P, Hu L-L, Chen L, Shao J-L, Zhang L, Kong X-Y, Cai Y-D, Chou K-C: **Deciphering the effects of gene deletion on yeast longevity using network and machine learning approaches**. *Biochimie* 2012, **94**(4):1017-1025.
143. Yang Z-K, Niu Y-F, Ma Y-H, Xue J, Zhang M-H, Yang W-D, Liu J-S, Lu S-H, Guan Y, Li H-Y: **Molecular and cellular mechanisms of neutral lipid accumulation in diatom following nitrogen deprivation**. *Biotechnology for biofuels* 2013, **6**(1):1.
144. Reginald H. Garrett CMG: **Biochemistry, 5th Edition**; 2013.
145. Carpinelli EC, Telatin A, Vitulo N, Forcato C, D'Angelo M, Schiavon R, Vezzi A, Giacometti GM, Morosinotto T, Valle G: **Chromosome scale genome assembly and transcriptome profiling of *Nannochloropsis gaditana* in nitrogen depletion**. *Molecular plant* 2014, **7**(2):323-335.
146. Simionato D, Block MA, La Rocca N, Jouhet J, Maréchal E, Finazzi G, Morosinotto T: **The response of *Nannochloropsis gaditana* to nitrogen starvation includes *de novo* biosynthesis of triacylglycerols, a decrease of chloroplast galactolipids, and reorganization of the photosynthetic apparatus**. *Eukaryotic cell* 2013, **12**(5):665-676.
147. Urzica EI, Vieler A, Hong-Hermesdorf A, Page MD, Casero D, Gallaher SD, Kropat J, Pellegrini M, Benning C, Merchant SS: **Remodeling of membrane lipids in iron-starved *Chlamydomonas***. *Journal of Biological Chemistry* 2013, **288**(42):30246-30258.
148. 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***. *The Plant Cell* 2012, **24**(9):3708-3724.
149. Tovar‐Méndez A, Miernyk JA, Randall DD: **Regulation of pyruvate dehydrogenase complex activity in plant cells**. *European Journal of Biochemistry* 2003, **270**(6):1043-1049.
150. Guo K, Li L: **Differential 12C-/13C-isotope dansylation labeling and fast liquid chromatography/mass spectrometry for absolute and relative quantification of the metabolome**. *Analytical chemistry* 2009, **81**(10):3919-3932.
151. Guo K, Li L: **High-performance isotope labeling for profiling carboxylic acid-containing metabolites in biofluids by mass spectrometry**. *Analytical chemistry* 2010, **82**(21):8789-8793.
152. Peng J, Li L: **Liquid–liquid extraction combined with differential isotope dimethylaminophenacyl labeling for improved metabolomic profiling of organic acids**. *Analytica chimica acta* 2013, **803**:97-105.
153. Friis RMN, Glaves JP, Huan T, Li L, Sykes BD, Schultz MC: **Rewiring AMPK and mitochondrial retrograde signaling for metabolic control of aging and histone acetylation in respiratory-defective cells**. *Cell reports* 2014, **7**(2):565-574.
154. Kato M, Li J, Chuang JL, Chuang DT: **Distinct structural mechanisms for inhibition of pyruvate dehydrogenase kinase isoforms by AZD7545, dichloroacetate, and radicicol**. *Structure* 2007, **15**(8):992-1004.
155. Miller JW, Uden PC: **Characterization of nonvolatile aqueous chlorination products of humic substances**. *Environmental science & technology* 1983, **17**(3):150-157.
156. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M: **KEGG as a reference resource for gene and protein annotation**. *Nucleic acids research* 2016, **44**(D1):D457-D462.
157. Ge F, Huang W, Chen Z, Zhang C, Xiong Q, Bowler C, Yang J, Xu J, Hu H: **Methylcrotonyl-CoA carboxylase regulates triacylglycerol accumulation in the model diatom *Phaeodactylum tricornutum***. *The Plant Cell* 2014, **26**(4):1681-

1697.
158. Lane N: **Life Ascending: The Ten Great Inventions of Evolution.** . *New York: WW Norton & Co* 2009.
159. Wagner A: **Arrival of the Fittest (first ed.)**. *New York: Penguin Group* 2014:p. 100.
160. Booth G: **Naphthalene Derivatives**: Wiley-VCH Verlag GmbH & Co. KGaA; 2000.
161. Green CR, Wallace M, Divakaruni AS, Phillips SA, Murphy AN, Ciaraldi TP, Metallo CM: **Branched-chain amino acid catabolism fuels adipocyte differentiation and lipogenesis**. *Nature chemical biology* 2016, **12**(1):15-21.
162. Peng C, Uygun S, Shiu S-H, Last RL: **The impact of the branched-chain ketoacid dehydrogenase complex on amino acid homeostasis in Arabidopsis**. *Plant physiology* 2015, **169**(3):1807-1820.
163. Araújo WL, Ishizaki K, Nunes-Nesi A, Larson TR, Tohge T, Krahnert I, Witt S, Obata T, Schauer N, Graham IA: **Identification of the 2-hydroxyglutarate and isovaleryl-CoA dehydrogenases as alternative electron donors linking lysine catabolism to the electron transport chain of Arabidopsis mitochondria**. *The Plant Cell* 2010, **22**(5):1549-1563.
164. Ishizaki K, Larson TR, Schauer N, Fernie AR, Graham IA, Leaver CJ: **The critical role of Arabidopsis electron-transfer flavoprotein: ubiquinone oxidoreductase during dark-induced starvation**. *The Plant Cell* 2005, **17**(9):2587-2600.
165. Angelovici R, Fait A, Fernie AR, Galili G: **A seed high‐lysine trait is negatively associated with the TCA cycle and slows down Arabidopsis seed germination**. *New Phytologist* 2011, **189**(1):148-159.
166. Vorapreeda T, Thammarongtham C, Cheevadhanarak S, Laoteng K: **Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: involvement in the lipid production of oleaginous yeast and fungi**. *Microbiology* 2012, **158**(1):217-228.
167. Flynn K, Butler I: **Nitrogen-sources for the growth of marine microalgae-role of dissolved free amino-acids**. *Marine Ecology Progress Series* 1986, **34**(3):281-304.
168. Huo Y-X, Cho KM, Rivera JGL, Monte E, Shen CR, Yan Y, Liao JC: **Conversion of proteins into biofuels by engineering nitrogen flux**. *Nature biotechnology* 2011, **29**(4):346-351.
169. Ma Y-H, Wang X, Niu Y-F, Yang Z-K, Zhang M-H, Wang Z-M, Yang W-D, Liu J-S, Li H-Y: **Antisense knockdown of pyruvate dehydrogenase kinase promotes the neutral lipid accumulation in the diatom *Phaeodactylum tricornutum***. *Microbial cell factories* 2014, **13**(1):1.
170. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Molecular biology and evolution* 1987, **4**(4):406-425.
171. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap**. *Evolution* 1985:783-791.
172. Zuckerkandl E, Pauling L: **Evolutionary divergence and convergence in proteins**. *Evolving genes and proteins* 1965, **97**:97-166.
173. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Molecular biology and evolution* 2011, **28**(10):2731-2739.

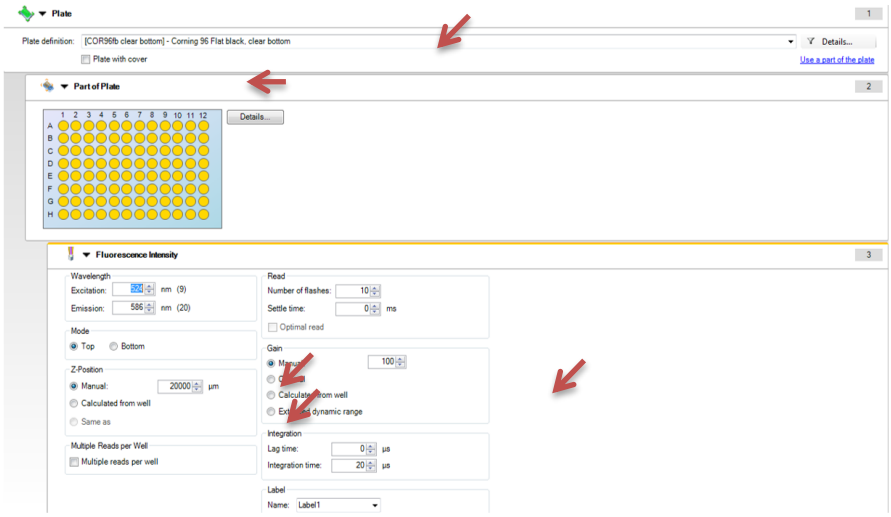
Appendix

Appendix 1 Primer sequence to do RACE PCR and real-time PCR.

Primer name	Primer Sequence	
RACE or sequence confirmation primers		
r_DtCuAO_R1	TCACATTGTTGCCGCAGACAT	RACE PCR for 5' end
r_DtCuAO_R2	ACACGAAAATCTGCATGACCT	RACE PCR for 5' end (nested)
r_DtCuAO_F1	AATGCGTTCTATGCAAAGGAGCA	RACE PCR for 3' end
r_DtCuAO_F2	TGTCATGCCTGTGGAGACTGT	RACE PCR for 3' end (nested)
Dtmccb_F1	ATGGGTGCGGTGCTTTGGGCT	
Dtmccb_R2	ACAGGGAACACTGTTAGCAAAG	
rDtACAD_R1	TCATGCCTAGTTTGTTGAGCT	RACE PCR for 5' end
rDtACAD_R2	ATTGGTGCACCACATTTTGG	RACE PCR for 5' end (nested)
rDtACAD_F1	CCAAAATGTGGTGCACCAAT	RACE PCR for 3' end
rDtACAD_F2	AGCTCAACAAACTAGGCATGA	RACE PCR for 3' end (nested)
qPCR primers		
qDtCuAO_F1	CATTGGCATACGGATTCTTG	
qDtCuAO_R1	GTCCGTATTGATCCGGCTAT	
qDtIVD_F1	TCTGTGCTGTGCTGAAACC	
qDtIVD_R1	GGCAAGACAGAGCCTGAAA	
qPDK_F	ATGAGCCTCCCAAAGATGAC	
qPDK_R	CGCGTAAGCATATCGACCTA	
DtTUB_F	CAGATGTGGGATGCCAAGAACAT	
DtTUB_R	GTTCAGCATCTGCTC ATCCACCT	
Genotyping PCR primers		
ble_F:	AAGCTGACCAGCGCCGTTC	
ble_R:	CCACGAAGTGCACGCAGTT	

Appendix 2 Protocol of quantitative measurement of neutral lipids in microalgae using Nile red.

(A) Tecan Infinite M200 PRO Instrument Parameters



(B) Materials

1. Prepare Nile red Stock:

Nile red Stock (100 µg/mL)	
Nile red powder	Nile red powder
Acetone	Acetone

* Nile red Stock was stored in amber bottle at 4ºC.

Nile red Working Solution (25 µg/mL)	
Nile red Stock (100 µg/mL)	Nile red Stock (100 µg/mL)
Acetone	Acetone

2. Prepare Triolein Stock:

Triolein T7140 Sigma (Glyceryl Trioleate) 99%

Mr = 885.43 g/mol

Density = 0.91 g/mL

Triolein Stock (5 mg/mL)	
Triolein	11 µl
Isopropanol	1989 µl

* store Triolein Stock in amber bottle at -20ºC.

3. Microalgal samples

4. Culture medium

(C) Methods

To construct the Triolein Standard Curve:

1. Prepare the 2 mL Mixture accordingly and vortex to homogenize.

In every 1980 µl culture medium, add the following components:

Triolein Concentration (µg/mL)	**0**	**2.5**	**5**	**7.5**	**10**	**12.5**	**15**	**20**	**25**	**30**	**35**	**40**
Isopropanol (µl)	20	19	18	17	16	15	14	12	10	8	6	4
Triolein Stock (µl)	0	1	2	3	4	5	6	8	10	12	14	16

2. From the Mixture, take out 200 µl and transfer to the 96-well black, clear bottom plate.

3. Read the plate at the aforementioned parameters.

4. To every sample well, add 2 µl Nile red Working Solution and pipette to mix.

5. Incubate the plate at room temperature, 5 minutes, in the dark.

6. Read the plate at the aforementioned parameters.

To quantify using the **Direct Staining** Method (without triolein spike):

*NOTE: remember to plot Triolein Standard Curve together with cell samples.

1. Measure the samples OD_{750}.

2. Spin down cells at 3,000 × g for 5 minutes.

3. Re-suspend cells in culture medium (to remove impurities) to density of OD_{750} = 0.1 to 0.5.

4. Transfer 200 µl cell samples, culture medium blank, and Triolein Standards (see before "To construct the Triolein Standard Curve") together to the same 96-well black, clear bottom plate.

5. Read the plate at the aforementioned parameters.

6. To every sample well, add 2 µl Nile red Working Solution and pipette to mix.

7. Incubate the plate at room temperature, 5 minutes, in the dark.

8. Read the plate at the aforementioned parameters.

To quantify using the **Standard Addition** Method (with triolein spike):

*NOTE: it is not necessary to plot Triolein Standard Curve.

1. Measure the samples OD_{750}.

2. Spin down cells at 3,000 × g for 5 minutes.

3. Re-suspend cells in culture medium (to remove impurities) to density of OD_{750} = 0.3.

4. Prepare the 2 mL Mixture accordingly and vortex to homogenize.

Aliquot 1980 µl of the prepared samples into 12 falcon tubes, add the following components:

Triolein Concentration (µg/mL)	**0**	**2.5**	**5**	**7.5**	**10**	**12.5**	**15**	**20**	**25**	**30**	**35**	**40**
Isopropanol (µl)	20	19	18	17	16	15	14	12	10	8	6	4

Triolein Stock (µl)	0	1	2	3	4	5	6	8	10	12	14	16

5. From the Mixture, take out 200 µl and transfer to the 96-well black, clear bottom plate.

6. Read the plate at the aforementioned parameters.

7. To every sample well, add 2 µl Nile red Working Solution and pipette to mix.

8. Incubate the plate at room temperature, 5 minutes, in the dark.

9. Read the plate at the aforementioned parameters.

*NOTE:

1. The experiment is better to be performed in dark condition, especially during the Nile red staining step.

2. Fluorescence intensity of the samples should be always measured before and after Nile red addition in order to subtract the intrinsic fluorescence value of the samples.

3. Above is the method for *Dunaliella tertiolecta* neutral lipid measurement. If applying to others, especially to those with rigid cell walls, additional treatments or different concentration of Nile red may need to be considered during Nile red staining.

4. Technical replicates are highly advisable.

(D) Calculations

*NOTE: Fluorescence intensity of the microalgal samples before Nile red addition is subtracted from the final fluorescence readings for subsequent calculations.

Direct Staining Method (without triolein spike):

1. All fluorescence readings of every microalgal sample are normalized to the culture medium blank.

2. Substitute the normalized fluorescence values of microalgal samples into the equation of the standard curve to get the neutral lipid contents.

Standard Addition Method (with triolein spike):

1. All fluorescence readings of every triolein spiked microalgal sample are normalized to the microalgal sample without triolein addition.

2. Plot the graph and get the best-fitted trendline. The x axis intercept of the trendline will directly indicate the neutral lipid content of the microalga.

Appendix 3 Run summary of second round of D9 duplicate samples on the Illumina MISEQ platform.

Sample name	Sequencing stats and pre-alignment QA/QC of raw data /pre-alignment QA/QC of post-trimming data					Post-alignment QA/QC after alignment	
	Total reads	Avg. read length	Avg. read quality	% N	% GC	Total reads	% Aligned
D9-1_S1_L001_R1	3,239,491/ **3,239,245**	133.7/ **133.6**	37.68/ **37.69**	0/ **0**	47.43/ **47.42**	3,239,245	68.94
D9-2_S2_L001_R1	3,515,548/ **35,15,368**	133.58/ **133.5**	37.74/ **37.75**	0/ **0**	46.49/ **46.48**	35,15,368	70.48
WT-1_S3_L001_R1	2,223,016/ **2,222,563**	139.2/ **139.0**	37.41/ **37.44**	0/ **0**	46.13/ **46.12**	2,222,563	52.5
WT-2_S4_L001_R1	2,898,342/ **2,898,030**	134.79/ **134.7**	37.59/ **37.61**	0/ **0**	47.41/ **47.4**	2,898,030	62.8

Appendix 4 D9 and WT *Dunaliella tertiolecta* KEGG pathway analysis results of RNA-Seq second run.

The top hit pathways in pathway analysis with p value <= 0.05.

Pathway Name	Enrichment Score	Enrichment p-value	Pathway ID
Metabolic pathways	21.8674	3.18E-10	kegg_pathway_21
Photosynthesis - antenna proteins	18.5174	9.08E-09	kegg_pathway_57
Photosynthesis	16.8733	4.70E-08	kegg_pathway_92

Biosynthesis of secondary metabolites	15.6338	1.62E-07	kegg_pathway_89
Porphyrin and chlorophyll metabolism	8.56963	0.000189783	kegg_pathway_69
Carbon metabolism	7.30144	0.000674564	kegg_pathway_31
Carbon fixation in photosynthetic organisms	5.84625	0.00289073	kegg_pathway_56
Alanine, aspartate and glutamate metabolism	5.60665	0.00367337	kegg_pathway_82
Energy Metabolism	4.76187	0.00854963	kegg_pathway_52
Biosynthesis of amino acids	4.63744	0.00968246	kegg_pathway_51
Pentose phosphate pathway	4.52073	0.0108811	kegg_pathway_33
Nitrogen metabolism	3.71457	0.0243658	kegg_pathway_18
FA biosynthesis	3.65338	0.0259033	kegg_pathway_67

Appendix 5 Detailed steps of transcriptome construction and analyses.

(a) Trinity (v2.2.0) for assembly of transcriptome.

Usage:

```
$TRINITY_HOME/Trinity --seqType fq \
--CPU 96 --max_memory 2048G \
--SS_lib_type RF --no_bowtie --no_cleanup \
--left  $INPUT/left_data1.fastq.gz $INPUT/left_data2.fastq.gz
$INPUT/left_data3.fastq.gz
--right  $INPUT/right_data1.fastq.gz $INPUT/right_data2.fastq.gz
$INPUT/right_data3.fastq.gz
--output $OUTPUT/Trinity.fasta
```

Note: we used 96 threads as each node of the data center is equipped with 48 CPU cores, and each core has 2 hardware threads. 2048 GB memory RAM was allocated to accommodate to the assembler's intensive memory use. seqType fq for fastq format data, SS_lib_type RF for paired-end of sequencing library, left_data.fastq.gz and right_data.fastq.gz for a pair of left and right sequencing files in fastq format. The output file was generated in Trinity.fasta under OUTPUT directory.

(b) mpiBLASTX (v1.6.0) for comparison of homologous sequences.

Usage:

```
### Set the number of cores (cpus) and memory that will be used for this job
#PBS -l select=40:ncpus=24:mem=60gb
mpirun -np 960 mpiblast -p BLASTX -d combined.faa -i
$INPUT/Dt_merged.fa
--use-virtual-frags --removedb \
-o $OUTPUT/blast_output.txt
```

Note: where -np 960 stands for using 960 (40 nodes) cores, -p BLASTX stands for using BLASTX software under blast suites, -d combined.faa stands for a combined protein reference file from all plants and bacterial after mpiformatdb, -i $INPUT/Dt_merged.fa stands for the inquiry file of the merged *D. tertiolecta* contigs from outputs of Trinity, --use-virtual-frags enable workers to cache database fragments in memory instead of local storage, --removedb removes the local copy of the database from each node before terminating execution. The BLASTX output was generated into blast_output.txt under OUTPUT directory.

(c) Dt_v11.fa and Dt_v11.2.fa transcript files and their annotation information were generated via optimized Bag2D to filter the redundant

contigs and get the top-hits within the criteria.

(d) RSEM (v1.2.29) for mapping the sequencing reads.

Usage:

Step 1: prepare reference

$RSEM_HOME/rsem-prepare-reference --bowtie2 $INPUT/Dt_v11.2.fasta $INPUT/prepare_reference

Step 2: calculate expression

$RSEM_HOME/rsem-calculate-expression --paired-end \

$INPUT/left_data_x1.fastq.gz $INPUT/right_data_x1.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_x1.results -p 48

$INPUT/left_data_x2.fastq.gz $INPUT/right_data_x2.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_x2.results -p 48

$INPUT/left_data_x3.fastq.gz $INPUT/right_data_x3.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_x3.results -p 48

$INPUT/left_data_wt1.fastq.gz $INPUT/right_data_wt1.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_wt1.results -p 48

$INPUT/left_data_wt2.fastq.gz $INPUT/right_data_wt2.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_wt2.results -p 48

$INPUT/left_data_wt3.fastq.gz $INPUT/right_data_wt3.fastq.gz \

--bowtie2 $INPUT/prepare_reference $OUTPUT/expression_genes_wt3.results -p 48

Step 3: generate_data_matrix

$RSEM_HOME/rsem-generate-data-matrix \

$OUTPUT/expression_genes_x1.results.genes.results \

$OUTPUT/expression_genes_x2.results.genes.results \

$OUTPUT/expression_genes_x3.results.genes.results \

$OUTPUT/expression_genes_wt1.results.genes.results \

$OUTPUT/expression_genes_wt2.results.genes.results \

$OUTPUT/expression_genes_wt3.results.genes.results \

>$OUTPUT/GeneMat_x.txt

Note: where reads are aligned to the non-redundant transcript sequences using bowtie2 and read count per transcript is estimated by RSEM, using paired-end of sequencing data of treated samples - x1.fastq, x2.fastq, x3.fastq (biological triplicates), and control samples - wt1.fastq, wt2.fastq, wt3.fastq (biological triplicates). -p 48 stands for using 48 threads. A data matrix file of GeneMat_x.txt was generated under OUTPUT directory.

(e) EbSeq (in RSEM package) for gene differential expression analysis.

Step 4: run_EbSeq

$RSEM_HOME/rsem-run-ebseq $OUTPUT/GeneMat_x.txt 3,3 $OUTPUT/GeneMat_x.results

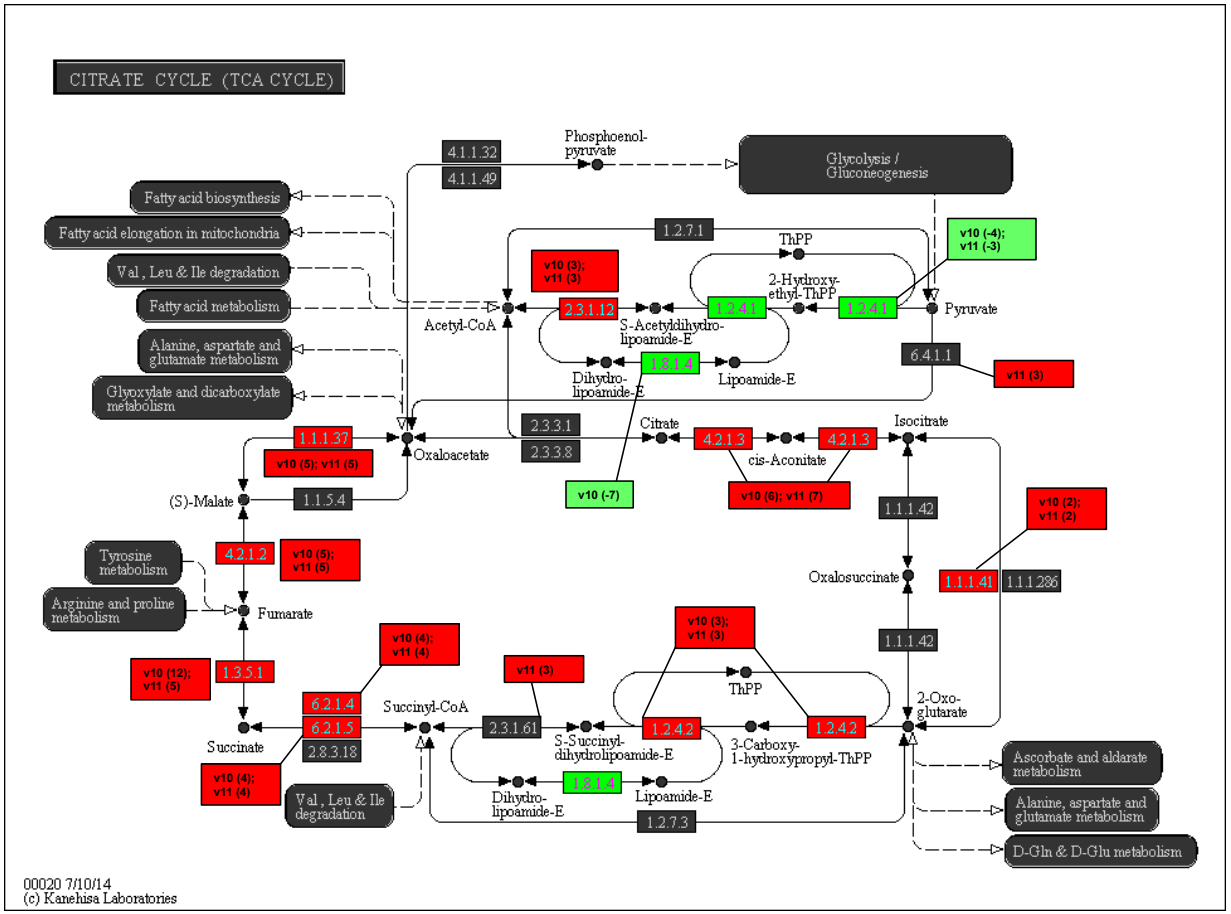
Step 5: FDR0.05

$RSEM_HOME/rsem-control-fdr $OUTPUT/GeneMat_x.results 0.05 $OUTPUT/GeneMat_x.de.txt

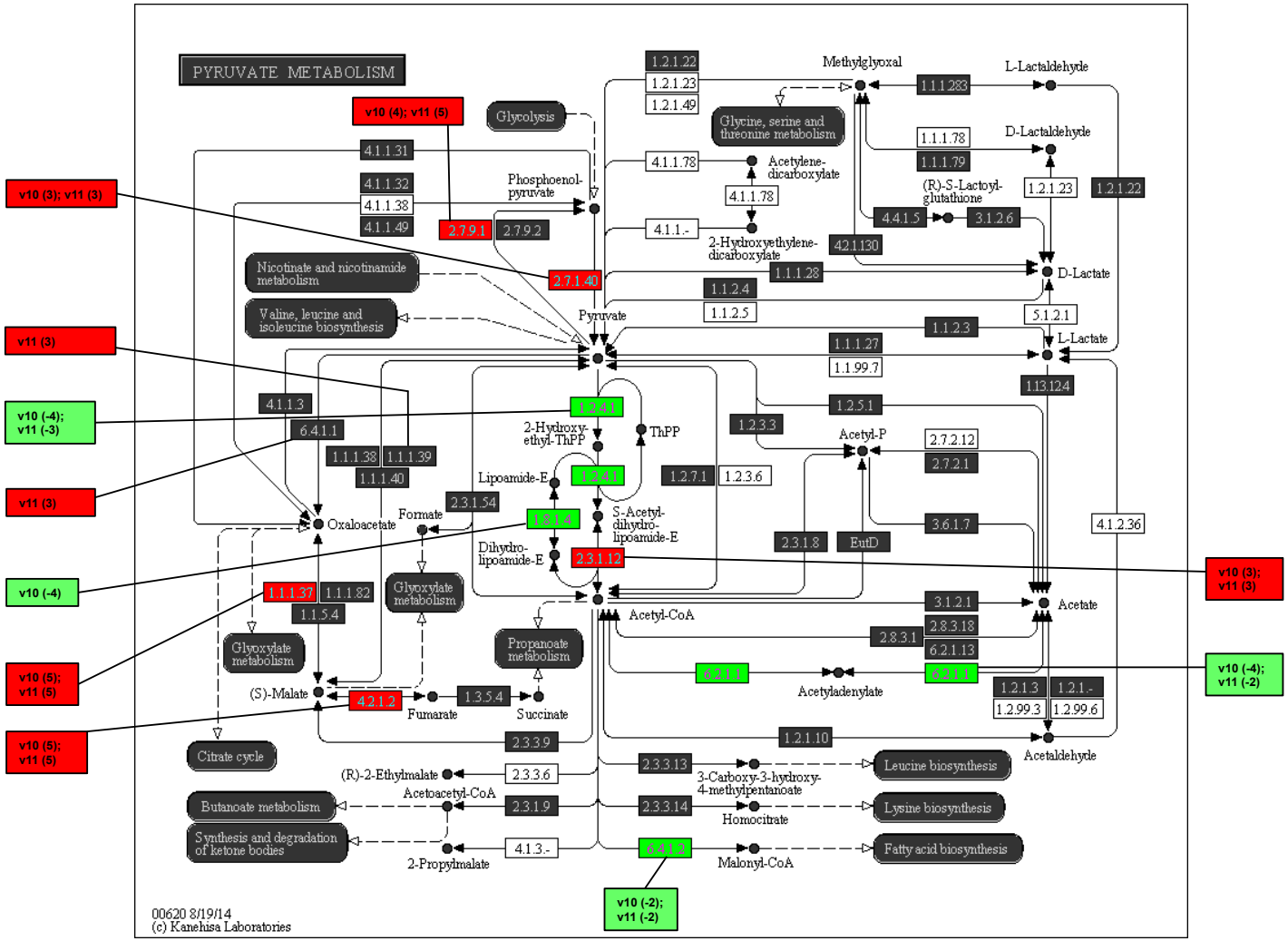
Note: based on the read count, differential expression changes with false discovery rate (FDR) < 0.05 was generated in GeneMat_x.de.txt under OUTPUT directory.

Appendix 6 Integration of significant genes hit in KEGG biological pathway analysis from Dt_v10 and Dt_v11.

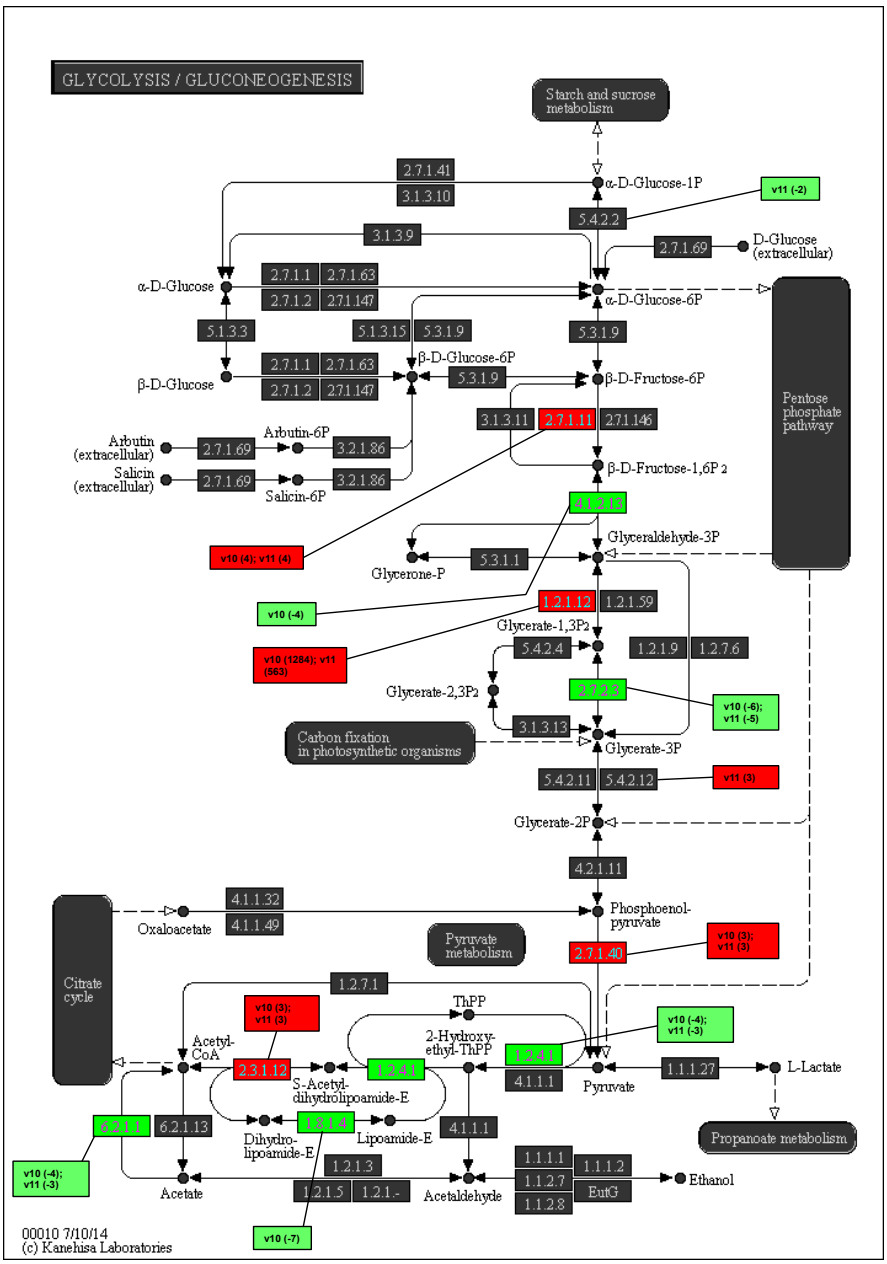
a

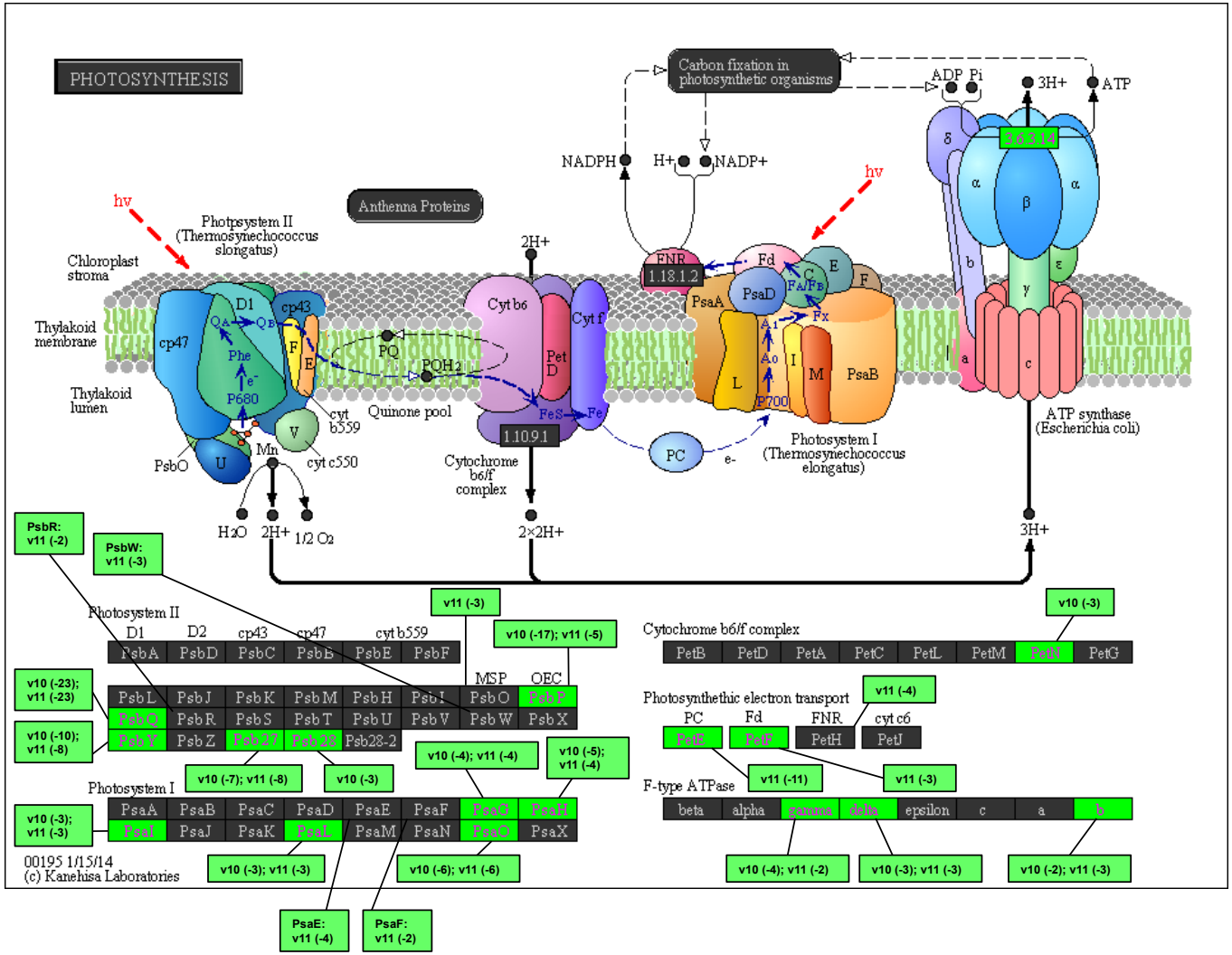


b


PYRUVATE METABOLISM
v10 (4); v11 (5)
Glycolysis
v10 (3); v11 (3)
v11 (3)
v10 (-4); v11 (-3)
v11 (3)
v10 (-4)
v10 (5); v11 (5)
v10 (5); v11 (5)
Nicotinate and nicotinamide metabolism
Valine, leucine and isoleucine biosynthesis
Phosphoenol-pyruvate
Pyruvate
Oxaloacetate
Formate
Lipoamide-E
Dihydro-lipoamide-E
S-Acetyl-dihydro-lipoamide-E
2-Hydroxy-ethyl-ThPP
ThPP
Acetyl-CoA
Acetyl-P
Acetate
Acetyladenylate
Glyoxylate metabolism
Glyoxylate metabolism
Propanoate metabolism
(S)-Malate
Fumarate
Succinate
Citrate cycle
(R)-2-Ethylmalate
Acetoacetyl-CoA
Butanoate metabolism
Synthesis and degradation of ketone bodies
2-Propylmalate
Methylglyoxal
L-Lactaldehyde
D-Lactaldehyde
Glycine, serine and threonine metabolism
Acetylene-dicarboxylate
2-Hydroxyethylene-dicarboxylate
(R)-S-Lactoyl-glutathione
D-Lactate
L-Lactate
Acetaldehyde
3-Carboxy-3-hydroxy-4-methylpentanoate
Homocitrate
Malonyl-CoA
Leucine biosynthesis
Lysine biosynthesis
Fatty acid biosynthesis
v10 (-2); v11 (-2)
v10 (3); v11 (3)
v10 (-4); v11 (-2)
00620 8/19/14
(c) Kanehisa Laboratories

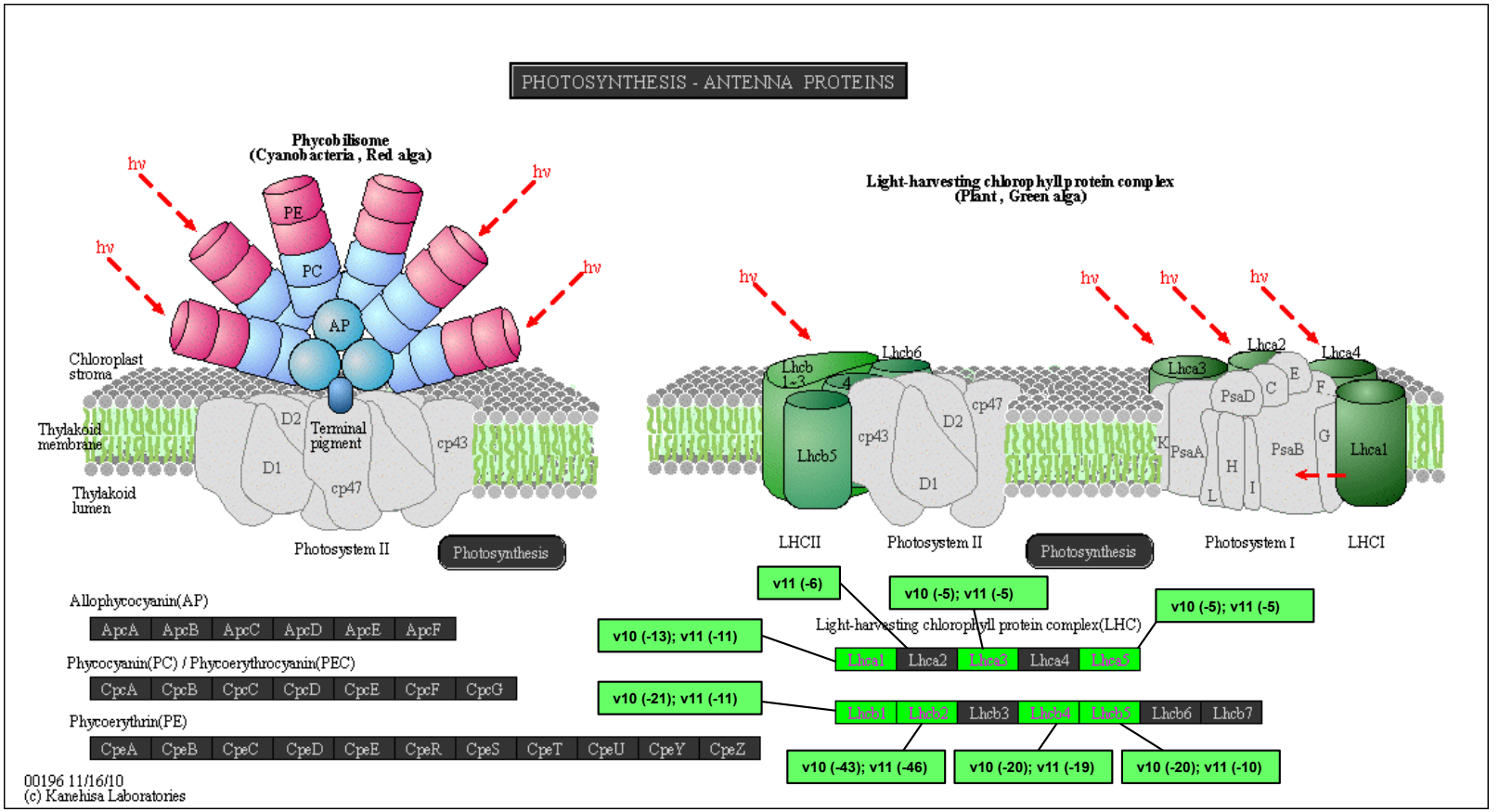
c


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
D-Glucose (extracellular)
α-D-Glucose
α-D-Glucose-6P
β-D-Glucose-6P
β-D-Glucose
β-D-Fructose-6P
Pentose phosphate pathway
Arbutin (extracellular)
Arbutin-6P
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P 2
Glyceraldehyde-3P
Glycerone-P
Glycerate-1,3P2
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
Glycerate-3P
Glycerate-2P
Phosphoenol-pyruvate
Oxaloacetate
Pyruvate metabolism
Citrate cycle
ThPP
2-Hydroxy-ethyl-ThPP
Acetyl-CoA
S-Acetyl-dihydrolipoamide-E
Dihydro-lipoamide-E
Lipoamide-E
Pyruvate
L-Lactate
Propanoate metabolism
Acetate
Acetaldehyde
Ethanol
00010 7/10/14
(c) Kanehisa Laboratories


PHOTOSYNTHESIS
Carbon fixation in photosynthetic organisms
ADP Pi
3H+
ATP
NADPH
H+
NADP+
hv
Anthenna Proteins
Photpsystem II (Thermosynechococcus elongatus)
Chloroplast stroma
Thylakoid membrane
Thylakoid lumen
2H+
FNR
1.18.1.2
Fd
PsaA
PsaD
PsaB
Cyt b6
Cyt f
Pet D
PQ
PQH2
Quinone pool
cyt b559
cyt c550
Mn
PsbO
U
V
FeS
Fe
1.10.9.1
Cytochrome b6/f complex
PC
e-
Photosystem I (Thermosynechococcus elongatus)
ATP synthase (Escherichia coli)
H2O
2H+
1/2 O2
2×2H+
3H+
PsbR: v11 (-2)
PsbW: v11 (-3)
v11 (-3)
v10 (-17); v11 (-5)
v10 (-3)
Photosystem II
D1 D2 cp43 cp47 cyt b559
PsbA PsbD PsbC PsbB PsbE PsbF
MSP OEC
PsbL PsbJ PsbK PsbM PsbH PsbI PsbO PsbP
PsbQ PsbR PsbS PsbT PsbU PsbV PsbW PsbX
PsbY PsbZ Psb27 Psb28 Psb28-2
v10 (-23); v11 (-23)
v10 (-10); v11 (-8)
v10 (-7); v11 (-8)
v10 (-3)
v10 (-4); v11 (-4)
v10 (-5); v11 (-4)
Photosystem I
PsaA PsaB PsaC PsaD PsaE PsaF PsaG PsaH
PsaI PsaJ PsaK PsaL PsaM PsaN PsaO PsaX
v10 (-3); v11 (-3)
v10 (-3); v11 (-3)
v10 (-6); v11 (-6)
PsaE: v11 (-4)
PsaF: v11 (-2)
Cytochrome b6/f complex
PetB PetD PetA PetC PetL PetM PetG
Photosynthethic electron transport
PC Fd FNR cyt c6
PetE PetF PetH PetJ
v11 (-4)
v11 (-11)
v11 (-3)
F-type ATPase
beta alpha gamma delta epsilon c a b
v10 (-4); v11 (-2)
v10 (-3); v11 (-3)
v10 (-2); v11 (-3)
00195 1/15/14
(c) Kanehisa Laboratories

d

e


PHOTOSYNTHESIS - ANTENNA PROTEINS
Phycobilisome
(Cyanobacteria , Red alga)
hv
PE
PC
AP
Chloroplast stroma
Thylakoid membrane
Thylakoid lumen
D2
Terminal pigment
D1
cp47
cp43
Photosystem II
Photosynthesis
Light-harvesting chlorophyll protein complex
(Plant , Green alga)
Lhcb6
Lhcb5
Lhca3
Lhca2
Lhca4
Lhca1
PsaD
PsaA
PsaB
LHCII
Photosystem II
Photosynthesis
Photosystem I
LHCI
Allophycocyanin(AP)
ApcA ApcB ApcC ApcD ApcE ApcF
Phycocyanin(PC) / Phycoerythrocyanin(PEC)
CpcA CpcB CpcC CpcD CpcE CpcF CpcG
Phycoerythrin(PE)
CpeA CpeB CpeC CpeD CpeE CpeR CpeS CpeT CpeU CpeY CpeZ
v11 (-6)
v10 (-5); v11 (-5)
v10 (-13); v11 (-11)
Light-harvesting chlorophyll protein complex(LHC)
Lhca1 Lhca2 Lhca3 Lhca4 Lhca5
v10 (-21); v11 (-11)
Lhcb1 Lhcb2 Lhcb3 Lhcb4 Lhcb5 Lhcb6 Lhcb7
v10 (-43); v11 (-46)
v10 (-20); v11 (-19)
v10 (-20); v11 (-10)
00196 11/16/10
(c) Kanehisa Laboratories

Appendix 7

(a) List of the top 10 most upregulated and downregulated genes in nitrogen-deprived *D. tertiolecta* cells from Dt_v11 database RSEM analysis.

Dt_name	ncbi_proteinID	Hypothetical function	KEGG gene name	KO	Fold-change
Locus_6333_4Transcript_1/1_Confidence_1.000_Length_1596	XP_001703199.1	glyceraldehyde 3-phosphate dehydrogenase, dominant splicing variant[Chlamydomonas reinhardtii]	cre:CHLREDRAFT_140618	K00134	563.404
TRINITY1_DN8534_c0_g2_i1	XP_002951533.1	hypothetical protein VOLCADRAFT_92124 [Volvox carteri f. nagariensis]	vcn:VOLCADRAFT_92124		101.033
TRINITY2_DN24779_c1_g4_i1	XP_013904340.1	putative MFS transporter, AGZA family, xanthine/uracil permease[Monoraphidium neglectum]			79.8633
TRINITY1_DN42814_c0_g1_i2	XP_002957274.1	hypothetical protein VOLCADRAFT_107613 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_107613		77.4615
Locus_3038_1Transcript_1/7_Confidence_0.267_Length_1420	XP_005645091.1	urea active transporter-like protein [Coccomyxa subellipsoideaC-169]	csl:COCSUDRAFT_30678		59.3786

Locus_5_2Transcript_33/37_Confidence_0.114_Length_7316	WP_027628829.1	group II intron reverse transcriptase/maturase [[Clostridium]cellobioparum]			44.9467
Locus_220_10Transcript_1/1_Confidence_1.000_Length_5406	WP_040977597.1	protein A*, partial [Necropsobacter massiliensis]			42.3118
TRINITY1_DN44339_c2_g2_i1	XP_012575125.1	PREDICTED: uncharacterized protein LOC101508115 [Cicer arietinum]	cam:101508115		39.195
TRINITY1_DN50530_c0_g1_i1	XP_001693482.1	predicted protein [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_182699	K13412	33.0293
TRINITY1_DN39349_c1_g1_i8	XP_005643341.1	hypothetical protein COCSUDRAFT_20416 [Coccomyxa subellipsoideaC-169]	csl:COCSUDRAFT_20416		29.9753
Locus_1040_4Transcript_1/1_Confidence_1.000_Length_888	XP_005647526.1	flavodoxin IsiB [Coccomyxa subellipsoidea C-169]	csl:COCSUDRAFT_53474		-3588.486699
TRINITY2_DN41486_c0_g1_i1	XP_007141246.1	hypothetical protein PHAVU_008G179700g [Phaseolus vulgaris]	pvu:PHAVU_008G179700g		-548.7540539
TRINITY1_DN48035_c0_g1_i1	XP_001694585.1	multicopper ferroxidase [Chlamydomonas	cre:CHLREDRAFT_184156	K14735	-529.139724

		reinhardtii]		6
TRINITY2_DN20149_c28_g4_i1	XP_012857134.1	PREDICTED: uncharacterized protein LOC105976409 [Erythrantheguttata]		-436.9063535
TRINITY1_DN41800_c6_g1_i1	WP_015230410.1	extracellular nuclease [Dactylococcopsis salina]		-204.7812322
Locus_316_3Transcript_1/6_Confidence_0.250_Length_794	XP_002945683.1	hypothetical protein VOLCADRAFT_102653 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_102653	-154.3255126
Locus_2406_2Transcript_5/9_Confidence_0.522_Length_1918	XP_001702318.1	alkaline phosphatase [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_196484	-134.3874686
TRINITY1_DN41800_c6_g1_i3	WP_014201896.1	hypothetical protein [Owenweeksia hongkongensis]		-116.4235162
TRINITY1_DN40017_c0_g1_i1	XP_001694065.1	predicted protein [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_173173	-68.0318389
TRINITY1_DN48116_c0_g1_i1	XP_002948144.1	hypothetical protein VOLCADRAFT_88469 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_88469	-54.30237736

(b) List of the top 10 most upregulated and downregulated genes in nitrogen-deprived *D. tertiolecta* cells from Dt_v11 database Partek analysis.

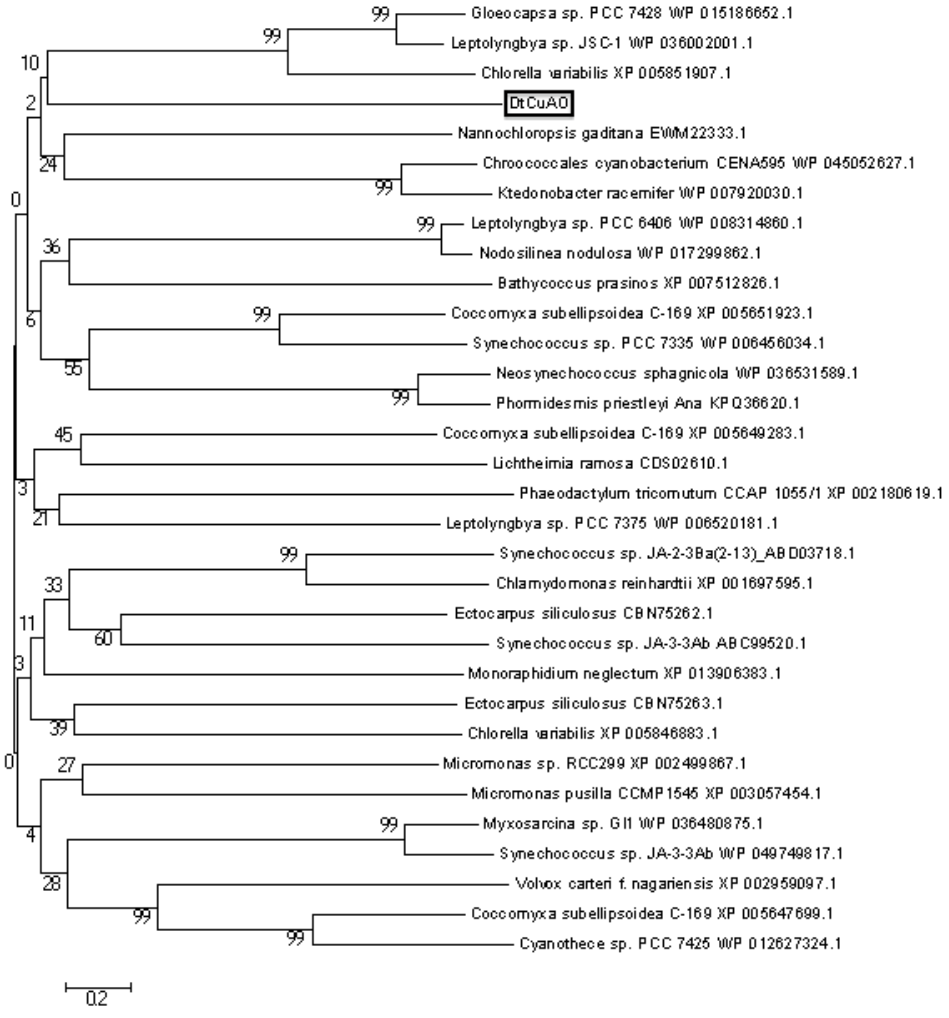
Chromosome	ncbi_proteinID	Hypothetical function	KEGG gene name	KO	Fold-change
TRINITY2_DN24779_c1_g4_i1	XP_013904340.1	putative MFS transporter, AGZA family, xanthine/uracil permease[Monoraphidium neglectum]			70.192
Locus_3038_1Transcript_1/7_Confidence_0.267_Length_1420	XP_005645091.1	urea active transporter-like protein [Coccomyxa subellipsoideaC-169]	csl:COCSUDRAFT_30678		65.8144
TRINITY1_DN42814_c0_g1_i2	XP_002957274.1	hypothetical protein VOLCADRAFT_107613 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_107613		64.9783
TRINITY1_DN39349_c1_g1_i17	WP_051512250.1	hypothetical protein [Skermanella stibiiresistens]			36.7832
TRINITY1_DN50530_c0_g1_i1	XP_001693482.1	predicted protein [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_182699	K13412	30.5754
comp31127_c5_seq1	XP_013903980.1	hypothetical protein MNEG_2999 [Monoraphidium neglectum]			25.7024
TRINITY2_DN21046_c1_g2_i3	XP_002958237.1	hypothetical protein VOLCADRAFT_108038 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_108038	K00264	22.1114
comp30738_c2_seq10	XP_013905754.1	UreA carboxylase [Monoraphidium neglectum]			15.8758
TRINITY2_DN3665_c0_g1_i1	XP_002957311.1	hypothetical protein VOLCADRAFT_98356	vcn:VOLCADRAFT_98356		15.5396

		[Volvox carteri f. nagariensis]			
Locus_894_7Transcript_11/11_Confidence_0.429_Length_1685	XP_002950414.1	hypothetical protein VOLCADRAFT_74587 [Volvox carteri f. nagariensis]	vcn:VOLCADRAFT_74587		14.7821
TRINITY1_DN48035_c0_g1_i1	XP_001694585.1	multicopper ferroxidase [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_184156	K14735	-574.15
Locus_316_3Transcript_1/6_Confidence_0.250_Length_794	XP_002945683.1	hypothetical protein VOLCADRAFT_102653 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_102653		-347.481
TRINITY1_DN41800_c6_g1_i1	WP_015230410.1	extracellular nuclease [Dactylococcopsis salina]			-198.531
Locus_2406_2Transcript_5/9_Confidence_0.522_Length_1918	XP_001702318.1	alkaline phosphatase [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_196484		-151.186
TRINITY2_DN43168_c0_g1_i1	WP_052372817.1	copper oxidase [Nocardia otitidiscaviarum]			-71.1026
TRINITY1_DN48116_c0_g1_i1	XP_002948144.1	hypothetical protein VOLCADRAFT_88469 [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_88469		-69.728
Locus_1_6Transcript_563/2655_Confidence_1.000_Length_1409	XP_001692802.1	predicted protein [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_147777		-63.4771
TRINITY2_DN24849_c20_g1_i1	XP_002952704.1	light-harvesting chlorophyll a/b-binding protein [Volvox carteri f.nagariensis]	vcn:VOLCADRAFT_109843	K08913	-50.4259
TRINITY1_DN14170_c1_g2_i1	XP_013901948.1	protein kinase A			-49.6937

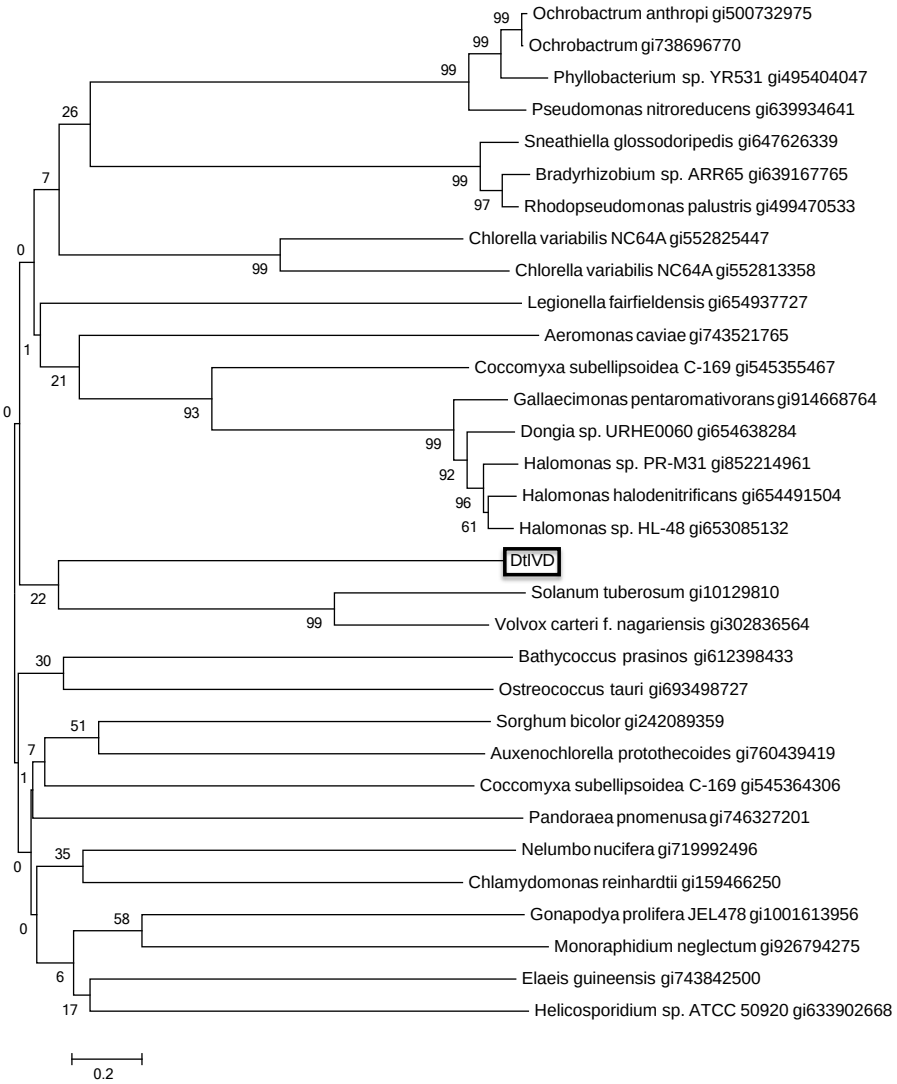
		[Monoraphidium neglectum]		
TRINITY1_DN41450_c1_g2_i1	XP_001695893.1	gamete-specific protein, partial [Chlamydomonas reinhardtii]	cre:CHLREDRAFT_205626	-45.1142

Appendix 8 Evolutionary relationships of taxa of 3 predicted genes.

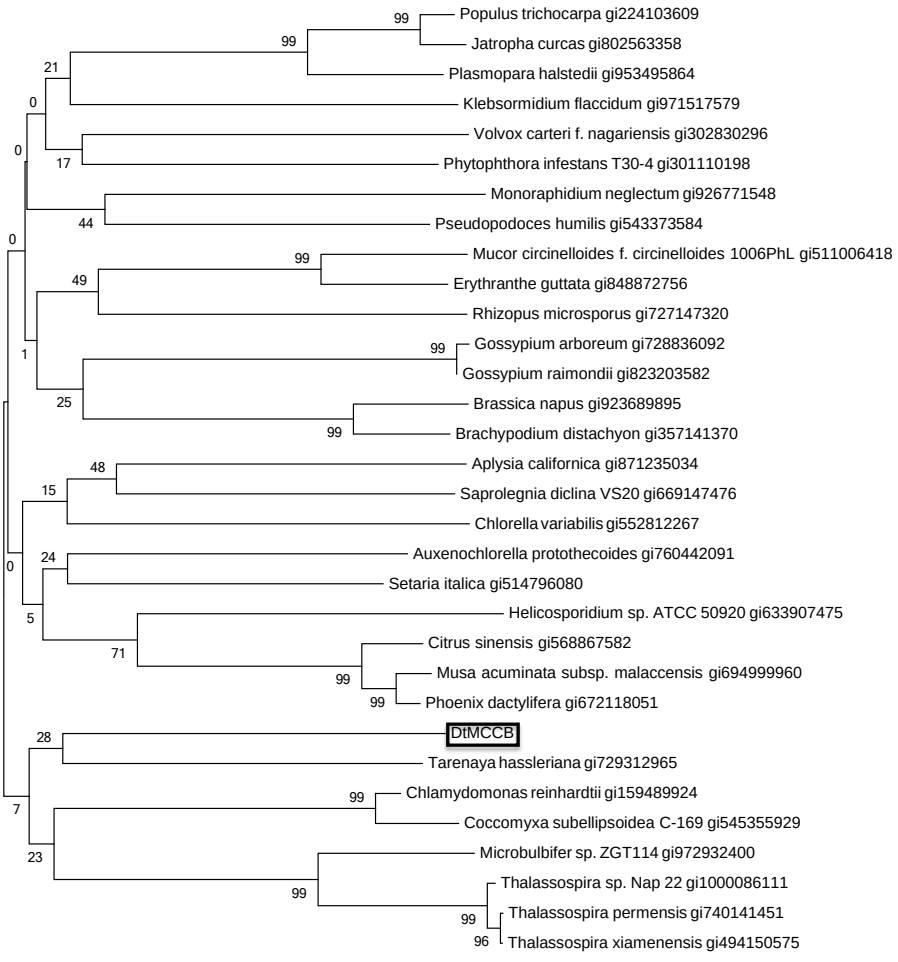
The evolutionary history was inferred using the Neighbor-Joining method [170]. (a) *DtCuAO* gene, the optimal tree with the sum of branch length = 32.94110683 is shown. There were a total of 415 positions in the final dataset. (b) *DtIVD* gene, the optimal tree with the sum of branch length = 28.60417585 is shown. There were a total of 294 positions in the final dataset. (c) *DtMCCB* gene, the optimal tree with the sum of branch length = 27.00169906 is shown. There were a total of 156 positions in the final dataset. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches [171]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [172] and are in the units of the number of amino acid substitutions per site. The analysis involved 32 amino acid sequences. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA5 [173].



a


Ochrobactrum anthropi gi500732975
Ochrobactrum gi738696770
Phyllobacterium sp. YR531 gi495404047
Pseudomonas nitroreducens gi639934641
Sneathiella glossodoripedis gi647626339
Bradyrhizobium sp. ARR65 gi639167765
Rhodopseudomonas palustris gi499470533
Chlorella variabilis NC64A gi552825447
Chlorella variabilis NC64A gi552813358
Legionella fairfieldensis gi654937727
Aeromonas caviae gi743521765
Coccomyxa subellipsoidea C-169 gi545355467
Gallaecimonas pentaromativorans gi914668764
Dongia sp. URHE0060 gi654638284
Halomonas sp. PR-M31 gi852214961
Halomonas halodenitrificans gi654491504
Halomonas sp. HL-48 gi653085132
DtlVD
Solanum tuberosum gi10129810
Volvox carteri f. nagariensis gi302836564
Bathycoccus prasinos gi612398433
Ostreococcus tauri gi693498727
Sorghum bicolor gi242089359
Auxenochlorella protothecoides gi760439419
Coccomyxa subellipsoidea C-169 gi545364306
Pandoraea pnomenusa gi746327201
Nelumbo nucifera gi719992496
Chlamydomonas reinhardtii gi159466250
Gonapodya prolifera JEL478 gi1001613956
Monoraphidium neglectum gi926794275
Elaeis guineensis gi743842500
Helicosporidium sp. ATCC 50920 gi633902668
0.2

b



c

Appendix 9 cDNA sequence of the studied genes and their corresponding sequence of translated amino acid.

(a) DtCuAO, (b) DtIVD, (c) DtMCCB, and (d) DtPDK.

a

>DtCuAO

ATGCCCGAGGGCCTCAAGGGCCTGAAGGTCATGCAGATTTTCGTGTACCTGAAGT
CTTACCCTGATGACAACCAGTATGCCCACCCTCTGGACATTATGCCGTTTGTGGAT
GTGATTTTGGGCAAGGTGGTTCGGATTGACATGCATGACACGCCGCCAGTGATTA
ACAAGGAGGACAACAACTACCACAGCGCCCTCTTTGATCCCGCAAAAATGAGGA
CAGACATCAAGCCACTGGATGTGATCCAGCCTGAGGGGCCAAGCTTCAATGTCTG
CGGCAACAATGTGAAGTGGTGCAAGTGGAGCTTCCACGTGGGATTCAATGCAAG
AGAGGGACTCGTGCTCAGTGACGTGTCGTACGATGGTCGCCCAATCTTGTATCGC
TTCAGCTTTGTGGAGATGGCGGTCCCGTATGCAGAACCTCGGGAGCCCTACGTGA
GGAAGTGCGCATTCGACATCGGTGATTACGGCTTTGGCAACACTGCAAACTCCCT
GCACCTCGGATGTGATTGCCTCGGTGCCATCCACTACTTCAATGGGGTTCTCTCAA
ACAGCAAAGGCGAGCCAGTGGATGTGCCCAAGGTTGTGTGCATGCATGAGGAGG

ACGTGGGCTTGCTGTGGAAGCATGTTGAACCCAGGACCGGCCACAATGAGAGCA
GGCGAAGCAGGCGGCTTGTTGTTAGCTTCATCTCAACAGTGGTCAACTACGAGTA
TGCTTTCTACTTCTACTTCTACCTTGACGGCACGTTTGAAGTGGAGGTGAAGCTG
ACCGGAGAACTTTCAACGAATGTGCTGTCCCCAGGGGAGGAAAATCCTGAATTT
GGCACCTTGGTCATGCCTTATGTGAATGCGCAGCACCATCAGCACATGTTCTGCGT
CCGTATTGATCCGGCTATTGATGATAAAAATGGCGGGAAAGATGTTGTCATTACCG
AGGTGAACTGTGAGCCAGTGCCTCCAGGCCCCAAGAATCCGTATGCCAATGCGTT
CTATGCAAAGGAGCAGCCGCTGCTGACAGTTAAGGAGGCTCAGAGAGACATAAG
CCCTGCCCACAACAGAGTATGGAAACTGACAAATCCGAGCTGCCTGAACCCCATT
ACCAAGAAACCTGTGTCCTACAAGCTGATGCCGCACGCCTCCCCGCCGCTCATGG
CACACTCAGAGAGCCTTGTTGCAAAGAAGGGCTACTTTGCCACCAAGAATCTATG
GGTTACGCCTTTTGCACAAGACCAGCGCTTCCCAGCTGGAGATTTTGTGTTCAAT
GAGAAGACATGTCACGGATTGAGTCAGTGGACAAAGGAGGATACAAGCCTGGAG
GGAGCAGACCCGGTGGTGTGGTATTCCTTTGGTGTGACACACGTGGTTCGGATTG
AGGACTTCCCTGTCATGCCTGTGGAGACTGTCGGCTTCCAACTGAAGCCCGTGGG
ATTTTTCACGGGCAACCCAGGCGTTGATGTGCCCCCCAGCTGCAACAAAGCCAGC
GTCTTGTGCAGGTCCTGTGTGTCTCAGAATGGCCACACATGA

>DtCuAO

MPEGLKGLKVMQIFVYLKSYPDDNQYAHPLDIMPFVDVILGKVVRIDMHDTPP
VINKEDNNYHSALFDPAKMRTDIKPLDVIQPEGPSFNVCGNNVKWCKWSFHVG
FNAREGLVLSDVSYDGRPILYRFSFVEMAVPYAEPREPYVRKCAFDIGDYGFGNT
ANSLHLGCDCLGAIHYFNGVLSNSKGEPVDVPKVVCMHEEDVGLLWKHVEPR
TGHNESRRSRRLVVSFISTVVNYEYAFYFYFYLDGTFEVEVKLTGELSTNVLSPG
EENPEFGTLVMPYVNAQHHQHMFCVRIDPAIDDKNGGKDVVITEVNCEPVPPG
PKNPYANAFYAKEQPLLTVKEAQRDISPAHNRVWKLTNPSCLNPITKKPVSYKL
MPHASPPLMAHSESLVAKKGYFATKNLWVTPFAQDQRFPAGDFVFNEKTCHGL
SQWTKEDTSLEGADPVVWYSFGVTHVVRIEDFPVMPVETVGFQLKPVGFFTGN
PGVDVPPSCNKASVLCRSCVSQNGHT

b

>DtIVD_cDNA

ATGCCCATGAGCCTGTGGCAAGAGATGGGGTGCATGGGGCTGCTTGGAGTAACG
GCGCCAGAGGAGTATGGCGGCCTGGGATTGGGCTATCGCGAACACTGTGTGGCCA
TGGAAGTAATTAGCGCTGCATCAGGCTCTGTAGGCCTGTCGTATGGAGCACACTCT
AACCTATGCGTCAACCAGATTGTGCGCAATGGAACCGCAGATCAAAAAGCCCGTT
ACCTGCCAAAGCTGATATCAGGAGAACACATCGGAGCACTGGCAATGTCTGAGG
CAGGCAGTGGAAGCGACGTTGTGAGCATGAAACTGAGAGCAACGCGCAGGAGC
GACCACTTTGTGCTCAACGGCACCAAAATGTGGTGCACCAATGGGCCAAAGGCG
GATACCTTGATAGTGTACGGCAAGACAGAGCCTGAAAAAGGAGCACATGGCATC
ACAGCATTCTGCATTGAAAAGGGCATGAGGGGTTTCAGCACAGCACAGAAGCTC
AACAAACTAGGCATGAGGGGGAGCGACACTGCTGAGCTTGTCTTTGAGAATTGC
GAAGTGCCTATGTCGAATGTTCTGGGTGAGATCAATGGAGGTGTTGGCGTGATGA
TGAGTGGCCTGGACTTGGAGAGGCTCGTCCTGGCCGCCGGGCCTGTTGGACTTAT
GCAGGCCTGCTTGGACGTGGCCGTACCATACGCAACGCAACGCAAGCAGTTTGG
TCAGGCAATTGGGGAGTTCCAGCTCATCCAAGGAAGACTTGCAAACATGTACAG
CCACCTCCATGCGAGCCGAGCATTCATGCTCAGTGTAGCTGACGCTGCAGATGCT
GGACGGCCAGATCGCAAGGATTGTGCCTCAGTTATCCTCTTTGCAGCTGAGTCTG
CCACTGAGACTGCCCTAGATGCCATTCAGATCCTTGGCGGAAATGGTTACATACA
AGAGTATCCCACAGGAAGGTACCTGCGTGACGCCAAGCTGTATGAAATAGGAGCT
GGCACCAGTGAAATAAGGAGGTTCTTGATTGCCAGGGAGCTGTTGAAGGAGTGT
CGTGAGGGCAAGGCC**TGA**

>DtIVD_amino_acid

MPMSLWQEMGCMGLLGVTAPEEYGGLGLGYREHCVAMEVISAASGSVGLSYG
AHSNLCVNQIVRNGTADQKARYLPKLISGEHIGALAMSEAGSGSDVVSMKLRAT

RRSDHFVLNGTKMWCTNGPKADTLIVYGKTEPEKGAHGITAFCIEKGMRGFST
AQKLNKLGMRGSDTAELVFENCEVPMSNVLGEINGGVGVMMSGLDLERLVLA
AGPVGLMQACLDVAVPYATQRKQFGQAIGEFQLIQGRLANMYSHLHASRAFML
SVADAADAGRPDRKDCASVILFAAESATETALDAIQILGGNGYIQEYPTGRYLRD
AKLYEIGAGTSEIRRFLIARELLKECREGKA

c

>DtMCCB_cDNA

ATGACCACCTTGGATACTTTACCACCTGGGCAAACAAGCAATCTTGCTCGGCGCC
AACAAGTCTCTGCTCACATGCATGCCTTGCTGCAAGAAATGGGTCAGGCTCTGCA
GCACATTCGGGAAGGAGGCGGGCATCATGCTGTGGAGCGGCACCGGCGGCGCAA
CAAAATGCTGCCTCGTGAGCGCATTGATGCTATAATGGATGCTGGGTCCCCATTCC
TTGAGTTCAGCCCACTTGCTGGGCATGGAATGTACGGTGCTGAGCAGGTCCCTGC
TGGAGGCATTGTGACAGGCATTGGGCAAGTGCATGGTCGGCTGGTGGCCATCGCA
GCCAATGATGCAACGGTAAAAGGAGGAAGCTACTACCCAATCACGGTGAAGAAG
CACCTGCGCCTTCAAGAAATTGCAGCTTCCTGCCGCTTGCCTTGCTTGTACCTCGT
GGATTCTGGGGGCGCAAATCTGCCCAGACAGGCAGATGTGTTCCCAGACAAAGA
GCACTTTGGCCGCATATTTTACAATCAAGCGAACATGTCTGCCGCCGGCATCCCTC
AGATTGCCCTGGTCCTAGGATCTTGCACTGCAGGAGGCGCGTATGTGCCAGCAAT
GGCTGATGAGTCTGTCATTGTGAAAGGCAATGGCACTGTCTTCTTAGGTGGCCCT
CCCCTGGTCAGGGCAGCCACAGGAGAAGTGGTGGGGCAAGAGGAGCTGGGTGG
TGCAGCTCTACATTGTGGCACCAGTGGCGTGACAGACCATTTTGCGCAGGATGAG
GCGCACGCGTTAGCACTGGCACGCAGTATCATTGCCTCAAGTTGCAGCACTGCCA
GGCATTCCAGGTCTGCCAGCATGGCAGGCATGCCATTGCCTTCTGCATGCGCGTG
GGATGAGCCTCTCTATCCAGCAGAAGAGTTGAGAGCGCTTGGGGAGGTGCCTGG
TGCTGAAAGTAGCAAAGGACAGGGTGGCGGCCCAATGGACCCGCGTGCCATTTG
TGCAAGGATTTTGGATGGCTCTCGCTTTGATGAATTCAAAGCCAACTATGGCACA
ACCCTTGTCACAGGGTTTGGTACACTGTATGGCCAGCCTGTAGGTGTTGTGGCAA
ATATGGGGGTCCTGCACAGTGCGTCGGCCCTCAAGGGTGCTCATTTTGTGCAGCT
TTGCGCACAGCGTGCCATACCTCTGCTCTTCTTGCAGAACATCACAGGCTTTATGG
TTGGCCGCTCTGCGGAGGCTGGGGGCATTGCAAAAGATGGGGCCAAGATGGTCA
TGGCAGTGGCATGTGCTAAGGTGCCAAAGATAACCGTGATTTGGGGGGGCAGCTT
CGGTGCTGGCAACTACAGTATGTCTGGCCGCGCATACAGCCCTGACTTTCTGTAC
ACATGGCCTAATGCTCGTGTAGGCGTGATGGGTGGAGACCAAGCGGCCCAAGTCC
TTGCTCAGGTGGAGCTTGAGAAGCGTGTGCGCACAGAAGGCAAGAAAGATGGTG
GAGAAGCGTGGAGCACAGAGGACCAAGCTGCATTCAAAAAGGGCATTGCAGAC
AAGCTCGATGCTGAAGCATCGCCATGGTTTGCCAGTGCACGGCTTTGGGATGATG
GTGTAATTGACCCGGCTGACACAAGGCGTGTATGTGGGCTTAGCCTTGCTGCGGC
TATGCAGAAATTCGAACACAATGGGCAAGCTCAAATCAGGCAAAGTGGTACTCCG
CAGCAATATGGTGTCTTCCGTATG**TAG**

>DtMCCB_amino_acid

MTTLDTLPPGQTSNLARRQQVSAHMHALLQEMGQALQHIREGGGHHAVERH
RRRNKMLPRERIDAIMDAGSPFLEFSPLAGHGMYGAEQVPAGGIVTGIGQVHG
RLVAIAANDATVKGGSYYPITVKKHLRLQEIAASCRLPCLYLVDSGGANLPRQA
DVFPDKEHFGRIFYNQANMSAAGIPQIALVLGSCTAGGAYVPAMADESVIVKGN
GTVFLGGPPLVRAATGEVVGQEELGGAALHCGTSGVTDHFAQDEAHALALARS
IIASSCSTARHSRSASMAGMPLPSACAWDEPLYPAEELRALGEVPGAESSKGQGG
GPMDPRAICARILDGSRFDEFKANYGTTLVTGFGTLYGQPVGVVANMGVLHSA
SALKGAHFVQLCAQRAIPLLFLQNITGFMVGRSAEAGGIAKDGAKMVMAVACA
KVPKITVIWGGSFGAGNYSMSGRAYSPDFLYTWPNARVGVMGGDQAAQVLAQ
VELEKRVRTEGKKDGGEAWSTEDQAAFKKGIADKLDAEASPWFASARLWDDG
VIDPADTRRVCGLSLAAAMQKFEHNGQAQIRQSGTPQQYGVFRM

d

>DtPDK_cDNA

ATGCAATCCGTGTTCCGAAGAGTGGTGGTGGCCCAGTTTGCCAAAGGGCGCCAC
CCTGCAAGCAGCTGGCTCAAGGCTCTGCAGGCGTCTTCGTACAGCAACGTGTCA
GCAGCAGCAACCGCCTACAGCAACAGCAATGCAGCGGTGGCCAGCAGCAGCAC
CGTGGCGAAGGGTGTTCGGTTGATGGACGTGCACGAGCAGTCCCTGGTTGACGA
CATCTTCTCGCATGCGCTCAAAAAGCAAACAGGCGTCAGCTTGAAGTACATGCTG
GATTTTGGAGCCAACCCCATCGAGCGCCAGCTCATCCTGTCTGCGCAGTTCCTGC
ACAAGGAGCTTCCAGTCAGGCTGGCGCACCGGGTAGCCGAGCTGGAGAACCTGC
CGTATGGGCTGTCTGCCAAGTCAAGCATCTTGAAGGTACGAGACTGGTATGTGGA
CTCTTTCAAAGAGTTGCGGCAGTTCCCTCCTATCACCAACGCCCAAGATGAAGCC
AACTTCACGATGCTGTTGCGCTCCATCTACCACCGCCATGCAAACGTGGTGCCTG
TTATGGCCAAGGGTGTATCAGAGCTGCGCATGGAGCTTATCAAGCAGCAGTCCGC
GGAGCGGGTGCTGTCCGAAATGCCGGAGGTGCACCAATTCTTGGATGGCTTCTAC
CTGTCGCGTATCGGCATCCGCATTCTCATAGGACAACACATTGCCCTGCATGAGCC
TCCCAAAGATGACCACATCGGCCTCATCTGCACCGCGTGCTCACCCGTCCAGGTT
GCCCAGGATGCCATCAATGATGCTAGGTCGATATGCTTACGCGAGTATGGCAGCTC
TCCAGAAGTCATGGTGTACGGCGCCCCTGAGTTTGTGTTCCCCTACGTACCCTCG
CATTTACACCACATGGTGTTTGAGCTTGTGAAGAACAGTTTGCGCGCGGTACAAG
ACAAATATGAGGATGCGTTAGAGGACTCGCCGCCTATAAGGGTGGTGGTGGCGGA
GGGTATGGAGGACGTGACCATCAAGGTGTCCGATGAAGGAGGTGGCATCCCACG
ATCAGGCATGCCCAACATCTGGACCTACCTGTACTCAACAGCCAATCACCCAGTG
CAGCTTGCTGAGCAAGATATACAGCCAACAATTGAGGACAGTCCTGTAGTTTTAG
CAGGCTATGGGTACGGCCTCCCTATCAGCCGGCTCTATGCGCGCTACTTTGGAGGG
GACCTGCAAATCATTTCCATGGAAGGGTACGGCACCGACGCCTACTTGCATTTAA
ATCGGCTGGGCAACTCGCAAGAGCCCTTGCCATAG

>DtPDK_amino_acid

MQSVFRRVVVAQFAKGRHPASSWLKALQASSYSNVSAAATAYSNSNAAVASSSTV
AKGVRLMDVHEQSLVDDIFSHALKKQTGVSLKYMLDFGANPIERQLILSAQFL
HKELPVRLAHRVAELENLPYGLSAKSSILKVRDWYVDSFKELRQFPPITNAQDE
ANFTMLLRSIYHRHANVVPVMAKGVSELRMELIKQQSAERVLSEMPEVHQFLD
GFYLSRIGIRILIGQHIALHEPPKDDHIGLICTACSPVQVAQDAINDARSICLREYG
SSPEVMVYGAPEFVFPYVPSHLHHMVFELVKNSLRAVQDKYEDALEDSPPIRVV
VAEGMEDVTIKVSDEGGGIPRSGMPNIWTYLYSTANHPVQLAEQDIQPTIEDSPV
VLAGYGYGLPISRLYARYFGGDLQIISMEGYGTDAYLHLNRLGNSQEPLP

List of Publications/ Submitted Manuscripts

- Lina Yao, Hui Shen, Nan Wang, Jaspaul Tatlay, Liang Li, Tin Wee Tan, Yuan Kun Lee. (2016) Elevated acetyl-CoA by amino acid recycling fuels microalgal neutral lipid accumulation in exponential growth phase for biofuel production. Plant Biotechnology Journal (IF: 6.090). DOI:10.1111/pbi.12648.

- Lina Yao, Tin Wee Tan, Yi-Kai Ng, Kenneth Hon Kim Ban, Hui Shen, Huixin Lin, Yuan Kun Lee. (2015) RNA-Seq transcriptomic analysis with Bag2D software identifies key pathways enhancing lipid yield in a high lipid-producing mutant of the non-model green alga *Dunaliella tertiolecta*. Biotechnology for Biofuels (IF: 6.444). 8.1 (2015): 1. DOI: 10.1186/s13068-015-0382-0.

- Lina Yao, Kenneth Wei Min Tan, Tin Wee Tan, Yuan Kun Lee. (2016) Exploring the transcriptome of non-model oleaginous microalga Dunaliella tertiolecta through high-throughput sequencing and high performance computing. BMC Bioinformatics (IF: 2.435). (Under 2nd review)

- Siti Radiah Binte Safie, Yi Kai Ng, Lina Yao, Yuan Kun Lee. (2016) Growth kinetics and trancriptomic profile of microalga *Dunaliella tertiolecta* in relation to light absorption and carbon fixation in changing light conditions. (In preparation for submission)

- Xiaonian Ma, Lina Yao, Bo Yang, Feng Chen, Jin Liu, Yuan Kun

Lee. (2016) RNAi mediated silencing of pyruvate dehydrogenase kinase promoted triacylglycerol biosynthesis in *Nannochloropsis salina*. (In preparation for submission)