MORPHOLOGICAL, BIOCHEMICAL AND TRANSCRIPTOMIC CHARACTERISATION OF *Chlorella sorokiniana* AND *Chlorella zofingiensis* DURING NORMAL AND STRESS CONDITIONS

By

SITI NOR ANI BINTI AZAMAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

July 2017

This thesis is lovingly dedicated to

My late father Azaman bin Dio
and my mother Aminah binti Jusoh

My beloved husband,
Wan Amirul Faiz bin Wan Muhammad

My dearest kids,
Wan Nurin Auni
and Wan Aisyah Inara

Who leads me with the light of their endless love, support and encourage me throughout my life.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

MORPHOLOGICAL, BIOCHEMICAL AND TRANSCRIPTOMIC CHARACTERISATION OF *Chlorella sorokiniana* AND *Chlorella zofingiensis* DURING NORMAL AND STRESS CONDITIONS

By

SITI NOR ANI BINTI AZAMAN

July 2017

Chairman : Yeap Swee Keong, PhD
Faculty : Institute of Bioscience

Chlorella has been identified as one of the most interesting microalgae species, which has high nutritional values, high growth rate, and is able to produce a wide range of metabolites in response to environmental changes. The objectives of this study are to characterise the morphology and biochemical contents and to identify the genes and miRNAs involved in regulating the production of carotenoids and lipids in *Chlorella sorokiniana* and *Chlorella zofingiensis* when cultured under high light intensity combined with glucose supplementation. In this study, stress was introduced to the *Chlorella* cultures by adding 2% glucose and increasing the light intensity from 10 to 100 µmol photons m^{-1} s^{-1}. Then, the pigments, total phenolic contents, and antioxidant activities of both *Chlorella* species were evaluated. The results showed that both strains grew larger when cultured under stress condition. Although the total carotenoid content was increased under stress condition, reduction of the pigment and total phenolic contents associated with lower antioxidant activity were also recorded. Subsequently, the transcriptome of *C. sorokiniana* was sequenced using Illumina paired-end sequencing, and 198,844,110 raw reads with the length of 100 bp were produced. After pre-processing, ~95% of high quality reads were *de novo* assembled using Trinity software into 18,310 contigs. Analysis of differential gene expression by DESeq2 package showed that a total of 767 genes were upregulated and 948 genes were downregulated in stress conditions. Then, miRNAs that regulate the genes during normal and stress conditions of both *C. sorokiniana* and *C. zofingiensis* were profiled and analysed using CLC Genomic Workbench and OmiRas. From both analysis pipelines, the known and predicted novel miRNAs were identified. Although most of the identified miRNAs were not functionally determined, this study suggests that they were species-specific, which may have roles in regulating genes during stress condition. In conclusion, identifying the genes and the regulation of various metabolite productions under different growth conditions are useful for further strain enhancement of the microalgae.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

PENCIRIAN MORFOLOGI, BIOKIMIA DAN TRANSKRIPTOMIK BAGI *Chlorella sorokiniana* DAN *Chlorella zofingiensis* SEMASA KEADAAN NORMAL DAN TEKANAN

Oleh

SITI NOR ANI BINTI AZAMAN

Julai 2017

Pengerusi : Yeap Swee Keong, PhD
Fakulti : Institut Biosains

Chlorella telah dikenalpasti sebagai salah satu spesies microalga yang paling menarik dan mempunyai nilai pemakanan dan kadar pertumbuhan yang tinggi, serta dapat menghasilkan pelbagai metabolit sebagai tindak balas kepada perubahan persekitaran. Objektif kajian ini adalah untuk mencirikan kandungan morfologi dan biokimia dan mengenal pasti gen dan miRNA yang terlibat dalam mengawal pengeluaran karotenoid dan lipid dalam *Chlorella sorokiniana* dan *Chlorella zofingiensis* apabila dibiakkan di bawah keamatan cahaya tinggi yang digabungkan dengan penambahan glukosa. Dalam kajian ini, tekanan telah diberikan kepada kultur *Chlorella* dengan menambahkan 2% glukosa dan meningkatkan keamatan cahaya dari 10 ke 100 µmol foton m^{-1} s^{-1}. Kemudian, pigmen, jumlah kandungan fenol, dan aktiviti bahan antioksida bagi kedua-dua spesies *Chlorella* telah dinilai. Keputusan menunjukkan bahawa kedua-dua strain tumbuh lebih besar apabila dikultur dalam keadaan tekanan. Walaupun jumlah kandungan karotenoid meningkat dibawah keadaan tekanan, penurunan pigmen dan jumlah kandungan fenol yang dikaitkan dengan aktiviti bahan antioksidan yang rendah juga direkodkan. Seterusnya, transkriptom *C. sorokiniana* telah dijujukkan menggunakan teknologi penjujukan hujung berpasangan Illumina, dan sebanyak 198,844,110 jujukan nukleotid mentah dengan panjang 100 bp telah dihasilkan. Setelah dipraproses, ~95% bacaan berkualiti tinggi dihimpunkan secara *de novo* menggunakan perisian Trinity menjadi 18,310 kontig. Analisis pembezaan pengekspresan gen oleh pakej DESeq2 menunjukkan sebanyak 767 gen dikawalatur menaik dan 948 gen dikawalatur menurun dalam keadaan tekanan. Kemudian, miRNA yang mengawal gen semasa keadaan normal dan tekanan kedua-dua *C. sorokiniana* dan *C. zofingiensis* diprofilkan dan dianalisis menggunakan CLC Genomic Workbench dan OmiRas. Berdasarkan kedua-dua saluran kaedah analisis ini, miRNA sedia diketahui serta miRNA ramalan novel telah dikenal pasti. Walaupun kebanyakan miRNA yang dikenal pasti ini tidak ditentukan fungsinya, kajian ini mencadangkan bahawa ia adalah khusus kepada spesies tersebut

yang mungkin mempunyai peranan tertentu dalam mengawalatur gen semasa keadaan tekanan. Kesimpulannya, mengenal pasti gen dan pengawalaturan pengeluaran pelbagai metabolit dalam keadaan pertumbuhan yang berbeza adalah sangat berguna bagi peningkatan strain mikroalga ini.

ACKNOWLEDGEMENTS

ALHAMDULILLAH, all praises to Allah for the strengths and His blessing in completing this thesis. First and foremost, I would like to express my gratitude to my supervisor Dr Yeap Swee Keong. Performing the projects described herein and writing this dissertation would not have been possible without his continuing support and guidance. I would like to thank other members of my committee, Prof Dr Fatimah Md Yusoff, Dr Tan Sheau Wei, Dr Norio Nagao, Prof David Hornby and Dr Qaiser I. Sheikh for their guidance and suggestions. Not forget to my collaborator from Columbia Asia University, Dr Darren C. J. Wong. Through meaningful discussions, they helped me to refine and expand upon my projects, resulting in better work. I thank each of them for their patience and their belief in my abilities. Their knowledge in a broad range of disciplines helped me gain a new perspective and direction on this research project. A special deeply thank to my parents and my family members have been a great support to me through their mutual encouragement. My husband has been incredibly patient and supportive throughout my graduate career. These works are dedicated to him because he is both my motivation and my reason to work hard and to try my best in everything I do. I also extend my thanks to my siblings, Siti Nor Ain, Muhammad Nor Adli and Siti Norhani for their support and sacrifice through my entire life. I would like to extend my special thanks to Ministry of Education that supported my work financially. Last but not least, I am also very grateful for the logistical and personnel support from the Institute of Bioscience and the University of Sheffield.

Thank you so much.

I certify that a Thesis Examination Committee has met on 14 July 2017 to conduct the final examination of Siti Nor Ani binti Azaman on her thesis entitled "Morphological, Biochemical and Transcriptomic Characterisation of *Chlorella Sorokiniana* and *Chlorella Zofingiensis* During Normal and Stress Conditions" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Rozita bt Rosli, PhD
Professor
Institute of Bioscience
Universiti Putra Malaysia
(Chairman)

Ho Chai Ling, PhD
Professor
Faculty of Biotechnology and Biomolecular Sciences
Universiti Putra Malaysia
(Internal Examiner 1)

Daniel James Gilmour, PhD
Lecturer
University of Sheffield
United Kingdom
(External Examiner 1)

John A Bryant, PhD
Professor
University of Exeter
United Kingdom
(External Examiner 2)

_____

NOR AINI BINTI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Yeap Swee Keong, PhD
Associate Professor
China-ASEAN College of Marine Sciences
Xiamen Universiti Malaysia
(Chairman)

Fatimah Md Yusoff, PhD
Professor
Institute of Bioscience
Universiti Putra Malaysia
(Member)

Tan Sheau Wei, PhD
Research Officer
Institute of Bioscience
Universiti Putra Malaysia
(Member)

Norio Nagao, PhD
Research Officer
Institute of Bioscience
Universiti Putra Malaysia
(Member)

David P. Hornby, PhD
Professor
Department of Molecular Biology and Biotechnology
University of Sheffield
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date**:**

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly acknowledgeg;
- ownership of intellectual property from the thesis is as stipulated in the Memorandum of Agreement (MoA), or as according to the Universiti Putra Malaysia (Research) Rules 2012, in the event where the MoA is absent;
- permission from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) are required prior to publishing it (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, Guide to Thesis Preparation 37 Lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012.
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Siti Nor Ani binti Azaman (GS35310JD)

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____
Name of Chairman of Supervisory Committee: Yeap Swee Keong

Signature: _____
Name of Member of Supervisory Committee: Fatimah Md Yusoff

Signature: _____
Name of Member of Supervisory Committee: Tan Sheau Wei

Signature: _____
Name of Member of Supervisory Committee: Norio Nagao

Signature: _____
Name of Member of Supervisory Committee: David P. Hornby

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AAE	Ascorbic acid equivalent
acetyl-ACP	Acetyl-acyl carrier protein
ADP	Adenosine diphosphate
AGPAT	1-acylglycerol-3-phosphate O-acyltransferase
ath-miR	*Arabidopsis thaliana*-microRNA
ATP	Adenosine tryphosphate
BBM	Bold basal medium
BHA	Butylated hydroxyanisole
BHT	Butylated hydroxytoluene
BLAST	Basic local alignment search tool
B-tub	Beta-tubulin
BWA	Burrows-wheeler aligner
CAL	Calmodulin
CAM	Crassulacean acid metabolism
CBF/DREBs	C-repeat-binding factor/dehydration-responsive element
CD	Chloroplast division
CDKA	Cyclin dependent kinase A
Cdna	Complementary DNA
CO2	Carbon dioxide
Cq	Quantification cycle
DAG	Diglycerides or diacylglycerol
DEGs	Differentially expressed genes
DGAT1	Diacylglycerol O-acyltransferase
DMAPP	Dimethylallyl pyrophosphate
DNA	Deoxyribonucleic acid
DOXP	Deoxyxylulose-5-phosphate
DPPH	2,2-diphenyl-1-picrylhydrazyl
DREs	Dehydration-responsive element
dsRNA	Double-stranded RNA
EFA	Essential fatty acids
EST	Expressed sequence tag
FADH	Flavin adenine dinucleotide
FDR	False discovery rate
FRAP	Ferric-reducing antioxidant power
GAE	Gallic acid equivalent
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GC content	Guanine-cytosine content
Gdna	Genomic DNA
GO	Gene ontology
GPAT	Glycerol-3-phosphate acyltransferase
GtRNAdb	Genomic tRNA database
HA	Hatching autospore
HCl	Hydrochloric acid
HKG	Housekeeping gene
HSP90	Heat shock proteins 90
IPP	Isopentenyl pyrophosphate
ITS	Internal transcribed spacer

KAAS	KEGG Automatic Annotation Server
KEGG	Kyoto encyclopedia of genes and genomes
Kmer	Short DNA sequence consisting of a fixed number (K) of bases
KO	KEGG orthology
KOALA	KEGG Orthology And Links Annotation
log_2FC	Log 2 fold change
MA	Mature autospore
MAGs	Monoglycerides or monoacylglycerol
Mbp	Mega base pairs
MEP pathway	Non-mevalonate pathway or Methylerythritol 4-phosphate
miRBase	MicroRNA database
miRNA	MicroRNA
mRNA	Mesengger RNA
N content	If a sequencer is unable to make a base call with sufficient confidence then it will normally substitute an N rather than a conventional base call
NaCl	Sodium chloride
NADH	Nicotinamide adenine dinucleotide
NaOH	Sodium hydroxide
NCBI	National center for biotechnology information
NGS	Next generation sequencing
NTC	Non-template control
OD	Optical density
ORF	Open reading frame
PC	Phosphatidylcholine
PCA	Principal component analysis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PGM	Phosphoglucomutase
piRNA	Piwi-interacting RNA
PPAP2	Phosphatidate phosphatase
PSII	Photosystem II
PUFA	Polyunsaturated fatty acid
Repbase	Repetitive DNA database
rfam	Non-coding RNA (ncRNA) database
RNA	Ribonucleic acid
RNA-SEQ	RNA sequencing
ROS	Reactive oxygen species
RPL19	Ribosomal protein L19
rRNA	Ribosomal RNA
RT-QPCR	Quantitative reverse transcription PCR
SAGE	Serial analysis of gene expression
SEM	Scanning electron microscope
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNAs
snoRNABase	Small nucleolar RNA database
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
spp	Several species
SSU	Small subunit

TAE	Tris base, acetic acid and EDTA
TAG	Triglycerides or triacylglycerol
TCA cycle	Tricarboxylic acid or citric acid cycle
TEM	Transmission electron microscope
TPC	Total phenolic content
tRNA	Transfer RNA
UBQ	Ubiquitin
UCP	Universal chlorophyte primer
UV	Ultraviolet
UV/VIS	Ultraviolet /visible
UV-B	Ultraviolet B
VST	Variance stabilization normalization

CHAPTER 1

GENERAL INTRODUCTION

Microalgae have been identified as good source of lipids for biofuels, protein and health metabolites including polyphenols, vitamins and antioxidants (Mostafa 2012). *Chlorella* sp. has received considerable attention, in view of its relatively high nutritional value, its capacity to modify its metabolites in response to changes in its growth medium, relatively rapid rates of reproduction and the possession of a thick cell wall that protect its nutrients (Iwamoto 2004). The microalgae *Chlorella* can produce an unusually wide range of metabolites during both normal growth and growth under stress (de Morais et al. 2015). It is well known that different microalgae produce different metabolites during stress (Skjanes et al. 2013).

In this study, two species of microalgae *Chlorella sorokiniana* and *Chlorella zofingiensis* were selected and evaluated based on their potential for producing high value metabolites under different culture conditions. The relative concentrations and profiles of various metabolites in microalgae are similar under optimal growth. However, the metabolite profile undergoes considerable changes under sub-optimal growth. Different algae species use different methods for managing these changes in the environment. Depending on their ability to handle various stresses, the microalgae produce different secondary metabolites to increase their chances of survival (Skjanes et al. 2013).

Numerous studies concerning the effects of stress upon the metabolome of microalgae have been documented in the literature (Ip and Chen 2005a; Lemoine and Schoefs 2010). However, our understanding of how microalgae respond to physiological stress at the molecular level is largely confined to model organisms, and the relevant pathways in microalgae have not been fully documented. Furthermore, the size and the appearance of microalgae can change profoundly depending on the environmental conditions and associated levels and types of stress. There is a clear need to improve the set of molecular methods and tools for working with microalgae. Previous studies suggested that, when the microalgae *Chlorella* was grown under different growth condition, the colour of the *Chlorella* changed from green to red or yellow, as a consequences of contributed by the pigment production (Ip et al. 2004; Del Campo et al. 2004; Ip and Chen 2005b; Cordero et al. 2011).

Phenotypic details surrounding the influence of stress on the production of pigments and metabolites that contribute to microalgae survival, can be understood by performing comparative transcriptome profiling (Fu et al. 2014; Sun et al. 2015). Transcriptomic study is an appropriate tool which provides an initial, broad view of the regulation of secondary metabolite biosynthetic pathways during a microalgal stress response. So far, the studies have focused on growth experiments and metabolite content screening, and have yielded

very limited information in respect of gene expression, in microalgae under normal and stress conditions (Doan et al. 2011; Perez-Garcia et al. 2011; Baba and Shiraiwa 2013; Goiris et al. 2015). With the discovery of specific classes of RNA molecules as gene regulatory in addition to the more widely accepted protein based transcriptional regulatory (Bartel 2004), transcriptome sequencing can provide a valuable approach for obtaining microalgae functional genomics information. However, transcriptomes relating to the biosynthetic pathways of lipids, polyphenols and vitamin D of *Chlorella* species remain to be profiled. Therefore, it is important to study the transcriptomic profile as well as the metabolite composition of *Chlorella.* It is important and timely to determine the true potential of these species and to support the potential for genetic engineering of these microalgae as they become an increasing focus for their development as alternative source of biofuel, food and health supplements.

In this study, *C. sorokiniana* and *C. zofingiensis*, two *Chlorella* species that exhibit relatively fast growth rates and have potential to produce high yields of secondary metabolites, including secondary pigments, during mixothropic growth were chosen. By performing systematic studies including both physiological (morphological and biochemical) and molecular response of both *Chlorellas* towards normal and stress conditions, this study would provide necessary information for understanding the molecular basis of some important metabolite biosynthesis pathways such as lipids and carotenoids.

1.1 Statement of the hypotheses

The hypotheses of this study are:

1. High light intensity combined with glucose supplementation that increase pigments and lipids production would alter antioxidant activity in *C. sorokiniana* and *C. zofingiensis*.

2. Transcriptomic analysis would reveal the upregulation of genes involve in carotenoid and lipid biosynthesis pathways of *C. sorokiniana* and *C. zofingiensis* when cultured under high light intensity combined with glucose supplementation.

3. Small RNA transcriptomic analysis would reveal the specific miRNA involve in regulating the production of carotenoid and lipid in *C. sorokiniana and C. zofingiensis* when cultured under high light intensity combined with glucose supplementation.

1.2 Research objectives

The objectives of this study are:

1. To characterise the morphology and biochemical contents (such as pigments, phenolic and antioxidant) of *C. sorokiniana* and *C. zofingiensis* under normal and high light intensity combined with glucose supplementation stress conditions.

2. To identify the genes involved in regulating the production of carotenoids and lipids in *C. sorokiniana and C. zofingiensis* when cultured under high light intensity combined with glucose supplementation through RNA-sequencing technique.

3. To identify the miRNAs involved in regulating the production of carotenoids and lipids in *C. sorokiniana and C. zofingiensis* when cultured under high light intensity combined with glucose supplementation through small-RNA sequencing technique.

CHAPTER 2

LITERATURE REVIEW

2.1 Microalgae *Chlorella*

Chlorella is one of the archetypical forms of coccoid green alga, which is placed under the phylum Chlorophyta and class *Trebouxiophyceae* and *Chlorella*ceae (Krienitz et al. 2004; Bock et al. 2011; Krienitz et al. 2015). Like many other eukaryotic green algae, *Chlorella* shares the main properties of eukaryote organisms such as double-bounded nucleus, mitochondria and chloroplast (Graham and Wilcox 1999). *Chlorella* species can be found in several environmental conditions ranging from what is considered moderate to extreme. They are commonly found in the water, freshwater lakes and ponds, marine or brackish waters. Some *Chlorella* species can be found in hot springs where temperature is greater than 100°C (Sakai et al. 1995) are not uncommon. In addition, they also can be found in places that experience extreme weather conditions such as the Arctic and Antarctica (Ahn et al. 2012; Shukla et al. 2013).

2.1.1 Morphological features of *Chlorella*

Chlorella has a spherical or ellipsoidal structure that is often subcategorised into four groups; spherical; ellipsoidal; spherical or ellipsoidal; and globular to sub-spherical cells (Iwamoto 2004). All groups of *Chlorella* have thin cell walls and the size of individual cells is between 2 to 12 µm depending on the culture conditions (Huss and Sogin 1990; Nemcova and Kalina 2000). Most of the *Chlorella* cells do not have flagella. A single cell of C*hlorella* can divide and subdivide into four cells every 16 to 20 hours through a non-motile reproductive phenomenon which gives the cells the name autospores (Hoshina and Fujiwara 2013). *Chlorella* contains a parietal (or cup) shaped chloroplast bounded by a double membrane (Tomaselli 2004). The chloroplasts possess thylakoids (membrane sacs involved in the photosynthetic light reaction). Unlike higher plants, most algal chloroplasts contain pyrenoids. The pyrenoid is surrounded by a starch sheath and functions as secondary storage product (Nemcova and Kalina 2000). A diagram outlining a typical microalgal cell such as *Chlorella* is shown in **Figure 2.1**.

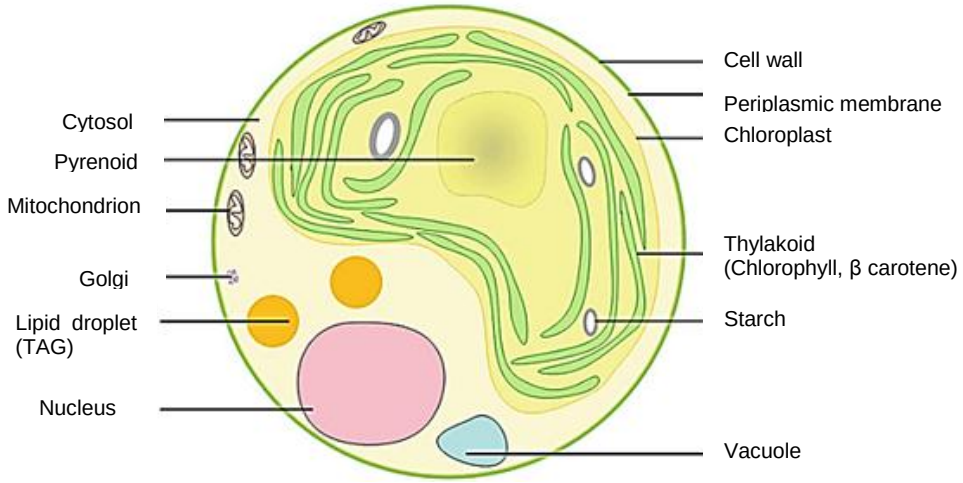


Figure 2.1: Typical organelle found in green microalgae (Pignolet, 2013)**.**

The cell wall structure of *Chlorella* has been shown to be both electron dense and electron transparent, comprising a layered structure (Nemcova and Kalina 2000). The thickness of the cell wall depends on the growth rate: in an intensively growing culture, the cell walls are typically thinner. The thickness of cell wall and its sugar composition have been previously used to distinguish different type of *Chlorella* species. For example, a study carried out by Yamamoto et al. (2004, 2005) revealed that the cell wall of C*hlorella* type species (such as *C. vulgaris, C. sorokiniana and C. labophora*) consisted of an electron dense thin layer with a thickness of 17-22 nm, while that of *Parachlorella kessleri* was electron transparent with a recorded thickness of 54-59 nm. In addition, the characteristics of cell wall based on sugar composition can be divided into two types; glucose-mannose and glucosamine. *Chlorella* which have glucose-mannose type cell wall are *C. saccharophila*, *C. ellipsoidea* and *C. fusca*, while *Chlorella* which have glucosamine type cell wall are *C. vulgaris*, *C. sorokiniana* and *C. kessleri* (Takeda 1988).

2.1.2 Biochemical composition of *Chlorella*

The biochemical components of *Chlorella* can be divided into three main groups; protein, carbohydrate and lipids. The dominant biochemical component of *Chlorella* species is protein, which made up to 60% of *Chlorella* cells, followed by carbohydrate and lipids about 20% each (Becker 2007). In addition, pigments (or carotenoids), vitamins and antioxidants have also been considered as primary components of *Chlorella* microalgae (Pignolet et al. 2013; Safi et al. 2014).

The proportions of protein and carbohydrate depend on the choice of media used to culture the microalgae, such as nitrate, key components in this respect are the concentrations of nitrate, phosphate and copper ions (Chia et al. 2013).

Carbohydrate and protein content of the microalgae can be manipulated by optimising the culture condition. The nutritional quality of algal protein is a function of the content, proportion and availability of its amino acids. Thus, different species of microalgae have different nutritional value as a result of the quantity and quality of their protein composition. The term protein efficiency ratio (PER), which is an expression of the weight gained by a test animal per unit of protein consumed and measurement of nitrogen retained for growth or maintenance, is the common method used to determine the quality of protein from different sources of microalgae (Becker 2007).

Carbohydrates in microalgae comprise a group of reducing sugars and polysaccharides. The polysaccharides found in microalgae exhibit structural variability and influence the biological functions of each species. Carbohydrates also serve as a form of energy storage for the cells (Safi et al. 2014). Physiologically, the polysaccharides in microalgae have been divided into three classes: structural, energy and glycocalyx. Structural polysaccharides are generally associated with cell walls; whereas energy polysaccharides are usually found as starch; and the glycocalyx polysaccharides are implicated in cellular communication and recognition (Pignolet et al. 2013). Among these three types of polysaccharides, starch is the most abundant in *Chlorella* species. In microalgae, starch is usually located in the chloroplast and is composed of amylose and amylopectin (Tomaselli 2004).

There is a range of lipids found inside the microalgae *Chlorella*, such as monoglycerides (MAGs), diglycerides (DAGs), triglycerides (TAGs), hydrocarbon and fatty acid. Most of them are accumulated either during stationary phase or under adverse environmental conditions (Hu et al. 2008; Pignolet et al. 2013). The main component of TAGs is fatty acids, which act as building blocks for the formation of various types of lipids inside the *Chlorella* cells. The common fatty acids found in *Chlorella* are 14:0, 16:0, 16:1, 16:2, 16:3, 18:0, 18:1, 18:2, and α-18:3 (Petkov and Garcia 2007).

The pigment profile of *Chlorella* has been shown to be similar to that of higher plants. There are two types of pigment present in *Chlorella*, i.e. primary and secondary carotenoids. The primary carotenoids usually found in *Chlorella* are α-carotene, β-carotene, neoxanthin, lutein, violaxanthin, antheraxanthin and zeaxanthin. The secondary carotenoids in *Chlorella* are equinenone, hydroxyequinenone, canthaxanthin and astaxanthin (Gouveia et al. 1996; Guedes et al. 2011). Production of pigments in *Chlorella* is a function of the stress imposed by environmental conditions. For example, Gouveia (1996) reported that *C. vulgaris* synthesized β-carotene, lutein and chlorophyll-a immediately before stationary phase and the concentration of these carotenoids was found to be reduced in contrast with increasing production of astaxanthin and cantaxanthin under high light intensity, salinity stress and nitrogen starvation. In another study, *C. sorokiniana* produced higher levels of lutein when the culture was exposed to high level of UV radiation (Cordero et al. 2011). Astaxanthin production by *C. zofingiensis* was also found to be increased when cultured under stress, although the stress factors were

different (Ip et al. 2004; Ip and Chen 2005a; Li et al. 2009). Besides the components discussed above, *Chlorella* are also rich in antioxidants and vitamins. For example, *Chlorella* contains large quantities of folate, vitamin B-12 and iron. Extracts of *Chlorella* have been shown to be effective in treating anaemia as well as reducing the risk of proteinuria and oedema in a cohort of pregnant woman in Japan (Nakano et al. 2010).

2.2 Factors that affect biochemical composition of microalgae

The steady state levels of the majority of metabolites in microalgae remain fairly constant under normal growth conditions. However, the content changes during sub-optimal conditions, such as nutritional stress. Moreover, environmental factors such as temperature, light, pH as well as nutrients not only affect photosynthetic and growth rates of the microalgae but influence the cellular metabolites and their composition. Sub-optimal conditions are generally referred to as external stress in respect of microalgae (Juneja et al. 2013; Skjanes et al. 2013). Different microalgae species use different mechanisms to regulate biochemical and metabolic adjustments to establish a new steady state of growth that promotes their survival (Hering et al. 2006). Thus, these sub-optimal conditions are commonly manipulated in order to obtain the desired content of microalgae for industrial purposes (Sharma et al. 2012; Skjanes et al. 2013). Some common factors that are known to affect microalgae metabolism and biochemical composition are discussed below:

2.2.1 Temperature

Temperature is one of the most important environmental factors that influence microalgal growth rate, biochemical composition, nutrient requirements and cell size. The range of temperature over which microalgae can grow is wider than for most organisms from 15 to 40°C depending on the strain. Cell growth at optimal temperatures results in a minimal cell size. Under non-optimal temperature, changes in cytoplasmic viscosity reduce the efficiency of carbon and nitrogen utilization (Raven and Geider 1988). Temperature also plays an important role in photo-inhibition process following these mechanisms: (i) low carbon dioxide fixation efficiency due to low temperature results with reduces electron transport at a given photon flux rate; (ii) low temperature inhibits the formation of active oxygen species, which leads to a reduction in photo-inhibition by protecting PSII (photosystem II); and (iii) low temperature inhibits the synthesis of D1 protein during photo-inhibition, which result with blocking the PSII repair cycle (Juneja et al. 2013).

Changes of temperature also lead to the modification in the level of unsaturated fatty acids in the lipid membrane (Thompson 1996). The fluidity of cell membrane is determined by the ratio of saturated and unsaturated fatty acids. Higher unsaturated fatty acid content leads to a high fluid membrane, while higher saturated fatty acids content leads to a rigid membrane. At low temperature, the saturated fatty acids are compressed and dense resulting in rigid cell membrane. The increasing level of unsaturated fatty acids at low

temperature will increase membrane fluidity. This was evidenced by *Duniella salina* which increased their level of unsaturated fatty acids when the temperature of growth was reduced from 30 to 12°C (Lynch and Thompson 1982). Together with greater fluidity, high levels of unsaturated fatty acids also increased the stability of cellular membranes (particularly thylakoid membranes), which in turn protect the photosynthesis machinery from photo-inhibition (Nishida and Murata 1996).

Temperature has also been shown to have an impact on starch content in algal cells. Starch is synthesized from phosphorylated metabolites arising from the dark reactions of photosynthesis. Increased temperatures lead to degradation of the starch produced mediated by α-amylase and α-glucan phosphorylase (Nakamura 1983). Nakamura and Miyachi (1982) have reported when the temperature was shifted from 20°C to 38°C in *C. vulgaris* culture, a siginificant reduction in starch and increment of sucrose were shown. They also reported the enhancement of starch degradation with temperature rise. For example, only 10% starch degradation was recorded when temperature was shifted to 28°C compared to 50% degradation when shifted to 40°C.

Temperature also has significant effect on the carotenoid production in microalgae. Carotenoids in algal cells are involved in photosynthesis by protecting the chlorophyll from photo-damage through light absorption process. In addition, they also involve in the energy transfer process and protect the reaction centre from auto-oxidation. An increase of cultivation temperature promoted the carotenoid accumulation in microalgal cells (Liu and Lee 2000; Tripathi et al. 2002). For example, Tjahjono et al. (1994) reported an increase of astaxanthin production up to 3-fold when the cultivation temperature was increased from 20°C to 30°C. Meanwhile, Liu and Lee (2000) found that the total secondary carotenoids was the highest at 35°C compared to 20°C. At higher temperature, an increased of biosynthetic enzyme activity (Liu and Lee 2000) and formation of active free oxygen radical leads to the production of high carotenoid content (Tjahjono et al. 1994).

2.2.2 Light

Solar energy (light) is the energy source for photosynthesis. Light intensity affects microalgal growth and the biochemical composition of cells through the process of photo-acclimation and photo-adaptation. In photo-adaptation, the microalgae change their fatty acid compositions, pigment composition, growth rate, and dark respiration rate (Vonshak and Torzillo 2004). For example, an increase in light intensity led to an increase in chlorophyll a and other light-harvesting pigments (such as primary carotenoids, chlorophyll b, chlorophyll c, and phycobiliproteins) (Hu 2004). On the other hand, when microalgae were exposed to a higher light intensity, chlorophyll a and other pigments which are directly involved in photosynthesis were reduced, while the secondary carotenoids, which serve as photo-protective agents, such as zeaxanthin, β-carotene and astaxanthin were found to be increased. The changes in ultrastructural, physiological and biophysical properties are due to photo-acclimation that accompanies changes in cell volume, cell number and density

of thylakoid membranes (Dubinsky and Stambler 2009). Light intensity greater than saturating limits, may discrupt the chloroplast lamellae and inactivate enzymes that are involved in carbon dioxide fixation, resulting in photo-inihibition (Brody and Vatter 1959). Changes in carbon and nitrogen flux have been shown to lead to the accumulation of secondary carotenoids in a special structure typified by the plastoglobuli of plastids or cytoplasmic lipid bodies, which protect the photosynthetic machinery from excess light.

Furthermore, the application of high light intensities can lead to oxidative damage of polyunsaturated fatty acid (PUFA). Typically, low light intensity induced the formation of polar lipids, whereas high light intensity decreases total polar lipid content. Membrane polar lipids usually associate with the chloroplast, when the chloroplast was exposed to a higher light intensity, simultaneous increase in the amount of TAGs was shown. The production of TAG under high light conditions might serve as a protective mechanism for the cell (Dongre et al. 2014). This is because, electron acceptors needed by the photosynthetic machinery might be depleted under high light conditions, so an increase in fatty acid which in turn were stored as TAG, potentially helps the cell to re-generate its electron acceptor pool (Sharma et al. 2012).

2.2.3 pH

pH is a major determining factor influencing the relative concentrations of the carbonaceous species in water. Alterations of pH changed the distribution of carbon dioxide species and carbon availability for algae growth as well as altering the availability of trace metals and essential nutrients (Chen and Celia 1994). At higher pH, the carbon for microalgae is available in the form of carbonates. Higher pH also lowered the affinity of microalgae to free carbon dioxide. In the photosynthesis process, the carbon dioxide used in the reaction reduced the carbon dioxide partial pressure and led to increase in pH. Alkaline pH also made the cell wall of mother cells flexible, thus prevent it from rupture and inhibit the release of autospore, which give time to the cells to complete the cell cycle. The flexibility of cell walls is due to the decrease of membrane-associated polar lipids when cultured under alkaline pH (Guckert and Cooksey 1990).

Cultivation of algae under acidic condition might alter the nutrient uptake and induce metal toxicity. It happens when the microalgae adapted to the changes of external environmental pH. Thus, more energy is required to pump proton out of the cells to maintain the internal pH in response to changes of external pH. This might be the mechanism for maintaining cellular metabolism by certain acid tolerant algae, thus the algal growth was not drastically affected by the acidic condition. For example, *Euglena mutabilis* is an acid-tolerant alga with a wider internal pH range compare to acid-intolerant algae. When the external pH is at 3, *E. mutabilis* can reduce the internal pH to 5 and increase the internal pH up to 8 when the external pH was above 9 (Lane and Burris 1981).

2.2.4 Salinity

Salinity is another important factor that affects the biochemical composition of microalgae. The photosynthesis process of microalgae was also inhibited by osmotic stress. The inhibition of photosynthesis in algae may be associated with the inhibition of PSII activity (León and Galván 1999, Murata et al. 2007). The salinity stress generally does not accelerate the photodamage to PSII directly, but it inhibited the repair of photodamaged PSII by suppressing the synthesis of protein required for repair mechanism. In this case, several possible mechanisms that led to these conditions were; (i) high concentrations of NaCl, which inactivate the translational machinery (or ribosomes) *in vitro*. Therefore, an increase in the intracellular concentration of NaCl, due to the influx of NaCl, might inhibit protein synthesis; (ii) Rubisco enzymes are inactivated by high concentration of NaCl, which consequently inhibit the carbon dioxide fixation and enhance the generation of ROS, resulting in the inhibition of protein synthesis; and (iii) the increase in intracellular concentrations of NaCl inactivates ATP synthase and decrease the intracellular ATP levels, which later reduce protein synthesis (Murata et al. 2007).

In addition, high levels of salt also affect lipid content in microalgae. Exposure to a higher salinity level increased the microalgae lipid content. For example, *Dunaliella* exhibited an increase in saturated and monosaturated fatty acids with an increase of NaCl concentration from 0.4 to 4 M (Xu and Beardall 1997).

2.2.5 Nutrient

Generally, the growth rate of the microalgae is proportional to the uptake rate of the most limiting nutrient under optimal temperature and pH (Juneja et al. 2013). The most important macronutrients for growth and metabolisms of microalgae are nitrogen and phosphate. Nitrogen is required for the formation of protein and nucleic acids, while phosphate, a key component of the nucleic acid backbone is also required for the formation of ATP; the energy “currency” of all cells. Additionally, phosphorus is also a key component of phospholipids. Limitation of these key nutrients drastically alters the metabolic pathway of the organisms. For example, the starvation of nitrogen and phosphorus shifted the lipid metabolism from membrane lipid synthesis to neutral lipid storage resulted with increasing of total lipid content (Fan et al. 2014).

Other important nutrients that support the growth of microalgae are carbon, hydrogen and oxygen. Carbon is one of the important non-mineral nutrients that must be supplied to the microalgae culture. Carbon can be utilised in the form of carbon dioxide, carbonate and bicarbonate. As with any photosynthetic, carbon based life forms, the latter forms of carbon are essential not only for photosynthesis but are required for microalgal respiration, energy source and raw material for the formation of additional cells etc. Reduced supplies of carbon fixed by algae leads to reduction in microalgal growth rates (Juneja et al. 2013). In addition, hydrogen and oxygen are the components of water which

is the medium in which all microalgae grow. All the nutrients and essentials ions that support microalgal growth are supplied in water to the microalgae. Interestingly, microalgae can utilise water to produce biohydrogen and oxygen through biophotolysis (Yu and Takahashi 2007).

2.2.6 Interacting environmental factors

Although numerous reports have shown that the cell composition of microalgae can be affected by a single chemical or physical factor, the effectiveness of such treatments is usually poor and slow. In a study by Cifuentes et al. (2001) on the effect of salinity on nine *Dunaliella* strains, it was found that none of the strains turned from green to orange or red. However, when these authors applied a combination of light, high salt and high temperature to the culture, an increase in β-carotene levels was observed in all the tested strains. Another example of the synergistic effect of a combination of multiple chemical and physical factors on astaxanthin formation in *Haematococcus pluvialis* was observed by Steinbrenner and Linden (2001). Based on their findings, a sustained increased in pigment synthesis was observed when a combination of two stress factors (high light plus salt, high light plus iron or salt plus iron), as compared to a single stress factor that produce only a moderate induction of astaxanthin synthesis.

2.3 Application of microalgae

Microalgae have been widely used for different applications, such as biofuels, human nutrition, cosmetics, animal feed, wastewater treatment and agrochemical applications. Microalgae exhibit three fundamental properties that can be manipulated for technical and commercial advantages: i) they are genetically diverse group of organisms with a wide range of physiological and biochemical characteristics, so that they can naturally produce many different and unusual fats, sugars and bioactive compounds; ii) the ability to perform photosynthesis makes them relatively easy to take up and incorporate the stable isotopes C^{13}, N^{15} and H^{2} from relatively inexpensive inorganic molecules such as carbon dioxide $^{13}CO_2$, nitrate $^{15}NO_3$, and water $^{2}H_2O$ to produce high value organic product such as amino acids, carbohydrate and nucleic acids iii) they constitute a large, unexplored group of organisms that provide a virtually untapped source of products. Thus, cultivation of microalgae is known to be the most profitable business in biotechnological industry (Avagyan 2008; Varfolomeev and Wasserman 2011). They produce less waste, ecologically pure growth with energy and resource saving process. Here, we describe the various commercial and industrial applications of microalgae (Spolaore et al. 2006).

2.3.1 Microalgae for human food

Microalgae are rich in carbohydrates, proteins, vitamins and minerals. Green microalgae have been used as a food supplement in Asian countries like China, Korea and Japan for hundreds of years. Nowadays, they are consumed

throughout the world for their nutritional value. Some of the most popular green microalgae, which are widely used and commercialised are *Chlorella vulgaris*, *Haematococcus pluvialis*, *Dunaliella salina*, *Spirulina maxima* and *Spirulina platensis*. These microalgae are known as excellent source of proteins, polyunsaturated fatty acids, pigments, vitamins and phenolics. The biomass of these microalgae is generally marketed as tablets, capsules and liquids, which are used as nutritional supplement (Spolaore et al. 2006; Priyadarshani and Rath 2012).

Besides consuming the microalgae biomass as a food supplement, the by-product or extract from microalgae is of major economic importance: it is used in the food industry as a food colorant and in the preservation of fruits and vegetables. For example, *Dunaliella salina* that is rich in β-carotene is used as a source of orange dye and as a Vitamin C supplement, and *Haematococcus pluvialis* is popular as a natural source of astaxanthin and lutein, which is used to treat certain owing to its high antioxidant activity (Spolaore et al. 2006; Vílchez et al. 2011). Furthermore, Rodriguez-Garcia and Guil-Guerrero (2008) have reported that the antioxidant activity of *Chlorella vulgaris* is higher than other commonly used antioxidants, which have great potential to replace synthetic antioxidants [such as Butylated hydroxyanisole (BHA) and Butylated hydroxytoluene (BHT)] in food preservation.

2.3.2 Microalgae for pharmaceuticals

Microalgae have received significant interest from the pharmaceutical sector, since they can produce natural, bioactive compounds that are otherwise synthetically difficult to produce (de Morais et al. 2015). The cell extract and cell-free extract of various microalgae (such as *Chlorella vulgaris*, *Chlamydomonas pyrenoidosa*, *Spirulina platensis*, *Dunaliella salina*, *Anabaena oryza*, *Nostoc ellipsosporum* and others) have been demonstrated to possess antioxidant, antimicrobial, antibacterial, antifungal, anticancer and antiviral activities (Mostafa 2012). These biological activities have been attributed to compounds belonging to several chemical classes including indoles, terpenes, acetogenins, phenols, fatty acids and volatile halogenated hydrocarbons (antimicrobial activity), phenolic compounds, saponins and alkaloids (antifungal activity), chlorellin, polyketides and lipopolysacharides (antibacterial activity), sulfated polysaccharides and sulphated flavones (antiviral activity), and cyanovirin (anticancer activity) that present in microalgae (Borowitzka 1995; Mayer and Hamann 2005; Mostafa 2012).

Moreover, essential fatty acids (EFAs) such as linoleic, arachidonic, linolenic and linoleic acids which can be found in microalgae play important roles in cellular and tissue metabolism, including regulation of membrane fluidity, electron and oxygen transport, as well as thermal adaptation. A numbers of studies have shown that consumption of PUFAs from seafood and plant sources may reduce the risk of coronary heart disease (Mozaffarian et al. 2005; Eussen et al. 2010). Furthermore, microalgal preparations such as *Spirulina* and *Chlorella* formulated in capsules or tablets, or even in food and beverages,

are known to have therapeutic values in treating hypercholesterolemia, hyperlipidaemia and atherosclerosis (Petkov and Garcia 2007).

2.3.3 Microalgae for cosmetics

Microalgae produce a wide variety of secondary metabolites with multiple bio-activities. In the cosmetics industry, component from microalgae are used as thickening agents, water binding agents and antioxidants. The most popular microalgal species in the skin care market are *Arthospira* and *Chlorella* (Stolz and Obermayer 2005). Furthermore, microalgal extract can also be found in anti-aging cream, refreshing or re-generating care products, anti-irritant in peelers, sun protection and hair care products. Organic metabolites that produce by certain microalgae (such as sporopollenin, scytonemin and mycosporin-like amino acids), which can absorb UV light are useful in UV radiation protection, thus proving suitable as a sunscreen ingredient (Priyadarshani and Rath 2012).

2.3.4 Microalgae for biofuels

Microalgae have long been recognised as potential sources for biofuel production owing to their high oil content and rapid biomass generation. Microalgae that can produce a significant amount of oil are known as oleaginous species (Hu et al. 2008). **Table 2.1** lists the oils produced by different oleaginous species of microalgae. The use of microalgae for biofuel production offers several advantages over other biomaterials, including robust oil productivity; do not require farmland for cultivation as well as low water supply; and cause low environmental impact (Xu et al. 2006). Furthermore, oil from microalgae can be processed into a broad spectrum of products including biodiesel, green diesel and gasoline replacements, bioethanol, methane, bio-oil and biochar through a various conversion processes such a trans-esterification, hydrothermal conversion, fermentation, anaerobic digestion and thermochemical conversion (Priyadarshani and Rath 2012).

Table 2.1: Oil content of microalgae (Chisti 2007)**.**

Microalgae	Oil content (% dry weight)
Botryococcus braunii	25–75
Chlorella sp.	28–32
Crypthecodinium cohnii	20
Cylindrotheca sp.	16–37
Nitzschia sp.	45–47
Phaeodactylum tricornutum	20–30
Schizochytrium sp.	50–77
Tetraselmis suecia	15–23

2.3.5 Microalgae for aquaculture

Microalgae are common natural food for many animals in aquaculture. Indeed microalgae are generally considered to be the best food source for aquaculture (Hemaiswarya et al. 2011). Microalgae feeds are used for the culture of larvae and juvenile shells and finfish and for raising the zooplankton required for feeding the juvenile animals (Benemann 1992). The most frequently used microalgae species in aquaculture are *Chlorella*, *Chaetoceros*, *Isochrysis*, *Nannochloropsis, Pavlova*, *Phaeodactylum*, *Tetraselmis*, and *Thalassiosira* (Hemaiswarya et al. 2011). They are supplied to the aquaculture organisms directly or indirectly. Direct feeding is associated with nutrition as a sole component, or as a food additive to supplement the basic nutrients. On the other hand, indirect feeding involves providing the microalgae to the zooplankton (such as daphnia, artemia and rotifers), which in turn is consumed as feed for fish (Spolaore et al. 2006; Priyadarshani and Rath 2012). In addition, microalgae such as *Dunaliella salina*, *Haematococcus pluvialis* and *Spirulina* have also been used successfully as a source of natural pigments to culture prawns, salmonids fish and ornamental fish (Gupta et al. 2007).

2.3.6 Microalgae for wastewater treatment

Another application of microalgae is in wastewater treatment. It has been employed successfully for the tertiary wastewater treatment mainly to remove nitrogen, phosphorus and heavy metals in aquatic wastewater treatment (Hammouda et al. 1995; Wang et al. 2010b; Abdel-Raouf et al. 2012). Besides providing the core of the biotreatment methodology applied to waste water, microalgae concentrates are also used to produce valuable biofuel feedstock, which can be used for several purposes. For example, Wang et al. (2010) who studied the performance of green algae *Chlorella* sp. on the wastewater sample from different stage of treatment process found that most of the metal ions were removed very efficiently, at the same time the process generated algal biomass which could be used for other applications. Another study done by Dalrymple et al. (2013) has proven that the algal biomass used in a wastewater treatment plant can be further manipulated to generate valuable products, especially biodiesel and biofuel. In their study, the sample from two different wastewater treatment plants of Tampa Bay cities in Florida were analysed and it was shown that it can produce approximately 270,000 gallons/year of biodiesel and 415,000 kg/year of biogas with 80% conversion efficiency, which would be able to sustain approximately 450 cars and power approximate 500 homes for a year, respectively.

2.4 Growth efficiency of microalgae

2.4.1 Autotrophic growth

The most common procedure for the cultivation of microalgae is autotrophic growth. Owing to the ability of microalgae to perform photosynthesis, they can harness environmentally available light energy and carbon dioxide to make their own food. Many microalgae are especially efficient solar energy convertor

and can be cultivated in naturally illuminated environments (such as open ponds) or artificially, in bioreactors (Perez-Garcia et al. 2011). Although they can use inorganic carbon and energy directly from the environment, the autotrophic culture mode has several disadvantages, which include long cultivation times and low biomass yields. Moreover, microalgae grown autotrophically in bioreactors require large surface area and shallow depth to provide sufficient light exposure by growing close to the surface of light source. Furthermore, maintenance of both outdoor and indoor bioreactors can be both tedious and expensive. The demanding space requirements and the constant need for light, make autotrophic cultivation of microalgae an expensive process (Knothe 2010). Furthermore, critics often raise concerns about the adverse environmental impact that could follow from the usage of large quantities of inorganic nutrients (Dhull et al. 2014).

2.4.2 Heterotrophic growth

Heterotrophic growth of microalgae occurs when photosynthesis is not possible or when light is experimentally withdrawn and/or energy producing reactions occur in the dark. In this case, the microalgae obtain their carbon through reduction of carbon from other organisms or alternative organic processes (Crane and Grover 2010). The advantages of heterotrophic growth for microalgae are: i) it is a simple process with cheaper and simpler bio-reactor design, ii) Easier scaling-up without the problems of surface area and the requirements for large land masses. iii) Higher growth rates and biomass yields, and iv) Possible to manipulate biomass composition by changing the culture medium for a specific metabolic and biosynthetic pathways (Barclay et al. 1994; Rosenberg et al. 2008; Zheng et al. 2012; Perez-Garcia and Bashan 2015).

Heterotrophic cultures also have several major limitations: i) Microalgae species that can grow heterotrophically are very limited in respect of metabolic diversity, ii) addition of organic substances such as glucose and glycerol require additional cost and energy, iii) more susceptible to contamination, iv) Excessive levels of organic substrate can lead to growth inhibition; and v) Inability to produce light-induced metabolites (Borowitzka 1999).

The metabolic products that have been obtained from heterotrophic cultivation of microalgae are lipids, polyunsaturated fatty acids, pigments and carotenoids. Generally, lipid production from heterotrophic growth is up to four-fold higher than under autotrophic growth based on the types of sugar used (either glucose, glycerol, or sucrose) (Li et al. 2014). Pigments and carotenoids such as lutein and astaxanthin are produced when using glucose as carbon source associated with nitrogen starvation (Ip and Chen 2005a; Ip and Chen 2005b). Meanwhile polyunsaturated fatty acid production through heterotrophic growth can be induced by lowering the growth temperature (Jiang and Chen 2000).

2.4.3 Mixotrophic growth

Mixotrophic cultivation is the growth mode where microalgae simultaneously use organic or inorganic sources and light in different combinations (Crane and Grover 2010). This kind of growth makes microalgae more physiologically flexible because it may meet both carbon and energy needs by organic or inorganic sources and light simultaneously (Chen et al. 2011). Mixotrophic growth offers several advantages, such as: i) Higher growth rates compare to autotrophic and heterotrophic mode by reducing growth cycles and high biomass production; ii) A longer exponential growth phase; iii) A reduction of the photo-inhibitory effect, iv) Flexibility for switching between growth modes (such as from heterotrophic to autotrophic mode), and iv) Protection from photo-oxidative damage stimulated by accumulating oxygen in enclosed photo-bioreactors (Kröger and Müller-Langer 2011).

2.5 Genomics studies of microalgae

Genomics or genome science is the study of the complete set of genes and non-coding nucleotide sequences within a genome (the complete set of information required to generate and sustain the life of an organism). Genomics also includes the study of genome structure, content and evolutionary relationship between other genomes. In broad terms, genome science also incorporates the analysis of gene expression as specified by the transcriptome of a specific cell type (especially in the case of multi-cellular organisms). The first requirement of any genome study is obtaining either the complete DNA sequence or the sequences of expressed genes, usually referred to as the transcriptome. The data obtained from nucleotide sequencing are evaluated using a range of bioinformatics techniques, which leads to assembly, phenotypic analysis and the comparative evaluation of the structure and function of the particular genome under study (Brown 2006).

Before the introduction of Next Generation Sequencing (NGS) technology, algae have been poorly studied owing to the limitations in genetic and molecular tools available for gene functions. Since the introduction of NGS, discoveries of algal genomes sequencing was dramatically accelerated. Some of the algal genomes that have been discovered are *Ostreococcus tauri* (Derelle et al. 2002), *Ostreococcus lucimarinus* (Palenik et al. 2007), *Chlamydomonas reinhardtii* (Merchant et al. 2007), *Micromonas pusilla* (Worden et al. 2009), *Chlorella variabilis* (Blanc et al. 2010) and *Volvox carteri* (Prochnik et al. 2010). The development of genomics has already made a major contribution to fundamental research on microalgae in the fields of functional biology, global ecology and the evolution of organisms (Armbrust et al. 2004).

Microalgae genomes can be structurally complex and vary considerably with respect to genome size (in bp). For example, the genome size of *Osteococcus tauri* is 12.6 Mbp, while other algal genomes can be up to hundreds and thousands of Mbp, such as *Emiliana heuxleyi* (168 Mbp), or *Karenia brevis*

(estimated at 10,000 Mbp) (Cadoret et al. 2012). Large genome sizes can make full genome sequencing a significant challenge. In view of this, many researchers have turned to transcriptome sequencing to build gene catalogues. Transcriptome sequencing was employed by several researchers to study *Pseudocattonella farcimen* (Dittami et al. 2012), *Chlorella vulgaris* (Guarnieri et al. 2011)*, Dunaliella salina* (Zhao et al. 2011) and *D. tertiolecta* (Rismani-Yazdi et al. 2011). In addition, there are several major transcriptome sequencing projects are currently under way; one example is Marine Microbial Eukarytoic Transcriptome Project (http://marinemicroeukaryotes.org/project _organisms).

2.6 Transcriptomic analysis of microalgae

Transcriptomics is the study of the transcriptome: a complete set of transcripts or a collection of RNA molecules derived from genomes of a cell produced from transcription under defined conditions. Comparison of transcriptomes allows the identification of genes that are differentially expressed in different cell populations, or in response to different environmental conditions (Adams 2008). A typical transcriptome can be classified into two main categories; coding and non-coding. Coding RNAs are RNA molecules that encode one or more proteins. The process of protein production (or more precisely the proteome) from a transcriptome is called translation. Only one type of RNA is involved in the translation process; messenger RNA (mRNA). Generally, mRNA comprises around 4% of the total RNA of certain cells. Many mRNA sequences have very short half-lives, which are readily degraded once they are synthesized (Alberts et al. 2002).

The second type of transcriptome was non-coding RNAs or known as functional RNAs. They were called non-coding RNAs because they do not act as templates for protein translation. However, these non-coding RNAs play a key role in cell physiology. Non-coding RNAs consist of ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), small nuclear RNAs (snRNAs) and small nucleolar RNAs (snoRNAs). Small RNA molecules can exert a potent effect on regulating development and metabolic processes inside cells: for example they may regulate gene expression through mRNA cleavage or translational repression (Brown 2006). Generally, in multicellular organisms, three major classes of endogenous small RNAs have been discovered: microRNAs (miRNAs), Piwi-associated small RNAs (piRNAs) and small interfering RNAs (siRNAs). miRNA are processed from RNA transcripts that form specific stem-loop structures. Small RNAs regulate gene expression based on the degree of sequence complementarity between miRNAs and their target mRNA. piRNAs are a newly discovered class of small RNAs that are associated with Piwi-family proteins. siRNAs are produced from long dsRNAs that originate from various sources including products of bidirectional transcription or RNA dependent RNA polymerase synthesis. siRNAs play various roles in post transcriptional regulation of gene expression in *cis* or in *trans*, suppression of transposable elements and formation of heterochromatin (Carrington and Ambros 2003; Bartel 2004).

Discovery of the small RNAs system in unicellular organism was found in 2007 by Zhao and colleagues using high throughput sequencing method. They discovered the first small RNAs system in unicellular model organism *Chlamydomonas reinhardtii* (Zhao et al. 2007b). Based on the computational analysis of the potential of the small RNAs to fold into microRNA precursors, they identified 68 candidate microRNAs. They then used a method developed in the plant *Arabidopsis* to predict the RNA targets of 18 of the microRNAs. Those targets are genes that are involved in cellular metabolism and physiological processes. They found evidence that four of the 18 targets were cleaved by the microRNA, which means the *Chlamydomonas* miRNAs can direct the cleavage of their targets with miRNA-complementary motifs and, presumably, act as regulatory molecules in growth and development (Molnar et al. 2009). The discovery of microRNAs in algae suggests that they evolved in a unicellular ancestor of both algae and higher plants. *Chlamydomonas reinhardtii* has large and complex genome, thus microRNAs might be present in other unicellular organisms that have complex genomes too.

The production of transcriptome ready for nucleotide sequencing is summarised as follows. Polyadenylated mRNAs are converted into cDNAs and subsequently amplified by real-time quantitative PCR (RT-qPCR). This basic methodology has been modified to determine the level of different types of RNAs present in the initial cell extract. Further modifications have been introduced to analyse cellular transcriptomes under certain experimental or physiological conditions. For examples, the expressed sequence tag based methods (EST and SAGE), hybridisation based gene microarray or chip technology, and NGS based RNA sequencing (RNA-seq) technology (Fryera et al. 2002) have all been deployed in transcriptome analysis. Among these methods, next generation sequencing (NGS) technique has emerged as the technique of choice, providing advantages in terms of throughput, specificity and sensitivity. Through this method, many important genes have been identified, including rare-disease-causing genes (Boycott et al. 2013), small RNAs in cancer research as well as in plant and animal research (Lu et al. 2005; Zhao et al. 2007b). **Table 2.2** provides a comparative summary of NGS with other methods.

Table 2.2: Comparison of transcriptomic analysis techniques.

Technology	Microarray	SAGE or EST sequencing	RNA-seq (NGS)
Technology specifications			
• Principle	Hybridization	Sanger sequencing	High-throughput sequencing
• Resolution	From several to 100 bp	Single base	Single base
• Throughput	High	Low	High
• Reliance on genomic sequence	Yes	No	In some cases
• Background noise	High	Low	Low
Application			
• Simultaneously map transcribed regions and gene expression	Yes	Limited for gene expression	Yes
• Dynamic range to quantify gene expression level	Up to a few-hundred fold	Not practical	>8,000-fold
• Ability to distinguish different isoforms	Limited	Yes	Yes
• Ability to distinguish allelic expression	Limited	Yes	Yes
Practical issues			
• Required amount of RNA	High	High	Low
• Cost for mapping transcriptomes of large genomes	High	High	Relatively low

Source: (Wang, Gerstein and Snyder 2009)

2.7 Challenges in Next Generation Sequencing

Despite the various advantages of using NGS, it also introduces significant computational and management challenges owing to higher error rates, shorter read lengths and unprecedented volumes of data. With the increasing amounts of NGS data recent years, NGS users have faced several informatics challenges, including the development of efficient methods to store, retrieve and process large amounts of data (Baker 2010). This is a critical barrier that must be overcome to reduce errors in image analysis, base-calling and to remove low-quality read datas. Reassuringly, several innovative computational tools and bioinformatics approaches have been developed to deal with data analysis and clinical translation (Magi et al. 2010). Based on a survey conducted by Bio IT World (Proffitt 2013), more than 50% of the 232 respondents believe that the biggest challenges for transforming NGS from lab to the clinic are data analytics and data management. For example, multiple identities of an identical genetic component such as SNPs result in ambiguities and confusion for researchers (Koboldt 2014).

2.8 System biology of microalgae

Systems biology is the analysis of relationship between genotype and phenotype of a given biological system or group of systems (Hong and Lee 2015). Data from high throughput omics research, such as genomics, transcriptomics, proteomics and metabolomics are critical to systems biology research. The data from omics experiments are often used to construct *in silico* models of biochemical reaction networks for the analysis of prediction of cellular behaviour under environmental and genetic perturbations. As mentioned in the previous topic, microalgae are one of significant interests to the Biotechnology community, for the synthesis of various industrially useful products, such as hydrocarbons and polysaccharides. These organisms exhibit, rapid growth characteristics, are amenable to large scale fermentation, and are potentially suitable for sustainable process development. Therefore, microalgae appear to represent a great opportunity for bioprospecting.

There are a number of web-based resources and databases which have recently emerged to assist in the analysis and management of various high throughput technologies including NGS, these include AlgaePath, BioCyc and AFAT (Lopez et al. 2011; Latendresse et al. 2012; Wolf 2013; Zheng et al. 2014). AlgaePath integrates genetic information, biochemical pathways, and NGS datasets for the green algae *Chlamydomonas reinhardtii* and *Neodesmus* sp. UTEX 2219–4. AlgaePath can be used to identify transcript abundance profiles and pathway information. The analysis using AlgaePath can be performed using five query page: i) gene search, ii) pathway search, iii) differentially expressed genes (DEGs) search, iv) gene group analysis, and v) co-expression analysis (Zheng et al. 2014). BioCyc integrates metabolic pathways databases from various sequenced organisms, whereas AFAT integrates multiple annotation databases from environmental data-mining for algal genomics into a centralizd system (Lopez et al. 2011; Latendresse et al. 2012).

2.9 Summary

Microalgae represent a promising source of natural pigments, lipid and other important biological metabolites which can potentially replace the synthetically produced metabolites. Production of pigments and lipids in *Chlorella* species has been intensively investigated in recent years. For example, *Chlorella zofingiensis* has been proven to become an alternative producer for astaxanthin production, which was previously dominated by *Haematococcus pluvialis* and *Dunaliella salina* (Ip et al. 2004; Ip and Chen 2005a; Li et al. 2009; Mulders et al. 2014). Meanwhile, *Chlorella sorokiniana* show high lutein production when exposed to high light intensity using random mutagenesis strategies (Cordero et al. 2011). Future exploration using *Chlorella* species for metabolite production mainly carotenoids and lipids may base on genetic engineering of this microalga for enhancement of productivity and purity, development of sustainable and cost-effective bioreactor system and exploration of biorefinery-based integrated production strategies. Advancements in these areas will greatly increase the production capacity and lower the production cost, allowing the *Chlorella* species to become a competitive source of natural carotenoids and lipids. To achieve these goals, the “omic” analysis, mainly transcriptomic which involve directly in regulating the biosynthesis pathways of carotenoids and lipids production, will benefit the understanding and the development of new molecular toolbox especially for microalgae for more efficient manipulation of the production strain.

CHAPTER 3

MORPHOLOGICAL AND BIOCHEMICAL CHARACTERISATION OF *Chlorella sorokiniana* AND *Chlorella zofingiensis* AND THE EFFECT OF STRESS ON THESE MICROALGAE

3.1 Introduction

Microalgae have been identified as a promising source of bioactive metabolites including lipids, carotenoids, polyphenols and vitamins warranting sustainable exploitation for energy, food and health applications (Chacón-Lee and González-Mariño 2010; Mostafa 2012). Studies of microalgae especially *Chlorella* is very important to the global scientific community, mainly because a significant number of beneficial content contribute to the various industries (Spolaore et al. 2006). The chemical compounds synthesized by microalgae are usually classified into primary and secondary metabolites based on the chemical class, functional groups and biosynthetic origin (Kumar et al. 2014). Primary metabolites are essential in maintaining the physical integrity and involved in various metabolic process mainly for the survival of the cells whereas secondary metabolites are not vital for the survival of cells. Both of these metabolites are pool of antioxidants such as phenols, carotenoids, terpenoids and flavonoid derivatives (Cardozo et al. 2007).

Many studies suggested that biological composition of microalgae such as carbohydrates, proteins, minerals and bioactive compounds have potential medicinal value which influences the nutritional value (Guedes et al. 2011; Mostafa 2012; de Morais et al. 2015). Bioactive compounds discovered from the methanol extract of red, green and brown algae such as polyphenols, flavonols, glycosides, catechin and phlorotannins have been reported to have strong antioxidant activity (Machu et al. 2015). In addition, carotenoids from *Chlorella* such as cantahanthin, astaxanthin and other non-provitamin A, have the ability to stimulate immune system and are potentially used to treat various cancers, coronary heart diseases, premature ageing and arthritis (Plaza et al. 2009; Obulesu et al. 2011).

Previous studies have demonstrated the possibility of controlling the composition of microalgae by changing the growth medium and introducing different stress conditions, as different physiological and biochemical contents are produced by the same microalgae under different growth conditions (Iwamoto 2004). For example, the colour of the *Chlorella* microalgae has been shown to change from green to red or yellowish, based on the pigment produced when grown under different conditions (Del Campo et al. 2004; Ip and Chen 2005b; Cordero et al. 2011).

In this study, two species of microalgae *Chlorella*, *C. sorokiniana* and *C. zofingiensis*, were selected because of their abilities to produce valuable

antioxidants from pigments and lipids such as lutein, astaxanthin and tocopherols which have potential applications in the pharmaceutical and health industry (Matsukawa et al. 2000; Brányiková et al. 2011; Liu et al. 2014; Mulders et al. 2014; Safafar et al. 2015; Liu et al. 2016). These strains are characterised by their high growth rates and high tolerances of various temperatures used during culture. These characteristics are expected to offer significant advantages for use in large-scale production bioreactors. Although *Chlorella* sp. are frequently consumed as a health supplement, most studies of *C. sorokiniana* and *C. zofingiensis* have focused on the profiling of their biochemical contents such as lipids and carotenoids, but less studies correlate with their antioxidant capacities (Del Campo et al. 2004; Ip and Chen 2005; Wang and Chen 2008; Liu et al. 2014). Thus, in the present study, the morphological and biochemical characteristics of *C. sorokiniana* and *C. zofingiensis* grown under normal and stress conditions were evaluated which focusing primarily on their antioxidant activities, mainly contributed by pigments and phenolic content. The antioxidant contents were assessed using DPPH radical scavenging and ferric-reducing antioxidant power (FRAP) assays.

The purpose of this study was therefore to investigate the physiological state in terms of morphology and biochemical content (mainly antioxidant contributed by pigments and phenolic compounds) under normal culture condition of *C. sorokiniana* and *C. zofingiensis*. Furthermore, the effect of stress (high light intensity and glucose addition during nitrogen-limited condition) on the physiological state of both *Chlorella spp*. and the comparison between normal and stress condition were also investigated.

3.2 Materials and methods

3.2.1 Source of microalgae and growth conditions

The *Chlorella sorokiniana* (NIES-2168) and *Chlorella zofingiensis* (ATCC30412) used in this study were obtained from the Marine Biotechnology Lab at the Faculty of Agriculture, Universiti Putra Malaysia, which had originally obtained them from UTEX and NIES (the culture collection of algae at the University of Texas and the National Institute of Environmental Studies, Japan).

The microalgae were cultured in Bold's Basal Medium (BBM) (PhytoTechnology Laboratory®, USA). The stock medium was diluted from 50x to 1x using sterile distilled water and adjusted to pH 6.6 using 1M NaOH. Sterilisation was performed by autoclaving for 20 min at 121°C. The ingredients of BBM are listed in **Table 3.1.** In the beginning of microalgae pre-culture preparation, the microalgae were first inoculated with 10% (v/v) of an exponentially growing algal culture and allow to grow under continuous light with an intensity of approximately 10 µmol photons m^{-1} s^{-1} with a shaking speed of 30 rpm at 27°C.

Table 3.1: Composition of Bold's basal medium.

Name	Molecular formula	50x (mg/L)	1x (mg/L)
Boric Acid	H_3BO_3	571	11.42
Calcium Chloride Anhydrous	$CaCl_2.2H_2O$	943.6	18.87
Cobalt Nitrate hexahydrate	$Co(NO_3)_2.6H_2O$	24.5	0.49
Cupric sulfate pentahydrate	$CuSO_4.5H_2O$	78.5	1.57
EDTA, Disodium Salt	EDTA-Na_2	3181	63.62
Ferrous Sulfate heptahydrate	$FeSO_4.7H_2O$	249	4.98
Magnesium Sulfate heptahydrate	$MgSO_4.7H_2O$	3750	75
Manganese Chloride tetrahydrate	$MnCl_2.4H_2O$	72	1.44
Sodium Molybdate	Na_2MoO_4	2984	59.68
Potassium Hydroxide	KOH	1550	31
Potassium Phosphate, Dibasic	K_2HPO_4	3750	75
Potassium Phosphate, Monobasic	KH_2PO_4	8750	175
Sodium Chloride	NaCl	1250	25
Sodium Nitrate	NaNO3	12500	250
Zinc Sulfate heptahydrate	$ZnSO_4.7H_2O$	441	8.82

3.2.2 Microscopic analysis of microalgae

3.2.2.1 Light microscope

The microalgae were investigated using a light microscope FSX100 (Olympus, New Zealand) using 40x bright field objective lenses. The cultures were diluted if needed and directly applied to the microscope slides.

3.2.2.2 Transmission electron microscope (TEM)

For transmission electron microscopy, microalgae were fixed in 4% (v/v) gluteraldehyde for 12 h at 4°C. Then, the fixed cells were washed three times using 0.1 M sodium cacodylate buffer for 10 min. After post fixation in 1% (v/v) osmium tetroxide for 2 h at 4°C, the cells were washed again and dehydrated in a serial dilution of acetone (35% to 100%) for 15 min. The cells were then infiltrated, and the beam capsule filled with resin mixture and polymerized in oven at 67°C. Semithin sections (1 µm thick) were cut using an ultramicrotome (Leica-Reichert Ultracut S, Austria). The sections were then stained with 2% (w/v) toluidine blue and viewed under a light microscope in order to select the region of interest prior to ultrathin sectioning. Ultrathin sections were cut and mounted onto 200-mesh copper grids and stained with uranyl acetate (2% uranyl acetate in 50% ethanol) and lead citrate (1.33 g lead nitrate and 1.76 g sodium citrate in 1 N of NaOH) for 10 min. The stained sections were finally examined using a transmission electron microscope (TEM) H7100 (Hitachi, Japan) at 80 kv.

3.2.2.3 Scanning electron microscopy (SEM)

For scanning electron microscopy analysis, the microalgae were treated as transmission electron microscope analysis until the step of dehydration with serial concentration of acetone (**section 3.2.2.2**). After that, the samples were coated with albumin on a small handmade baskets made from aluminium foil with a diameter of 1cm. The samples were transferred into specimen basket for critical point dried at 42ºC under pressure of 85 bar using a critical point dryer (Leica EM CPD030). The samples were mounted on stubs using double-sided tape or colloidal silver and then sputter coated with gold at 20 mA for 180 s using sputter coater (BAL-TEC SCD005). The samples were examined under a scanning electron microscope (SEM) JSM-6400 (JEOL, Japan) at 15 kv.

3.2.3 Determination of microalgae growth curve

The optical density (OD_{750}) of the microalgae was measured every two days using a Pharmacia Biotech Ultrospec 3000 pro spectrophotometer (Biochrom Ltd., UK) with sterile medium acting as a blank. Futhermore, the number of cells was also determined using a haemocytometer. The microalgae cultures were diluted before counting where necessary. The final concentration of cells was obtained by multiplying the cell number and dilution factor (df), together, (i.e. cell number/mL = X_m x 10^4 x df, where X_m is number of microalgae, and df is dilution factor). The growth curve of microalgal cells was determined by plotting a graph of cell number/mL or absorbance at 750 nm, against time (in days).

3.2.4 Determination of microalgal growth kinetics

3.2.4.1 Specific growth rate

The specific growth rate was calculated according to Kumar et al. (2014):

$$Specific\ growth\ rate, \mu = \frac{\ln(N_2 - N_1)}{t_2 - t_1}$$

Where, N_2 and N_1 were the number of cells per mL at days t_2 and t_1 respectively. The specific growth rate was taken in exponential phase.

3.2.4.2 Generation time

The generation time or doubling time of microalgae was calculated as follows:

$$Generation\ time, t_d = \frac{ln\ 2}{\mu}$$

Where, μ is specific growth rate of microalgae.

3.2.5 Effect of light stress on microalgae culture

3.2.5.1 Experimental design

The cells were first cultivated in Bold's Basal Medium (BBM) at 27°C under continuous light at a light intensity about 10 µmol photons m^{-1} s^{-1} with a shaking speed of 30 rpm. Three flasks for each *Chlorella* species were prepared for pre-cultured microalgae. The triplicates of pre-culture microalgae were allowed to grow until the mid-logarithmic phase, reaching approximately 8-10 x 10^6 cells/mL on day 11 and 2.5-3.0 x 10^6 cells/mL on day 15 for *C. zofingiensis* and *C. sorokiniana*, respectively. Then, both microalgae cultures were further divided into two flasks, and each flask contained 100 mL of 2 x 10^6 cells/mL. One flask of cells was cultured under normal condition and the other was cultured under stress condition. The same culture sources were used to reduce the variability of the cultures in subsequent comparisons.

The normal condition was the same as the pre-culture condition, whereas the stress culture condition consisted of a higher light intensity than the normal condition (100 µmol photons m^{-1} s^{-1}) and the addition of 2% of glucose (Ip and Chen 2005b; Sun et al., 2008). Microalgae cultured under both conditions were allowed to grow for 7 days once the stress condition was initiated at the mid-logarithmic phase. This was to induce the formation of secondary carotenoids under nitrogen limited condition. The experiments were conducted in a shaking incubator, and a conical flask was used as the growth chamber; the white fluorescence light source was located above the cultures. All experiments were repeated independently in triplicates. The microalgae were harvested by separating the pellet from the medium by centrifugation at 8,000 x g for 10 min. The pellet was then flash-frozen using liquid nitrogen and stored at -20°C prior to use.

3.2.5.2 Preparation of ethanol extract

The harvested microalgae were freeze-dried and ground using a mortar and pestle. A 0.2 g sample of ground microalgae was extracted for 24 h in 10 mL ethanol at room temperature. The extraction was repeated twice and the extract was filtered through Whatmann paper. Each filtrate was concentrated to dryness under reduced pressure using a speedVac Concentrator 5310 (Eppendorf, Germany). Finally, the dry extracts were lyophilized and stored at -20°C for further analysis.

3.2.5.3 Determination of pigment content

The pigment contents of the microalgae were determined using a method described by Lichtenthaler and Buschmann (2001). Briefly, the extracted sample was dissolved in 95% ethanol and filtered through two-layers of cheese cloths and centrifuged at 2500 rpm for 10 min. The supernatant was separated and the absorbance was measured at 400-700 nm. According to Lichtenthaler and Buschmann (2001), chlorophyll a exhibits a maximum absorbance at 664

nm, chlorophyll b at 648 nm and total carotenoids at 470 nm. The amount of these pigments was calculated according to the following formulae:

$$C_a\ (\mu g/mL) = 13.36\ A_{664.1} - 5.19\ A_{648.6}$$

$$C_b\ (\mu g/mL) = 27.43\ A_{648} - 8.12\ A_{664.1}$$

$$C_{(x+c)}\ (\mu g/mL) = (1000\ A_{470} - 2.13C_a - 97.64\ C_b) / 209$$

Where, C_a: chlorophyll a, C_b: chlorophyl b, $C_{(x+c)}$: total carotenoid]

3.2.5.5 Statistical Analysis

The data were expressed as means of ± standard deviation (SD). Statistical analysis was carried out by using SPSS software (Version 21). The optical density, cell numbers and dry biomass concentration were tested statistically using one-way analysis of variance (ANOVA) and post-hoc test. The significant level was set at $p < 0.05$.

3.2.5.5 Determination of total phenolic content

The total phenolic contents (TPC) of the ethanolic extracts were estimated using the Folin-Ciocalteau method (Taga et al. 1984). 100 µL of the aliquot sample was mixed with 2.0 mL of 2% Na_2CO_3 and allowed to stand for 2 min at room temperature. After incubation, 100 µL of 50% Folin-Ciocalteau's phenol reagent was added and the reaction mixture was mixed thoroughly and allowed to stand for 30 min at room temperature in the dark. The absorbance of all sample solutions was measured at 750 nm using a UV/VIS spectrophotometer. The blank consisted of all reagents and solvents without samples. Gallic acid was used as a positive control and was prepared in concentrations from 0.001 mg/mL to 1.0 mg/mL. The phenolic contents of the samples were expressed as gallic acid equivalent (GAE). Results were presented as means of experiments performed in triplicate ± standard deviation.

3.2.5.6 Determination of DPPH radical scavenging activity

Free radical scavenging activity was measured using 2,2-diphenyl-1-picrylhydrazyl (DPPH) according to the method described in Cox et al. (2010). Briefly, 2.0 mL of an aliquot of the test samples was added to 2.0 mL of 0.16 mM DPPH methanolic solution. The mixture was vortexed for 1 min and then incubated at room temperature for 30 min in the dark. The absorbance of all the sample solutions was measured at 517 nm using a UV/VIS spectrophotometer. The ability to scavenge the DPPH radical activity was calculated using the following equation:

$$\left[1 - \left(\frac{Sample - Sample\ blank}{Control}\right)\right] \times 100$$

Where: *sample* is the absorbance of the test sample containing the DPPH solution, *sample blank* is the absorbance of the sample without the DPPH solution and control is the absorbance of the DPPH solution without the sample. In this study, ascorbic acid was used as a positive control. Results were presented as means of experiments performed in triplicate ± standard deviation.

3.2.5.7 Determination of ferric reducing antioxidant power assay

The ferric-reducing antioxidant power (FRAP) assay of the ethanolic extract was carried out according to Hajimahmoodi et al. (2010). Briefly, the FRAP reagent containing 5 mL of a 10mM 2,4,6-tripyridyl-S-triazine (TPTZ) solution in 40 mM HCl with 5 mL of 20 mM $FeCl_3$ and 50 mL of 0.3 M acetate buffer (pH 3.6) was freshly prepared and incubated at 37°C. A 100 µL extract of each sample was mixed with the FRAP reagent and incubated at 37°C for 10 min before being measured at 593 nm. When necessary, the extracted samples were appropriately diluted with ethanol. Ascorbic acid of known concentration was used as a positive control and the final results were expressed as micromolar of ascorbic acid equivalent (AAE) (µM) per mg sample. Results were presented as means of experiments performed in triplicate ± standard deviation.

3.3 Results and discussion

3.3.1 Characterisation of microalgae *Chlorella*

3.3.1.1 Light microscopic observation of microalgae

The light micrograph of *Chlorella sorokiniana* and *Chlorella zofingiensis* are shown in **Figure 3.1.** Both *Chlorella* species showed either spherical or slightly ellipsoidal single cells, with a relatively thin cell wall. The obvious difference between *Chlorella sorokiniana* and *zofingiensis* is the size: *C. zofingiensis* is twice as big as *C. sorokiniana*. The diameter of *C. zofingiensis* cells ranged from 2 – 11 µm, while the diameter of *C. sorokiniana* cells ranged from 2 – 5 µm.

These microalgae reproduce asexually by autosporulation in which the autospores are discharged from an aperture that develops by a localized dissolution of the parent cells wall. The process of division is divided into several stages, which begin with chloroplast division, maturation of autospore and hatching of autospore (Yamamoto et al. 2005). Based on **Figure 3.1**, cells in different ontogenetic stages can be observed clearly using phase contract microscopy technique.

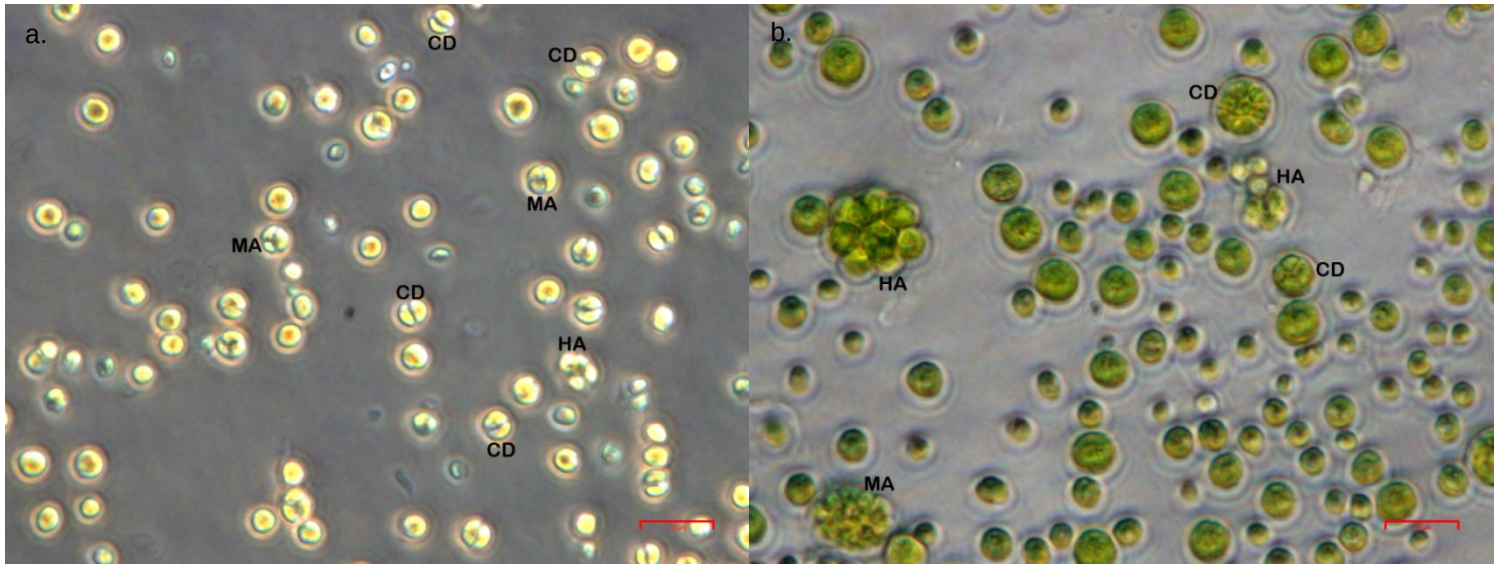


Figure 3.1: Morphology of microalgae *Chlorella* observed under light microscope using phase contrast technique. The above figure shows (a) *C. sorokiniana* culture and (b) *C. zofingiensis* culture. Different ontogenetic stages of the cells were labelled accordingly: CD is chloroplast division; MA is mature autospore; and HA is hatching autospore. Scale bar = 10µm.

3.3.1.2 Ultrastructural morphology of *Chlorella* revealed by SEM

Scanning electron microscopy showed that *C. sorokiniana* has a smooth cell wall, while *C. zofingiensis* has rough cell wall that is covered by an irregular network of subtle ribs (**Figure 3.2**). Interestingly, *C. zofingiensis* cell wall contains mucilage. This can be seen from the SEM micrograph of *C. zofingiensis*, where a tail-like structure can be seen between two cells that located side by side or non-detachable membrane seen between two *C. zofingiensis* cells at higher magnification. In the electron micrography the dividing cells of *C. sorokiniana* were also observed, as evidenced by the presence of a cleavage furrow on the surface of the cell membrane.

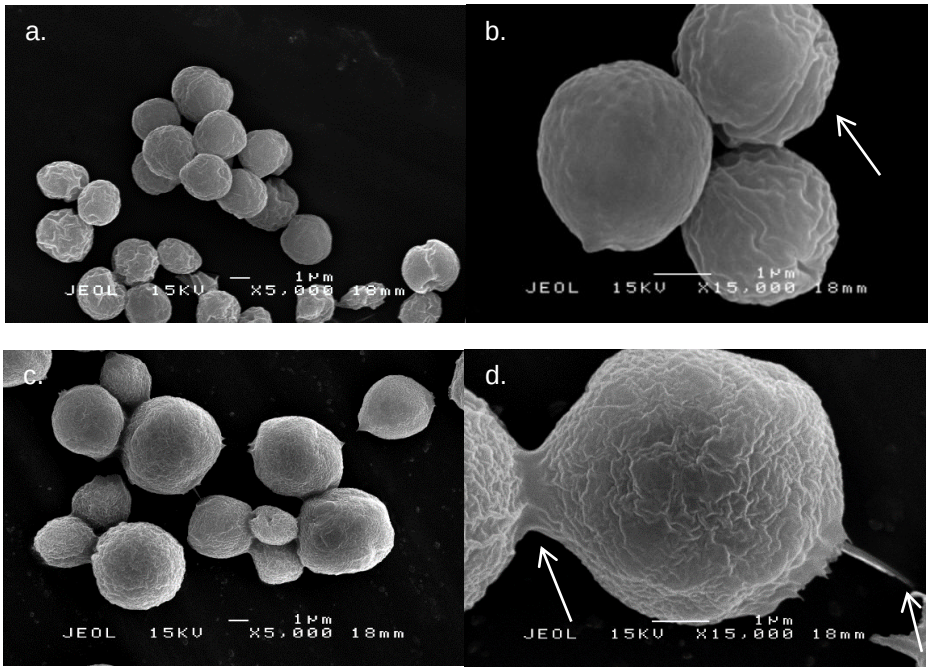


Figure 3.2: Scanning electron micrographs of *C. sorokiniana* (a and b) and *C. zofingiensis* (c and d). Arrow in *C. sorokiniana* SEM micrograph indicates cleavage furrow of dividing cell, while arrow in *C. zofingiensis* SEM micrograph indicates tail-like structure or mucilage.

3.3.1.3 Ultrastructural morphology of *Chlorella* revealed by TEM

The ultrastructure of *Chlorella* species contained the typical and common characteristics of all *Chlorella* including a nucleus, nucleolus, chloroplast, mitochondrion, starch and vacuole. The ultrastructural morphology of both *C. sorokiniana* and *C. zofingiensis* revealed by TEM are shown in **Figure 3.3**. The nucleus and nucleolus of *C. sorokiniana* are primiarily located at the side of cells, while in *C. zofingiensis* the nucleus and nucleolus are located in the centre of the cells. The chloroplast of *Chlorella* is either parietal or cup-shaped

with a clearly visible pyrenoid covered by starch grain. Each cell of *C. sorokiniana* contains pyrenoid matrix and is penetrated by a double thylakoid, which divides the pyrenoid into two almost identical halves. For *C. zofingiensis*, there is no evidence for the presence of pyrenoid. Many thylakoids are present and the thylakoid lamellae are arranged in several parallel rows in either a tightly packed stack (in *C. sorokiniana*) or as a loose stack (in *C. zofingiensis*). The starch granules found in *C. sorokiniana* are larger in size but fewer in number compared to the starch granules in *C. zofingiensis*, which are smaller and highly abundant.

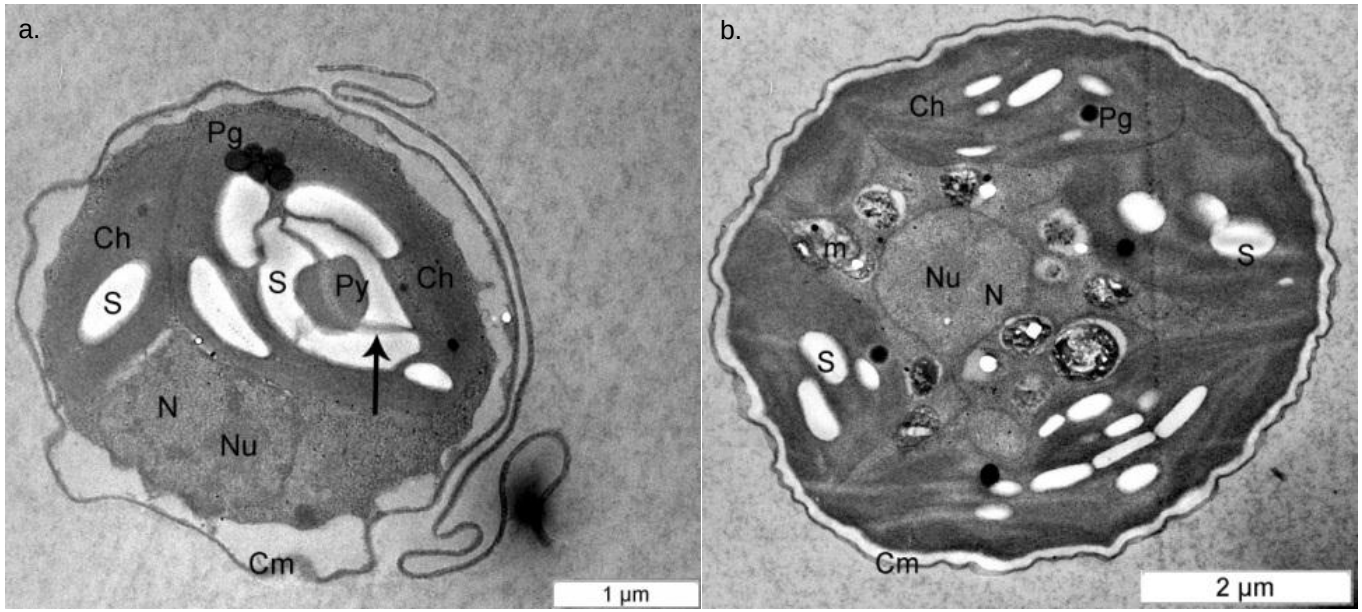


Figure 3.3: Transmission electron micrograph of *C. sorokiniana* (arrow shows the thylakoid penetrating the pyrenoid matrix) (a); and *C. zofingiensis* (b). S is starch granule, N is nucleus, Nu is nucleolus, Ch is chloroplast, Py is pyrenoid, Pg is plastoglobule, Cm is cell membrane.

3.3.1.4 Growth profile of microalgae

Growth profiles of both *Chlorella* species are shown in **Figure 3.4**. Under general optimal microalgae culture conditions, the microalgae clearly shown four of five phases of algal growth in batch cultures; lag, log (or exponential), stationary and declining growth phase. Both species showed short lag phase between two to four days of culture before started the exponential phase. The specific growth rate, division and generation time of each *Chlorella* species are shown in **Table 3.2**.

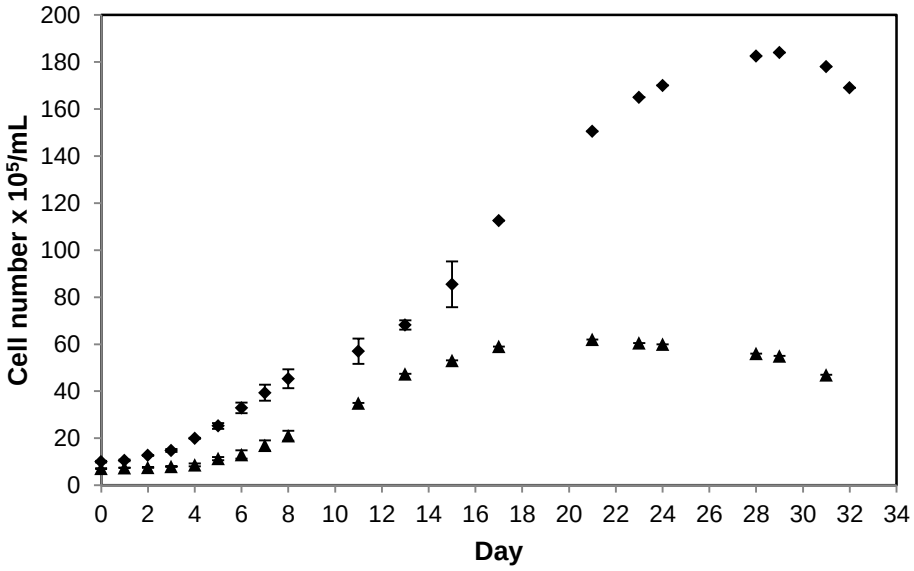


Figure 3.4: Growth curves of *C. sorokiniana* (♦) and *C. zofingiensis* (▲) under normal condition. Experiments were carried out in batch flask cultures for triplicates. Bars represent standard deviation.

Table 3.2: Growth kinetic of *C. sorokiniana* and *C. zofingiensis.*

Species	Specific growth rate (μ, day^{-1})	Generation time (td)
C. sorokiniana	1.99	0.35
C. zofingiensis	2.49	0.29

Specific growth rate for *C. sorokiniana* was taken from t1=13 to t2=21, meanwhile for *C. zofingiensis* t1=7 to t2=13.

Based on the growth curve shown in **Figure 3.4**, *C. sorokiniana* took longer time compared to *C. zofingiensis* to reach stationary phase, which is about 28 days, meanwhile *C. zofingiensis* took about 15 to 17 days to reach stationary phase. This was due to the difference in their generation time, in which *C.*

sorokiniana has longer generation time compared to *C. zofingiensis* (**Table 3.2**). Moreover, the specific growth rate of *C. zofingiensis* was higher than *C. sorokiniana*. Although they took different length of time to reach stationary phase, both showed declining growth starting from day 29 onwards. This indicates that specific requirements for cell divisionis limiting at later stage. In this phase, culture biomass is often very high, growth limiting factors such as nutrients, carbon dioxide, light begin to deplete making the growth rate decreases. At this stage, the cells may undergo dramatic biochemical changes. Besides reducing the protein content, nitrogen limitation cause changes in lipid and carbohydrate contents (Negi et al. 2016). Furthermore, shading effect due to light limitation may result in increasing pigment content of most species and shifts in fatty acid composition (Carvalho et al. 2011).

3.3.2 Effect of stress on morphology and biochemical content of *Chlorella*

Environmental factors such as temperature, light, pH and nutrients not only affect photosynthesis and the growth rate of the microalgae, but also influence the activity of cellular metabolome and its composition. Stress is defined as an environmental condition that results in a metabolic imbalance that requires biochemical and metabolic adjustments before a new steady state of growth can be established (Vonshak and Torzillo 2004). Owing to the effect of a combination of environmental factors, it is likely that significant changes to the microalgal physiology would be observed. Thus in this study a combination of different environmental stresses were applied, in order to monitor their effect(s) on the microalgal composition. The microalgae were induced with high light intensity and by the addition of glucose under nutrient limited conditions. Under this condition, both *Chlorella* species could produce high amount of secondary pigments and accumulates high lipid contents which may contribute to a high antioxidant content of the microalgae.

3.3.2.1 Morphological changes

It was observed that the stresses applied triggered morphological changes in the *Chlorella* cells. The most apparent aspect of these changes was impact on the cell size: both *Chlorella* species grew larger when cultured under stress conditions (**Figure 3.5**). The scanning electron micrograph clearly shows that the size of the cells increased due to an increase in the cell contents (**Figure 3.6**). **Figure 3.7** shows the distribution of cell sizes under different culture conditions for *C. sorokiniana* and *C. zofingiensis*. On average, the size of *C. sorokiniana* under normal conditions was 2 to 5 µm. When stresses were introduced, the size of *C. sorokiniana* increased slightly to 2 to 8 µm. Compared to *C. sorokiniana*, *C. zofingiensis* doubled the size of its cells when stresses were introduced to the culture, from about 2 to 8 µm to 3 to 13 µm.

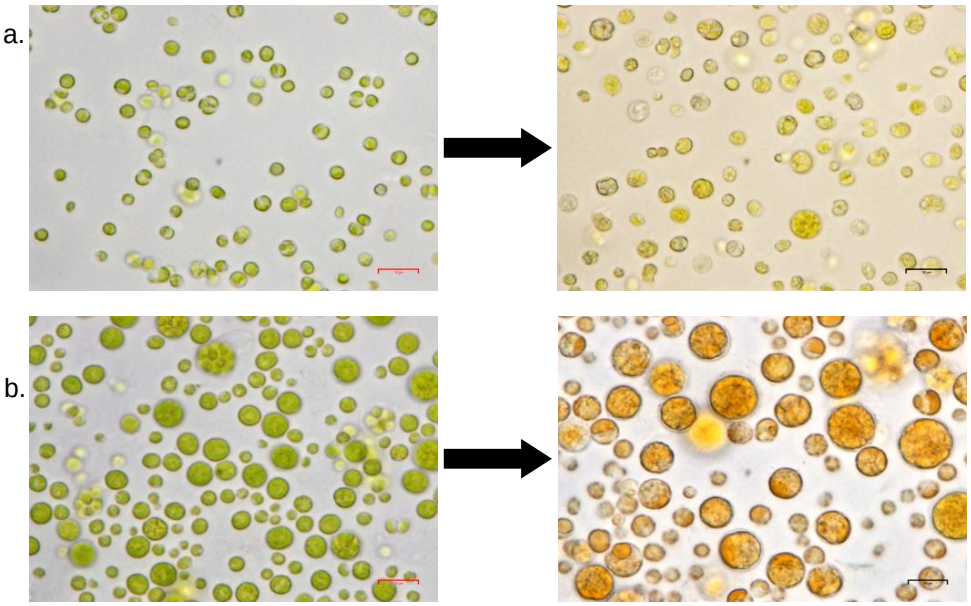


Figure 3.5: Morphology of (a) *C. sorokiniana* and (b) *C. zofingiensis* under light microscope changes from normal to stress condition (left to right). Both *Chlorella* spp. grew larger when cultured under stress condition. Changes of colour of the cells also witnessed, *C. sorokiniana* changes from green to yellowish while *C. zofingiensis* changes from green to orange red. Scale bar =10µm.

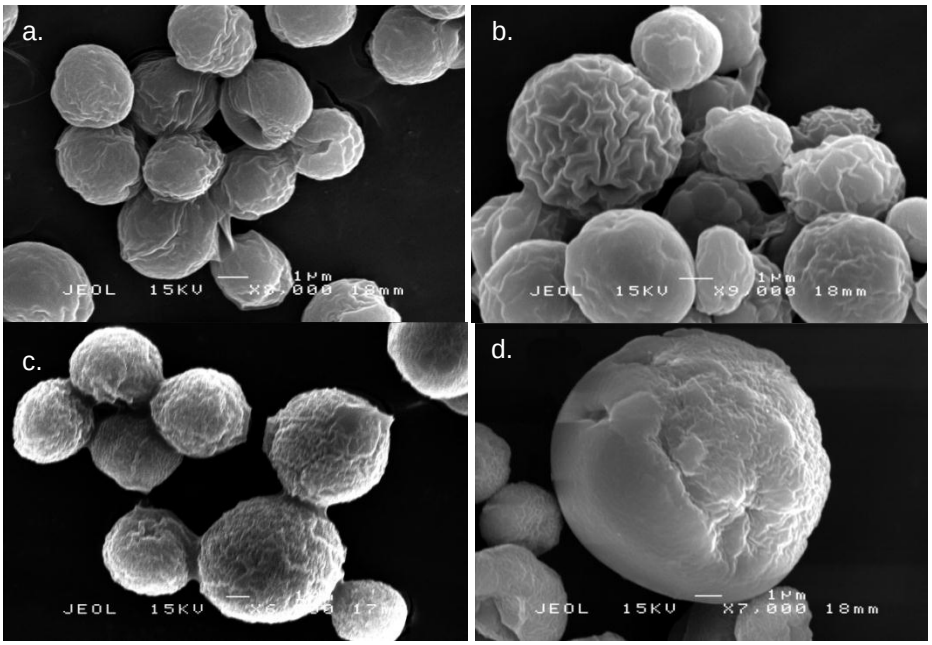


Figure 3.6: Scanning electron micrograph of *C. sorokiniana and C. zofingiensis* under normal and stress induced conditions. (a) *C. sorokiniana* in normal condition, (b) *C. sorokiniana* in stress induced condition, (c) *C. zofingiensis* in normal condition, and (d) *C. zofingiensis* in stress induced condition. The enlargment of cells are clearly seen on both *C. sorokiniana* and *C. zofingiensis*.

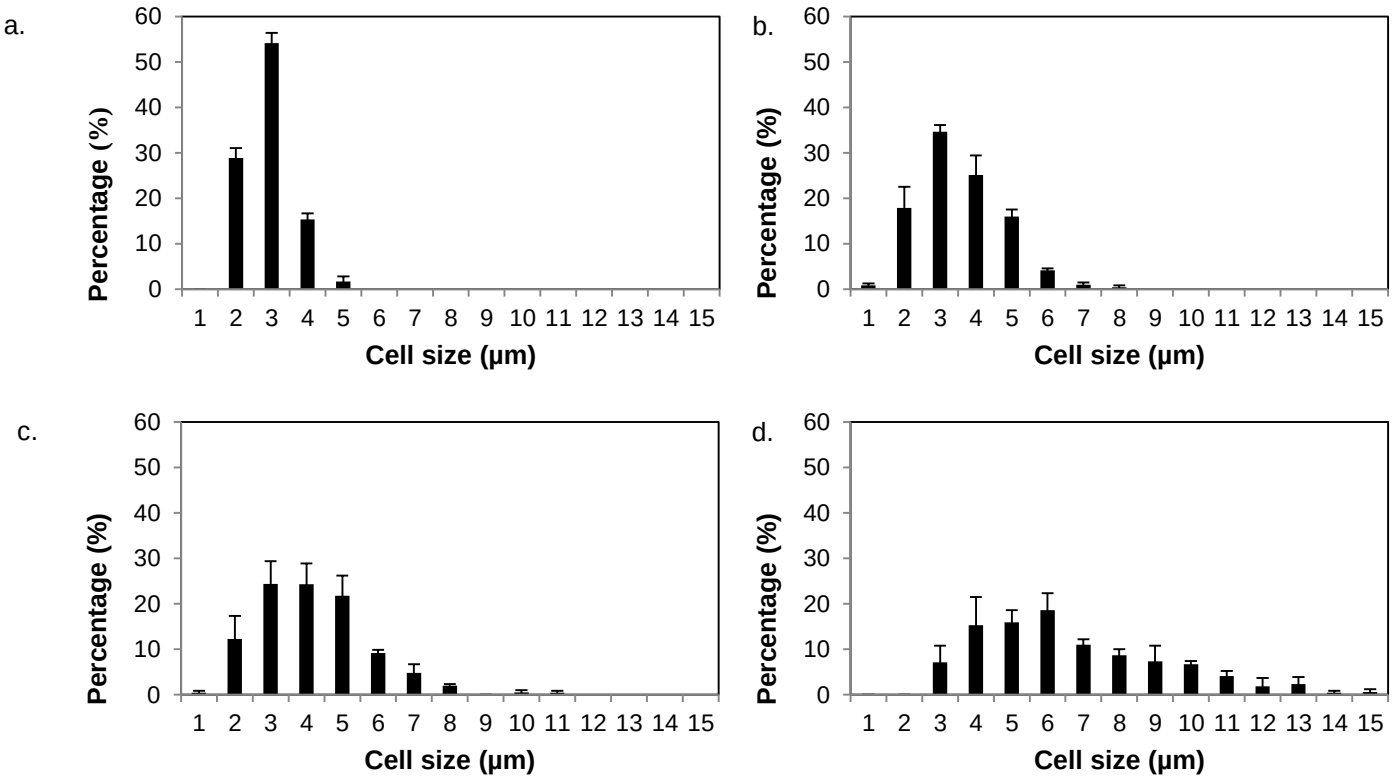


Figure 3.7: Cell size distribution of *C. sorokiniana* and *C. zofingiensis* in normal and stress induced conditions. (a) *C. sorokiniana* in normal condition, (b) *C. sorokiniana* in stress induced condition, (c) *C. zofingiensis* in normal condition, and (d) *C. zofingiensis* in stress induced condition.

Several researchers studying the effect of different stressors on cell morphology have also reported that the size of the microalgae increased when stress is introduced (Latasa and Berdalet 1994; George et al. 2014; Chokshi et al. 2015). Most of them found that the size of the cells increased between 1 and 2 fold, regardless of microalgae species. This result was supported by the SEM ultrastructure evaluation (**Figure 3.6**). However, electron micrographs also showed that though *C. sorokiniana* had a lower number of chloroplasts, these chloroplasts were larger than those of *C. zofingiensis* under normal condition (**Figure 3.8**). Wan et al. (2011) reported that *C. sorokiniana* expresses acetyl-coA carboxylase at higher levels in the cytosol than in the chloroplast under stress culture condition, suggesting that this species is less dependent on photosynthetically fixed carbon for lipid synthesis. Thus, fewer chloroplasts are needed to achieve sufficient amount of lipid accumulation. Future studies should compare the expression levels of acetyl-coA carboxylase in the cytosol and the chloroplasts in *C. zofingiensis* to further support this idea. Besides chloroplasts, *C. sorokiniana* also has a larger starch granule and plastoglobule compared to *C. zofingiensis* (**Figure 3.8**). When cultured under stress conditions, the lipid body and starch granule (accumulated in the middle of the cells surrounded by the lipid body) were predominant in the content of both *C. sorokiniana* and *C. zofingiensis*, which is similar to other reported strains of microalgae (Siaut et al. 2011; Yao et al. 2012; George et al. 2014). However, the biosynthesis of the starch and the lipid body was not necessarily proportionate and their accumulation was strain-dependent and varied according to the medium and culture conditions (Siaut et al. 2011; Takeshita et al. 2014).

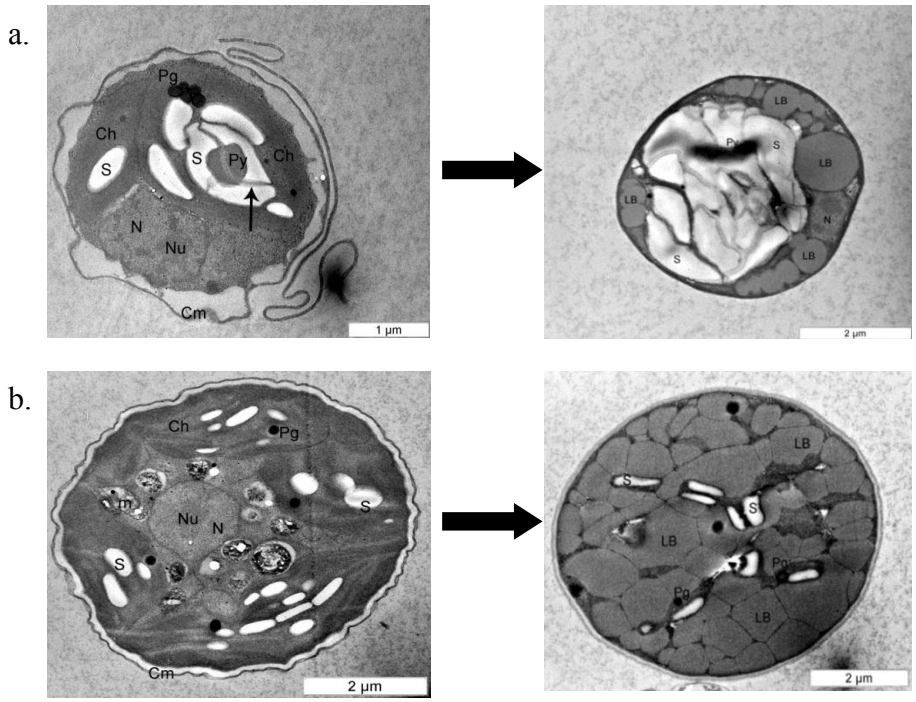


Figure 3.8: Transmission electron micrograph of *C. sorokiniana* (arrow shows the thylakoid penetrating the pyrenoid matrix) (a) and *C. zofingiensis* (b) changes from normal to stress condition (left to right). S is starch granule, N is nucleus, Nu is nucleolus, Ch is chloroplast, Py is pyrenoid, Pg is plastoglobule, LB is lipid bodies, Cm is cell membrane.

3.3.2.2 Pigments

All oxygenic photoautotrophic organisms like plants and green algae contain pigments such as chlorophyll, to capture and absorb energy from light to produce food through photosynthesis. There are several types of chlorophyll molecules, i.e. chlorophyll a, chlorophyll b and chlorophyll c. Chlorophyll a functions as a light-harvesting antenna in the main reaction centre, while chlorophyll b and c are accessory pigments that form light harvesting complexes, extending the range of light absorption (Lodish et al. 2000; Masojıdek et al. 2004). Besides chlorophyll, microalgae also contain both primary and secondary carotenoids. Primary carotenoids, such as β-carotene, lutein and zeaxanthin usually accumulate in the chloroplast while secondary carotenoids such as astaxanthin, canthaxanthin and adonixanthin are found in lipid bodies outside the chloroplast (Grünewald et al. 2001). The pigments contents of *C. sorokiniana* and *C. zofingiensis* during normal and stress conditions are shown in **Table 3.3**.

Based on the result presented in **Table 3.3**, chlorophyll a was the most abundant pigments found in both microalgae under normal conditions, followed

by chlorophyll b and total carotenoid. This is probably because chlorophyll a is the principal pigment in microalgae, while chlorophyll b is an accessory pigment that collects the energy to pass on to chlorophyll a. Between these two microalgae, *C. sorokiniana* had higher chlorophyll a content (17.929 µg/mg) compared to *C. zofingiensis* (15.690 µg/mg). This chlorophyll content is higher than other *Chlorella* species reported so far; such as 13.8 µg/mg in *C. vulgaris* (Goiris et al. 2015) and 1.455 µg/mg in *Chlorella* sp. (Safafar et al. 2015). When the cells were cultured under the stress condition, the chlorophyll a and b contents dropped by 80% compared with those of cells cultured under normal condition. These reductions in the chlorophyll contents under the stress condition were correlated with alleviation of photoinhibition under high light intensity. Besides, the addition of glucose made the microalgae less dependent to photosynthesis to produce energy leading to reduction of photosynthetic pigment such as chlorophyll (Liu et al. 2009). This was further evidence by the changes of colour of the cells, whereby a discolouration from greenish to a pale green or yellowish green was observed in *C. sorokiniana*, while *C. zofingiensis* became red or orange when cultured under stress-inducing conditions (**Figure 3.9**).

Table 3.3: Pigments present in the microalgae *C. sorokiniana* and *C. zofingiensis* under normal and stress conditions.

Culture	Sample	Chlorophyll a (µg/mg sample)	Chlorophyll b (µg/mg sample)	Total Carotenoids (µg/mg sample)
Normal	*C. sorokiniana*	17.929 ± 0.027^{a}	6.436 ± 0.040^{c}	3.882 ± 0.017
	C. zofingiensis	15.690 ± 0.003^{b}	7.311 ± 0.054^{d}	4.005 ± 0.046
Stress	*C. sorokiniana*	2.593 ± 0.005	1.127 ± 0.013^{e}	5.256 ± 0.217^{g}
	C. zofingiensis	2.598 ± 0.025	1.431 ± 0.051^{f}	5.805 ± 0.012^{h}

Data are given as mean ± standard deviation of triplicate experimental culture. In the column with a superscript letter is significant difference to each other (ANOVA, Post Hoc test, $p < 0.05$).

On the other hand, the total carotenoid content of both microalgae were increased between 30 - 40%, which is comparable with the carotenoid content reported in the literature (between 2 to 7 mg/g) (Matsukawa et al. 2000; Ip et al. 2004; Safafar et al. 2015; Goiris et al. 2015). As shown in **Table 3.3**, the amount of total carotenoid content per dry weight of the sample was almost the same under both conditions; this result was in contrast to those for chlorophylls a and b. This was due to the role of secondary carotenoids during stress condition, which protect the cells under higher light intensity to avoid photoinhibition effects (Zakar et al. 2016). According to Vonshak and Torzillo (2004), cells that undergo photoacclimation to a higher irradiance produce relatively higher amount of carotenoid compared to chlorophyll a. The changes in carotenoid content were also shown in this study by the discolouration of *C. sorokiniana* from greenish to a pale green or yellowish green and of *C. zofingiensis* from green to red or orange (**Figure 3.9**). Based on the previous studies, the type of carotenoid commonly produced by *C. sorokiniana* is lutein (Matsukawa et al. 2000; Cordero et al. 2011), while *C. zofingiensis* produces

astaxanthin (Liu et al. 2014). The yellow and red colour of *C. sorokiniana* and *C. zofingiensis* cultures in **Figure 3.9** depicting the colour of lutein and astaxanthin (Gupta et al. 2007).

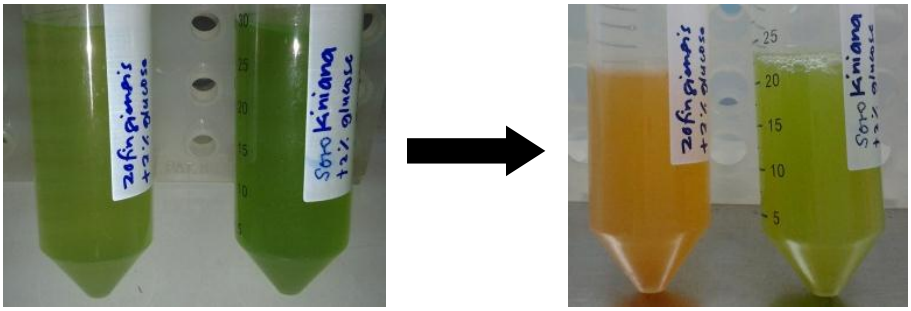


Figure 3.9: The changes in microalgae culture before (left) and after (right) induction of stress. Both *C. sorokiniana* and *C. zofingiensis* changes the colour from green to pale green and orange.

3.3.2.3 Total phenolic content

Phenolic compounds that are commonly found in plants, including microalgae, have been reported as having a wide range of biological activities including antioxidant properties. Being among the most important antioxidants, phenolic compounds have the ability to donate a hydrogen atom or an electron to form stable radical intermediates. The Folin-Ciocalteu method was used to study the total phenolic content of the microalgae (Ndhlala et al. 2010). The total phenolic contents of ethanol extracts of both *Chlorella* species during normal and stress conditions are presented in **Figure 3.10**. The total phenolic content under normal conditions was higher than stress conditions for both species, at 73.7 µg/mg gallic acid aquivalent (GAE) for *C. sorokiniana* and 40.8 µg/mg GAE for *C. zofingiensis*. The phenolic content of both microalgae were high compared to the other microalgae reported in the literature (Goiris et al. 2012; Hemalatha et al. 2013; Saranya et al. 2014; Safafar et al. 2015). For example, Saranya et al. (2014) and Goiris et al. (2012) who compared biochemical contents of different microalgae, found that *Isochrysis* sp. which had the highest phenolic content with only 2.5 mg/g GAE and 4.57 mg/g GAE. In another study, Ali et al. (2014) screened different microalgae for their carotenoid and phenolic content found *Chlorella* sp. possessed a high phenolic content (39.1 mg/g GAE).

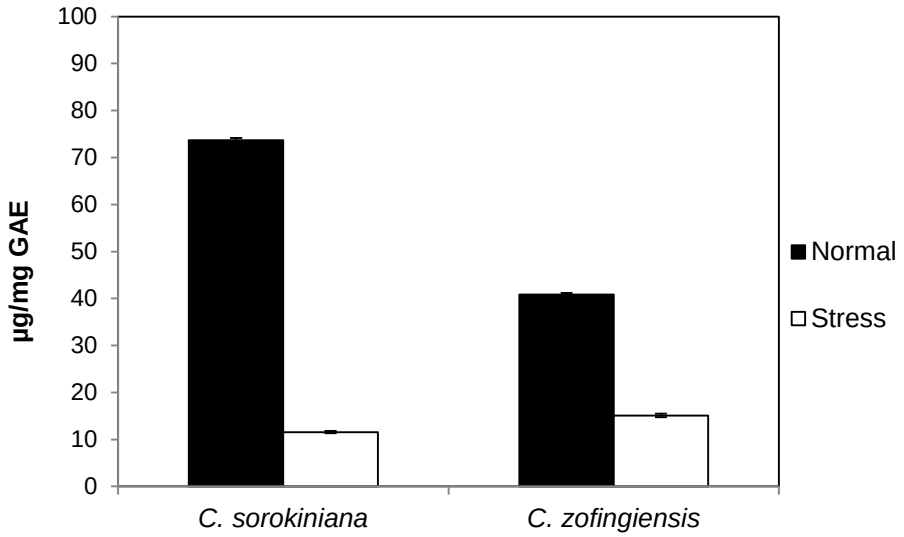


Figure 3.10: Total phenolic content of *C. sorokiniana* and *C. zofingiensis* under normal (black) and stress (white) conditions are compared. Error bars designate standard deviation from the average of three technical replicates.

Meanwhile, when the cells were cultured under stress condition, the total phenolic content reduced about 84.4 % for *C. sorokiniana* and 63 % for *C. zofingiensis*, becoming 11.56 µg/mg GAE and 15.10 µg/mg GAE, respectively. Although this results were contradicted with the previous findings which reported that *Chlorella* species would produce high phenolic content under stresses condition (Shetty and Sibi 2015), factors such as harvesting under nutrient-limited condition and different extraction technique might contribute to this effect (Goiris et al. 2012). As reported by Goiris et al. (2015) the production of carotenoid and phenolic content were reduced under nutrient-limited conditions, but the other contents such as ascorbic acid and tocopherols increased.

3.3.2.4 Antioxidant activity

The DPPH radical scavenging activities (%) of *C. sorokiniana* and *C. zofingiensis* under normal and stress conditions are presented in **Figure 3.11.** Extracts from both strains of microalgae possessed the ability to scavenge DPPH under normal and stress conditions to the same degree. *C. zofingiensis* showed slightly stronger scavenging activity in normal and stress conditions compared to *C. sorokiniana* with 13% and 14% scavenging activity under normal and stress conditions, respectively, while *C. sorokiniana* showed 11.5% and 13.7% of scavenging activity in normal and stress condition, respectively.

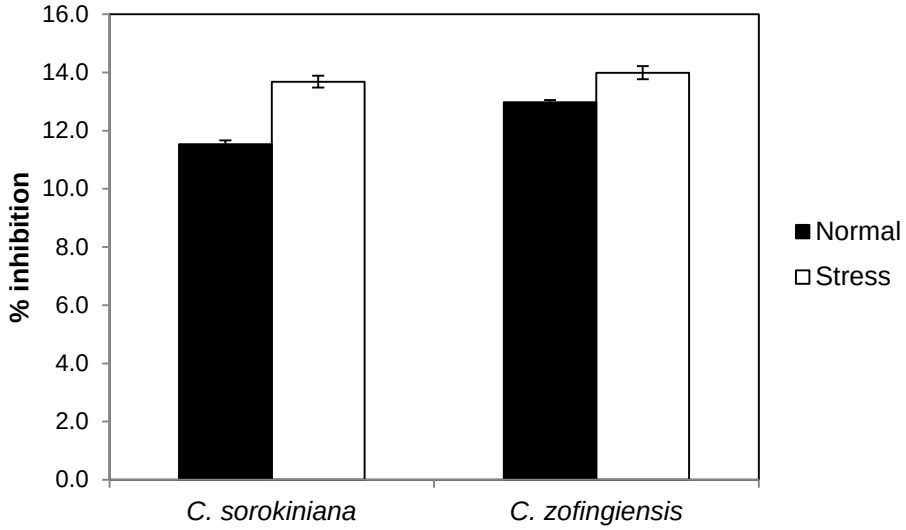


Figure 3.11: Percentage of DPPH radical scavenging activity by *C. sorokiniana* and *C. zofingiensis* under normal (black) and stress (white) conditions are compared. Error bars designate standard deviation from the average of three technical replicates.

Generally, the percentage of scavenging activity obtained in this study was within the range obtained in different microalgae reported in previous research (9 to 35%) (Hemalatha et al. 2013; Saranya et al. 2014; Safafar et al. 2015). For example Hemalatha et al. (2013) who investigated the antioxidant properties of different microalgae such as *Navicula clavata*, *Chlorella marina* and *Dunaliella salina*, reported the DPPH scavenging activity of selected microalgae at between 9% and 24%. Meanwhile, Saranya et al. (2014) and Safafar et al. (2015) showed the highest DPPH scavenging activity from a methanolic extract of *Isochrysis galbana* and *Chlorella sorokiniana* to be about 34% each. There are several factors that might affect DPPH scavenging activity such as limitation of DPPH when assessing the carotenoid samples and steric accessibility of DPPH assay (Prior et al. 2005). Some of carotenoid molecules may exhibit absorbance peaks at 517 nm which could overlap with the DPPH absorbance spectrum (Arnao 2000; Pérez-Jiménez et al. 2008). Besides, small molecules such as carotenoids have a higher apparent antioxidant capacity due to their better access to the DPPH radical site compared to the other type of antioxidant molecules.

For the FRAP assay, the reducing power of both *Chlorella* species was high under normal condition compared to stress conditions (**Figure 3.12**), whereby *C. zofingiensis* showed a higher reducing power activity compared to *C. sorokiniana*. The FRAP value of *C. zofingiensis* and *C. sorokiniana* under normal condition were 9.29±0.029 µM/mg ascorbic acid equivalent (AAE) and 6.13±0.097 µM/mg AAE, respectively. On the other hand, both species showed

a FRAP value of about 1 µM/mg AAE under stress condition. FRAP detects antioxidants that act through single electron transfer but cannot detect compounds that act as radical quenchers by hydrogen atom transfer (Prior et al. 2005), which is mainly carried out by phenols that have the ability to transfer single electron.

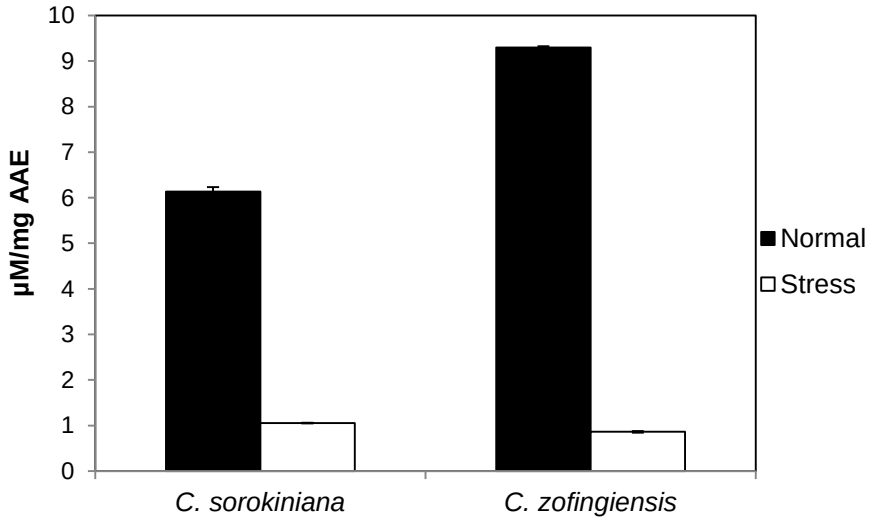


Figure 3.12: Ferric reducing antioxidant power of *C. sorokiniana* and *C. zofingiensis* under normal (black) and stress (white) conditions are compared. Error bars designate standard deviation from the average of three technical replicates.

Based on the result obtained, the trend of the antioxidant activity detected using the FRAP assay was proportional to the level of phenolic content. Thus, the FRAP assay results reflect the antioxidant activity from phenolic content. Although the total phenolic content of *C. zofingiensis* was lower than *C. sorokiniana*, the antioxidant activity detected by FRAP assay was high. This might be due to the presence of different kinds of antioxidant compounds similar to phenols, such as flavonoids and vitamin C (Ndhlala et al. 2010). More interestingly, the reducing activity of these microalgal extracts was higher compared to the literature, as most of the previously evaluated microalgae, such as *Navicula clavata*, *Chlorella marina*, *Dunaliella salina Chaetoceros calcitrans*, *Chlorella salina* and *Isochrysis galbana* had a reducing activity below 1 mg/g AAE (Uma et al. 2011; Hemalatha et al. 2013; Saranya et al. 2014). There are several factors that affect the yield of chemical extraction from microalgae, such as temperature, extraction time, types of solvent used and the chemical composition of the samples. The antioxidant in microalgae could have different properties, thus the antioxidant capacity of microalgae are strongly influenced by the extraction solvent (Goh et al. 2010; Prior et al. 2005). Though most microalgae studies used methanol, acetone and hexane as their extraction solvent (Uma et al. 2011; Hemalatha et al. 2013; Saranya et al. 2014), many industrial extraction methods especially for food and pharmaceutical applications, such as subcritical water extraction SWE (Herrero

et al. 2006) and accelerated solvent extraction (ASE) (Herrero 2005) used ethanol as their extraction solvent.

From the result of the DPPH and FRAP assay, the DPPH assay detected more antioxidant activity in stress sample compared to the FRAP assay. This may be due to several reasons. First, there is a DPPH limitation when assessing carotenoid samples, whereby some of carotenoid molecules may exhibit absorbance peaks at 517 nm which could overlap with the DPPH absorbance spectrum (Arnao 2000; Pérez-Jiménez et al. 2008). Second, in the DPPH assay steric accessibility is the major determinant of the reaction mechanisms, hence small molecules such as carotenoid have a higher apparent antioxidant capacity due to their better access to the DPPH radical site (Prior et al. 2005). On the other hand, some protein and thiol antioxidants, such as glutathione cannot be measured by the FRAP assay (Ndhlala et al. 2010). This might also affect the antioxidant activity detected by FRAP assay under stress conditions, where glutathione was produced in greater quantities (Cheng and He 2014).

Comparing the result obtained for both *Chlorella* species, *C. zofingiensis* showed a slightly higher antioxidant activity than *C. sorokiniana*. Even though the total phenolic content of *C. sorokiniana* was higher than that of *C. zofingiensis*, the FRAP result shown by *C. zofingiensis* was higher, indicating that, *C. zofingiensis* possessed a more potent antioxidant activity compared to *C. sorokiniana*. This occurrence suggests that different species of microalgae respond to stress conditions in different ways, leading to the production of different classes of antioxidants. In addition, different classes of antioxidant responded in different ways under different culture conditions. For example, the antioxidant activity of carotenoids and phenolic compounds was not the same under normal and stress conditions, whereby the antioxidant activity of phenolic compound was high under normal condition but low under stress condition, while total carotenoid antioxidant activity was high under both normal and stress conditions.

3.4 Conclusions

This work shows the different responses of *Chlorella sorokiniana* and *Chlorella zofingiensis* to different culture conditions. The antioxidant activities measured by the DPPH and FRAP assays reflect the total carotenoid and phenolic contents of the microalgae under normal and stress conditions. Phenolic compounds were produced at higher levels under normal conditions, while total carotenoids were produced at the same degree under both normal and stress conditions. This indicates that different metabolites were produced under different culture conditions. In addition, the levels of antioxidants in both *Chlorella* species also differed with regard to the phenolic and carotenoid contents. Hence, future studies profiling the carotenoids and polyphenols using HPLC and LC-MS are of high priority, as these works will improve the understanding of the detailed changes of these important metabolites under the stress related condition. Thus, the manipulation of microalgae culture condition is very important for the production of desired metabolites in a specific

microalgae species, which would affect their commercial application, especially the downstream processes, as different products require different processing methods and system setups.

CHAPTER 4

DE NOVO TRANSCRIPTOME ANALYSIS OF *Chlorella sorokiniana* IN RESPONSE TO STRESS CONDITION

4.1 Introduction

Phenotypic details surrounding the influence of stress on the production of pigments and metabolites that contributed to microalgae survival can be understood by performing comparative transcriptome profiling (Fu et al. 2014; Sun et al. 2015). Transcriptome would generate a comprehensive overview of what genes are active at various physical, biochemical, and developmental conditions (Adams 2008). Thus, by collecting and comparing transcriptomes of different types of cells or tissues, researchers can gain a deeper understanding of what constitutes a specific cell type and how changes in transcriptional activity may reflect the production of different metabolites in microalgae.

The applications of NGS technology have accelerated the research progress on new microalgae species by providing an improved understanding of essential components in a particular cellular process. Full transcriptome analysis using NGS technology is particularly powerful since it can provide a global overview of the response of *Chlorella* cultures to a various environmental states. It can potentially link the expression of specific sets of transcripts with phenotypic response. This emerging tool had been widely applied to investigate various study such as comparative analysis of autotrophic and heterotrophic growth of *A. protothecoides* (Gao et al. 2014), analysis of carbonic anhydrase expression of *C. pyrenoidosa* under salt stress (Sun et al. 2014), analysis of starch and lipid synthesis in *C. pyrenoidosa* (Fan et al. 2015), and analysis of *C. sorokiniana* growth under high carbon dioxide concentrations (Sun et al. 2016).

In the previous chapter, the changes on biochemical contents and the morphology of microalgae *Chlorella* under normal and stress conditions were examined. From the results obtained, it was clearly seen that the production of secondary carotenoid pigments contribute to a higher scavenging activity of DPPH analysis when cultured under stress condition. Thus, in order to understand how the stress (light and glucose under nutrient limited condition) induces *C. sorokiniana* to synthesis carotenoid and lipid that have antioxidant activitiy, *de novo* transcriptome analysis was used to identify the genes and their expressions involved in the biosynthesis pathways of carotenoids and lipids.

So, the purpose of this study was therefore to investigate the genes involved in regulating the production of carotenoids and lipids in *Chlorella sorokiniana* when cultured under high light intensity and glucose addition during nitrogen-limited condition through RNA-sequencing technique. In this study, six

transcriptome datasets from normal and stress conditions were *de novo* assembled, annotated, and differential genes expression was analyzed as well. Then, RT-qPCR was used to validate the expression level of genes based on the RNA-sequencing data.

4.2 Materials and methods

4.2.1 Experimental design

The same samples as mentioned in section **3.2.1** were used in this experiment. Briefly, the cells were first cultivated in Bold's Basal Medium (BBM) at 27°C under continuous light at a light intensity about 10 µmol photons m^{-1} s^{-1} with 30 rpm shaking speed. Then, the cultures were separated into two flasks, one for normal growth and the other one for stress induced growth. The normal growth flask was maintained under the same conditions as the initial culture. For induction of stress responses, the cultures were transferred to a higher light intensity at 100 µmol photons m^{-1} s^{-1} and 2% glucose (w/v) was added (Ip and Chen 2005; Sun et al. 2008). Both cultures from normal and stress induced conditions were harvested 7 days after inoculation. The total samples obtained from the experiments were 6 samples from each of *Chlorella* species; 3 samples of normal *C. sorokiniana* and 3 samples of stressed *C. sorokiniana*; and 3 samples of normal *C. zofingiensis* and 3 samples of stressed *C. zofingiensis*.

4.2.2 RNA extraction

Total RNA was extracted using TRIzol method. Briefly, 100 mg frozen tissue was ground using a prechilled mortar and pestle into a fine powder. 1 mL TRIzol reagent (Invitrogen, California) was added to the ground tissue and homogenisation continued until no visible debris remain. The homogenised samples were incubated at room temperature for 5 min. Chloroform with 0.2 times the volume of Trizol solution was added to the solution and the mixture was vortexed vigorously for about 15 sec and incubated at room temperature for 2 to 3 minutes. After centrifugation, the aqueous layer was transferred to a column of RNeasy from a mini RNA isolation kit (Qiagen, Germany) for further purification. The residual DNA was eliminated by performing a column DNase digestion at 37°C for 30 minutes. The integrity of the extracted RNA was determined by gel electrophoresis and its concentration was measured using a biospectrometer (Eppendorf, Germany).

4.2.3 Determination of RNA concentration and quality

4.2.3.1 Determination of RNA concentration using BioSpectrometer

The concentration and quality of RNA preparation were determined by biospectrometer (Eppendorf, Germany) at a wavelength of 260 nm (A_{260}). The degree of contamination in the samples was estimated by measuring the A_{260}/A_{280} ratio and A_{260}/A_{230} ratio. A_{260}/A_{280} ratios higher than 1.8 are indicative

of limited protein contaminations, whereas low A_{260}/A_{230} ratios are indicative of residual contamination by organic compounds such as phenol, sugars or alcohol (Fleige et al. 2006).

4.2.3.2 Determination of RNA concentration using Qubit 2.0 Fluorometer

The concentration of RNA measured using Qubit 2.0 Fluorometer is based on the binding of fluorescent dyes to the RNA molecule. In this study, Qubit RNA HS (High Sensitivity) assay kit (Life Technologies, Germany) was used to determine the concentration of RNA in samples. Briefly, the RNA samples were diluted to a final concentration between 250 pg/µL and 100 ng/µL using the buffer provided before loaded to Qubit 2.0 Fluorometer.

4.2.3.3 Determination of RNA integrity using BioAnalyser

The extracted RNA was analysed using the Agilent High sensitivity DNA assay kit on the 2100 bioanalyzer (Agilent Technologies, Germany) in compliment with the RNA 6000 Nano LabChip kit. The assays used were Eukaryote total RNA Nano for all the normal and stress samples with three biological replicates. Data analysis was performed in accordance with the Agilent protocol.

4.2.4 RNA-seq library construction for Illumina HiSeq2000

The library for RNA-sequencing was prepared using NEBNext Ultra Directional RNA Library Prep kit for Illumina (Cat. No. E7420) from New England Biolabs (NEB, UK). The libraries were prepared following the instruction and recommendation of the manufacturer's manual. In the first step, 1 µg mRNA was isolated from extracted total RNA using NEBNext Oligo $d(T)_{25}$ beads, followed by fragmentation and priming using NEBNext First Strand Synthesis Reaction Buffer (Cat. No. E7421A) and Random Primer mix (Cat. No. E7422A). The fragmented RNA was reverse-transcribed with ProtoScript II Reverse Transcriptase (Cat. No. E7423A) at 25°C for 10 min followed by 42°C and 70°C for 15 min each reactions in the presence of Murine RNase Inhibitor (Cat. No. E7424A) and 0.1µg/µL actinomycin D to inhibit the generation of second-strand products.

The second-strand cDNA was synthesized using NEBNext Second Strand Synthesis Enzyme Mix (Cat. No E7425A) and its reaction buffer (Cat. No. E7426A) for 1 hr at 16°C in the presence of dUTPs. The end repair of double-strand cDNAs was carried out with NEBNext End Prep Enzyme Mix (Cat. No. E7371A) with incubation at 20°C and 65°C for 30 min each. After that, the second-strand of the library construct was ligated to sequencing adapters for Illumina by Blunt/TA Ligase Master Mix (Cat. No. E7373A) and diluted NEBNext Adaptor (Cat. No. E7335) at 20°C for 15 min. Following adapter ligation, the second strand library construct was enriched by PCR amplification using NEBNext High-Fidelity PCR Master Mix for 15 cycles.

The purification of reaction products between each step was performed with AMPure XP Beads (Beckman Coulter, UK). The molarity and size of the libraries was assessed by Agilent high sensitivity chip on 2100 Bioanalyzer (Agilent Technologies, Germany). Sequencing was performed on the Illumina HiSeq 2000.

4.2.5 RNA-seq data analysis

RNA-seq data analysis begins with the input of the raw read files and the reference files. However, the *Chlorella* species used in this study does not have a reference genome, thus *de novo* assembly of data was carried out and further used to construct the transcriptome. The RNA-Seq data analysis workflow used in this study is illustrated in **Figure 4.1**.

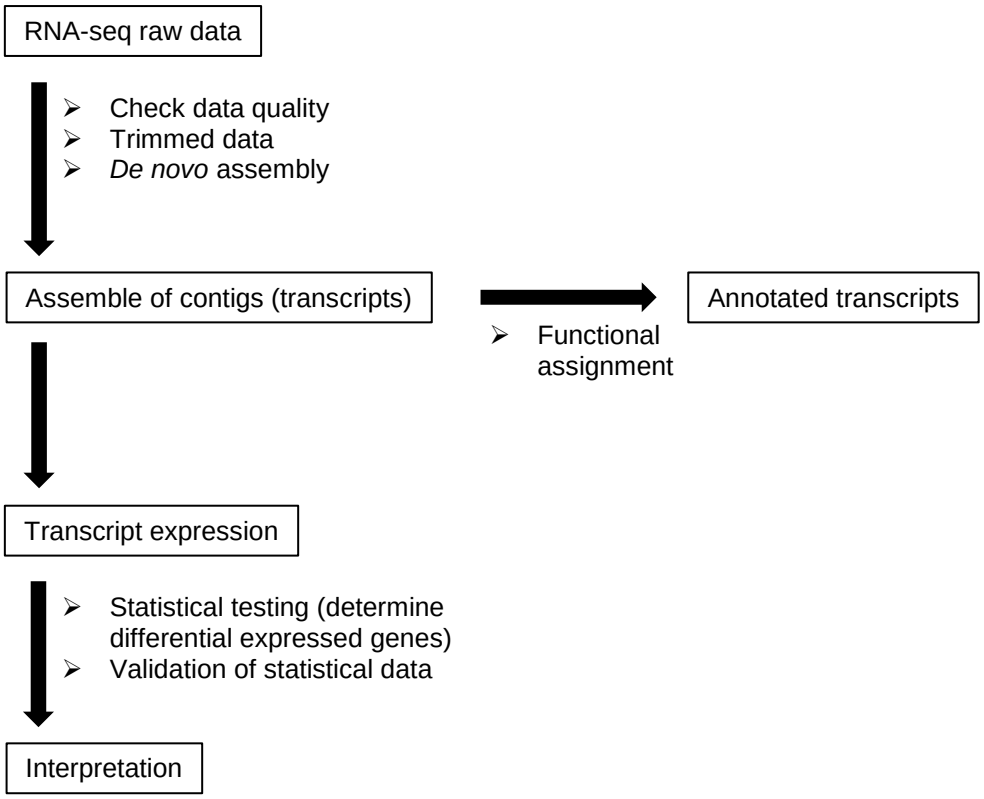


Figure 4.1: RNA-Seq data analysis workflow used in this study (Yang et al. 2013)**.**

4.2.5.1 Quality assessment and pre-processing

First, RNA-Seq reads of the samples were screened for good quality reads through FastQC. FastQC is Java software that provides tools to perform a QC study in raw high throughput sequence data. FastQC interprets the raw sequence data in terms of graphs and tables to assess the data, which are then transformed into an HTML report. The analysis performed by FastQC on RNA-seq data were per base sequence quality, per sequence quality scores, per base sequence content, per base GC content, per base N content, sequence length distribution, sequence duplication levels, overrepresented sequences and Kmer content. These analyses served to confirm that the data was a sufficient quality and that there were no problems or inappropriate biases.

If there were some abnormalities found within the raw data, data can be trimmed in silico using CLC-Bio Workbench (Version 6.0.1, CLC Bio, Denmark) (Gallardo-Escárate et al. 2014). This was carried out to remove the abnormalities region of 100 bp paired-end reads such as low-quality reads, adaptor sequences, and poor-quality bases. The trimming program was set as follow: removal of low quality score below 0.001 (Q30); removal of ambiguous nucleotide with maximum 2 nucleotides; removal of 1 nucleotide from both 3' and 5' end; and search the adapter at both strand using specified index. Once the trimmed data were obtained, the reads were processed again using FastQC to ensure the data meet the necessary quality standard for further processing.

Then, the filtered paired-end reads of each sample were pooled, separated into their respective orientation and collapsed further to retain only the unique non-redundant reads using the FASTX-toolkit (http://hannonlab.cshl.edu/fastx_ toolkit/).

4.2.5.2 *De novo* assembly and alignment

After trimming the raw sequence reads, Illumina sequences (in FastQ format) were used for *de novo* assembly by Trinity (version 2.0.6) on default settings (Grabherr et al. 2011a). Using Trinity software, the RNA-seq reads were processed using three independent software modules: Inchworm, Chrysalis and Butterfly which distributes the sequences data into many individual de Bruijn graphs. After *de novo* assembly of a reference transcript and/or the reference, transcriptomes were obtained and the samples of RNA-seq reads were aligned back to the reference transcript using Burrows-Wheeler Aligner (BWA) (Li and Durbin 2010). The BWA default values for mapping were used, except for number of threads (-t)=8 and maximum number of alignments (sample -n)=40. Under these settings, read pairs mapping to multiple equally best positions are placed randomly. Properly paired reads with a mapping quality of at least 20 (-q=20) were selected.

Gene predictions and protein coding potential of the *de novo* assembly transcripts were determined using AUGUSTUS (version 2.5.5) (Stanke et al. 2006). *Chlamydomonas reinhardtii* genes and protein sequences, as well as gene structure were used to train the AUGUSTUS gene models to get more accurate ORF prediction. In this case only predicted genes protein-coding (contains Met; a start codon Methionine) genes were allowed.

The alignment of filtered paired-end reads against the predicted protein-coding genes was performed using Subread aligner (Liao et al. 2013) and counting of reads (read summarization) was performed with Feature Counts (Liao et al. 2014) with default settings.

4.2.5.3 Gene annotations

General annotation of the transcriptome assemblies was performed using the Mercator pipeline (Lohse et al. 2014) using the default settings with an additional mappings against the JGI *Chlamydomonas* reference database and InterPro domains and MapMan BINs (Thimm et al. 2004) functional category assignments. Moreover, the transcriptome we also annotated with other database resources such as the non-redundant (nr) protein database, SwissProt or UniProt, Refseq, KEGG (Kyoto Encyclopedia of Genes and Genomes) and GO (Gene Ontology). Two BLAST programs were used to align contigs obtained from the assemblies: BLASTX for protein sequences and BLASTN for nucleotide sequences. The significant hits were been filtered with an E-value threshold of 10^{-5}.

4.2.5.4 Differential expression analysis

Differential gene expression analysis was carried out with DESeq2 (Love et al. 2014). The transcript abundance was estimated using DESeq2 using the variance stabilization normalization (VST). Genes were deemed significantly different between treatments at an FDR < 0.05. Enrichment of functional categories (MapMan BINs) was determined by Fisher's exact test adjusted with Bonferonni correction for multiple testing corrections using the Corto tool (Giorgi et al. 2013). MapMan BIN categories were deemed significantly enriched at an adjusted P-value < 0.05. Genes with adjusted P-value (FDR) ≤ 0.05 and fold change ≥ 2 were regarded as differentially expressed, which represents a 100% linear fold change i.e. $\log_2 2 = 1$ or 100%. Alignment, read summarization, statistical calculations, and graphing were performed in R (https://www.r-project.org/).

4.2.6 Real-time reverse transcription PCR

Quantitative real-time PCR (RT-qPCR) was carried out to validate the results of the RNA-seq. Total RNA was extracted using TRIzol method (**Section 4.2.2**). Specific primers were designed from the sequences of contigs obtained from

RNA-seq analysis. **Table 4.1** and **Table 4.2** list all genes (housekeeping and target) used in the RT-qPCR analysis.

4.2.6.1 cDNA synthesis

The reverse transcription step for the preparation of the cDNA library was performed using SensiFAST cDNA Synthesis Kit (Bioline, UK) according to the manufacturer's instructions. Three-hundred nanograms of the RNA template from each sample were converted into cDNA in 20 µL. The control was prepared by using a similar amount of total RNA that subjected to the cDNA synthesis reaction without the presence of reverse transcriptase enzyme. A 300 ng cDNA template pool produced from the reverse transcription reaction was subsequently used to asses housekeeping genes and target genes transcript levels in real-time PCR assays.

4.2.6.2 Selection of housekeeping gene for RT-qPCR studies

10 housekeeping genes that are most commonly used in algal gene expression analysis were chosen for our study from the literature (Chung et al. 2008; Fischer et al. 2007; Boldt et al. 2008; Rosic et al. 2011; Dong et al. 2012; Guo and Ki 2012; Liu et al. 2012a; Shi et al. 2013; Adelfi et al. 2014). The primers and details of each housekeeping genes are presented in **Table 4.1**. These genes were subjected to homology search in *C. sorokiniana* transcriptome, and their homologs were used for primer design. The functional of the obtained sequences were checked using BLASTN search against GenBank EST database. Primer pairs were designed using Primer Designer program in Clone Manager software (Clone Manager 7, Version 7.11, Sci Ed Central). The amplicon size ranged from 143 bp for RPL19 and 308 bp for SAM gene.

Table 4.1: Details of primers and amplicon for housekeeping gene from *C. sorokiniana* transcriptome.

No.	Housekeeping gene	Primer sequence (5' - 3')	Amplicon size (bp)
1.	Ribosomal protein L19 (RPL19)	Fwd: GTCTGGCTGGACCCCAATGA Rev: GCCTCTGCTGCGGTGC	143 bp
2.	Beta-Tubulin (B-tub)	Fwd: AAATCGTTCACGTTCAGGCCG Rev: GAACACTTGGCCGTAGGGG	248 bp
3.	Alpha-tubulin (A-tub)	Fwd: CACAGTTTACCCGTCTCCCC Rev: TCAGCCGGTTAAGGTTGGTG	185 bp
4.	Ubiquitin (UBQ)	Fwd: CGAGGGGGCATGCAGA Rev: ATTGCCACCACGAAGACGAA	240 bp
5.	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	Fwd: GGCAAGCCGATCAAGGTG Rev: CATCCTGGGTCTTGGAATCGT	229 bp
6.	Calmodulin (CAL)	Fwd: GGATGGGAATGGGACTATCG Rev: CCTCGTCGGTGAACTTCTTG	191 bp
7.	Cyclin dependent kinase A (CDKA)	Fwd: AGCGCTACTACGACGACGTG Rev: GGACGATGCTTGGCGTAGTG	157 bp
8.	Heat shock proteins 90 (HSP90)	Fwd: GGGCAAGAGTCGCAACATCC Rev: TGTCCTCACGGTCCGCAATC	236 bp
9.	Proliferating cell nuclear antigen (PCNA)	Fwd: TGGGCCTGAACCTGAACAAC Rev: TTGATGCCCTCCTTCGTCAC	301 bp
10.	S-adenosyl methionine synthetase (SAM)	Fwd: TACCCGTACATGGCCAACCC Rev: TGGCCCTGCTGGTAGAACTG	308 bp

4.2.6.3 Gene expression validation using RT-qPCR

In order to validate the gene expression levels obtained from RNA-seq study, several genes were randomly chosen for RT-qPCR analysis. The primers for the selected genes are presented in **Table 4.2**. The RT-qPCR reactions were performed using KAPA SYBR FAST one-step RT-qPCR Kit (KAPA Biosystem, USA). Three independent biological replicates from normal and stress samples were analysed for each gene. The primers for each gene were designed using Primer Designer program in Clone Manager software (Clone Manager 7, Version 7.11, Sci Ed Central). The cycling program was set at 95°C for 5 min, 40 cycles of 95°C for 20 sec (denaturation) and 60°C for 20 sec (annealing). The relative expression levels of the genes were quantified using CFX Manager 3.1 software (BioRad, USA). The expression levels of the targeted genes were normalized to the selected housekeeping genes as described by Vandesompele et al. (2002).

Table 4.2: Details of primers and amplicon for target gene from *C. sorokiniana* transcriptome.

No.	Target gene	Primer sequence (5' - 3')	Amplicon size (bp)
1.	Stearyl-ACP desaturase, Δ9-D (DESA1)	Fwd: TGCTGGGTCGCTTCCTCTAC Rev: CCGTTGTTCAGCTCCGAGTG	397 bp
2.	Malate dehydrogenase (MDH)	Fwd: TCACCAAGGAGGAGATTGAG Rev: CGCCTGTCTTAATGTTGGTC	342 bp
3.	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD)	Fwd: CTGGTGGCAATTCACGACTC Rev: TCAGTGACCTCCAGCTTCTC	263 bp
4.	Ribose 5-phosphate isomerase A (rpiA)	Fwd: GCGAGAAGATGGTGGAGATG Rev: GGCGTCTCAAAGTACAGGTC	253 bp
5.	Acetyl Coa carboxylase (ACACA)	Fwd: GCATGGCCGACCAGTTTGTG Rev: CCGGTCGTAGACATCAGGTG	347bp
6.	TAG lipase (TagL)	Fwd: GCTGCTGGAACACCAGATGC Rev: GGCGCTGTCAGGTAGTTGAG	357bp
7.	1,4-glucan branching enzyme (glgB)	Fwd: GAGCGCGATGACTTTGGCAC Rev: TGGCGTTGTAGCCCAGCTTC	358 bp
8.	Phosphofruktokinase (pfkA)	Fwd: CAAGCCCATCACCCTTACCC Rev: CTCCACAGCCGTCTCAAACC	298 bp

4.3 Results and discussion

4.3.1 Library preparation

The sequencing libraries with an average distance size approximately 300 bp were successfully constructed (**Figure 4.2**). The samples were send for sequencing using Illumina platform on HiSeq2000.

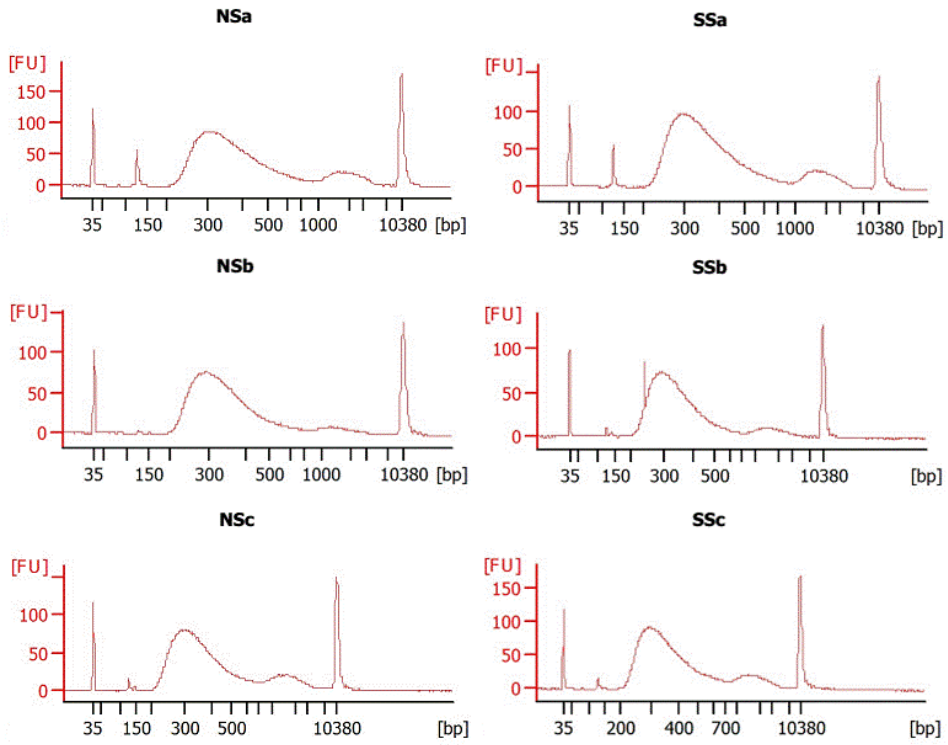


Figure 4.2: Assessment of library quality on Bioanalyser using Agilent DNA high sensitivity chip. The electropherogram shows a DNA distribution with a peak size approximately 300 bp. Normal samples are: NSa, NSb and NSc, while Stress samples are SSa, SSb and SSc. For normal samples, NS stand for *C. sorokiniana* in normal condition, while a, b and c are biological replicates. For stress sample, SS stand for *C. sorokiniana* in stress condition, while a, b and c are biological replicates.

4.3.2 Sequencing and assembly

A total of 198,844,110 raw sequence reads was obtained from both normal and stress induced samples of *C. sorokiniana* with three biological replicates each. Then, these raw reads were analysed using FastQC to check the quality score across all bases from the sequencing data. The raw data were next processed to identify adapter contamination and other abnormalities using trimming

programme on CLC-BIO genomic workbench. A total of 11,338,204 reads which corresponds to 5.7% of total sequence were trimmed using an exact matching strategy. After this cleaning step, a total of 187,505,906 (94.3%) “high quality” clean reads with an average length of 75 bp was obtained from three biological replicates of each samples (**Table 4.3**).

Meanwhile, for *C. zofingiensis*, a total of 215,171,544 raw sequence reads was obtained from both normal and stress samples including each replicates. After removing adapter and other abnormalities, a total of 201,081,012 clean reads with an average length of 84 bp was obtained from three biological replicates of normal and stress *C. zofingiensis* samples. All cleans reads from *C. sorokiniana* and *C. zofingiensis* were assembled *de novo* using Trinity assembler (Grabherr et al. 2011b). However, *C. zofingiensis* data could not be used for further analysis due to difficulties in the assembly steps. According to Roth et al. (2017) who successfully assembled the chromosome level genome and transcriptome of *C. zofingiensis* reported that, this microalga has complex transcriptome, gene order and strands within blocks are scrambled. Thus, in the subsequent sections only *C. sorokiniana* result will be presented and discussed.

The assembly program for *C. sorokiniana* produced 18,310 assembled transcripts (or unique consensus sequences (contigs)) with a contig read length ranging from 165 to 16,695 bp with an average and median length of 1191 bp and 990 bp, respectively. From these transcripts, AUGUSTUS determined that 9341 transcripts from the transcriptome were successfully annotated while the rest might be false prediction or that they are unique to the *Chlorella* species. From total, only 1850 transcripts were significant with FDR ≤ 0.05, that include 933 annotated transcripts. Compared to other *Chlorella* that had been sequenced, the predicted transcript of this *C. sorokiniana* was close to *C. variabilis* transcript, which predicted about 9791 protein-encoding-genes (Eckardt 2010). Generally, the genomic size of sequenced chlorophyte green algae (such as *Chlorella protothecoides* sp.0710 (Gao et al. 2014), *Chlorella variabilis* NC64A (Blanc et al. 2010), *Coccomyxa subellipsoidea* C-169 (Blanc et al. 2012), *Chlamydomonas reinhardtii*, *Volvox carteri*, *Micromonas pusilla* CCMP1545 and *Ostreococcus tauri*) were ranged between 12 Mb to as high as 138 Mb. The transciptome represents a small part of the genome. In complex organism, the non-protein-coding gene that were transcribed were more complex, in which they produce more than one variant of mRNA due to an alternative splicing, RNA editing, and alternative transcription initiation and termination. Therefore, the transcriptome captures a level of complexity that the simple genome sequence does not (Adams, 2008).

Table 4.3: Sequencing throughput and trimming results for *C. sorokiniana* RNA-seq data. Two samples: Normal and Stress, with three biological replicates of 100 bp reads were generated on the Illumina HiSeq 2000 platform.

Sample	Total sequences		Average length	
	Before	After	Before	After
Normal replicate 1	39,445,456	37,216,505	100	73
Normal replicate 2	26,450,502	25,428,583	100	79
Normal replicate 3	33,209,952	30,146,997	100	74
Stress replicate 1	37,433,950	36,094,256	100	75
Stress replicate 2	33,381,782	31,635,371	100	75
Stress replicate 3	28,922,468	26,984,194	100	75
Total	198,844,110	187,505,906	100	75

Table 4.4: Summary of *de-novo* assembly result for *C. sorokiniana* RNA-seq data.

Subject	Number of reads	Length (bp)
Total number of high quality assembled reads	104,487,087	
Number of transcripts (contigs)	18,310	
N50 (bp)		1,446
N90 (bp)		654
Longest transcript (bp)		16,695
Number of transcripts >5 Kb	82	
Number of transcripts >10 Kb	8	

4.3.3 Annotation of gene function

In order to assign accurate annotation information, the transcript sequence were annotated through multiple databases including; the NCBI database Non-redundant protein sequences (NR), the manually annotated and curated protein sequence database (Swiss Prot), InterPro, Kyoto Encyclopedia of Genes and Genomes Ortholog database (KO), Gene Ontology (GO), Mercator Mapman, and RefSeq. Annotation results of unigenes are shown in **Table 4.5**. The result obtained from annotated to NR database showed the highest match in which 99.8% (or 18,276 genes) of *C. sorokiniana* unigenes had similar identity with sequence published in NR database. There were 50% to 55% matches when the assembled unigenes were annotated to the other databases such as InterPro, Swiss Prot, Mercator Mapman and Refseq. Whereas, annotation to KEGG and GO databases produced 38.1% and 43.1% matches each.

Table 4.5: Summary of the functional annotation of the assembled *C. sorokiniana* transcriptome using different databases.

Public Database	No. of annotated genes	% of annotated genes
NR	18276	99.8
InterPro	10019	54.7
UniProt/Swiss Prot	10259	56.0
KO	6971	38.1
GO	7896	43.1
Mercator Mapman	9341	51.0
Refseq	10187	55.6

Based on the NR annotation result, the E-value frequency distribution analysis revealed that 70.3% of the matched sequences had strong homology with E-value ≤ 1.0E-60, while the remaining 29.7% fell into range of 1.0E-60 to 1.0E-5 (**Figure 4.3**). Furthermore, all the annotated sequences had similarity more than 70%. Based on the homologous species identified among the annotated unigenes, 71.6% of the unigenes sequence matched to *Chlorella variabilis*, followed by *Auxenochlorella protothecoids* (5.5%), *Chlamydomonas reinhardtii* (3.8%), *Coccomyxa subellipsoidea* (1.7%) and *Volcox carteri* (1.2%). About 16.3% of the unigenes had similarity matched with other types of viridiplantae.

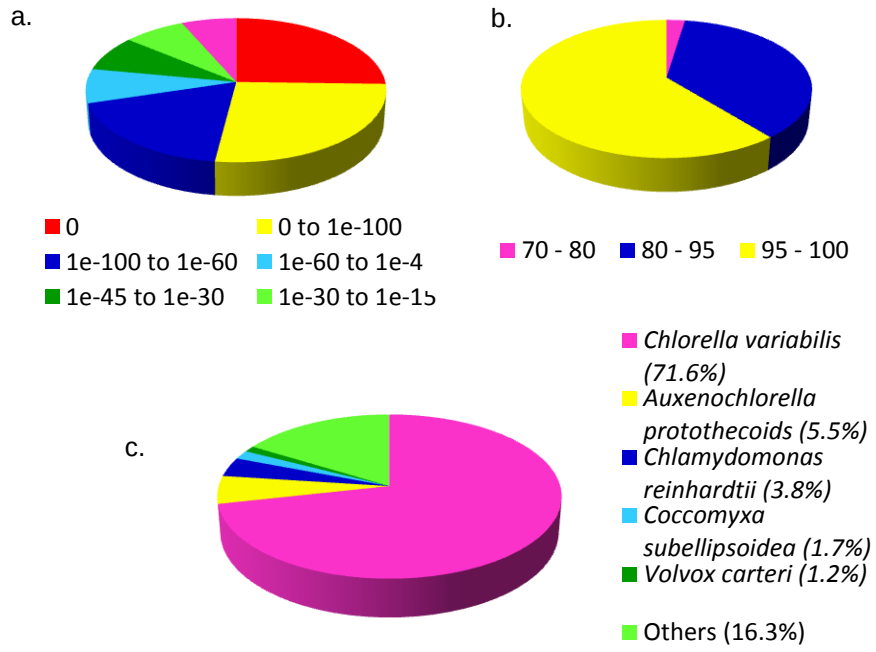


Figure 4.3: Transcript homology searches against the NR database. a) The proportional frequency of the E-value distribution. b) The proportional frequency of the sequence similarity distribution. c) The proportional species distribution of *C. sorokiniana* transcriptome among other viridiplantae.

In the subsequent functional analysis, the functional annotation was done based on the KO (KEGG Orthology) database using KOALA tools (at KEGG Web site; http://www.kegg.jp) that assign the KO identifiers (K numbers) to the transcript by BLAST and GHOSTX searchers. In this study, of 18,310 transcripts that aligned with KO database, only 6971 transcript (38.1%) were annotated and had significant matches to different functional categories (**Figure 4.4**). The highest number of genes identified was from genetic information processing category, which are 3007 genes (43%). The second highest functional categories of unigenes identified are carbohydrate metabolism (637 genes), followed by environmental information processing (563 genes), amino acid metabolism (377 genes), lipid metabolism (329 genes), energy metabolism (304 genes), cellular processes (311 genes) and metabolism of cofactor and vitamins (231 genes). The remaining 17% of KEGG annotated unigenes were from other categories.

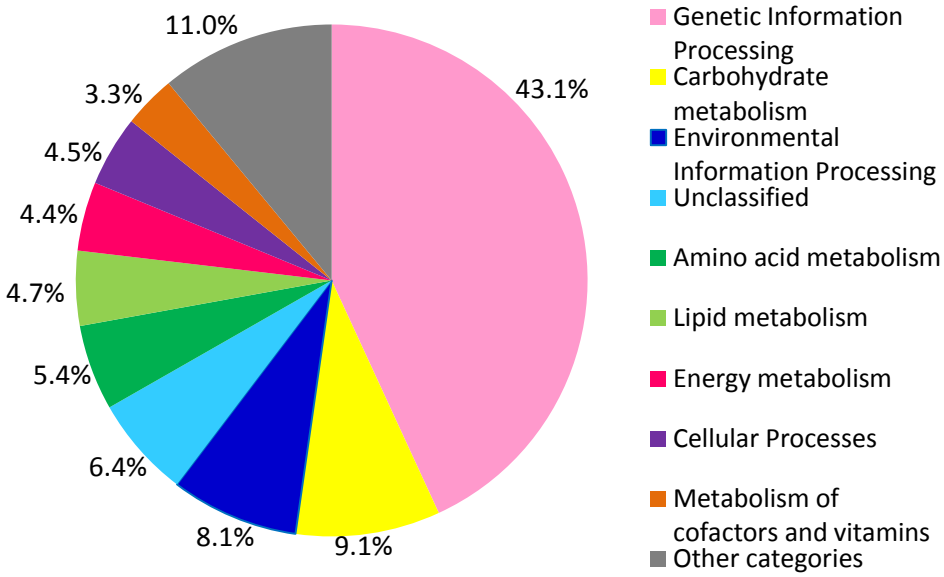


Figure 4.4: Annotation of *C. sorokiniana* transcripts based on functional categories KO database using KOALA tools at KEGG Web site (http://www.kegg.jp/blastkoala/).

Under pathway reconstruction of KO, each transcript from different functional categories was further elucidated into different metabolic pathways. All the metabolic pathways were then divided into four main pathway modules such as metabolism, genetic information processing, environmental information processing and cellular processes (**Figure 4.5**). According to this pathway assignment, 5018 unigenes (27.4% of all unigenes) were classified to 23 KEGG pathways.

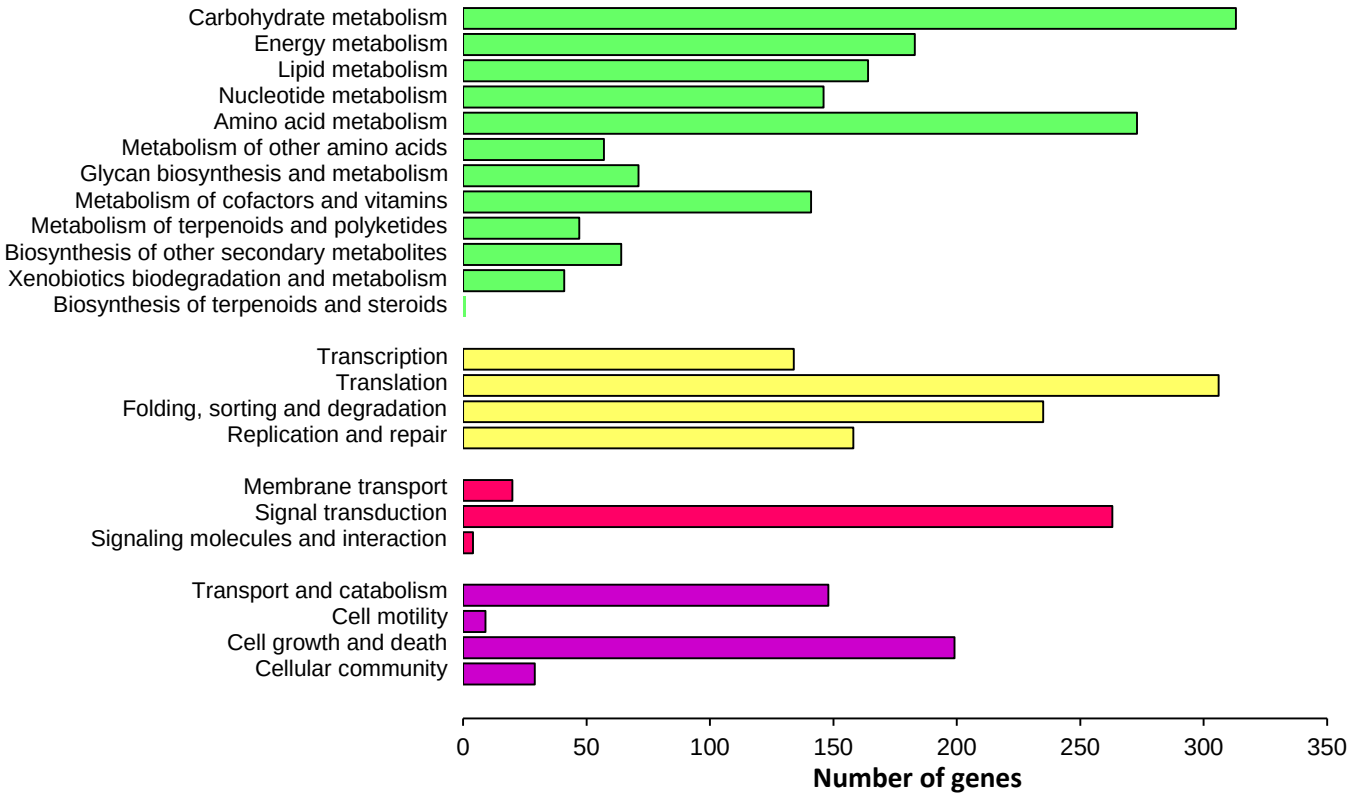


Figure 4.5: Functional classification and pathway assignment of unigenes by KEGG Orthology (KO). The results are summarized in five main pathways modules: Metabolism (green); Genetic Information Processing (yellow); Environmental Information Processing (red); and Cellular Processes (purple). The y-axis indicates the name of the KEGG metabolic pathways. The x-axis indicates the number of genes annotated under the pathway.

4.3.4 Pathway description and gene discoveries

From the annotation result, the transcripts coding for all the enzymes related to the major metabolic pathways in *C. sorokiniana* were identified. The completeness of these reconstructed pathways indicates that the gene function assignments were biologically meaningful and the EC number has been correctly assigned to the annotated sequences (**Table 4.6**). Other KEGG pathways with multiple transcript hits encoding for a nearly complete pathway include nucleotide and amino acid metabolism (such as serine and threonine biosynthesis, lysine biosynthesis, and histidine biosynthesis) and cofactor and vitamin biosynthesis (such as shikimate pathway, phenylalanine biosynthesis, coenzyme A biosynthesis and biotin biosynthesis).

Table 4.6: Essential metabolic pathways annotated in the *C. sorokiniana* transcriptome.

Pathway	Number of enzyme found	Number of known enzymes
Reductive pentose phosphate (Calvin-Bensen cycle)	11	12
Glycolysis/gluconeogenesis	10	10
Citrate cycle (TCA cycle)	9	10
Fatty acid biosynthesis	6	6
Triacylglycerol (TAG) biosynthesis	4	4
Starch biosynthesis	4	4
Isoprenoid biosynthesis (MEP pathway)	7	7

Chlorella has been known as one of the best microalgae for production of biofuel as well as for biopharmaceuticals product (Petkov and Garcia 2007; Mostafa 2012). Several researchers also reported that manipulation of different environmental factors could affect biochemical production from microalgae (Vonshak and Torzillo 2004; Hu 2004; Juneja et al. 2013; Fan et al. 2014). In the previous chapter, *C. sorokiniana* showed several fold increase in carotenoid production when cultured under stress-related conditions, which demonstrate the existence of a carotenoid biosynthesis pathway in this microalgae. Besides, the accumulation of lipid bodies and starch granule from microscopic results also proved that this microalga can be an alternative source for biofuel production. Although metabolic transition had been identified in the accumulation of pigment, lipid and starch in *Chlorella* species (Li et al. 2015), lack of genetic information and differential expression of key metabolic pathway hampered the manipulation and improvement of this microalga. Elucidation of these important metabolic pathways (such as carotenoid, lipid and starch) in this microalga could facilitate genetic modification to improve its productivity and reduce the reliance on model organism. Thus, among several essential metabolic pathways identified from the annotation results, pathways related to carotenoid, lipid and starch biosynthesis will be described further.

4.3.4.1 Carotenoid biosynthesis

Carotenoids are among the best-known antioxidant originating from microalgae and play an important role in protecting the microalgal system. These pigment molecules directly quench singlet oxygen, thereby preventing free radical reactions (Vachali et al. 2012; Safafar et al. 2015). Previous studies have revealed that carotenoids contribute significantly to the total antioxidant capacity of microalgae (Takaichi 2011; Goiris et al. 2012). Thus, microalgae have become an alternative source of carotenoids such as astaxanthin from *Haematococcus* and β-carotene from *Dunaliella* that are used in the food and pharmaceutical industries (Spolaore et al. 2006). Previous studies have proved that *Chlorella* strain can be better carotenoid producer compared to the current producers (Del Campo et al. 2004; Ip et al. 2004; Ip and Chen, 2005). Several carotenoids that can be found in microalgae are lycopene, β-carotene, α-carotene, and their derivatives such as lutein, astaxanthin, fucoxanthin and zeaxanthin (Takaichi, 2011).

Although microalgae had common carotenoid biosynthesis pathway with land plant, little is known about microalgal-specific carotenoid pathway. Some of carotenogenesis enzymes and genes from microlage that were reported in the literature are CrtB (phytoene synthase), CrtP (diapolycopene oxygenase), CrtL-b (lycopene cyclase), CrtR-b (β-carotene hydroxylase), ZEP (zeaxanthin epoxidase), VDE (violaxanthin de-epoxidase), and CrtW (β-carotene ketolases) found in *Chlorella*, *Chlamydomonas*, *Dunaliella* and *Haematococcus* (Takaichi, 2011). Pathway analysis and sequence information of specific gene will faciltitate further understanding and exploitation of carotenoid biosynthesis in microalgae. In this study, several important genes and enzymes involve in carotenoid biosynthesis pathway were found in *C. sorokiniana* transcriptome. Besides lutein production, this *C. sorokiniana* also has potential to produce other pigments, such as β-carotene, astaxanthin and their derivatives (**Figure 4.6**).

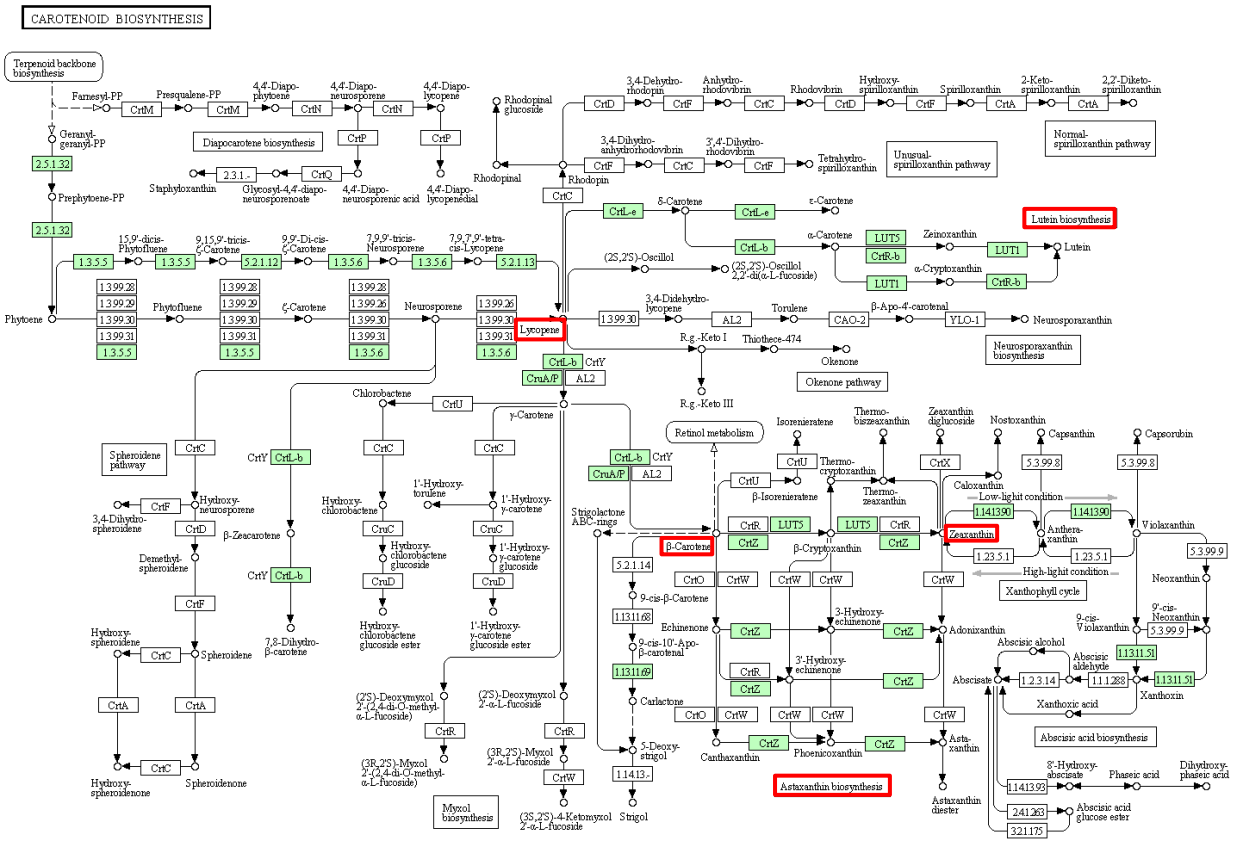


Figure 4.6: Carotenoid biosynthesis pathways reconstructed based on annotation of *C. sorokiniana* transcriptome to the KAAS (or KEGG database). The green box indicates the gene encoding enzyme of particular step was identified in the *C. sorokiniana* transcriptome. Red box shows pigments that can be produced from *C. sorokiniana*.

Based on the KEGG pathway assignment of the functionally annotated sequences, 14 enzymes involve in biosynthesis of different pigments were successfully identified in *C. sorokiniana* transcriptome (**Table 4.7**).

Table 4.7: Enzymes involved in the carotenoid biosynthesis pathway identified by annotation of *C. sorokiniana* transcriptome.

No.	Enzyme	EC number	Enzyme name	Gene name	Transcript found
1.	Phytoene synthase	EC 2.5.1.32, EC 2.5.1.99	PSY	crtB	1
2.	15-cis-phytoene desaturase	EC 1.3.5.5	PDS	crtP	2
3.	ζ-carotene isomerase	EC 5.2.1.12	ZISO	Z-ISO	1
4.	Prolycopene isomerase	EC 5.2.1.13	CrtISO	crtH	1
5.	β-carotene 3-hydroxylase	EC 1.14.13.129	CrtR-b	crtZ	1
6.	β-ring hydroxylase	EC:1.14.-.-	LUT5	CYP97A3	1
7.	Lycopene cyclase CruA	EC 5.5.1.19	CruA	cruA	1
8.	Lycopene β-cyclase	EC 5.5.1.19	CrtL-b	crtY	2
9.	Lycopene ε-cyclase	EC 5.5.1.18	CrtL-e	crtL2	1
10.	Carotene ε-monooxygenase	EC 1.14.99.45	LUT1,	CYP97C1	1
11.	Zeaxanthin epoxidase	EC 1.14.15.21	ZEP	ABA1	1
12.	9-cis-epoxycarotenoid dioxygenase	EC 1.13.11.51	NCED	-	1
13.	Carlactone synthase	EC 1.13.11.69 EC 1.13.11.70	CCD8	-	1
14.	ζ-carotene desaturase	EC 1.3.5.6	ZDS,	crtQ	2

In order for these carotenoid biosynthesis pathways start producing pigments, the synthesis of important precursor from non-mevalonate (MEP/DOXP) pathway is very important. In this case, all the key enzymes involve in the synthesis of important precursors for pigments biosynthesis, which are isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (**Figure 4.7**) were identified. The non-identified enzyme for mevalonate kinase (MVK, EC 2.7.1.36) confirmed that this pathway occurs in the plastids, not in the cytoplasm, which is similar the other photosynthetic green algae (Paniagua-Michel et al. 2012). This pathway starts with condensation of pyruvate and glyceraldehyde-3-phosphate which yields 1-deoxy-D-xylulose 5-phosphate (DOXP) by DOXP synthase (Dxs, EC 2.2.1.7). The subsequent step is transformation of DOXP to 2-C-methyl-D-erythritol 4-phosphate (MEP) by DOXP reductase (Dxr, EC 1.1.1.267). After that, MEP is converted into cyclic diphosphate by IspD (EC 2.7.7.60), IspE (EC 2.7.1.148) and IspF (EC

4.6.1.12). The final steps of the non-mevalonate pathway are catalysed by IspG (EC 1.17.7.1, EC 1.17.7.3) followed by IspH (EC 1.17.1.2) which producing IPP and DMAPP (Wanke et al. 2001; Zhao et al. 2013). Furthermore, the transcript for enzyme that code for isopentenyl-diphosphate delta-isomerase (idi, EC 5.3.3.2) which an important checkpoint for isoprenoid biosynthesis that catalyses the interconversion of IPP and DMAPP (Berthelot et al. 2012) was also found in *C. sorokiniana* transcriptome. **Table 4.8** list all the enzyme involved in the non-mevalonate pathway and the transcripts found result from the annotation of *C. sorokiniana* transcriptome to KO database.

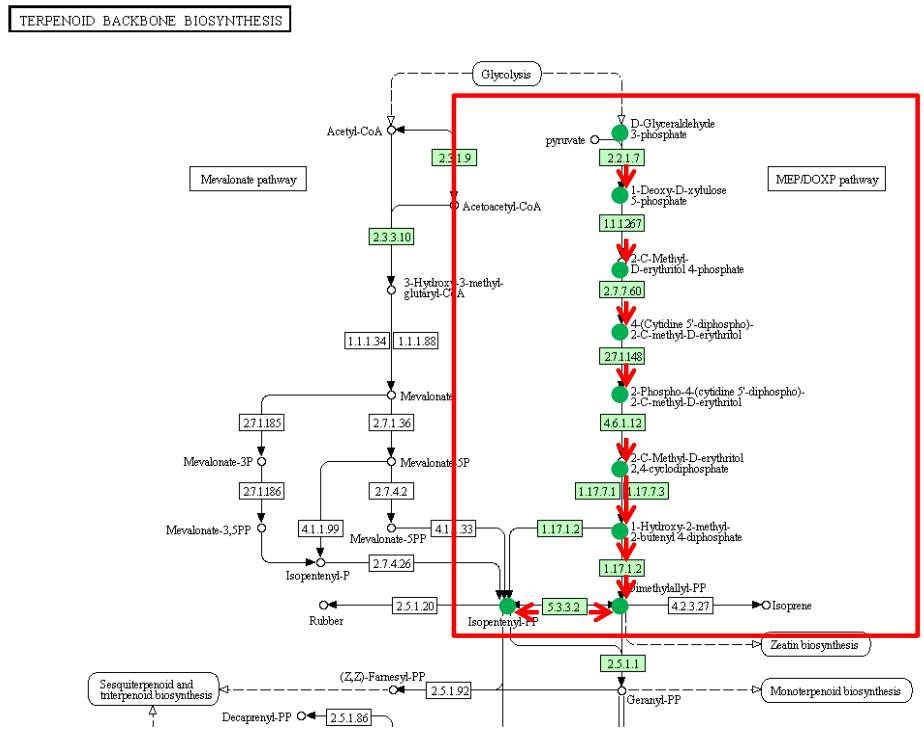


Figure 4.7: Biosynthetic pathways for the generation of the isoprenoid building units, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) (highlighted in red box) reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Green dot represents biomolecules, while green box highlighted the enzymes found in the *C. sorokiniana* transcriptome.

Table 4.8: Enzymes involved in the non-mevalonate pathway identified by annotation of *C. sorokiniana* transcriptome.

Step	Enzyme	EC number	Enzyme name	Gene name	Transcript found
1.	DOXP synthase	EC 2.2.1.7	Dxs	-	5
2.	DOXP reductase	EC 1.1.1.267	Dxr	IspC	2
3.	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	EC 2.7.7.60	YgbP	IspD	1
4.	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	EC 2.7.1.148	CDP-ME kinase, YchB	IspE	1
5.	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	EC 4.6.1.12	MEcPP synthase, YgbB	IspF	2
6.	HMB-PP synthase	EC 1.17.7.1 , EC 1.17.7.3	GcpE	IspG	1
7.	HMB-PP reductase	EC 1.17.1.2	LytB	IspH	1
8.	Isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	idi, IDI	-	1

Once, IPP was converted to geranyl-PP (GPP) by geranylgeranyl pyrophosphate synthase (GGPS, EC 2.5.1.29), geranylgeranyl pyrophosphate (GGPP) was produced and acted as the substrate for phytoene synthase (PSY, EC 2.5.1.32). Following the formation of phytoene (PE), four steps were needed to produce lycopene (LYC) which involved sequential catalysis of phytoene desaturase (PDS, EC 1.3.5.5), ζ-carotene isomerase (ZISO, EC 5.2.1.12), and prolycopene isomerase (CrtISO, EC 5.2.1.13). Lycopene is an important intermediate for production of various carotenoids such as lutein, β-carotene, zeaxanthin, cantaxanthin and astaxanthin. In this study the identification of transcript coding for lycopene β-cyclase (CrtL-b, EC 5.5.1.19) and lycopene ε-cyclase (CrtL-e, EC 5.5.1.18) sufficiently confirmed the presence of β-carotene and α-carotene in *C. sorokiniana*, which is similar in *Hematococcus pluvialis* (Steinbrenner and Linden 2001).

Although the ability for high lutein production by *C. sorokiniana* had been proved in previous study (Del Campo et al. 2004; Cordero et al. 2011), the details of lutein biosynthetic pathway was undetermined. In this study, lutein was produced from conversion of α-carotene by beta-ring hydroxylase (LUT5, EC:1.14.-.-), β-carotene 3-hydroxylase (CrtZ, EC: 1.14.13.129) and carotene ε-monooxygenase (LUT1, EC 1.14.99.45). On the other hand, β-carotene is hydroxylated by CrtZ and LUT5 to form zeaxanthin through β-cryptoxanthin. CrtZ which plays an important precursor for astaxanthin formation was found in this *C. sorokiniana* transcriptome suggesting that this microalga has the same ability with *C. zofingiensis* for astaxanthin production. Most of what is known about astaxanthin biosynthesis in algae comes from studies of *H. pluvialis*,

which lack of sequence genome. Biosynthesis of astaxanthin through carotenoid biosynthesis pathway had been reviewed by Rismani-Yazdi et al. (2011) and Li et al. (2014), however no reports found for chlorophycea algae. It is thought that β-carotene is exported from the chloroplast into lipid droplets in *H. pluvialis* where astaxanthin is synthesized by the introduction of two keto-groups catalyzed by a di-iron betaketolase (BKT), which is followed by the introduction of two hydroxyl groups catalyzed by a hydroxylase (CHYB) (Guarnieri et al. 2011). In this study, the annotation result of *C. sorokiniana* transcriptome suggests that astaxanthin is produced from hydroxylation of β-carotene to zeaxanthin (by LUT5 or cytochrome P450) and canthaxanthin (by CrtZ), which was similar to what had been hypothesized in *C. zofingiensis* (Li et al. 2014).

4.3.4.2 Fatty acid biosynthesis, elongation and degradation

The ability to accumulate significant amount of neutral lipids and rapid biomass production made microalgae as a potential sources of biofuel and other valuable products of much interest. In most microalgae, fatty acids are produced during optimal growth conditions for esterification of glycerol-based membrane lipids that consist of glycosylglycerides and phosphoglycerides, whereas, neutral triacylglycerols (TAGs) are produced during stress conditions. In the previous chapter, *Chlorella sorokiniana* showed sequential accumulation of starch and lipids when cultured under stress condition, which was similar to what had been reported in the literature (Li et al. 2015). This confirmed that, *C. sorokiniana* can be an alternative source for lipid biosynthesis production. Based on the functional annotation of *C. sorokiniana* transcriptome, genes that encode for the enzymes involved in the biosynthesis and degradation of fatty acids were successfully identified. List of enzymes involve in the fatty acid biosynthesis and its transcript found in *C. sorokiniana* transcriptome when annotated to the KEGG database are shown in **Table 4.9**.

Fatty acid biosynthesis on *C. sorokiniana* starts with carboxylation of acetyl-CoA into malonyl-CoA which catalyse by acetyl-CoA carboxylase (ACC, EC 6.4.1.2). After that, malonyl-CoA will be converted to malonyl-ACP by malonyl-CoA ACP transacylase (FAS or FabD, EC 2.3.1.39), which then used as carbon donor for the subsequent elongation reaction. During elongation reaction, ketoacyl-ACP synthase (KAS) will catalyse a series of condensation reaction of malonyl-ACP and acyl-ACP (or acetyl-CoA) acceptor. The first condensation reaction catalysed by beta-ketoacyl-ACP synthase III (KAS III or FabH , EC 2.3.1.180) that will produce four-carbon product, acetoacetyl-ACP. The other condensation enzyme like KASI and KASII are responsible for producing various lengths of carbon chain and their elongation. In the transcriptome of *C. sorokiniana* we only detected KASII and KASIII, but not KASI, which is similar to most of higher plant. Both KASII and KASIII were extensively found in various plants such as *Arabidopsis thaliana*, *Spinach oleracea*, *Pisum sativum, Cuphea wrightii* and *Allium porrum* (Yang et al. 2016). KASIII plays an important role as a rate-limiting enzyme in TAG accumulation and KASII could cause significant changes of the fatty acid

composition in the conversion of C16 to C18 in TAG biosynthesis (Yang et al. 2016).

The product of KASII and KASIII condensation reaction produced beta-ketoacyl-ACP. After condensation, these products will be reduced by 3-ketoacyl-ACP reductase (FAS2 or FabG, EC 1.1.1.100), dehydrate by 3-hydroxyacyl ACP dehydrase (FabZ, EC 4.2.1.17) and further reduced by Enoyl-ACP reductase (ENR or FabI, EC 1.3.1.9). These steps will be repeated 6 times until the 16-carbon-acyl-ACP is formed. Finally, fatty acyl-ACP thioesterase (FatA, EC 3.2.1.14) will cleave the thioester linkage holding 16C-acyl-CoA and ACP, freeing the palmitic acid (16:0). Another cycle of condensation, reduction, dehydration and further reduction will produce stearic acid (18:0). These C16 and C18 trienoic fatty acids could be used as the precursors for the synthesis of cellular membranes, long-chain polyunsaturated fatty acids (LC-PUFAs) and storage neutral lipids (mainly TAGs) (Rismani-Yazdi et al., 2011). All transcripts code for the enzyme involved in the fatty acid biosynthesis found in *C. sorokiniana* transcriptome are shown in **Figure 4.8.**

Table 4.9: Enzyme involved in fatty acid biosynthesis identified by annotation of the *C. sorokiniana* transcriptome.

Step	Enzyme	EC Number	Enzyme name	Gene name	Transcript found
1.	Acetyl Coa carboxylase	EC 6.4.1.2	ACACA		3
2.	Acetyl CoA:ACP transacylase	EC 2.3.1.38, EC 2.3.1.39	KASIII	FabH	1
3.	Malonyl CoA:ACP transacylase	EC 2.3.1.39	FAS	FabD	2
4.	3-ketoacyl-ACP synthase	EC 2.3.1.179	KASII	FabF	4
5.	3-ketoacyl-ACP reductase	EC 1.1.1.100	FAS2	FabG	3
6.	3-hydroxyacyl ACP dehydrase	EC 4.2.1.17	ECH	FabZ	1
7.	Enoyl-ACP reductase	EC 1.3.1.9	EAR	FabI	2
8.	Fatty acyl-ACP thioesterase A	EC 3.1.2.14	FATA	FatB	1

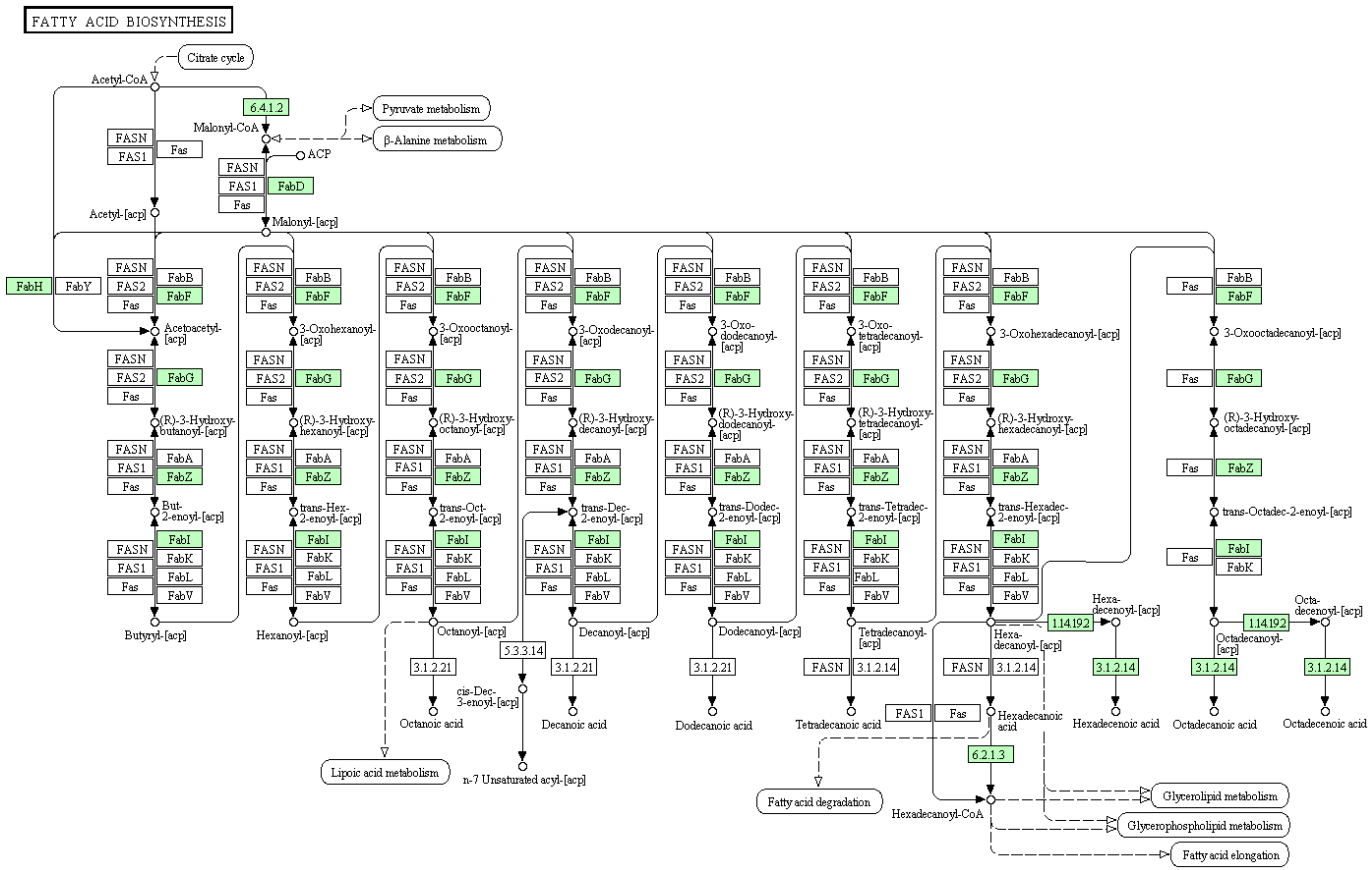


Figure 4.8: Fatty acid biosynthesis pathway reconstructed based on annotation of *C. sorokiniana* transcriptome to the KAAS (or KEGG database). The green box indicates the gene encoding enzyme of particular step was identified in the *C. sorokiniana* transcriptome.

Further elongation of fatty acid occurs in the mitochondrion and endoplasmic reticulum (ER) because the elongation enzyme (elongase) is only present in these organelles. From the annotation data, only one elongation enzyme was evident in mitochondria: mitochondrial trans-2-enoyl-CoA reductase (EC 1.3.1.38), with four elongation enzymes in the ER which includes: 3-ketoacyl-CoA synthase (EC 2.3.1.199); very-long-chain 3-oxoacyl-CoA reductase (EC 1.1.1.330); very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase (EC 4.2.1.134); and very-long-chain enoyl-CoA reductase (EC 1.3.1.93). The number of transcript found for each enzyme coded for elongation of fatty acid synthesis is shown in **Table 4.10**.

After the elongation reaction, if double bonds are introduced to the acyl group by desaturase enzyme, unsaturated fatty acid will be produced. In this study, the transcript encoding several desaturase enzymes such as stearyl-ACP desaturase (Δ^9D) (DESA1, EC 1.14.19.2) which desaturate stearic acid (18:0) to oleic acid (18:1), omega-6 fatty acid delta-12 desaturase ($\Delta^{12}(\omega^6)$D) (FAD2 and FAD6, EC 1.14.19.-) which desaturate oleic acid to form linoleic acid and, omega-3 fatty acid delta-15 desaturase ($\Delta^{15}(\omega^3)$D) (FAD8, EC 1.14.19.-) which further desaturate linoleic acid to form α-linoleic acid, were found. These enzymes were encoded by at least 1 transcript in *C. sorokiniana* transcriptome (**Table 4.10**). Interestingly, a desaturase enzyme that is found in animals, which is stearoyl-CoA delta-9 desaturase (EC 1.14.19.1) was detected in this microalgae transcriptome. Stearoyl-CoA-desaturase is an enzyme that catalyses the rate-limiting step in the synthesis of unsaturated fatty acids in human or animal while the rate limiting step for unsaturated fatty acid production in plant is catalysed by stearyl-ACP desaturase (Ntambi and Miyazaki 2004; Wu et al. 2009).

Table 4.10: Fatty acid elongation and desaturation that occur in mitochondrion and endoplasmic reticulum.

Step	Elongation enzyme	EC Number	Gene name	Transcript found
1.	Mitochondrial trans-2-enoyl-CoA reductase	EC 1.3.1.38	MECR, NRBF1	6
2.	3-ketoacyl-CoA synthase	EC 2.3.1.199	KCS	8
3.	17-β-estradiol 17-dehydrogenase / very-long-chain 3-oxoacyl-CoA reductase	EC 1.1.1.62 EC 1.1.1.330	KAR, IFA38	1
4.	Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase	EC 4.2.1.134	PHS1, PAS2	4
5.	Very-long-chain enoyl-CoA reductase	EC 1.3.1.93	TER, TSC13, CER10	3
6.	3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase	EC 1.1.1.35, EC 4.2.1.17, EC 5.1.2.3	fadJ	1

Step	Desaturase Enzyme	EC Number	Gene name	Transcript found
1.	Stearyl-ACP desaturase, Δ9-D	EC 1.14.19.2	DESA1	6
2.	Stearoyl-CoA desaturase (delta-9 desaturase),	EC 1.14.19.1	SCD, desC	2
3.	Omega-6 fatty acid desaturase (delta-12 desaturase)	EC 1.14.19.-	FAD6, desA	1
4.	Omega-6 fatty acid desaturase (delta-12 desaturase)	EC 1.14.19.-	FAD2	4
5.	Omega-3 fatty acid desaturase (delta-15 desaturase)	EC 1.14.19.-	FAD8, desB	1

In addition to synthesis, all the enzymes involved in β-oxidation or degradation of fatty acid into acetyl-CoA that occurs in mitochondria were also identified. The enzyme involved in β-oxidation of fatty acid and the transcript encoded by the *C. sorokiniana* transcriptome are shown in **Table 4.11**. The degradation of fatty acids is very important for energy production and is utilised by several different metabolic pathways, especially energy production in the citric acid (TCA) cycle (or Krebs cycle). Besides producing acetyl-coA, this process also produces NADH and $FADH_2$ which acts as electron carriers for the electron transport chain in glycolysis and TCA cycle (Fillmore et al. 2011).

Table 4.11: Enzymes involved in fatty acid degradation identified by annotation of the *C. sorokiniana* transcriptome.

Step	Fatty acid degradation	EC Number	Gene name	Transcript found
1.	Long-chain acyl-CoA synthetase	EC:6.2.1.3	ACSL, fadD	18
2.	Alcohol dehydrogenase	EC 1.1.1.1	frmA, ADH5, adhC	2
3.	Aldehyde dehydrogenase (NAD+)	EC:1.2.1.3	E1.2.1.3	7
4.	Alkane 1-monooxygenase	EC:1.14.15.3	alkB1_2	2
5.	Acyl-CoA oxidase	EC:1.3.3.6	ACOX1, ACOX3	7
6.	Acyl-CoA dehydrogenase	EC:1.3.8.7	ACADM, acd	1
7.	Acetyl-CoA C-acetyltransferase	EC:2.3.1.9	atoB	1
8.	Enoyl-CoA hydratase	EC:4.2.1.17	paaF, echA	1
9.	3-hydroxybutyryl-CoA epimerase	EC 5.1.2.3	fadJ	1
10.	Acetaldehyde dehydrogenase / alcohol dehydrogenase	EC:1.2.1.10, EC 1.1.1.1	adhE	4
11.	Acetyl-CoA acyltransferase 1	EC:2.3.1.16	ACAA1	2
12.	3-hydroxyacyl-CoA dehydrogenase	EC 1.1.1.35, EC 1.1.1.211	MFP2	1
13.	3,2-trans-enoyl-CoA isomerase, mitochondrial	EC:5.3.3.8	DCI	5
14.	Alcohol dehydrogenase	EC:1.1.1.1	yiaY	2

By identifying the transcript code for the important enzyme in this fatty acid biosynthesis, the keystone gene especially for the production of fatty acid that could be used for devising prediction and control strategies for microalgal communities in the future could be determined.

4.3.4.3 Starch metabolism

Starch is another major carbon and energy storage compound found in many microalgal cells. The production of starch in microalgae might become an alternative source for the production of a variety of biofuels (such as ethanol, butanol and hydrogen) which can overcome the pre-treatment processes disadvantages attributes by plant based starch and lignocellulosic materials (Chen et al. 2013). Excess glucose produced during photosynthesis is converted to starch for energy storage through a series of enzymatic steps. There are four main reactions involved in the synthesis of starch which occur in

the chloroplast: i) substrate activation; ii) chain elongation; iii) chain branching; and iv) chain debranching (Busi et al. 2014).

Based on the KEGG reconstruction assignment, four key enzymes for starch biosynthesis were identified in *C. sorokiniana* transcriptome. The biosynthesis of starch begins with the phosphoglucomutase enzyme (PGM, EC 5.4.2.2) which facilitates the interconversion of glucose-1-phosphate produce during photosynthesis process to glucose-6-phosphate. After that, glucose 1-phosphate adenylyltransferase (or ADP-glucose pyrophosphorylase, EC 2.7.7.27) catalyzes the reaction of α-D-glucose 1-phosphate and ATP to produce ADP-glucose and inorganic pyrophosphate. ADP-glucose pyrophosphorylase was known as the key enzyme regulating starch synthesis in animal and plants (Ballicora et al. 2003; Morell et al. 2008).

In the next step, ADP-glucose is converted to amylose by starch synthase (EC 2.4.1.21). Amylose is a linear chain of glucose residues connected by α-1,4-glycosidic bonds. It will undergo branching step by 1,4-α-glucan branching enzyme (EC 2.4.1.18) which introduce α-1,6 branch points to form amylopectin or starch. All transcripts that code for starch biosynthesis enzymes described above was identified in the transcriptome of *C. sorokiniana* and summarised in **Table 4.12**. On the other hand, we also identified transcript that code for debranching enzyme, or enzyme involved in the catabolism of starch such as α-amylase (EC 3.2.1.1), β-amylase (EC 3.2.1.2), 4-α-D-glucanotransferase (EC 2.4.1.25) and starch phosphorylase (EC 2.4.1.1) (Busi et al. 2014).

Table 4.12: Enzymes involved in starch biosynthesis and degradation identified by annotation of *C. sorokiniana* transcriptome.

Starch biosynthesis

Step	Enzyme	EC number	Enzyme name	Transcript found
1.	Phosphoglucomutase	EC:5.4.2.2	pgm	1
2	Glucose-1-phosphate adenylyltransferase or ADP glucose pyrophosphorylase	EC:2.7.7.27	glgC	5
3.	Starch synthase	EC:2.4.1.21	glgA	9
4.	1,4-α-glucan branching enzyme	EC:2.4.1.18	glgB	8

Starch degradation

Step	Enzyme	EC number	Enzyme name	Transcript found
1.	α-amylase	EC:3.2.1.1	amyA, malS	2
2.	β-amylase	EC:3.2.1.2	-	5
3.	Starch phosphorylase	EC:2.4.1.1	glgP, PYG	2
4.	4-α-D-glucanotransferase/ debranching enzyme	EC:2.4.1.25	malQ	14

From the annotation results, we could see the transcript found for debranching enzyme was higher compared to others. This is proportional with the findings of Hwang et al. (2013) in which they reported that malQ was highly expressed during stationary phase. In addition, this enzyme is very important for the formation of starch granule; the storage energy of microalgae. This is because the amylopectin produced by starch synthase is highly branched and form glycogen instead of granule inside the cytoplasm. 4-α-glucanotransferase or debranching enzyme are required to trim the amylopectin branched to form correct branching structure to allow its crystallisation to form the starch granule (Ball et al. 1996).

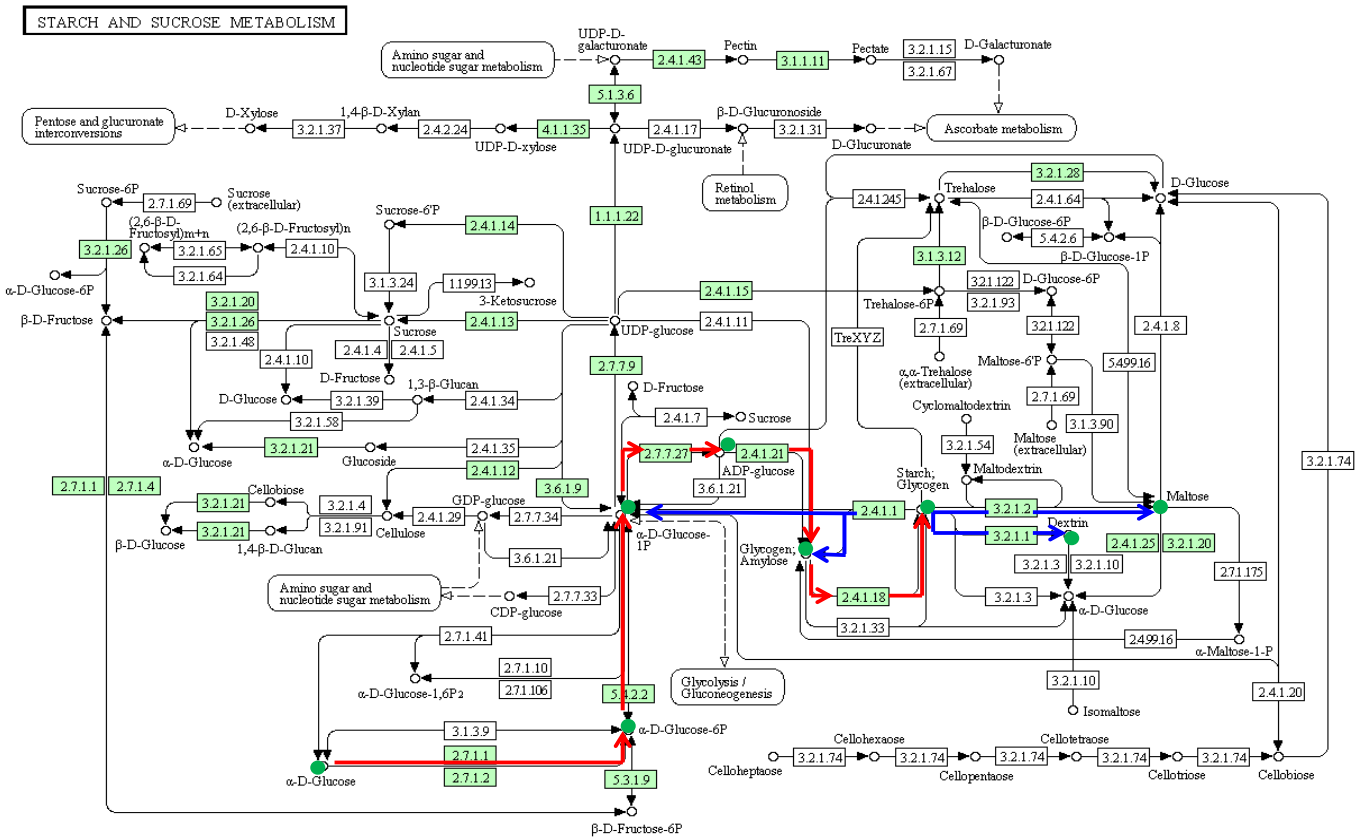


Figure 4.9: Starch biosynthesis and degradation pathway reconstructed based on the *de novo* assembly and annotation of *C. sorokiniana* transcriptome. Green dot represents biomolecules, while green box highlighted the enzymes found in the *C. sorokiniana* transcriptome. Red line shows starch biosynthesis pathway, while blue line shows starch degradation pathway.

4.3.5 Analysis of differential gene expression

Differential expression analysis was performed using DESeq2 package. In the beginning, PCA analysis and Heat Map analysis was done to assess overall similarity between samples (Love et al. 2015). The comparison was made on normal (N) and stress (S) sample with Sa, Sb and Sc were for replicates. **Figure 4.10** depicts a heat map matrix and PCA plot, respectively, revealed that NSb and NSc as well as SSb and SSc grouped tightly with one another distinguishing the normal (or control) from the stress treatments. However, NSa seems to be an outliers for the N group while SSa may also be an outlier for the S group. However, both NSa and SSa were still far from each other meaning that they were not related to teach other. This might be due to RNA extraction, library preparation and sequencing differences (Conesa et al. 2016).

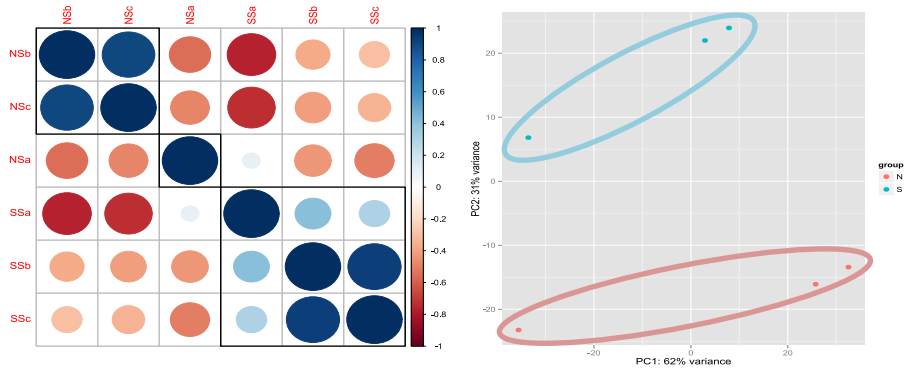


Figure 4.10: Heat map and PCA plot of normal (N) and stress (S) samples for *C. sorokiniana* RNA-seq differential expression analysis. Black box in heat map represents clustering. Its shows that stress sample clustered in one group, while in normal sample, NSa seems an outlier. PCA plot revealed that the normal and stress samples form distinct clusters, indicating that samples within each biologic group have more similarity. Red circle is normal samples; blue circle is stress samples.

In the subsequent analysis, DESeq2 estimates the dispersion of each gene and analyses whether there is differential expression between normal and stress conditions and the result obtained are presented in a MA-plot (**Figure 4.11**) which illustrates the dependence between the gene fold change log ratio and its means of counts. In this case, 933 transcripts were significantly differentially expressed among the treatments. In which 431 and 502 were upregulated and downregulated respectively in stress samples.

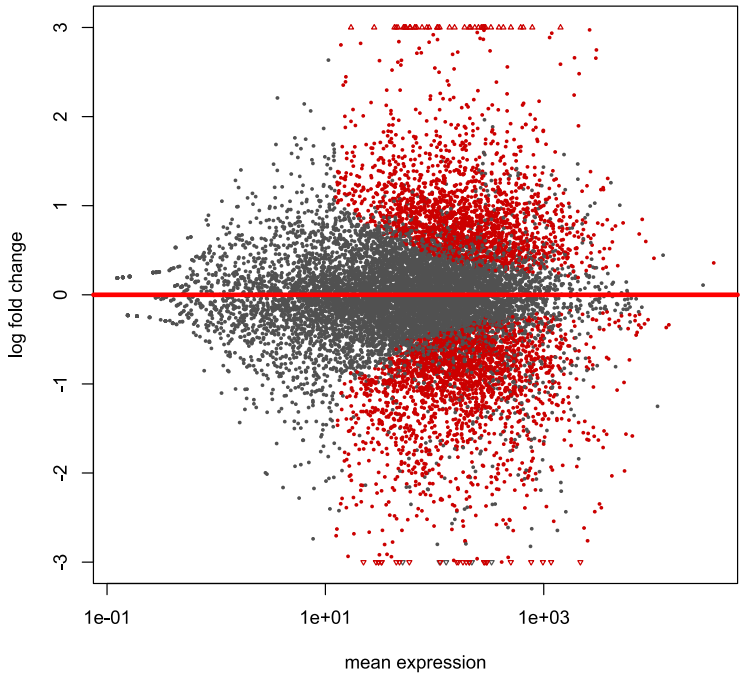


Figure 4.11: MA-plot contrasting gene expression levels between normal and stress condition based on normalized counts. The plot depicts the shrinkage of log_2 fold changes resulting from the incorporation of zero-centered normal prior. Red and grey points are DE and non-DE transcripts, respectively. Triangles depict shrunken log_2 fold change point a greater and/or lesser than -3 and 3 log_2 fold change. Each dot in the MA plot corresponds to a gene. The x-axis represented the mean expression which corresponds to the average counts adjusted by library size factors (normalization constant) of the sample, and y-axis represented the log_2 fold change, which describes how much a quantity changes from one condition to another (Wang et al. 2010a)**.**

4.3.6 Pathway analysis of differentially expressed genes

To reveal the differential expression profiles between normal and stress, the potential differentially expressed genes (DEGs) were analysed. When the threshold of FDR ≤ 0.05 and an absolute value of fold change ≥ 2 were used, a total of 479 genes showed significant differences, in which 165 were upregulated and 314 genes were downregulated (**Figure 4.12**). Based on the functional annotation of *C. sorokiniana* transcriptome to KEGG database, genes that upregulated during stress conditions involved in carbohydrate metabolism (such as glycolysis or gluconeogenesis, TCA cycle, starch metabolism, pyruvate metabolism, glyoxylate and dicarboxylate metabolism, and propanoate metabolism), amino acid metabolism, energy metabolism (such as oxidative phosphorylation, photosynthesis, carbon fixation pathways, methane metabolism and nitrogen metabolism), genetic information processing (such as translation and signal transduction), lipid metabolism (such as fatty acid biosynthesis and biosynthesis of unsaturated fatty acids), environmental information processing, metabolism of cofactors and vitamins (such as porphyrin and chlorophyll metabolism), biosynthesis of secondary metabolites and nucleotide metabolism.

Meanwhile, genes that downregulated during stress conditions involved in carbohydrate metabolism (such as glycolysis or gluconeogenesis, fructose and mannose metabolism, galactose metabolism, ascorbate and aldarate metabolism, pyruvate metabolism,and propanoate metabolism), genetic information processing (such as aminoacyl-tRNA biosynthesis and RNA degradation), environmental information processing (such as AMPK signaling pathway, cAMP signaling pathway and calcium signalling pathway), amino acid and nucleotide metabolism, and cellular processes (such as cell growth and death).

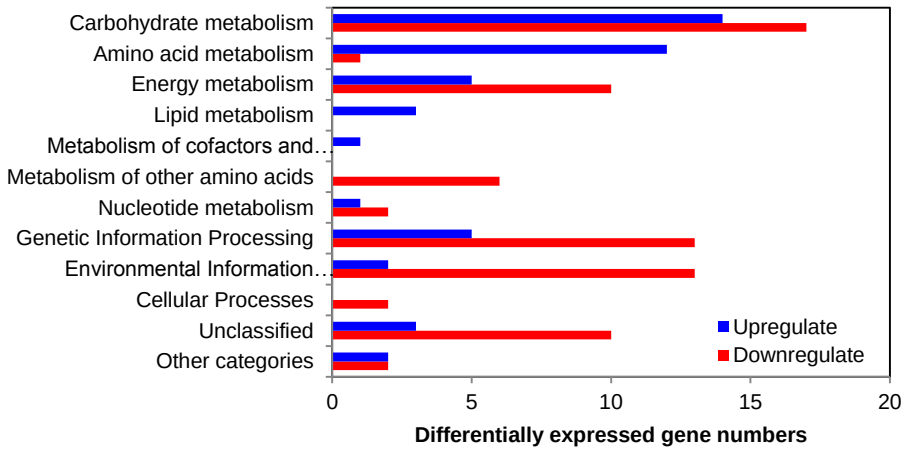


Figure 4.12: Functional categories of differentially expressed genes based on annotation with KEGG database. The blue bars represents the upregulated DEGs, the red bars represents the downregulated DEGs in stress conditions.

4.3.7 Selection of suitable housekeeping genes for *C. sorokiniana* using real-time qPCR

Ten potential house keeping genes (HKGs) were choosen based on previously reported genes that had been used as a controls in qPCR of algal gene expression studies (Chunget et al. 2008; Fischer et al. 2007; Boldt et al. 2008; Rosic et al. 2011; Dong et al. 2012; Guo and Ki 2012; Liu et al. 2012a; Shi et al. 2013; Adelfi et al. 2014). Generally, these genes were not differentially expressed, high abundance and has a minimal expression level variation in every sample analysed. In this case, minimal expression level variation indicated by a low coefficient of variation value (<0.5) of RNA-seq feature count data (McCarthy et al. 2012).

The suitable HKGs were chosen based on several criterions such as the abundances (or expression level) of these genes found in the samples based on the quantification cycle value (Cq) value, and the melting point analysis. A low Cq-value indicates high expression level; the converse implies low expression (Arya et al. 2005; Liu et al. 2012a; Bustin et al. 2015). Based on **Table 4.13** the Cq values ranged from 17 to 27. Among eight genes, RPL19, A-tub and UBQ showed high expression with low Cq-values (less than 20), whereas B-tub showed low expression (Cq-value 27). In the melting point analysis, the primers should produce one single peak from tested sample and should not produce peak in non-template control (NTC) sample (D'haene and Hellemans 2010). In this study, all seven primer combinations produce one single peak, except SAM (**Figure 4.13**). However, from Cq-value analysis (**Table 4.13**), UBQ, HSP90 and CDKA showed a melting peak in NTC reaction. This indicated the primers for UBQ, HSP90 and CDKA were not specific and they might generate non-specific PCR products. Thus, the most suitable genes to be HKGs for qPCR analysis are RPL, A-tub and PCNA.

Table 4.13: Summary of HKGs quantification Cq and melting point analysis data.

Target	Sample	Cq Mean	Cq Std. Dev	Melt Temperature (°C)
RPL	Unknown	17.11	0	86.5
	NTC	None	0	None
B-tubulin	Unknown	27.06	0	89.25
	NTC	None	0	None
A-tubulin	Unknown	19.37	0	86.5
	NTC	None	0	None
UBQ	Unknown	19.81	0	86.5
	NTC	35.12	0	81.5
SAM	Unknown	25.90	0	90.25
	NTC	None	0	None
HSP	Unknown	24.70	0	87.25
	NTC	35.72	0	76.5
CDKA	Unknown	21.84	0	88.5
	NTC	30.39	0	78.5
PCNA	Unknown	22.68	0	85.75
	NTC	None	0	None

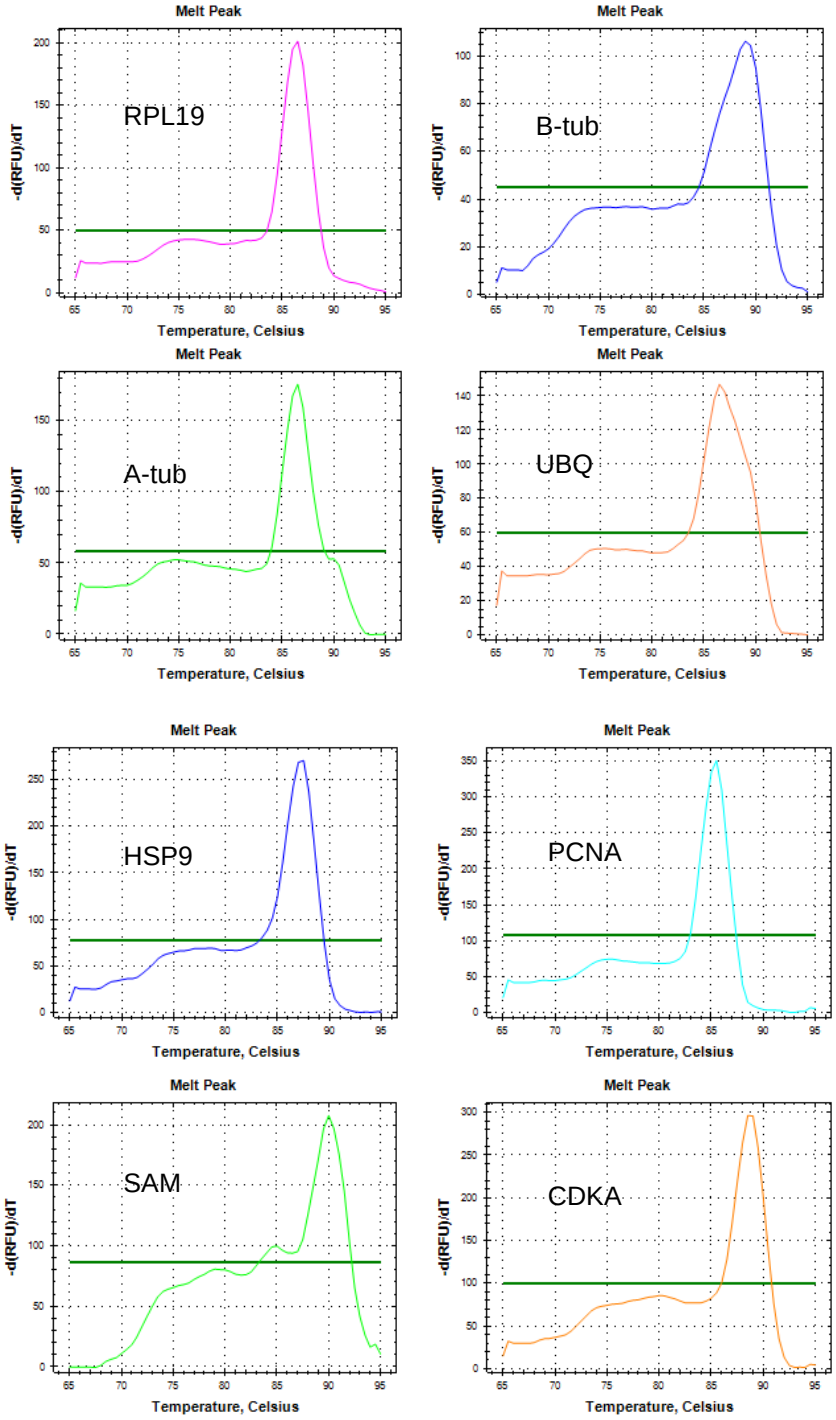


Figure 4.13: Melting curves of different HKGs primers used for RT-qPCR. All primer combinations show one single peak of melting curve except for SAM.

4.3.8 Validation of gene expression through quantitative real-time PCR

Quantitative real-time PCR (RT-qPCR) was used to validate 7 differentially expressed genes identified by RNA-seq. The genes were chosen from each of the important pathways in the transcriptomic analysis of RNA-seq data. For example stearyl-ACP desaturase (DESA1) from unsaturated fatty acid synthesis pathway, malate dehydrogenase (MDH) from TCA cycle, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (IspD) from MEP pathway, Acetyl-Coa carboxylase (ACACA) from fatty acid biosynthesis pathway, TAG lipase (TagL) from TAG biosynthesis pathway, 1,4-glucan branching enzyme (glgB) from starch biosynthesis pathway and phosphofruktokinase (pfkA) from glycolysis pathway. The primers used for each target genes and selected housekeeping genes with their efficiency and R^2 data are shown in **Table 4.14**. In RNA-seq sequencing results, the expression levels of DESA1, MDH and IspD were significantly upregulated, whilst the expression of ACACA, TagL, glgB and pfkA were significantly downregulated in stress conditions. The gene expression profile based on RT-qPCR was highly correlated with RNA-seq confirming the reliability and validity of the RNA-seq technique (**Figure 4.14**).

Table 4.14: Housekeeping and target genes and their primers used for RT-qPCR.

Gene symbol	Target gene	Forward/reverse primers (5' - 3')	Product size (bp)	Efficiency value (%)	R^2
RPL	Ribosomal protein L19	GTCTGGCTGGACCCCAATGA GCCTCTGCTGCGGTGC	143	90.4	0.999
A-tub	Alpha-tubulin	CACAGTTTACCCGTCTCCCC TCAGCCGGTTAAGGTTGGTG	185	97.6	0.996
PCNA	Proliferating cell nuclear antigen	TGGGCCTGAACCTGAACAAC TTGATGCCCTCCTTCGTCAC	301	90.3	0.987
DESA 1	Stearyl-ACP desaturase	TGCTGGGTCGCTTCCTCTAC CCGTTGTTCAGCTCCGAGTG	397	94.3	0.995
MDH	Malate dehydrogenase	TCACCAAGGAGGAGATTGAG CGCCTGTCTTAATGTTGGTC	342	105.3	0.989
IspD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	CTGGTGGCAATTCACGACTC TCAGTGACCTCCAGCTTCTC	263	99.3	0.981
ACACA	Acetyl Coa carboxylase	GCATGGCCGACCAGTTTGTG CCGGTCGTAGACATCAGGTG	347	94.1	0.966
pfkA	Phosphofruktokinase	CAAGCCCATCACCCTTACCC CTCCACAGCCGTCTCAAACC	298	91.5	0.991
TagL	TAG lipase	GCTGCTGGAACACCAGATGC GGCGCTGTCAGGTAGTTGAG	357	105.6	0.965
glgB	1,4-glucan branching enzyme	GAGCGCGATGACTTTGGCAC TGGCGTTGTAGCCCAGCTTC	358	98.2	0.948

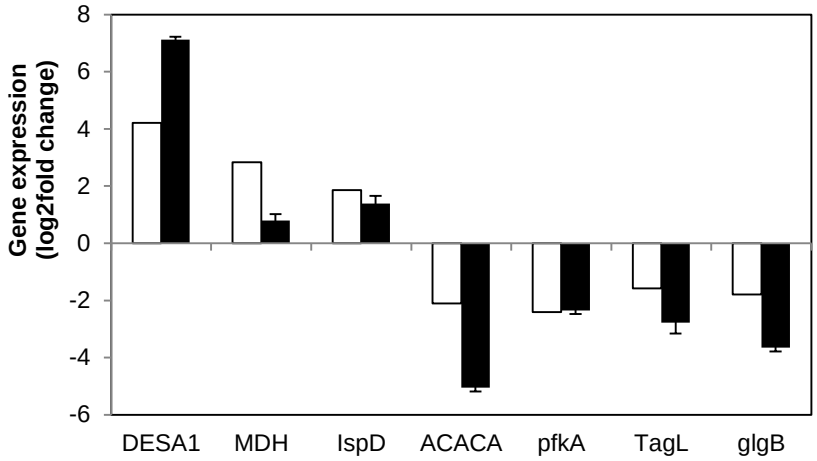


Figure 4.14: RT-qPCR validations of the gene expression patterns. The white bars represent the changes in transcript from RNA-seq, and the black bars represent the relative expression level estimated by RT-qPCR.

4.3.9 Regulation of high value metabolties and its relationship with other pathways during stress

Many environmental stresses such as drought, salinity, light, unfavourable temperatures, and their rapid fluctuations adversely affect the process of photosynthetic carbon metabolism in microalgae. It may alter the ultrastructure of the organelles, change the concentration of various pigments and metabolites (Juneja et al. 2013). The two important biological pathways in photosynthesis: photosynthetic electron transport system (PETs) and Calvin-Benson cycles were highly affected. PETS converts light energy into ATP and NADPH; meanwhile Calvin–Benson cycle also known as a photosynthetic carbon fixation cycle, fixes the CO_2 and utilise the organic products from photosynthesis.

In this study, genes that encode for the enzymes involved in these two main steps (PETs and Calvin-Bensen cycle) were upregulated (**Appendix F**). For example, petH (ferredoxin-NADP reductase); an important component of photosystem II which controls the photoexcitation energy transfer via phycobilisome and functions as a redox generator to trigger the activity of Fd-dependent enzymes during dark reaction, was found to be upregulated under stress condition (Okada 2009). Meanwhile, ndh1 (NADH: ubiquinone reductase); a component in photosystem I, which transfers electron from electron donor NADH to a ubiquinone molecule, with concomitant generation of a proton motive force used to synthesis ATP (Battchikova et al. 2011) was also upregulated. ATP and NADPH generated in this reaction were used to maintain the growth of microalgae (Kramer and Evans 2011).

TCA cycle; the intermediate pathway that connects photosynthesis with other metabolic pathways is known as central hub of metabolism of the cells (Sweetlove and Fernie, 2005). The TCA cycle was found to be upregulated under stress condition (**Appendix F**). The enhancement of TCA cycle increases ATP synthesis which supports lipid synthesis and increases the production of intermediates (such as malic acid) to generate substrates or carbon skeletons for fatty acid synthesis. Moreover, the last step in TCA cycle, which is catalysed by MDH (malate dehydrogenase) is not only related with respiration but also with several other processes such as photorespiration, ß-oxidation of fatty acids, and stress tolerance (Tomaz et al. 2010; Nunes-Nesi et al. 2013)

In addition to the TCA cycle, upregulation of malQ (debranching enzyme) in the starch catabolism pathway helps facilitate the breakdown of starch into glucose by remodelling the glycogen structure for further degradation (Berg et al. 2002a). Mutation of this gene gave rise to specific alterations in amylopectin structure and starch biosynthesis has been reported in *C. reinhardtii* (Colleoni et al. 1999). Regulation of glucose levels was carried out in glycolysis pathways. In this study, two of three important enzymes (phosphofructokinase (pfkA) and pyruvate kinase) of this pathway were downregulated (**Appendix F**). The downregulation of pyruvate kinase which converts phosphoenolpyruvate into pyruvate indicates that the acetyl-coA was abundant in the cells, in other words high amount of acetyl-coA will inhibit pyruvate kinase.

As discussed above, the metabolic pathways involved in the biosynthesis and degradation of energy rich molecules were closely related. Acetyl-CoA is an important intermediate in numerous physiological processes (Sasaki and Nagano 2004). It can be produced from different metabolic processes such as the degradation of fatty acids, ketone bodies, as well as amino acids and via pyruvate metabolism. Acetyl-coA levels influence pigment accumulation through incorporation into the MEP pathway. Under stress conditions, two enzymes (YgbP and GcpE) that are involved in the biosynthesis of isoprenoids were upregulated. GcpE was found to be an important enzyme in the terminal steps of the MEP pathway, where it catalyzes the production of isoprenoids and terpenoids (Kollas et al. 2002). Isoprenoids are essential precursors for pigment biosynthesis including lutein, astaxanthin, β-carotenes, as well as other bioactive compounds such as flavanoids, terpenes and phenolic compounds (Dewick 2002; Baba and Shiraiwa 2013). The upregulation of this compound increased the potential of this microalga to be used as a source for various commercial applications.

Fatty acid biosynthesis, especially unsaturated fatty acids was upregulated in stress condition (**Appendix F**). This was evidence by the upregulation of unsaturated fatty acid biosynthesis ensymes (such as DESA1, SCD and FAD2). Polyunsaturated fatty acids (PUFAs) play an important role in the structural components of cell membranes, and as antioxidants (PUFAs can counteract free radical formation during photosynthesis). As such, PUFA-rich TAGs might donate specific compounds necessary to rapidly reorganize

membranes through adaptive metabolic responses to sudden changes in environmental conditions (Fields et al. 2014). This process is related to the downregulation of two enzymes from TAG metabolism; glycerol-3-phosphate acyltransferase (GPAT) and TAG lipase (TGL4). GPAT is involved in the regulation of phospholipid content in membrane, meanwhile TGL4 plays a role in the conversion of TAG into DAG. The regulation of cell membrane composition is not only important during stress, but also in cellular responses to changes in inoculum volume (Lu et al. 2013).

4.4 Conclusion

Transcriptomic analysis of *C. sorokiniana* under stress conditions enabled the exploration of a broad diversity of genes and pathways. From the analysis 18,310 assembled transcripts were obtained. The functional annotation and classification of these transcripts had given better understanding of *C. sorokiniana* transcriptome and its behaviour under stress induced conditions. The findings point to several molecular mechanisms that potentially drive the overproduction of high value metabolites such as carotenoids and fatty acids under stress induced conditions, which include the upregulation of unsaturated fatty acid metabolism, biosynthesis of secondary metabolites as well as vitamin and cofactor metabolism. Furthermore, the expression profile based on RT-qPCR was similar with RNA-seq confirming the reliability and validity of the RNA-seq technique. Molecular genetic manipulation of this microalga might be an effective way to enhance their properties to make it suitable for commercial development.

CHAPTER 5

SMALL RNA TRANSCRIPTOMIC PROFILING OF *C. sorokiniana* AND *C. zofingiensis*

5.1 Introduction

Small non-coding RNAs (ncRNAs) genes have been found in numerous organisms which act as key regulators for development through modulation of the processing, stability, and translation of larger RNAs. Their ability to silence specific genes affects a wide range of biological functions, ranging from gene regulation during embryological development and cell differentiation, to genome rearrangement (He and Hannon 2004; Cerutti et al. 2011). There are three major classes of endogenous small RNAs have been discovered: microRNAs (miRNAs), Piwi-associated small RNAs (piRNAs) and small interfering RNAs (siRNAs). MiRNAs are single stranded non-coding regulatory RNAs that fold into imperfect stem-loop structure with approximate 20-24 nucleotides in size (Elsabahy et al. 2011).

Of the classes of small RNAs, the microRNA (miRNA) family is the most extensively characterized. MicroRNAs are estimated to occur at a frequency of approximately 0.5–1.5% of the total genes in the genome of an organism (Allen et al. 2006). MiRNAs play role in controlling the gene expression that involved in developmental, physiological or metabolic processes, or stress responses through mRNA cleavage or translational repression. The choice of control mechanism by miRNA is dependent on the degree of sequence complementary between miRNAs and their target, in which perfect pairing results in target mRNA cleavage whereas imperfect pairing results in repression. The other mechanism of miRNA mediated regulation is miRNA mediated mRNA decay (Wu et al. 2006).

In recent years, there has been an increase in the utilisation of deep sequencing of the transcriptome for the identification of differential expression gene as well as for the opportunity to discover novel transcript, including new alternative isoforms and miRNAs. In this study, the production of high value metabolites such as carotenoids and lipids resulting from stress induced condition on microalgae growth of *C. sorokiniana* and *C. zofingiensis* has increased interest in the characterisation of the genetic profiles of these microalgae. Although genes that play an important role in certain metabolic pathways were identified and used in non-model organism, the strategies were not effective as the model organism (Ekblom and Galindo 2011). This suggests that the factors and mechanisms responsible for the differences in metabolite deposition, which have not been elucidated completely, could aid in the development of new strategies to modulate carotenoid and lipid deposition and improve their commercial value.

Thus, this study aims to profile the miRNAs that involved in regulating the production of carotenoids and lipids in *C. sorokiniana* and *C. zofingiensis* when cultured under high light intensity and glucose addition during nitrogen-limited condition using small RNA sequencing technique.

5.2 Materials and methods

5.2.1 Experimental design

The samples used in this study were the same with previous chapter, thus the experimental design for this chapter are based on section **4.2.1**.

5.2.2 RNA extraction

RNA extraction technique used in this chapter was similar to the previous chapter (section **4.2.2** and **4.2.3**).

5.2.3 Library preparation

The library for small RNA-sequencing was prepared according to the NEBNext Multiplex Small RNA Library Prep Set for Illumina manual (Cat. No. E7300) from New England Biolabs (NEB, UK). The main steps for preparation of small RNA library based on this kit are: 1) 3' ligation, 2) primer hybridisation, 3) 5' ligation, 4) first strand synthesis, 5) PCR amplification, and 6) size selection.

In the first step, the extracted RNA was incubated with 3'SR Illumina adaptor for 2 min at 70°C followed by addition of ligation reaction buffer and ligation enzyme mix and allowed to stand for 1 hr at 25°C in a thermal cycler. After that, the mixture of 3'SR adapter and ligated RNA sequence was hybridised using SR reverse transcription primer. This was to ensure that the excess free 3'SR adapter became a double stranded DNA to prevent ligation with 5' adapter used in the subsequent ligation step. For 5' ligation step, the 5´ SR adaptor must be denatured prior to use by incubating the required volume of 5' SR adaptor in the thermal cycler at 70°C for 2 min. The ligation reaction was carried out by mixing the ligation mixture from previous step with 5´ ligation reaction buffer and 5' ligation enzyme mix, followed by incubation for 1 hr at 25°C.

The first-strand cDNA was synthesized by performing reverse transcription. In this step the ligation mixture was mixed with Murine RNase Inhibitor, ProtoScript II reverse transcriptase enzyme and first strand synthesis reaction buffer and incubated at 50°C for 60 min. Then, the cDNA was amplified using LongAmp Taq2X Master Mix including SR primer and index primer for 12 to 15 cycles.

Before the amplified or constructed cDNA was analyzed using bioanalyzer, it was purified using QIAQuick PCR purification kit (Qiagen, Germany). Then, the purified PCR products were assessed by DNA 1000 chip on 2100 bioanalyzer (Agilent Technologies, Germany). Meanwhile, size selection for small RNA libraries was performed by AMPure XP beads and 6% polyacrylamide gel. Finally, before the sequencing was performed in Illumina MiSeq, the small RNA library was analyzed again using Agilent high sensitivity chip on 2100 bioanalyzer (Agilent Technologies, Germany).

5.2.4. Quality assessment and pre-processing of small RNA-seq data

Raw reads obtained from Mi-Seq machine were first examined using FastQC following the step conducted for RNA-seq data (**Section 4.2.5.1**). Reads containing adapter contaminants were removed by data trimming program in the CLC Genomic Workbench. The Q30 and GC-content of the raw data were calculated and read lengths within a specific size range from 15 to 27 nt were chosen for analysis.

5.2.5 miRNA-seq data analysis by CLC Genomic Workbench

In the analysis of sequence reads using CLC Genomic Workbench, the reads were clustered together to group different types of small RNAs. Next, the sequences were mapped against the miRBase21 with two nucleotides mismatches allowed. To identify differentially expressed miRNAs between normal and stress samples, statistical analyses were conducted on the square-root transformed read-per-millions base pairs (RPM) of each miRNA. T-test was used to test for normal distribution of the data and homogeneity of variances, respectively.

5.2.6 miRNA-seq data analysis by OmiRas

After removing adapters and low quality reads from the original sequencing data, the FASTQ file of cleaned data was submitted to OmiRas tools (tools.genxpro.net/omiras), a free web server established for differential expression analysis of miRNA-Seq data between two groups. On OmiRas, bowtie processed mapping of short reads of each library, which are summarized to tags in a quantified FASTA format, on the *Arabidopsis thaliana* genome reference sequence (TAIR9), allowing at most two mismatches. Mapping locus annotations were given for each tag, based on information derived from miRbase v21 (www.mirbase.org), snoRNABase (www-snorna.biotoul.fr), Repbase (www.girinst.org/repbase), Genomic tRNA database (gtrnadb.ucsc. edu), and Rfam (www.sanger.ac.uk/resources/databases/rfam. html). Following normalization of the number of reads for each tag with the number of mapping loci, differential expression analysis was carried out between two groups for each miRNA according to the DESeq algorithm (bioconductor.

org/packages/release/bioc/html/DESeq.html) in the setting of false discovery rate (FDR)-corrected P-value at 0.05.

5.3 Results and discussion

5.3.1 General characteristics of the small RNA libraries

The small RNA cDNA Illumina sequencing library was constructed from RNA isolated from exponentially growing *C. sorokiniana* and *C. zofingiensis* under normal and stress conditions. The final gel purification step of the small RNA cDNA library construction showed the presence of the sizes expected for small RNAs amplified with the known linkers (**Figure 5.1**). The purified products were subsequently sequenced using the MiSeq using Illumina platform, resulting in 19,916,716 reads. After pre-processing the small RNA sequencing read to remove adapter sequence at 3' and 5' ends, the clean reads were further analysed using two analysis pipelines; CLC Genomic Workbench and OmiRas web server.

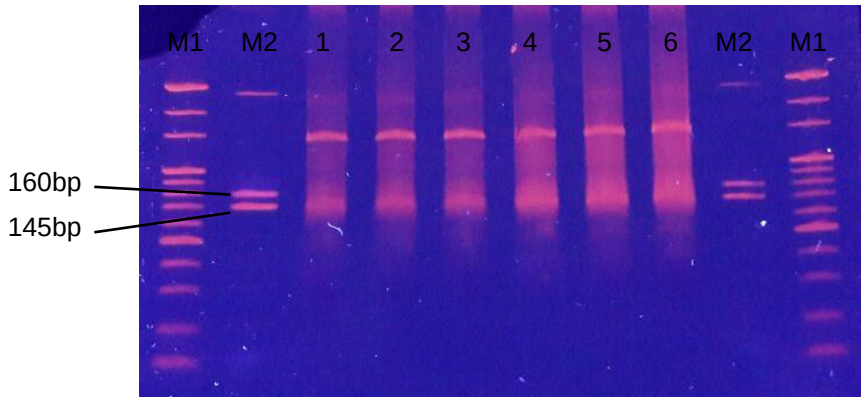


Figure 5.1: Polyacrylamide gel electrophoretic separation of final amplified small RNA products from *C. sorokiniana* and *C. zofingiensis* after library construction and amplification, prior to Illumina sequencing. The gel showing the presence of small RNA bands (143-153 bp). M1 and M2 indicate DNA marker 1 and DNA marker 2. Lane 1 to 6 is combination of small RNA library from normal and stress samples of *C. sorokiniana* according to the concentration. Lane 1 are samples of 0.3 to 0.4 ng/µL, Lane 2 are samples of 0.6 to 0.7 ng/µL, Lane 3 are samples of 0.8 to 1.5 ng/µL, Lane 4, 5 and 6 are samples of 5 ng/µL.

5.3.2 miRNA-seq data analysis by CLC Genomic Workbench

From total reads obtained from small RNA sequencing of *C. sorokiniana*, about 17% small RNAs were obtained from normal sample (NSa, NSb, and NSc), meanwhile about 24% small RNA were obtained from stressed sample (SSa,

SSb, and SSc) (**Table 5.1**). The number of small RNA found in *C. zofingiensis* almost similar to *C. sorokiniana* in which 15% small RNA was found in normal sample and 32% found in stressed sample (**Table 5.2**). When these small RNAs were annotated with *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* miRBase, only several miRNAs were matched. Most of the matched sequences were from *A. thaliana* compared to *C. reinhardtii*. The details of annotation results for small RNA of both *C. sorokiniana* and *C. zofingiensis* to the miRBase are shown in **Appendix J**. The summary of miRNA found in both culture conditions (normal and stress) for *C. sorokiniana* and *C. zofingiensis* are shown in **Table 5.3**.

Table 5.1: *C. sorokiniana* small RNA sequencing data and annotation to miRBase.

Name	Reads	Small RNAs	miRBase annotation	
			Arabidopsis thaliana	***Chlamydomonas reinhardtii***
NSa	688,187	77,873	9	0
NSb	290,634	47,708	4	1
NSc	83,609	19,585	1	0
SSa	606,565	132,844	12	0
SSb	560,213	132,837	10	0
SSc	435,796	119,735	10	1

Table 5.2: *C. zofingiensis* small RNA sequencing data and annotation to miRBase.

Name	Reads	Small RNAs	miRBase annotation	
			Arabidopsis thaliana	***Chlamydomonas reinhardtii***
NZa	528,500	84,686	8	0
NZb	773,365	116,211	5	0
NZc	839,393	124,015	16	1
SZa	2,064,458	454,912	32	4
SZb	425,502	159,527	4	0
SZc	427,558	161,050	7	0

From the annotation results, the most abundance miRNA found in both *C. sorokiniana* and *C. zofingiensis* in normal and stress conditions was miR5645d. The other miRNA found in both strains was miRNA169g. The miRNA that found in *C. sorokiniana* only were miR169m and miR5638b (in both conditions), miR159a, miR169a, miR909 and miR5017 (in normal condition), and miR404, miR408, miR824, miR829, miR831, miR844, miR1156 and miR5653 (in stress condition). Meanwhile, for *C. zofingiensis*, miR5645c and miR5658 were found in both conditions, miR908 and miR5655 were found in normal conditions, and miR160c, miR157d, miR169e, miR417, miR835, miR841b, miR858a, miR1144a, miR1151b, miR1158, miR1166, miR3932b, miR5640, miR5660, miR5998a, miR8170 and miR8183 were found in stress condition (**Table 5.3**).

Table 5.3: Summary of miRNA found in small RNA transcriptome of both *C. sorokiniana* and *C. zofingiensis* using CLC Genomic Workbench analysis.

miRNA	*C. sorokiniana*		*C. zofingiensis*	
	Normal	Stress	Normal	Stress
miR160c	-	-	-	1
miR157d	-	-	-	1
miR159a	1	-	-	-
miR169a	1	-	-	-
miR169e	-	-	-	1
miR169g	-	2	-	1
miR169m	19	22	-	-
miR404	-	1	-	-
miR408	-	1	-	-
miR417	-	-	-	1
miR824	-	2	-	-
miR829	-	1	-	-
miR831	-	1	-	-
miR835	-	-	-	1
miR841b	-	-	-	1
miR844	-	1	-	-
miR858a	-	-	-	1
miR908	-	-	1	-
miR909	1	-	-	-
miR1144	-	-	-	1
miR1151b	-	-	-	1
miR1156	-	1	-	-
miR1158	-	-	-	1
miR1166	-	-	-	1
miR3932b	-	-	-	1
miR5017	1	-	-	-
miR5638b	2	1	-	-
miR5640	-	-	-	1
miR5645c	-	-	1	2
miR5645d	1	11	24	16
miR5653	-	1	-	-
miR5655	-	-	1	-
miR5658	-	-	5	8
miR5660	-	-	-	8
miR5998a	-	-	-	1
miR8170	-	-	-	1
miR8183	-	-	-	1

Differential expression analysis was performed to identify the important miRNA that involved in the gene regulation during stress conditions. T-test was used to assess the differential expression of miRNAs in both *C. sorokiniana* and *C. zofingiensis* miRNA-seq. Unfortunately, the miRNA genes that are differentially expressed are with respect to function, therefore it was not possible to name the gene and predict its function. However, the highest fold change of differentially expressed miRNA genes for *C. sorokiniana* was up to 7.16 for upregulated miRNA gene, and 9.77 for downregulated miRNA genes. Meanwhile, for *C. zofingiensis* the upregulated and downregulated miRNA gene was up to 10.43 fold change (**Appendices H and I**).

In order to gain insight into the biological role of each miRNA, it is essential to know the full repertoire of their targets. However, this is not an easy task as there were very limited numbers of miRNA targets that have been experimentally validated so far. In addition, the study of miRNA from microalgae are still in the beginning and not many databases are available, which mostly dependent on the miRNA database from plant such plant miRNA database (PMRD). One of the popular approaches to identify the specific miRNA targets is through in-silico prediction using computational techniques (Jones-Rhoades et al. 2004, Kuhn et al. 2008). However, one miRNA could target multiple mRNAs and certain mRNA can participate in multiple signalling pathways. Due to these difficulties, another improved method suggested by Lin et al. (2017) based on computer simulation program was reported. Thus, comprehensive methods are required to define each of the predicted miRNA from microalgae identified in this study and further studies would be carried out.

5.3.3 miRNA-seq data analysis by OmiRas

In the analysis using OmiRas, the small RNA-seq data were annotated via several databases such as miRBase, snoRNABase, Repbase, GtRNAdb, and rfam which detect mature-miRNA, pre-miRNA, rRNA, snoRNA and tRNA. From the results obtained, the average number of mature miRNA detected in *C. sorokiniana* for normal and stressed samples were 5 and 12, respectively. Whereas, the average number of mature miRNA found in normal and stressed samples of *C. zofingiensis* were 10 and 13, respectively (**Appendix J**). By using OmiRas analysis software, the other types of small RNA also could be detected, such as pre-miRNA, rRNA, snoRNA and tRNA.

From the data obtained, several mature miRNAs were identified in both strains in all conditions such as ath-miR-850, ath-miR-5638b and ath-miR-5656. The other miRNA that were found in both strains either in normal or stressed samples were ath-miR-851-3p (found in stress conditions only) and ath-miR-2111a (found in both conditions of *C. zofingiensis* and and stress condition of *C. sorokiniana*). The miRNA that were found in *C. sorokiniana* only include ath-miR-169m, ath-miR-399b, miR-399f, and ath-miR-3933 (in both conditions), whereas ath-miR-156e, ath-miR-169g*, ath-miR-830 and ath-miR-831 were found only in stress condition. There were no miRNA detected only in normal condition of *C. sorokiniana*. On the other hand, the miRNA that had been found

in normal and stressed samples of *C. zofingiensis* were ath-miR-168b, ath-miR-172b*, ath-miR-859, ath-miR-5022, ath-miR-5646 and ath-miR-5660. Meanwhile, ath-miR-156c, ath-miR-395a, ath-miR-830*, ath-miR-5644, ath-miR-5649a, ath-miR-5649b and ath-miR-5652 were found in *C. zofingiensis* normal sample, and ath-miR-156f, ath-miR-167a, ath-miR-168a, ath-miR-171a, ath-miR-1886.1, ath-miR-4240, ath-miR-5630a, ath-miR-5630b, and ath-miR-5659 were found in *C. zofingiensis* stressed sample only (**Table 5.4**).

Table 5.4: The average number of mature miRNA found in small RNA transcriptome of both *C. sorokiniana* and *C. zofingiensis* using OmiRas analysis.

Mature miRNA	*C. sorokiniana*		*C. zofingiensis*	
	Normal	**Stress**	**Normal**	**Stress**
ath-miR-156c	-	-	1	-
ath-miR-156e	-	5	-	-
ath-miR-156f	-	-	-	7
ath-miR-167a	-	-	-	3
ath-miR-168a	-	-	-	9
ath-miR-168b	-	-	3	14
ath-miR-169g*	-	3	-	-
ath-miR-169m	290	344	-	-
ath-miR-171a	-	-	-	25
ath-miR-172b*	-	-	1	1
ath-miR-395a	-	-	1	-
ath-miR-399b	16	10	-	-
ath-miR-399f	5	12	-	-
ath-miR-830	-	2	-	-
ath-miR-830*	-	-	1	-
ath-miR-831	-	9	-	-
ath-miR-850	1	5	24	1
ath-miR-851-3p	-	1	-	1
ath-miR-859	-	-	5	23
ath-miR-1886.1	-	-	-	1
ath-miR-2111a	-	4	15	9
ath-miR-3933	3	7	-	-
ath-miR-4240	-	-	-	3
ath-miR-5022	-	-	3	5
ath-miR-5630a	-	-	-	1
ath-miR-5630b	-	-	-	1
ath-miR-5638b	17	79	12	5
ath-miR-5644	-	-	1	-
ath-miR-5646	-	-	2	1
ath-miR-5649a	-	-	1	-
ath-miR-5649b	-	-	1	-
ath-miR-5652	-	-	1	-
ath-miR-5656	12	86	32	17
ath-miR-5659	-	-	-	1
ath-miR-5660	-	-	2	44

Subsequently, the differential expression analysis was performed on *C. sorokiniana* and *C. zofingiensis* small RNA-seq data. One candidate miRNA (ath-mir-5645e) was significantly upregulated in stress condition of *C. sorokiniana* with 5.705 fold change (**Table 5.5**). Unfortunately, there were no significant differentially expressed miRNA found in *C. zofingiensis* based on the p-value more than 0.05 (**Table 5.6**).

Table 5.5: Differential expression analysis of *C. sorokiniana* small RNA-seq data performed by OmiRas.

Id	Type	Normal	Stress	Log_2fc	P-value
ath-miR-169m	mature_miRNA	59.444	50.702	-0.229	1.000
ath-miR-5638b	mature_miRNA	0.000	2.040	inf	1.000
ath-miR-5656	mature_miRNA	1.986	2.185	0.138	1.000
ath-miR-831	mature_miRNA	0.000	0.253	inf	1.000
ath-mir-169m	pre_miRNA	59.444	51.679	-0.202	1.000
ath-mir-2111a	pre_miRNA	0.000	1.230	inf	1.000
ath-mir-396a	pre_miRNA	0.409	0.000	-inf	1.000
ath-mir-396b	pre_miRNA	0.409	0.405	-0.012	1.000
ath-mir-401	pre_miRNA	0.000	0.405	inf	1.000
ath-mir-5634	pre_miRNA	0.000	0.318	inf	1.000
ath-mir-5638b	pre_miRNA	0.000	2.040	inf	1.000
ath-mir-5645e	pre_miRNA	0.352	18.341	5.705	0.001
ath-mir-5652	pre_miRNA	1.402	1.787	0.350	1.000
ath-mir-5656	pre_miRNA	1.986	2.185	0.138	1.000
ath-mir-5657	pre_miRNA	0.000	0.253	inf	1.000
ath-mir-824	pre_miRNA	0.000	0.405	inf	1.000
ath-mir-831	pre_miRNA	0.000	0.253	inf	1.000
ath-mir-834	pre_miRNA	0.176	0.000	-inf	1.000

Table 5.6: Differential expression analysis of *C. zofingiensis* small RNA-seq data performed by OmiRas.

Id	Type	Normal	Stress	Log_2fc	P-value
ath-miR-171a	mature_miRNA	0.000	1.388	inf	0.627
ath-miR-172b*	mature_miRNA	0.203	0.230	0.184	1.000
ath-miR-2111a	mature_miRNA	0.540	0.000	-inf	0.892
ath-miR-5638b	mature_miRNA	0.472	0.000	-inf	0.892
ath-miR-5656	mature_miRNA	1.286	0.645	-0.995	0.976
ath-miR-5660	mature_miRNA	0.000	2.545	inf	0.501
ath-miR-850	mature_miRNA	2.167	0.000	-inf	0.667
ath-miR-859	mature_miRNA	0.000	0.230	inf	1.000
ath-mir-171a	pre_miRNA	0.000	1.388	inf	0.627
ath-mir-172b	pre_miRNA	0.203	0.230	0.184	1.000
ath-mir-2111a	pre_miRNA	1.623	0.000	-inf	0.641
ath-mir-399d	pre_miRNA	0.000	3.979	inf	0.316
ath-mir-5634	pre_miRNA	0.000	0.230	inf	1.000
ath-mir-5638b	pre_miRNA	0.472	0.000	-inf	0.892
ath-mir-5645d	pre_miRNA	0.000	0.230	inf	1.000
ath-mir-5652	pre_miRNA	9.330	2.776	-1.749	0.667
ath-mir-5656	pre_miRNA	1.286	0.645	-0.995	0.976
ath-mir-5660	pre_miRNA	0.000	2.545	inf	0.501
ath-mir-835	pre_miRNA	2.099	0.000	-inf	0.667
ath-mir-850	pre_miRNA	2.167	0.000	-inf	0.667
ath-mir-853	pre_miRNA	0.000	0.460	inf	0.892
ath-mir-859	pre_miRNA	0.000	0.230	inf	1.000

5.3.4 Novel miRNA prediction

Several novel miRNA were identified through OmiRas analysis. Nine small RNA sequences folded into a perfect hairpin structure with folding energy <-20 kcal/mol (**Table 5.7**) were found in *C. sorokiniana* small RNA transcriptome. **Figure 5.2** shows the secondary structure predicted for novel *C. sorokiniana* miRNA.

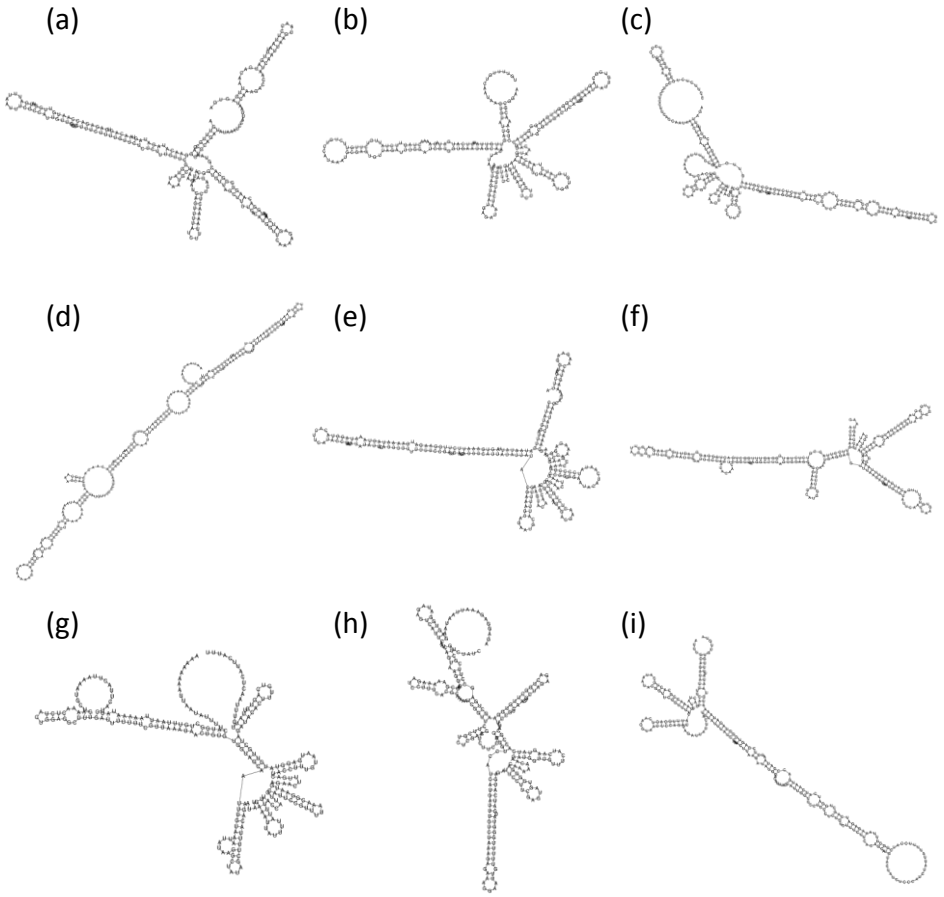


Figure 5.2: Secondary structure prediction of novel *C. sorokiniana* miRNA precursors, (a) 3122, (b) 2166, (c) 457, (d) 3238, (e) 384, (f) 2363, (g) 2684, (h) 3027, (i) 3183.

Table 5.7: Novel putative miRNAs in *C. sorokiniana* with their mapping coordinates and sequence of the predicted miRNA precursor (primary sequence) and mature sequence.

Id	Start	Stop	Strand	Energy	Primary sequence	Mature sequence
3122	20566411	20566678	-	-55.8	AGTTTGCGAGGAACATAGATAGTAA CAGTGAGAGACCAATGTTGGAACGA TTTAATTTCGCCATTGTTGTTCTGTT TCTTTTGTTCTCTCTTTGTTCGGTAT AAATGCTTAAGTACAAATGAGTTTTA TCTTTAAAACTCTCATGGTGTTAGCT TTTCACATGGTTCAAAAGAGATCAAT AATGGACAAGTGGGCACGTAACTGC AGATATAAATCCATTAATCAAAGACC AATCAATTAAGTAACCTTAATTTGAT GACAATTGAGTGA	ACAGTGAGAGACCAA
2166	11167827	11168094	-	-61.8	TGTTTTGTCAATAAGTGTATTAGTTG TTAAATTCATGATAGAAAACTAGTTC TATTGTCTTGTTCTGTCCATAAATGC GAGTGTTAGATTTGGATCGTGGTGA ATAGCAATATTATATCTAGTGATGAG ATAATCATTGGAGTTTCTACTGTGAT TTCTTGATCATAATGTTTCGATAATG GTGTGTATCTAACATGTAACATGTAA TGCGTTGACTCATGGCAATTATTGT GGCAGTTGGTGGGTTGATGGGTTAC CTGAAACGCGGAA	GTTGGTGGGTTGAT
3238	3009639	3009906	-	-59	AATAACAACCTTCTCGTAATATATCT	GTGAATTGCAGAAAC

					CAGGATACTTCGCAGCAACTTCATC ACAACACTGCAAAGCAGGAAGCAAA TAAAAACTTGTCTCAGAACAAAGGTT AAGAAAAAGAAGCCAGTGAGAGACT AGAAAAATTGCTATGTTCCCCTAATT GAGAAGTGCTAATGTATTCTGAGTT CCAAAATGAGCTGACCTGCAGGAAA AGACCATCAGTTTTCTGCATAATGTT GGCTTTGTGAATTGCAGAAACTTTCT TCCTTCCGTGAGTCT	
457	5972912	5973181	+	-49.4	CCTTCTCTGCTCTGTCTTCCCCGTG CAAAGCCTCCTCCTTCACCAATAAA CTACATTTATTAGCCTTTTATAAAAT CACTGGAATGGCGTTTAGCGTTTCT CACTTTAAGTAAGTTAAGTACCTCCT CCTTCTCGGATGTTACGAGCCAGTC AGGGAGCAACCGATCCAGCAAAGC GTGGTTAAGAGTTGGTGGTGGTTCA TCCTCATAATGTAGCCGTTGCTTCA GCTGTTTTTCTCTGAGCCGTAAGTT GTAATGAACGTATGTTAGA	TCTCTGAGCCGTAA
2363	4784899	4785166	-	-65	ACTAAGTTCCTTAGCGTCTCTATGG GAGTTTGGCTGACTGGGAGGTAGTT GACAAGTATGCTTATGGGGTTTGGC AATTGGATTCCTAGATACAACTCGG TTGGAGGCTTTAGTGGAGGTAGGAA AGGCATGGTTTTCCAAATGAGGAGG	GGAGTTTGGCTGAC

					TTTAAGACGGATTCGTGTATTCAACT AAGTGTGTTTGACATGAGTTTTAAAG GGAAGAATAACCATCAATGTAAGTG GTTGGGTAACGTTGATGTATATTTAG TGCATAATCTTAGTAAG	
384	6784701	6784970	+	-54.4	CTTATGGATTTGATAATATTAGTGAA ATTCTTGGAAGATGAAAATAGTAAAT AGGTACACTTCATACCTTAACATATA TTCCATTTTTTTTTTCCGACTGAATG ATTTCATTATATTATACGAAAGTAGA ATGCATTTTTGAAGTACAACGACGA CAAGAGACGAACTCTGTCGGGAACT AAGCCGGCGGAATACATCAATATCC CCGGAAACCTAAGCCCAAAAGAGG GAAGCAAACTATAAGCATAAAAGTA CAAGAAGGTAAGCTTA	TGAAATTCTTGGAAG
2684	16123977	16124244	-	-46.9	AAAAAAAATTAATATTTTCATGGCTG TTTAATTAAAAATATCTACTTACTTAA AATGCTGAAAACTCTACTGGAGGCT TTGATTTTTTTCGTTAAATAACTGTG TGGTAGCAATTTGTAAATTATAAGGC TATAGCTTTTACAGAATATGAAATTA TTTTATTTTTACCATTGCATTCCGTTT TAAACGGAATGAGAACTTTGACTAC CTTTGTAATGAGGTATGTTGCTAGAT ACACTTTGTCAGTGTTATTCTTTAAC	GCTGTTTAATTAAAAA

					TCATTCATTT	
3027	10354962	10355229	-	-60.6	AGAGGAGAAATTAATAAGAGTGGAT AGAGTCATTCAATCAGTAAGGGAAG CGAAAAAGATCCTTTCACGCAAAAA TCGGTTCTTCGTCTCGAAGCTCTTC CAGTGGTCCATCATGATTTGTCTAAA GAGAAAGGAATTGTTTTTGGGTATG ATCATGAGATTCCTTCGAGTAGTGG GGACAAAAACAAGATGGTTGTTTCA TCAAGTATCACCGCATCTCCGTCAA CTGCTAGAGACTGGCTTGTGGAAGC TGAGCCTTCGCTGTCTATC	CTGGCTTGTGGAAG
3183	8081968	8082235	-	-51.8	ATTTGCAGATATTTTCATTACTATTG AAACCGCACTGCATAATACGGTTCG GTTTTCATTTGGAGTCAAGGTAATTC CAGATTAGATTATTCATTTATTTATC CAAAAACTTAAGATATGGACATTTAG CATTGGTTCCTCCCTCTGTCTTTCTT CACAGCATGTACTATCTCCCATCTAT TATTTTCCATCGGCTTTAAAGAGATC AGCTGGATCATCAATGACCCATAAG CTCGAGTGGGTAGATAAAATGTAAT ATTTTGTGTTGTA	GAGTGGGTAGATAA

5.3.5 Role of miRNA

Based on Qin et al. (2014), most of the well-studied and functionally known miRNA are conserved miRNAs which designated with identification numbers from miR156 to miR408. The other miRNA which are unique to specific species were considered non-conserved miRNA. They are usually imprecisely processed, weakly expressed, and lack known functional targets. However, some of non-conserved miRNA may exhibit higher expression levels under particular conditions, resulting from environmental adaptations (Qin et al. 2014; Jia et al. 2014). In this study, most of the identified miRNA were from the non-conserved group of miRNA and were expressed at very low levels.

On the other hand, several miRNAs found in the *C. sorokiniana* and *C. zofingiensis* small RNA transcriptome analyses are involved in the UV-B radiation stress response. Previous studies have shown that the miRNA families that were upregulated during UV-B stress are miR156/157, miR159/319, miR160, miR165/166, miR167, miR169, miR170/171, miR172, miR393, miR398 and miR401 (Zhou et al. 2007). All these families except miR165/166, miR393, miR398 and miR401, were found in *C. sorokiniana* and *C. zofingiensis* small RNA transcriptome. These miRNAs have been found to target various genes encoding transcription factors that subsequently direct the expression of related genes (Guleria et al. 2011). This finding is consistent with the treatment given for stress condition in which, higher light intensity was introduced to the stress samples used in this study.

Based on the specific role of each miRNA involved in adaptive response on UV-B stress, the identification of miR156 which functions in regulating the carotenoids production was proportional with the result obtained from RNA-seq analysis (**Chapter 4**). From the RNA-seq results, several enzymes involved in the production of carotenoid precursor were upregulated [YgbP (or IspD) and GcpE (or IspG)], suggesting that they may contain the SQUAMOSA PROMOTER BINDING PROTEIN-LIKE (SPL) transcription factor that targeted by miR156 (Wei et al. 2012). Further evidence can be found from our result in **Chapter 3**, in which the carotenoid contents were increased by about 30 to 40 % in both *Chlorella* species when stress was induced to the culture.

The other miRNAs with known function found in this study are miR395 and miR399. These miRNA families were reported to be involved in the response of *Arabidopsis thaliana* to nutrient stresses. MiR395 was upregulated during salt stress, causing suppression of the respective factors (Guleria et al. 2011). Meanwhile, miR399 families were involved in the regulation the phosphate (Pi) homeostasis and targeting multiple sites on the 5′ untranslated region (UTR) of At2g33770 mRNA encoding a ubiquitin-conjugating E2 enzyme, a component in the ubiquitin-dependent protein degradation pathway (Chiou 2007). Besides being major components of nucleic acid and membrane phospholipids, phosphate also plays an important role in energy transfer and the regulation of enzyme reactions and metabolic pathways.

5.4 Conclusion

Small RNA transcriptome of *C. sorokiniana* and *C. zofingiensis* have been developed which focus on miRNA content. From the analysis, about 35 to 37 of conserved and non-conserved miRNA had been identified from both *Chlorella* species under normal and stress conditions. Besides, novel miRNAs from *C. sorokiniana* had also been identified, suggesting that a significant number of novel miRNAs remain to be discovered and characterised. These results may help to improve our understanding of regulatory mechanisms of miRNA in biological and metabolic process of *Chlorella* species.

CHAPTER 6

CONCLUSION AND FUTURE WORKS

6.1 Conclusion

In this study, transcriptional profiling of *C. sorokiniana* and *C. zofingiensis* has been carried out under normal and stress induced conditions. Stress factor used in this study aimed to induce production of carotenoids and lipid in these microalgae species. Prior to transcriptional profiling, biochemical analysis of *Chlorella* under normal and stress conditions was undertaken. In that **Chapter 3**, this study demonstrated that the microalgae showed different pattern of metabolite production under normal and stress conditions. Under stress conditions, both *Chlorella* spp. produced more carotenoid contents compared to phenolic content. Besides, the accumulation of lipid molecule predominant the microalgal cells were clearly shown by microscopic analysis. Different levels of metabolite content (phenols, pigments and lipids) as well as antioxidant content, under different culture conditions pave the way of further exploitation of this microalgae for various application. Furthermore, manipulation of microalgae culture conditions is critical for the production of desired metabolites in a specific microalgal species.

Following biochemical characterisation under normal and stress conditions, the transcriptome through RNA-seq analysis was profiled. Every candidate gene and enzyme involved in the core microalgal metabolic pathways was elucidated. The findings point to several molecular mechanisms that potentially drive the overproduction of high value metabolites such as carotenoid and fatty acid under stress induced conditions. The identified gene sequences and measured metabolic responses during stress condition can be used in future metabolic engineering studies to improve the carotenoid content and lipid production in this microalga that make it suitable for commercial development.

Subsequently, in order to achieve complete profiling of genetic information for transcriptome of *Chlorella* species under stress related condition, the identification of miRNA; a key gene expression regulator was examined through a small RNA-seq study. From the results obtained, most of the identified miRNA were from non-conserved miRNA. Although the processing and functions of miRNAs remain to be elucidated, this finding is a step forward in understanding the complex network of small RNAs operating to maintain the unique characteristics of microalgae. The corcordance of metabolite measurements and observed physiological response with transcriptomic data showed in this study may provide a more complete picture of gene regulation and expression which can be used in future genetic manipulation of this microalgae.

6.2 Future works

6.2.1 The improvement of transcriptome development

In this study, there were difficulties in developing the transcriptome of *C. zofingiensis.* It may require more troubleshooting of analysis strategies to find the problem associated with the de novo assembly steps. Further analysis could be carried out to identify the problem associated with the unsuccessful development of *C. zofingiensis* transcriptome, especially in the bioinformatics analysis part. It is advantageous, if *C. zofingiensis* transcriptome could be obtained. Comparative transcriptome between two *Chlorella* species under the same stress condition could be analysed, and the key features of controlling the regulation of metabolite production during particular stress could be elucidated. The strategies employed by Feldmesser et al. (2014) which used integrated technology of de novo and genome based assembly, and EST clustering in constructing *E. huxleyi* transcriptome might be one of strategies for improving the *Chlorella* transcriptome. Subsequently, two methods: DETONATE (Li and Dewey 2011) and TransRate (Smith-Unna et al. 2016) could be applied for assessing the contiguity of the transcriptome assembly.

6.2.2 Identification of novel gene encoding protein from non-annotated results

Furthermore, a systems biology approach is a useful method for building a comprehensive understanding of the biological roles of a gene in a cell. This approach works the best when the genome with a relatively complete gene annotation are available. Since, significant number of the *C. sorokiniana* and *C. zofingiensis* genes do not have well annotated protein domains or orthologous genes from closely related species, different strategy to reannotate these genes are necessary. Hence, the data can be used to further improve the annotation of *C. sorokiniana* and *C. zofingiensis* by clustering of genes with similar expression profiles.

6.2.3 Integration between transcriptomic and proteomic study

The Central Dogma of Molecular Biology suggests that the transcriptome is substantially correlated to the proteome. But the correlation could be compromised by the post-translational modifications and various protein stabilisation mechanisms found in Nature. Although this association has not been examined comprehensively in microalgae, previous studies in other organisms such as yeast suggest an important but modest correlation between the levels of transcripts and the levels of proteins. Hence, this transcriptome study may benefit from a study that examines proteome of *C. sorokiniana* and *C. zofingiensis* under the same conditions.

REFERENCES

Abdel-Raouf, N., A. Al-Homaidan & I. Ibraheem (2012) Microalgae and wastewater treatment. *Saudi Journal of Biological Sciences,* 19**,** 257-275.

Adams, J. (2008) Transcriptome: connecting the genome to gene function. *Nature Education,* 1**,** 195.

Adelfi, M. G., M. Borra, R. Sanges, M. Montresor, A. Fontana & M. I. Ferrante (2014) Selection and validation of reference genes for qPCR analysis in the pennate diatoms *Pseudo-nitzschia multistriata* and *P. arenysensis*. *Journal of Experimental Marine Biology and Ecology,* 451**,** 74-81.

Ahn, J.-W., K. Hwangbo, S. Y. Lee, H.-G. Choi, Y.-I. Park, J. R. Liu & W.-J. Jeong (2012) A new Arctic *Chlorella* species for biodiesel production. *Bioresource Technology,* 125**,** 340-343.

Alberts, B., A. Johnsen & J. Lewis (2002) *From RNA to Protein*. New York: Garland Science.

Allen, A. E., A. Vardi & C. Bowler (2006) An ecological and evolutionary context for integrated nitrogen metabolism and related signaling pathways in marine diatoms. *Current Opinion in Plant Biology,* 9**,** 264-273.

Armbrust, E. V., J. A. Berges, C. Bowler, B. R. Green, D. Martinez, N. H. Putnam, S. Zhou, A. E. Allen, K. E. Apt, M. Bechner, M. A. Brzezinski, B. K. Chaal, A. Chiovitti, A. K. Davis, M. S. Demarest, J. C. Detter, T. Glavina, D. Goodstein, M. Z. Hadi, U. Hellsten, M. Hildebrand, B. D. Jenkins, J. Jurka, V. V. Kapitonov, N. Kroger, W. W. Lau, T. W. Lane, F. W. Larimer, J. C. Lippmeier, S. Lucas, M. Medina, A. Montsant, M. Obornik, M. S. Parker, B. Palenik, G. J. Pazour, P. M. Richardson, T. A. Rynearson, M. A. Saito, D. C. Schwartz, K. Thamatrakoln, K. Valentin, A. Vardi, F. P. Wilkerson & D. S. Rokhsar (2004) The genome of the diatom Thalassiosira pseudonana: ecology, evolution, and metabolism. *Science,* 306**,** 79-86.

Arnao, M. B. (2000) Some methodological problems in the determination of antioxidant activity using chromogen radicals: a practical case. *Trends in Food Science & Technology,* 11**,** 419-421.

Arya, M., I. S. Shergill, M. Williamson, L. Gommersall, N. Arya & H. R. Patel (2005) Basic principles of real-time quantitative PCR. *Expert Review of Molecular Diagnostics,* 5**,** 209-219.

Avagyan, A (2008) Microalgae production development global prospects and profitable technology wasterwater purification by the use microalgae. In *Water and Wastewater International*.

Baba, M. & Y. Shiraiwa (2013) Biosynthesis of lipids and hydrocarbons in algae. In *Photosynthesis,* ed. Z. Dubinsky, 331-356. Croatia: InTech.

Baker, M. (2010) Next-generation sequencing: adjusting to data overload. *Nature Methods,* 7**,** 495-499.

Ball, S., H.-P. Guan, M. James, A. Myers, P. Keeling, G. Mouille, A. Buléon, P. Colonna & J. Preiss (1996) From glycogen to amylopectin: a model for the biogenesis of the plant starch granule. *Cell,* 86**,** 349-352.

Ballicora, M. A., A. A. Iglesias & J. Preiss (2003) ADP-glucose pyrophosphorylase: a regulatory enzyme for bacterial glycogen synthesis. *Microbiology and Molecular Biology Reviews,* 67**,** 213-225.

Barclay, W., K. Meager & J. Abril (1994) Heterotrophic production of long chain omega-3 fatty acids utilizing algae and algae-like microorganisms. *Journal of Applied Phycology,* 6**,** 123-129.

Bartel, D. P. (2004) MicroRNAs-genomics, biogenesis, mechanism, and function. *Cell,* 116**,** 281-297.

Battchikova, N., E.-M. Aro & P. J. Nixon (2011) Structure and physiological function of NDH-1 complexes in cyanobacteria. In *Bioenergetic Processes of Cyanobacteria,* eds. G. A. Peschek, C. Obinger & G. Renger, 445-467. Springer.

Becker, E. W. (2007) Micro-algae as a source of protein. *Biotechnology Advances,* 25**,** 207-10.

Benemann, J. R. (1992) Microalgae aquaculture feeds. *Journal of Applied Phycology,* 4**,** 233-245.

Berg, J., J. Tymoczko & L. Stryer (2002a) Glycogen breakdown requires the interplay of several enzymes. In *Biochemistry,* ed. W. H. Freeman. New York.

Berg, J., J. Tymoczko & L. Stryer (2002b) The glycolytic pathway is tightly controlled. In *Biochemistry,* ed. W. H. Freeman. New York.

Berthelot, K., Y. Estevez, A. Deffieux & F. Peruch (2012) Isopentenyl diphosphate isomerase: A checkpoint to isoprenoid biosynthesis. *Biochimie,* 94**,** 1621-1634.

Blanc, G., I. Agarkova, J. Grimwood, A. Kuo, A. Brueggeman, D. D. Dunigan, J. Gurnon, I. Ladunga, E. Lindquist, S. Lucas, J. Pangilinan, T. Proschold, A. Salamov, J. Schmutz, D. Weeks, T. Yamada, A. Lomsadze, M. Borodovsky, J. M. Claverie, I. V. Grigoriev & J. L. Van Etten (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biology,* 13**,** R39.

Blanc, G., G. Duncan, I. Agarkova, M. Borodovsky, J. Gurnon, A. Kuo, E. Lindquist, S. Lucas, J. Pangilinan & J. Polle (2010) The *Chlorella*

variabilis NC64A genome reveals adaptation to photosymbiosis, coevolution with viruses, and cryptic sex. *The Plant Cell Online,* 22**,** 2943-2955.

Bock, C., L. Krienitz & T. Proschold (2011) Taxonomic reassessment of the genus Chlorella (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea,* 11**,** 293-312.

Boldt, L., D. Yellowlees & W. Leggat (2008) Measuring *Symbiodinium* sp. gene expression patterns with quantitative real-time PCR. *Proceedings of the 11th ICRS***,** 118-122.

Borowitzka, M. A. (1995) Microalgae as sources of pharmaceuticals and other biologically active compounds. *Journal of Applied Phycology,* 7**,** 3-15.

Borowitzka, M. A. (1999) Commercial production of microalgae: ponds, tanks, tubes and fermenters. *Journal of Biotechnology,* 70**,** 313-321.

Boycott, K. M., M. R. Vanstone, D. E. Bulman & A. E. MacKenzie (2013) Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nature Reviews Genetics,* 14**,** 681-691.

Brody, M. & A. E. Vatter (1959) Observations on cellular structures of *Porphyridium cruentum*. *The Journal of Biophysical and Biochemical Cytology,* 5**,** 289-293.

Brown, T. A (2006) Genomes. 3-30. Garland science.

Busi, M. V., J. Barchiesi, M. Martín & D. F. Gomez-Casati (2014) Starch metabolism in green algae. *Starch-Stärke,* 66**,** 28-40.

Bustin, S., H. S. Dhillon, S. Kirvell, C. Greenwood, M. Parker, G. L. Shipley & T. Nolan (2015) Variability of the Reverse Transcription Step: Practical Implications. *Clinical chemistry,* 61**,** 202-212.

Cadoret, J.-P., M. Garnier & B. Saint-Jean (2012) Microalgae, functional genomics and biotechnology. *Advances in Botanical Research,* 64**,** 285-341.

Cagliari, A., R. Margis, F. dos Santos Maraschin, A. C. Turchetto-Zolet, G. Loss & M. Margis-Pinheiro (2011) Biosynthesis of triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology,* 2**,** 10.

Cardozo, K. H., T. Guaratini, M. P. Barros, V. R. Falcão, A. P. Tonon, N. P. Lopes, S. Campos, M. A. Torres, A. O. Souza & P. Colepicolo (2007) Metabolites from algae with economical impact. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology,* 146**,** 60-78.

Carrington, J. C. & V. Ambros (2003) Role of microRNAs in plant and animal development. *Science,* 301**,** 336-338.

Carvalho, A. P., S. O. Silva, J. M. Baptista & F. X. Malcata (2011) Light requirements in microalgal photobioreactors: an overview of biophotonic aspects. *Applied Microbiology and Biotechnology,* 89**,** 1275-1288.

Catalanotti, C., W. Yang, M. C. Posewitz & A. R. Grossman (2013) Fermentation metabolism and its evolution in algae. *Frontiers in Plant Science,* 4.

Cerutti, H., X. Ma, J. Msanne & T. Repas (2011) RNA-mediated silencing in algae: biological roles and tools for analysis of gene function. *Eukaryotic cell,* 10**,** 1164-1172.

Chacón-Lee, T. L. & G. E. González-Mariño (2010) Microalgae for "healthy" foods—Possibilities and challenges. *Comprehensive Reviews in Food Science and Food Safety,* 9**,** 655-675.

Chen, C.-Y., K.-L. Yeh, R. Aisyah, D.-J. Lee & J.-S. Chang (2011) Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: a critical review. *Bioresource Technology,* 102**,** 71-81.

Chen, C.-Y., X.-Q. Zhao, H.-W. Yen, S.-H. Ho, C.-L. Cheng, D.-J. Lee, F.-W. Bai & J.-S. Chang (2013) Microalgae-based carbohydrates for biofuel production. *Biochemical Engineering Journal,* 78**,** 1-10.

Chen, C. Y. & D. G. Celia (1994) Effects of pH on the growth and carbon uptake of marine phytoplankton. *Marine Ecology Progress Series,* 109**,** 83-94.

Cheng, D. & Q. He (2014) Assessment of environmental stresses for enhanced microalgal biofuel production—an overview. *Frontiers in Energy Research,* 2**,** 1-26.

Chia, M. A., A. T. Lombardi & G. Melao Mda (2013) Growth and biochemical composition of *Chlorella vulgaris* in different growth media. *Annals of the Brazilian Academy of Sciences,* 85**,** 1427-1438.

Chiou, T.-J. (2007) The role of microRNAs in sensing nutrient stress. *Plant, Cell & Environment,* 30**,** 323-332.

Chisti, Y. (2007) Biodiesel from microalgae. *Biotechnology advances,* 25**,** 294-306.

Chokshi, K., I. Pancha, K. Trivedi, B. George, R. Maurya, A. Ghosh & S. Mishra (2015) Biofuel potential of the newly isolated microalgae *Acutodesmus dimorphus* under temperature induced oxidative stress conditions. *Bioresource Technology,* 180**,** 162-171.

Chung, C.-C., S.-P. L. Hwang & J. Chang (2008) Nitric oxide as a signaling factor to upregulate the death-specific protein in a marine diatom, *Skeletonema costatum*, during blockage of electron flow in photosynthesis. *Applied and Environmental Microbiology,* 74**,** 6521-6527.

Cifuentes, A. S., M. A. González, I. Inostroza & A. Aguilera (2001) Reappraisal of physiological attributes of nine strains of *Dunaliella* (Chlorophyceae): growth and pigment content across a salinity gradient. *Journal of Phycology,* 37**,** 334-344.

Colleoni, C., D. Dauvillée, G. Mouille, M. Morell, M. Samuel, M.-C. Slomiany, L. Liénard, F. Wattebled, C. d'Hulst & S. Ball (1999) Biochemical characterization of the *Chlamydomonas reinhardtii* α-1, 4 glucanotransferase supports a direct function in amylopectin biosynthesis. *Plant Physiology,* 120**,** 1005-1014.

Conesa, A., P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo & X. Zhang (2016) A survey of best practices for RNA-seq data analysis. *Genome Biology,* 17**,** 19.

Cordero, B. F., I. Obraztsova, I. Couso, R. Leon, M. A. Vargas & H. Rodriguez (2011) Enhancement of lutein production in *Chlorella sorokiniana* (Chorophyta) by improvement of culture conditions and random mutagenesis. *Marine Drugs,* 9**,** 1607-1624.

Cox, S., N. Abu-Ghannam & S. Gupta (2010) An assessment of the antioxidant and antimicrobial activity of six species of edible Irish seaweeds. *International Food Research Journal* 17**,** 205-220.

Crane, K. W. & J. P. Grover (2010) Coexistence of mixotrophs, autotrophs, and heterotrophs in planktonic microbial communities. *Journal of Theoretical Biology,* 262**,** 517-527.

D'haene, B. & J. Hellemans (2010) The importance of quality control during qPCR data analysis. *International Drug Discovery***,** 18-24.

da Silva Gorgônio, C. M., D. A. G. Aranda & S. Couri (2013) Morphological and chemical aspects of *Chlorella pyrenoidosa*, *Dunaliella tertiolecta, Isochrysis galbana* and *Tetraselmis gracilis* microalgae. *Natural Science* 5**,** 783-791

Dalrymple, O. K., T. Halfhide, I. Udom, B. Gilles, J. Wolan, Q. Zhang & S. Ergas (2013) Wastewater use in algae production for generation of renewable resources: a review and preliminary results. *Aquatic Biosystems,* 9**,** 2.

de Morais, M. G., B. d. S. Vaz, E. G. de Morais & J. A. V. Costa (2015) Biologically active metabolites synthesized by microalgae. *BioMed Research International*.

Del Campo, J. A., H. Rodriguez, J. Moreno, M. A. Vargas, J. Rivas & M. G. Guerrero (2004) Accumulation of astaxanthin and lutein in *Chlorella zofingiensis* (Chlorophyta). *Applied Microbiology Biotechnology,* 64**,** 848-854.

Derelle, E., C. Ferraz, P. Lagoda, S. Eychenié, R. Cooke, F. Regad, X. Sabau, C. Courties, M. Delseny & J. Demaille (2002) DNA Libraries for Sequencing the Genome of *Ostreococcus tauri* (Chlorophyta, Pasinophyceae): The smallest Free-Living Eukarytoic Cell. *Journal of phycology,* 38**,** 1150-1156.

Dewick, P. M. (2002) The biosynthesis of C 5–C 25 terpenoid compounds. *Natural Product Reports,* 19**,** 181-222.

Dhull, N. P., R. Soni, D. K. Rahi & S. K. Soni (2014) Evaluation of autotrophic and mixotrophic regimen *Chlorella pyrenoidosa* cells in various wastes water for its biochemical composition and biomass production. *Peer Journal,* 2.

Dittami, S. M., I. Riisberg, U. John, R. J. Orr, K. S. Jakobsen & B. Edvardsen (2012) Analysis of expressed sequence tags from the marine microalga *Pseudochattonella farcimen* (Dictyochophyceae). *Protist,* 163**,** 143-161.

Doan, T. T. Y., B. Sivaloganathan & J. P. Obbard (2011) Screening of marine microalgae for biodiesel feedstock. *Biomass and Bioenergy,* 35**,** 2534-2544.

Dong, M., X. Zhang, X. Chi, S. Mou, J. Xu, D. Xu, W. Wang & N. Ye (2012) The validity of a reference gene is highly dependent on the experimental conditions in green alga *Ulva linza*. *Current Genetics,* 58**,** 13-20.

Dongre, S. K., S. Manglawat, P. Singh, M. Yadav & A. Tiwari (2014) Effect of environmental parameters enhancing the microalgal lipid as a sustainable energy source for biodiesel production-A Review. *International Journal of Biological & Pharmaceutical Research,* 5**,** 327-335.

Dubinsky, Z. & N. Stambler (2009) Photoacclimation processes in phytoplankton: mechanisms, consequences, and applications. *Aquatic Microbiol Ecology,* 56**,** 163-176.

Ducat, D. C. & P. A. Silver (2012) Improving carbon fixation pathways. *Current Opinion in Chemical Biology,* 16**,** 337-344.

Eckardt, N. A. (2010) The *Chlorella* genome: big surprises from a small package. *Plant Cell,* 22**,** 2924.

Ekblom, R., & Galindo, J. (2011). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107(1), 1–15. http://doi.org/10.1038/hdy.2010.152

Elsabahy, M., A. Nazarali & M. Foldvari (2011) Non-viral nucleic acid delivery: key challenges and future directions. *Current Drug Delivery,* 8**,** 235-244.

Eussen, S., O. Klungel, J. Garssen, H. Verhagen, H. van Kranen, H. van Loveren & C. Rompelberg (2010) Support of drug therapy using functional foods and dietary supplements: focus on statin therapy. *British Journal of Nutrition,* 103**,** 1260-1277.

Fan, J., Y. Cui, M. Wan, W. Wang & Y. Li (2014) Lipid accumulation and biosynthesis genes response of the oleaginous *Chlorella pyrenoidosa* under three nutrition stressors. *Biotechnology for Biofuels,* 7**,** 1-14.

Fan, J., K. Ning, X. Zeng, Y. Luo, D. Wang, J. Hu, J. Li, H. Xu, J. Huang & M. Wan (2015) Genomic Foundation of Starch-to-Lipid Switch in Oleaginous *Chlorella* spp. *Plant Physiology,* 169**,** 2444-2461.

Feldmesser, E., S. Rosenwasser, A. Vardi & S. Ben-Dor (2014) Improving transcriptome construction in non-model organisms: integrating manual and automated gene definition in *Emiliania huxleyi*. *BMC Genomics,* 15**,** 1-16.

Feng, P., Z. Deng, L. Fan & Z. Hu (2012) Lipid accumulation and growth characteristics of *Chlorella zofingiensis* under different nitrate and phosphate concentrations. *Journal of Bioscience and Bioengineering,* 114**,** 405-410.

Fields, M. W., A. Hise, E. J. Lohman, T. Bell, R. D. Gardner, L. Corredor, K. Moll, B. M. Peyton, G. W. Characklis & R. Gerlach (2014) Sources and resources: importance of nutrients, resource allocation, and ecology in microalgal cultivation for lipid accumulation. *Applied Microbiology and Biotechnology,* 98**,** 4805-4816.

Fillmore, N., O. A. Alrob & G. D. Lopaschuk (2011) Fatty Acid beta-Oxidation.

Fleige, S., & Pfaffl, M. W. (2006). RNA integrity and the effect on the real-time qRT-PCR performance. *Molecular aspects of medicine,* 27(2), 126-139.

Fischer, B. B., A. Krieger-Liszkay, É. Hideg, I. Šnyrychová, M. Wiesendanger & R. I. Eggen (2007) Role of singlet oxygen in chloroplast to nucleus retrograde signaling in *Chlamydomonas reinhardtii*. *FEBS Letters,* 581**,** 5555-5560.

Fryera, R. M., J. Randalla, T. Yoshidaa, L.-L. Hsiao, J. Blumenstock, K. E. Jensen, T. Dimofte, R. V. Jensen & S. R. Gullans (2002) Global Analysis of Gene Expression: Methods, Interpretation, and Pitfalls. *Experimental Nephrology,* 10**,** 64-74.

Fu, X., D. Wang, X. Yin, P. Du & B. Kan (2014) Time course transcriptome changes in *Shewanella algae* in response to salt stress. *PloS one,* 9**,** e96001.

Fučíková, K. & L. A. Lewis (2012) Intersection of *Chlorella, Muriella and Bracteacoccus*: Resurrecting the genus *Chromochloris* KOL et CHODAT (Chlorophyceae, Chlorophyta). *Fottea,* 12**,** 83-93.

Gao, C., Y. Wang, Y. Shen, D. Yan, X. He, J. Dai & Q. Wu (2014) Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics,* 15.

Gallardo-Escárate, C., Valenzuela-Muñoz, V., & Nuñez-Acuña, G. (2014). RNA-Seq analysis using *de novo* transcriptome assembly as a reference for the salmon louse *Caligus rogercresseyi*. *PloS one,* 9(4), e92239.

George, B., I. Pancha, C. Desai, K. Chokshi, C. Paliwal, T. Ghosh & S. Mishra (2014) Effects of different media composition, light intensity and photoperiod on morphology and physiology of freshwater microalgae *Ankistrodesmus falcatus*–A potential strain for biofuel production. *Bioresource Technology,* 171**,** 367-374.

Giorgi, F. M., C. Del Fabbro & F. Licausi (2013) Comparative study of RNA-seq-and Microarray-derived coexpression networks in *Arabidopsis thaliana*. *Bioinformatics,* 29**,** 717-724.

Goiris, K., K. Muylaert, I. Fraeye, I. Foubert, J. De Brabanter & L. De Cooman (2012) Antioxidant potential of microalgae in relation to their phenolic and carotenoid content. *Journal of Applied Phycology,* 24**,** 1477-1486.

Goiris, K., W. Van Colen, I. Wilches, F. León-Tamariz, L. De Cooman & K. Muylaert (2015) Impact of nutrient stress on antioxidant production in three species of microalgae. *Algal Research,* 7**,** 51-57.

Gouveia, L., V. Veloso, A. Reis, H. Fernandes, J. Novais & J. Empis (1996) Evolution of pigment composition in *Chlorella vulgaris*. *Bioresource Technology,* 57**,** 157-163.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury & Q. Zeng (2011a) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology,* 29.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury & Q. Zeng (2011b) Trinity-reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology,* 29**,** 644-652.

Graham, L. E. & L. W. Wilcox (1999) *Algae*. Upper Saddle River, New Jersey: Prentice Hall.

Grünewald, K., J. Hirschberg & C. Hagen (2001) Ketocarotenoid biosynthesis outside of plastids in the unicellular green alga *Haematococcus pluvialis*. *Journal of Biological Chemistry,* 276**,** 6023-6029.

Guarnieri, M. T., A. Nag, S. L. Smolinski, A. Darzins, M. Seibert & P. T. Pienkos (2011) Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga. *PLoS One,* 6**,** e25851.

Guckert, J. B. & K. E. Cooksey (1990) Triglyceride accumulation and fatty acid profile changes in *Chlorella* (Chlorophyta) during high pH-induced cell cycle inhibition. *Journal of Phycology,* 26**,** 72-79.

Guedes, A. C., H. M. Amaro & F. X. Malcata (2011) Microalgae as sources of carotenoids. *Marine Drugs,* 9**,** 625-644.

Guiry, M. D. (2012) How many species of algae are there? *Journal of Phycology,* 48**,** 1057-1063.

Guleria, P., M. Mahajan, J. Bhardwaj & S. K. Yadav (2011) Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics, Proteomics & Bioinformatics,* 9**,** 183-199.

Guo, R. & J.-S. Ki (2012) Evaluation and validation of internal control genes for studying gene expression in the dinoflagellate *Prorocentrum minimum* using real-time PCR. *European Journal of Protistology,* 48**,** 199-206.

Gupta, S., A. Jha, A. Pal & G. Venkateshwarlu (2007) Use of natural carotenoids for pigmentation in fishes. *Natural Product Radiance,* 6**,** 46-49.

Hall, T. A (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series*, 95-98.

Hammouda, O., A. Gaber & N. Abdelraouf (1995) Microalgae and wastewater treatment. *Ecotoxicology and Environmental safety,* 31**,** 205-210.

He, L. & G. J. Hannon (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nature Reviews Genetics,* 5**,** 522-531.

Hemaiswarya, S., R. Raja, R. R. Kumar, V. Ganesan & C. Anbazhagan (2011) Microalgae: a sustainable feed source for aquaculture. *World Journal of Microbiology and Biotechnology,* 27**,** 1737-1746.

Hemalatha, A., K. Girija, C. Parthiban, C. Saranya & P. Anantharaman (2013) Antioxidant properties and total phenolic content of a marine diatom, green microalgae. *Advances in Applied Science Research,* 4**,** 115-157.

Hering, D., R. K. Johnson & A. Buffagni (2006) Linking organism groups—major results and conclusions from the STAR project. In *The Ecological Status of European Rivers: Evaluation and Intercalibration of Assessment Methods,* eds. M. T. Furse, D. Hering, K. Brabec, A. Buffagni, L. Sandin & P. Verdonschot, 109-113. Netherlands: Springer.

Herranz, H. & S. M. Cohen (2010) MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes & Development,* 24**,** 1339-1344.

Hindák, F. (1982) Taxonomic position of the chlorococcal alga *Chlorella zofingiensis* Dönz 1934 (Chlorophyceae). *Algological Studies,* 40**,** 13-23.

Hirata, S., M. Taya & S. Tone (1996) Characterization of *Chlorella* cell cultures in batch and continuous operations under a photoautotrophic condition. *Journal of Chemical Engineering of Japan,* 29**,** 953-959.

Hong, S.-J. & C.-G. Lee (2015) Microalgal Systems Biology for Biofuel Production. In *Algal Biorefineries*, 3-21. Springer.

Hörtensteiner, S. (1999) Chlorophyll breakdown in higher plants and algae. *Cellular and Molecular Life Sciences,* 56**,** 330-347.

Hoshina, R. & Y. Fujiwara (2013) Molecular characterization of *Chlorella* cultures of the National Institute for Environmental Studies culture collection with description of *Micractinium inermum* sp. nov., *Didymogenes sphaerica* sp. nov., and *Didymogenes soliella* sp. nov.(Chlorellaceae, Trebouxiophyceae). *Phycological Research,* 61**,** 124-132.

Hu, Q (2004) Environmental Effects on Cell Composition. In *Handbook of Microalgal Culture: Biotechnology and Applied Phycology,* ed. A. Richmond, 83-93. Blackwell Publishing Ltd.

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

Huss, V. A. R. & M. L. Sogin (1990) Phylogenetic position of some *Chlorella* species within Chlorococales based on SSu rRNA. *Journal of Molecular Evolution,* 31**,** 432-442.

Hwang, S., K.-H. Choi, J. Kim & J. Cha (2013) Biochemical characterization of 4-α-glucanotransferase from *Saccharophagus degradans* 2-40 and its potential role in glycogen degradation. *FEMS Microbiology Letters,* 344**,** 145-151.

Ip, P.-F. & F. Chen (2005a) Employment of reactive oxygen species to enhance astaxanthin formation in *Chlorella zofingiensis* in heterotrophic culture. *Process Biochemistry,* 40**,** 3491-3496.

Ip, P.-F. & F. Chen (2005b) Production of astaxanthin by the green microalga *Chlorella zofingiensis* in the dark. *Process Biochemistry,* 40**,** 733-738.

Ip, P.-F., K.-H. Wong & F. Chen (2004) Enhanced production of astaxanthin by the green microalga *Chlorella zofingiensis* in mixotrophic culture. *Process Biochemistry,* 39**,** 1761-1766.

Iwamoto, H (2004) *Chapter 11. Industrial production of microalgal cell-mass and secondary products–major industrial species : Chlorella*. Oxford, UK. : Blackwell Publishing Ltd.

Jia, L., D. Zhang, X. Qi, B. Ma, Z. Xiang & N. He (2014) Identification of the conserved and novel miRNAs in mulberry by high-throughput sequencing. *PloS One,* 9**,** e104409.

Jiang, Y. & F. Chen (2000) Effects of temperature and temperature shift on docosahexaenoic acid production by the marine microalge *Crypthecodinium cohnii*. *Journal of the American Oil Chemists' Society,* 77**,** 613-617.

Jiang, Y. J. & K. R. Feingold (2011) The expression and regulation of enzymes mediating the biosynthesis of triglycerides and phospholipids in keratinocytes/epidermis. *Dermato-endocrinology,* 3**,** 70-76.

Jones-Rhoades, M. W., & Bartel, D. P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. *Molecular cell,* 14(6), 787-799.

Juneja, A., R. M. Ceballos & G. S. Murthy (2013) Effects of environmental factors and nutrient availability on the biochemical composition of algae for biofuels production: a review. *Energies,* 6**,** 4607-4638.

Khraiwesh, B., J.-K. Zhu & J. Zhu (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1819**,** 137-148.

Knothe, G. (2010) Biodiesel: current trends and properties. *Topics in Catalysis,* 53**,** 714-720.

Koboldt, D. (2014) New challenges of next-gen sequencing. *Sequencing and Genomics,* 2016.

Kollas, A.-K., E. C. Duin, M. Eberl, B. Altincicek, M. Hintz, A. Reichenberg, D. Henschker, A. Henne, I. Steinbrecher & D. N. Ostrovsky (2002) Functional characterization of GcpE, an essential enzyme of the non-mevalonate pathway of isoprenoid biosynthesis. *FEBS Letters,* 532**,** 432-436.

Kramer, D. M. & J. R. Evans (2011) The importance of energy balance in improving photosynthetic productivity. *Plant Physiology,* 155**,** 70-78.

Krienitz, L., E. H. Hegewald, D. Hepperle, V. A. Huss, T. Rohr & M. Wolf (2004) Phylogenetic relationship of *Chlorella* and *Parachlorella* gen. nov.(Chlorophyta, Trebouxiophyceae). *Phycologia,* 43**,** 529-542.

Krienitz, L., V. A. Huss & C. Bock (2015) *Chlorella*: 125 years of the green survivalist. *Trends Plant Sci,* 20**,** 67-9.

Kröger, M. & F. Müller-Langer (2011) Impact of heterotrophic and mixotrophic growth of microalgae on the production of future biofuels. *Biofuels,* 2**,** 145-151.

Kuhn, D. E., Martin, M. M., Feldman, D. S., Terry, A. V., Nuovo, G. J., & Elton, T. S. (2008). Experimental validation of miRNA targets. *Methods,* 44(1), 47-54.

Kumar, S., N. Gupta, S. Kumar, V. Yadav, A. Prakash & H. Gurjar (2014) Metabolites in plants and its classification. *World Journal of Pharmacy and Pharmaceutical Sciences,* 4**,** 287-305.

Kuzuyama, T. (2002) Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units. *Bioscience, Biotechnology, and Biochemistry,* 66**,** 1619-1627.

Lakeman, M. B., P. von Dassow & R. A. Cattolico (2009) The strain concept in phytoplankton ecology. *Harmful Algae,* 8**,** 746-758.

Lane, A. E. & J. E. Burris (1981) Effects of environmental pH on the internal pH of *Chlorella pyrenoidosa*, *Scenedesmus quadricauda*, and *Euglena mutabilis*. *Plant Physiology,* 68**,** 439-442.

Latasa, M. & E. Berdalet (1994) Effect of nitrogen or phosphorus starvation on pigment composition of cultured *Heterocapsa* sp. *Journal of Plankton Research,* 16**,** 83-94.

Latendresse, M., S. Paley & P. D. Karp (2012) Browsing metabolic and regulatory networks with BioCyc. In *Bacterial Molecular Networks,* eds. J. v. Helden, A. Toussaint & D. Thieffry, 197-216. Springer.

Lemoine, Y. & B. Schoefs (2010) Secondary ketocarotenoid astaxanthin biosynthesis in algae: a multifunctional response to stress. *Photosynthesis research,* 106**,** 155-177.

León, R. & F. Galván (1999) Interaction between saline stress and photoinhibition of photosynthesis in the freshwater green algae *Chlamydomonas reinhardtii*. Implications for glycerol photoproduction. *Plant Physiology and Biochemistry,* 37**,** 623-628.

Leopoldini, M., N. Russo & M. Toscano (2011) The molecular basis of working mechanism of natural polyphenolic antioxidants. *Food Chemistry,* 125**,** 288-306.

Li, B. & C. N. Dewey (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics,* 12.

Li, H. & R. Durbin (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics,* 26**,** 589-595.

Li, Y.-R., W.-T. Tsai, Y.-C. Hsu, M.-Z. Xie & J.-J. Chen (2014) Comparison of autotrophic and mixotrophic cultivation of green microalgal for biodiesel production. *Energy Procedia,* 52**,** 371-376.

Li, Y., J. Huang, G. Sandmann & F. Chen (2009) High-Light and Sodium Chloride Stress Differentially Regulate the Biosynthesis of Astaxanthin Inchlorella Zofingiensis(Chlorophyceae). *Journal of Phycology,* 45**,** 635-641.

Li, T., Gargouri, M., Feng, J., Park, J.-J., Gao, D., Miao, C., Gang, D. R, Chen, S. (2015). Regulation of starch and lipid accumulation in a microalga *Chlorella sorokiniana*. Bioresource technology, 180, 250-257.

Li, Q., Liu, J., Zhang, L., & Liu, Q. (2014). De Novo Transcriptome Analysis of an Aerial Microalga *Trentepohlia jolithus*: Pathway Description and Gene Discovery for Carbon Fixation and Carotenoid Biosynthesis. *PloS one,* 9(9), e108488.

Liao, Y., G. K. Smyth & W. Shi (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research,* 41**,** e108-e108.

Liao, Y., G. K. Smyth & W. Shi (2014) FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics,* 30**,** 923-930.

Lichtenthaler, H. K. & C. Buschmann (2001) Chlorophylls and Carotenoids: Measurement and Characterization by UV-VIS Spectroscopy. *Current Protocols in Food Analytical Chemistry*.

Lin, C.-K. E., Kaptein, J. S., & Sheikh, J. (2017). Differential expression of microRNAs and their possible roles in patients with chronic idiopathic urticaria and active hives. *Allergy & Rhinology,* 8(2), e67.

Liu, B.-H. & Y.-K. Lee (2000) Secondary carotenoids formation by the green alga *Chlorococcum* sp. *Journal of Applied Phycology,* 12**,** 301-307.

Liu, C., G. Wu, X. Huang, S. Liu & B. Cong (2012a) Validation of housekeeping genes for gene expression studies in an ice alga *Chlamydomonas* during freezing acclimation. *Extremophiles,* 16**,** 419-425.

Liu, J., Z. Sun, H. Gerken, Z. Liu, Y. Jiang & F. Chen (2014) *Chlorella zofingiensis* as an alternative microalgal producer of astaxanthin: biology and industrial potential. *Marine Drugs,* 12**,** 3487-515.

Liu, J., Z. Sun, Y. Zhong, J. Huang, Q. Hu & F. Chen (2012b) Stearoyl-acyl carrier protein desaturase gene from the oleaginous microalga *Chlorella zofingiensis*: cloning, characterization and transcriptional analysis. *Planta,* 236**,** 1665-1676.

Liu, J., Mao, X., Zhou, W., & Guarnieri, M. T. (2016). Simultaneous production of triacylglycerol and high-value carotenoids by the astaxanthin-

producing oleaginous green microalga *Chlorella zofingiensis*. *Bioresource technology,* 214, 319-327.

Lodish, H., A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore & J. Darnell (2000) Photosynthetic stages and light-absorbing pigments. In *Molecular Cell Biology,* ed. W. H. Freeman. New York: NCBI Bookshelf.

Lohse, M., A. Nagel, T. Herter, P. May, M. Schroda, R. Zrenner, T. Tohge, A. R. Fernie, M. Stitt & B. Usadel (2014) Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. *Plant, Cell & Environment,* 37**,** 1250-1258.

Lopez, D., D. Casero, S. J. Cokus, S. S. Merchant & M. Pellegrini (2011) Algal Functional Annotation Tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. *BMC Bioinformatics,* 12**,** 282.

Love, M., S. Anders & W. Huber (2014) Differential analysis of count data–the DESeq2 package. *Genome Biology,* 15**,** 1-54.

Love, M. I., S. Anders, V. Kim & W. Huber (2015) RNA-Seq workflow: gene-level exploratory analysis and differential expression. *F1000Research,* 4**,** 1070. http://www-huber.embl.de/ (last accessed.

Lu, C., S. S. Tej, S. Luo, C. D. Haudenschild, B. C. Meyers & P. J. Green (2005) Elucidation of the small RNA component of the transcriptome. *Science,* 309**,** 1567-1569.

Lu, S., J. Wang, Q. Ma, J. Yang, X. Li & Y.-J. Yuan (2013) Phospholipid metabolism in an industry microalga *Chlorella sorokiniana*: the impact of inoculum sizes. *PloS one,* 8**,** e70827.

Lynch, D. V. & G. A. Thompson (1982) Low temperature-induced alterations in the chloroplast and microsomal membranes of *Dunaliella salina*. *Plant physiology,* 69**,** 1369-1375.

Machu, L., L. Misurcova, J. Vavra Ambrozova, J. Orsavova, J. Mlcek, J. Sochor & T. Jurikova (2015) Phenolic content and antioxidant capacity in algal food products. *Molecules,* 20**,** 1118-1133.

Magi, A., M. Benelli, A. Gozzini, F. Girolami, F. Torricelli & M. L. Brandi (2010) Bioinformatics for next generation sequencing data. *Genes,* 1**,** 294-307.

Martinez, N. J., M. C. Ow, M. I. Barrasa, M. Hammell, R. Sequerra, L. Doucette-Stamm, F. P. Roth, V. R. Ambros & A. J. Walhout (2008) A *C. elegans* genome-scale microRNA network contains composite feedback motifs with high flux capacity. *Genes & Development,* 22**,** 2535-2549.

Masojıdek, J., M. Koblızek & G. Torzillo (2004) Photosynthesis in microalgae. In *Handbook of Microalgal Culture: Biotechnology and Applied Phycology,* ed. A. Richmond, 20. Blackwell Publishing.

Matsukawa, R., M. Hotta, Y. Masuda, M. Chihara & I. Karube (2000) Antioxidants from carbon dioxide fixing *Chlorella sorokiniana*. *Journal of Applied Phycology* 12**,** 263–267.

Mayer, A. M. & M. T. Hamann (2005) Marine pharmacology in 2001–2002: marine compounds with anthelmintic, antibacterial, anticoagulant, antidiabetic, antifungal, anti-inflammatory, antimalarial, antiplatelet, antiprotozoal, antituberculosis, and antiviral activities; affecting the cardiovascular, immune and nervous systems and other miscellaneous mechanisms of action. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology,* 140**,** 265-286.

McCarthy, D. J., Y. Chen & G. K. Smyth (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research***,** 1-10.

Merchant, S. S., S. E. Prochnik, O. Vallon, E. H. Harris, S. J. Karpowicz, G. B. Witman, A. Terry, A. Salamov, L. K. Fritz-Laylin & L. Maréchal-Drouard (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science,* 318**,** 245-250.

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

Molnár, A., F. Schwach, D. J. Studholme, E. C. Thuenemann & D. C. Baulcombe (2007) miRNAs control gene expression in the single-cell alga *Chlamydomonas reinhardtii*. *Nature,* 447**,** 1126-1129.

Morell, M. K., Z. Li, A. Regina, S. Rahman, C. d'Hulst & S. G. Ball (2008) Control of starch biosynthesis in vascular plants and algae. In *Control of Primary Metabolism in Plants,* eds. W. C. Plaxton & M. T. McManus, 258. Blackwell Publishing.

Moroney, J. V. & A. Somanchi (1999) How do algae concentrate CO_2 to increase the efficiency of photosynthetic carbon fixation? *Plant Physiology,* 119**,** 9-16.

Mostafa, S. S (2012) *Microalgal biotechnology: prospects and applications*. INTECH Open Access Publisher.

Mozaffarian, D., A. Ascherio, F. B. Hu, M. J. Stampfer, W. C. Willett, D. S. Siscovick & E. B. Rimm (2005) Interplay between different polyunsaturated fatty acids and risk of coronary heart disease in men. *Circulation,* 111**,** 157-164.

Mulders, K. J., Janssen, J. H., Martens, D. E., Wijffels, R. H., & Lamers, P. P. (2014). Effect of biomass concentration on secondary carotenoids and triacylglycerol (TAG) accumulation in nitrogen-depleted *Chlorella zofingiensis*. *Algal Research*, 6, 8-16.

Müller, S., L. Rycak, P. Winter, G. Kahl, I. Koch & B. Rotter (2013) omiRas: a Web server for differential expression analysis of miRNAs derived from small RNA-Seq data. *Bioinformatics,* 29**,** 2651-2652.

Murata, N., S. Takahashi, Y. Nishiyama & S. I. Allakhverdiev (2007) Photoinhibition of photosystem II under environmental stress. *Biochimica et Biophysica Acta (BBA)-Bioenergetics,* 1767**,** 414-421.

Naik, S. N., V. V. Goud, P. K. Rout & A. K. Dalai (2010) Production of first and second generation biofuels: a comprehensive review. *Renewable and Sustainable Energy Reviews,* 14**,** 578-597.

Nakamura, Y. (1983) Change in molecular weight distribution in starch when degraded at different temperatures in *Chlorella vulgaris*. *Plant Science Letters,* 30**,** 259-265.

Nakamura, Y. & S. Miyachi (1982) Effect of temperature on starch degradation in *Chlorella vulgaris* 11h cells. *Plant and Cell Physiology,* 23**,** 333-341.

Nakano, S., H. Takekoshi & M. Nakano (2010) *Chlorella pyrenoidosa* supplementation reduces the risk of anemia, proteinuria and edema in pregnant women. *Plant foods for human nutrition,* 65**,** 25-30.

Ndhlala, A. R., M. Moyo & J. Van Staden (2010) Natural antioxidants: fascinating or mythical biomolecules? *Molecules,* 15**,** 6905-6930.

Negi, S., A. N. Barry, N. Friedland, N. Sudasinghe, S. Subramanian, S. Pieris, F. O. Holguin, B. Dungan, T. Schaub & R. Sayre (2016) Impact of nitrogen limitation on biomass, photosynthesis, and lipid accumulation in *Chlorella sorokiniana*. *Journal of Applied Phycology,* 28**,** 803-812.

Nemcova, Y. & T. Kalina (2000) Cell wall development, microfibril and pyrenoid structure in type strains of *Chlorella vulgaris, C. kessleri, C. sorokiniana* compared with *C. luteoviridis* (Trebouxiophyceae, Chlorophyta). *Archiv für Hydrobiologie , Supplement,* 136**,** 95-106.

Nishida, I. & N. Murata (1996) Chilling sensitivity in plants and cyanobacteria: the crucial contribution of membrane lipids. *Annual Review of Plant Biology,* 47**,** 541-568.

Ntambi, J. M. & M. Miyazaki (2004) Regulation of stearoyl-CoA desaturases and role in metabolism. *Progress in Lipid Research,* 43**,** 91-104.

Nunes-Nesi, A., W. L. Araújo, T. Obata & A. R. Fernie (2013) Regulation of the mitochondrial tricarboxylic acid cycle. *Current opinion in plant biology,* 16**,** 335-343.

Obulesu, M., M. R. Dowlathabad & P. Bramhachari (2011) Carotenoids and Alzheimer's disease: an insight into therapeutic role of retinoids in animal models. *Neurochemistry International,* 59**,** 535-541.

Okada, K. (2009) PetH is rate-controlling in the interaction between PetH, a component of the supramolecular complex with photosystem II, and PetF, a light-dependent electron transfer protein. *Biochemical and biophysical research communications,* 389**,** 394-398.

Palenik, B., J. Grimwood, A. Aerts, P. Rouzé, A. Salamov, N. Putnam, C. Dupont, R. Jorgensen, E. Derelle & S. Rombauts (2007) The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences,* 104**,** 7705-7710.

Paniagua-Michel, J., Olmos-Soto, J., & Ruiz, M. A. (2012). Pathways of carotenoid biosynthesis in bacteria and microalgae. *Microbial Carotenoids from Bacteria and Microalgae: Methods and Protocols,* 1-12.

Perez-Garcia, O. & Y. Bashan (2015) Microalgal heterotrophic and mixotrophic culturing for bio-refining: From metabolic routes to techno-economics. In *Algal biorefineries*, 61-131. Springer.

Perez-Garcia, O., F. M. Escalante, L. E. de-Bashan & Y. Bashan (2011) Heterotrophic cultures of microalgae: metabolism and potential products. *Water Research,* 45**,** 11-36.

Pérez-Jiménez, J., S. Arranz, M. Tabernero, M. E. Díaz-Rubio, J. Serrano, I. Goñi & F. Saura-Calixto (2008) Updated methodology to determine antioxidant capacity in plant foods, oils and beverages: Extraction, measurement and expression of results. *Food Research International,* 41**,** 274-285.

Petkov, G. & G. Garcia (2007) Which are fatty acids of the green alga *Chlorella*? *Biochemical Systematics and Ecology,* 35**,** 281-285.

Pignolet, O., S. Jubeau, C. Vaca-Garcia & P. Michaud (2013) Highly valuable microalgae: biochemical and topological aspects. *Journal of Industrial Microbiol & Biotechnology,* 40**,** 781-96.

Plaza, M., M. Herrero, A. Cifuentes & E. Ibáñez (2009) Innovative natural functional ingredients from microalgae. *Journal of Agricultural and Food Chemistry,* 57**,** 7159-7170.

Prior, R. L., X. Wu & K. Schaich (2005) Standardized methods for the determination of antioxidant capacity and phenolics in foods and dietary supplements. *Journal of Agricultural and Food Chemistry,* 53**,** 4290-4302.

Priyadarshani, I. & B. Rath (2012) Commercial and industrial applications of micro algae–A review. *Journal of Algal Biomass Utilization,* 3**,** 89-100.

Prochnik, S. E., J. Umen, A. M. Nedelcu, A. Hallmann, S. M. Miller, I. Nishii, P. Ferris, A. Kuo, T. Mitros & L. K. Fritz-Laylin (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science,* 329**,** 223-226.

Proffitt, A (2013) Next generation sequencing survey. BIO-IT World.

Provan, J., S. Murphy & C. A. Maggs (2004) Universal plastid primers for Chlorophyta and Rhodophyta. *European Journal of Phycology,* 39**,** 43-50.

Qin, Z., C. Li, L. Mao & L. Wu (2014) Novel insights from non-conserved microRNAs in plants. *Frontiers in Plant Science,* 5**,** 586.

Raven, J. A. & R. J. Geider (1988) Temperature and algal growth. *New phytologist,* 110**,** 441-461.

Remias, D., U. Karsten, C. Lütz & T. Leya (2010) Physiological and morphological processes in the Alpine snow alga *Chloromonas nivalis* (Chlorophyceae) during cyst formation. *Protoplasma,* 243**,** 73-86.

Rismani-Yazdi, H., B. Z. Haznedaroglu, K. Bibby & J. Peccia (2011) Transcriptome sequencing and annotation of the microalgae *Dunaliella tertiolecta*: pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics,* 12**,** 148.

Rodriguez-Garcia, I. & J. L. Guil-Guerrero (2008) Evaluation of the antioxidant activity of three microalgal species for use as dietary supplements and in the preservation of foods. *Food Chemistry,* 108**,** 1023-1026.

Rosenberg, J. N., G. A. Oyler, L. Wilkinson & M. J. Betenbaugh (2008) A green light for engineered algae: redirecting metabolism to fuel a biotechnology revolution. *Current Opinion in Biotechnology,* 19**,** 430-436.

Rosic, N. N., M. Pernice, M. Rodriguez-Lanetty & O. Hoegh-Guldberg (2011) Validation of housekeeping genes for gene expression studies in *Symbiodinium* exposed to thermal and light stress. *Marine Biotechnology,* 13**,** 355-365.

Safafar, H., J. van Wagenen, P. Møller & C. Jacobsen (2015) Carotenoids, phenolic compounds and tocopherols contribute to the antioxidative properties of some microalgae species grown on industrial wastewater. *Marine Drugs,* 13**,** 7339-7356.

Safi, C., B. Zebib, O. Merah, P.-Y. Pontalier & C. Vaca-Garcia (2014) Morphology, composition, production, processing and applications of *Chlorella vulgaris*: A review. *Renewable and Sustainable Energy Reviews,* 35**,** 265-278.

Sakai, N., Y. Sakamoto, N. Kishimoto, M. Chihara & I. Karube (1995) *Chlorella* strains from hot springs tolerant to high temperature and high CO 2. *Energy Conversion and Management,* 36**,** 693-696.

Saranya, C., A. Hemalatha, C. Parthiban & P. Anantharaman (2014) Evaluation of antioxidant properties, total phenolic and carotenoid content of *Chaetoceros calcitrans*, *Chlorella salina* and *Isochrysis galbana International Journal of Current Microbiology and Applied Science,* 3**,** 365-377.

Sasaki, Y. & Y. Nagano (2004) Plant acetyl-CoA carboxylase: structure, biosynthesis, regulation, and gene manipulation for plant breeding. *Bioscience, biotechnology, and biochemistry,* 68**,** 1175-1184.

Sharma, K. K., H. Schuhmann & P. M. Schenk (2012) High Lipid Induction in Microalgae for Biodiesel Production. *Energies,* 5**,** 1532-1553.

Shi, X., H. Zhang & S. Lin (2013) Tandem repeats, high copy number and remarkable diel expression rhythm of form II RuBisCO in P*rorocentrum donghaiense* (dinophyceae). *PloS One,* 8**,** e71232.

Shukla, S. P., J. Kviderova, J. Tříska & J. Elster (2013) *Chlorella mirabilis* as a potential species for biomass production in low-temperature environment. *Frontiers in Microbiology,* 4.

Siaut, M., S. Cuiné, C. Cagnon, B. Fessler, M. Nguyen, P. Carrier, A. Beyly, F. Beisson, C. Triantaphylidès & Y. Li-Beisson (2011) Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: characterization, variability between common laboratory strains and relationship with starch reserves. *BMC Biotechnology,* 11**,** 7.

Skjanes, K., C. Rebours & P. Lindblad (2013) Potential for green microalgae to produce hydrogen, pharmaceuticals and other high value products in a combined process. *Critical Reviews in Biotechnology,* 33**,** 172-215.

Smith-Unna, R., C. Boursnell, R. Patro, J. Hibberd & S. Kelly (2016) TransRate: reference free quality assessment of de novo transcriptome assemblies. *Genome Research***,** gr. 196469.115.

Spolaore, P., C. Joannis-Cassan, E. Duran & A. Isambert (2006) Commercial applications of microalgae. *Journal of Bioscience and Bioengineering,* 101**,** 87-96.

Stanke, M., O. Keller, I. Gunduz, A. Hayes, S. Waack & B. Morgenstern (2006) AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Research,* 34**,** W435-W439.

Steinbrenner, J. & H. Linden (2001) Regulation of two carotenoid biosynthesis genes coding for phytoene synthase and carotenoid hydroxylase during stress-induced astaxanthin formation in the green alga *Haematococcus pluvialis*. *Plant Physiology,* 125**,** 810-817.

Stolz, P. & B. Obermayer (2005) Manufacturing microalgae for skin care. *Cosmetics and Toiletries,* 120**,** 99-106.

Sun, N., Y. Wang, Y.-T. Li, J.-C. Huang & F. Chen (2008) Sugar-based growth, astaxanthin accumulation and carotenogenic transcription of heterotrophic *Chlorella zofingiensis* (Chlorophyta). *Process Biochemistry,* 43**,** 1288-1292.

Sun, P., Y. Mao, G. Li, M. Cao, F. Kong, L. Wang & G. Bi (2015) Comparative transcriptome profiling of *Pyropia yezoensis* (Ueda) M.S. Hwang & H.G. Choi in response to temperature stresses. *BMC Genomics,* 16**,** 1-16.

Sun, X., W. Wang, J. Shen & N. Xu (2014) Transcriptome sequencing of the marine microalga, *Chlorella pyrenoidosa* (Chlorophyta), and analysis of carbonic anhydrase expression under salt stress. *Botanica Marina,* 57**,** 403-412.

Sun, Z., Y. F. Chen & J. Du (2016) Elevated CO2 improves lipid accumulation by increasing carbon metabolism in *Chlorella sorokiniana*. *Plant Biotechnology Journal,* 14**,** 557-566.

Sweetlove, L. J., & Fernie, A. R. (2005). Regulation of metabolic networks: understanding metabolic complexity in the systems biology era. *New phytologist, 168*(1), 9-24.

Taga, M. S., E. E. Miller & D. E. Pratt (1984) Chia seeds as a source of natural lipid antioxidants *Journal of the American Oil Chemists' Society,* 61.

Takaichi, S. (2011). Carotenoids in algae: distributions, biosyntheses and functions. *Marine Drugs,* 9(6), 1101-1118. doi: 10.3390/md9061101.

Takeda, H. (1988) Classification of *chlorella* strain by means of sugar components in cell wall. *Biochemical Systematics and Ecology,* 16**,** 367-371.

Takeshita, T., S. Ota, T. Yamazaki, A. Hirata, V. Zachleder & S. Kawano (2014) Starch and lipid accumulation in eight strains of six *Chlorella* species under comparatively high light intensity and aeration culture conditions. *Bioresource Technology,* 158**,** 127-34.

Takeuchi, K. & K. Reue (2009) Biochemistry, physiology, and genetics of GPAT, AGPAT, and lipin enzymes in triglyceride synthesis. *American Journal of Physiology-Endocrinology And Metabolism,* 296**,** E1195-E1209.

Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution,* 30**,** 2725-2729.

Thimm, O., O. Bläsing, Y. Gibon, A. Nagel, S. Meyer, P. Krüger, J. Selbig, L. A. Müller, S. Y. Rhee & M. Stitt (2004) Mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *The Plant Journal,* 37**,** 914-939.

Thompson, G. A. (1996) Lipids and membrane function in green algae. *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism,* 1302**,** 17-45.

Thompson, J. D., T. Gibson & D. G. Higgins (2002) Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics***,** 2.3. 1-2.3. 22.

Tjahjono, A. E., Y. Hayama, T. Kakizono, Y. Terada, N. Nishio & S. Nagai (1994) Hyper-accumulation of astaxanthin in a green alga *Haematococcus pluvialis* at elevated temperatures. *Biotechnology Letters,* 16**,** 133-138.

Tomaselli, L (2004) The Microalgal Cell. In *Handbook of microalgal culture: Biotechnology and applied phycology,* ed. A. Richmond, 3-19. Blackwell Publishing.

Tomaz, T., M. Bagard, I. Pracharoenwattana, P. Lindén, C. P. Lee, A. J. Carroll, E. Ströher, S. M. Smith, P. Gardeström & A. H. Millar (2010) Mitochondrial malate dehydrogenase lowers leaf respiration and alters photorespiration and plant growth in Arabidopsis. *Plant physiology,* 154**,** 1143-1157.

Tripathi, U., R. Sarada & G. Ravishankar (2002) Effect of culture conditions on growth of green alga—*Haematococcus pluvialis* and astaxanthin production. *Acta Physiologiae Plantarum,* 24**,** 323-329.

Tsang, J., J. Zhu & A. van Oudenaarden (2007) MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals. *Molecular Cell,* 26**,** 753-767.

Uma, R., V. Sivasubramanian & S. N. Devaraj (2011) Evaluation of in vitro antioxidant activities and antiproliferative activity of green microalgae, *Desmococcus olivaceous* and *Chlorococcum humicola*. *Journal of Algal Biomass Utilization,* 2**,** 82-93.

Vachali, P., P. Bhosale & P. S. Bernstein (2012) Microbial carotenoids. In *Microbial Carotenoids from Fungi,* ed. J.-L. Barredo, 41-59. Humana Press.

Vandesompele, J., K. De Preter, F. Pattyn, B. Poppe, N. Van Roy, A. De Paepe & F. Speleman (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology,* 3**,** 1-12.

Varfolomeev, S. & L. Wasserman (2011) Microalgae as source of biofuel, food, fodder, and medicines. *Applied Biochemistry and Microbiology,* 47**,** 789-807.

Vílchez, C., E. Forján, M. Cuaresma, F. Bédmar, I. Garbayo & J. M. Vega (2011) Marine carotenoids: biological functions and commercial applications. *Marine Drugs,* 9**,** 319-333.

Vonshak, A. & G. Torzillo (2004) Environmental stress physiology. In *Handbook of Microalgal Culture: Biotechnology and Applied Phycology,* ed. A. Richmond, 57-82. Blackwell Publishing.

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

Wang, L., M. Min, Y. Li, P. Chen, Y. Chen, Y. Liu, Y. Wang & R. Ruan (2010b) Cultivation of green algae *Chlorella* sp. in different wastewaters from municipal wastewater treatment plant. *Applied Biochemistry and Biotechnology,* 162**,** 1174-1186.

Wang, Y. & T. Chen (2008) The biosynthetic pathway of carotenoids in the astaxanthin-producing green alga *Chlorella zofingiensis*. *World Journal of Microbiology and Biotechnology,* 24**,** 2927-2932.

Wang, Z., M. Gerstein & M. Snyder (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10**,** 57-63.

Wanke, M., K. Skorupinska-Tudek & E. Swiezewska (2001) Isoprenoid biosynthesis via 1-deoxy-D-xylulose 5-phosphate/2-C-methyl-D-erythritol 4-phosphate (DOXP/MEP) pathway. *Acta Biochimica Polonica,* 48**,** 663-672.

Wei, S., M. Y. Gruber, B. Yu, M.-J. Gao, G. G. Khachatourians, D. D. Hegedus, I. A. Parkin & A. Hannoufa (2012) *Arabidopsis* mutant sk156 reveals complex regulation of SPL15 in a *miR156*-controlled gene network. *BMC Plant Biology,* 12**,** 1-17.

Westerhoff, P., Q. Hu, M. Esparza-Soto & W. Vermaas (2010) Growth parameters of microalgae tolerant to high levels of carbon dioxide in batch and continuous-flow photobioreactors. *Environmental technology,* 31**,** 523-532.

White, T. J., T. Bruns, S. Lee & J. Taylor (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR protocols: a guide to methods and applications,* 18**,** 315-322.

Wolf, J. B. (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-seq tutorial. *Molecular Ecology Resources,* 13**,** 559-72.

Worden, A. Z., J.-H. Lee, T. Mock, P. Rouzé, M. P. Simmons, A. L. Aerts, A. E. Allen, M. L. Cuvelier, E. Derelle & M. V. Everett (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science,* 324**,** 268-272.

Wu, L., J. Fan & J. G. Belasco (2006) MicroRNAs direct rapid deadenylation of mRNA. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4034-4039.

Wu, Q., T. Liu, H. Liu & G. Zheng (2009) Unsaturated fatty acid: metabolism, synthesis and gene regulation. *African Journal of Biotechnology,* 8.

Xu, H., X. Miao & Q. Wu (2006) High quality biodiesel production from a microalga *Chlorella protothecoides* by heterotrophic growth in fermenters. *Journal of Biotechnology,* 126**,** 499-507.

Xu, X.-Q. & J. Beardall (1997) Effect of salinity on fatty acid composition of a green microalga from an antarctic hypersaline lake. *Phytochemistry,* 45**,** 655-658.

Xu, Z., D. Zou & K. Gao (2010) Effects of elevated CO2 and phosphorus supply on growth, photosynthesis and nutrient uptake in the marine macroalga *Gracilaria lemaneiformis* (Rhodophyta). *Botanica Marina,* 53**,** 123-129.

Yamamoto, M., M. Fujishita, A. Hirata & S. Kawano (2004) Regeneration and maturation of daughter cell walls in the autospore-forming green alga *Chlorella vulgaris* (Chlorophyta, Trebouxiophyceae). *Journal of Plant Research,* 117**,** 257-64.

Yamamoto, M., I. Kurihara & S. Kawano (2005) Late type of daughter cell wall synthesis in one of the Chlorellaceae, *Parachlorella kessleri* (Chlorophyta, Trebouxiophyceae). *Planta,* 221**,** 766-75.

Yang, S., M. T. Guarnieri, S. Smolinski, M. Ghirardi & P. T. Pienkos (2013) De novo transcriptomic analysis of hydrogen production in the green alga *Chlamydomonas moewusii* through RNA-Seq. *Biotechnology for Biofuels,* 6**,** 17.

Yang, T., Xu, R., Chen, J., & Liu, A. (2016). β-Ketoacyl-acyl Carrier Protein Synthase I (KASI) Plays Crucial Roles in the Plant Growth and Fatty Acids Synthesis in Tobacco. *International Journal of Molecular Sciences*, 17(8), 1287. http://doi.org/10.3390/ijms17081287

Yao, C., J. Ai, X. Cao, S. Xue & W. Zhang (2012) Enhancing starch production of a marine green microalga *Tetraselmis subcordiformis* through nutrient limitation. *Bioresource Technology,* 118**,** 438-44.

Yu, J. & P. Takahashi (2007) Biophotolysis-based hydrogen production by cyanobacteria and green microalgae. *Communicating current research and educational topics and trends in applied microbiology,* 1**,** 79-89.

Zakar, T., Laczko-Dobos, H., Toth, T. N., & Gombos, Z. (2016). Carotenoids assist in cyanobacterial Photosystem II assembly and function. *Frontiers Plant Science,* 7.

Zhao, B., L. Ge, R. Liang, W. Li, K. Ruan, H. Lin & Y. Jin (2009) Members of miR-169 family are induced by high salinity and transiently inhibit the NF-YA transcription factor. *BMC Molecular Biology,* 10**,** 1-10.

Zhao, B., R. Liang, L. Ge, W. Li, H. Xiao, H. Lin, K. Ruan & Y. Jin (2007a) Identification of drought-induced microRNAs in rice. *Biochemical and Biophysical Research Communications,* 354**,** 585-590.

Zhao, L., W.-c. Chang, Y. Xiao, H.-w. Liu & P. Liu (2013) Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annual Review of Biochemistry,* 82**,** 497-530.

Zhao, R., Y. Cao, H. Xu, L. Lv, D. Qiao & Y. Cao (2011) Analysis of expressed sequence tags from the green alga *Dunaliella Salina* (Chlorophyta). *Journal of Phycology,* 47**,** 1454-1460.

Zhao, T., G. Li, S. Mi, S. Li, G. J. Hannon, X. J. Wang & Y. Qi (2007b) A complex system of small RNAs in the unicellular green alga *Chlamydomonas reinhardtii*. *Genes & Development,* 21**,** 1190-1203.

Zheng, H.-Q., Y.-F. Chiang-Hsieh, C.-H. Chien, B.-K. J. Hsu, T.-L. Liu, C.-N. N. Chen & W.-C. Chang (2014) AlgaePath: comprehensive analysis of metabolic pathways using transcript abundance data from next-generation sequencing in green algae. *BMC genomics,* 15**,** 196.

Zheng, Y., Z. Chi, B. Lucker & S. Chen (2012) Two-stage heterotrophic and phototrophic culture strategy for algal biomass and lipid production. *Bioresource Technology,* 103**,** 484-488.

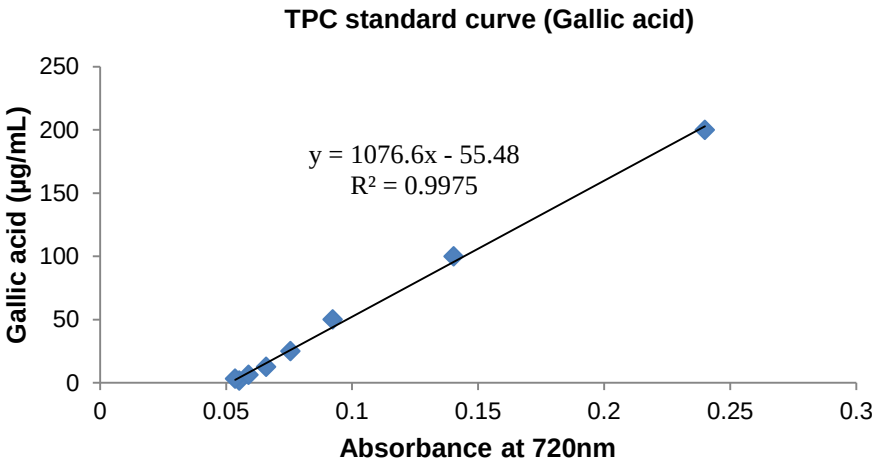
Zhou, X., G. Wang & W. Zhang (2007) UV-B responsive microRNA genes in *Arabidopsis thaliana*. *Molecular Systems Biology,* 3**,** 103.

Zhu, S., W. Huang, J. Xu, Z. Wang, J. Xu & Z. Yuan (2014) Metabolic changes of starch and lipid triggered by nitrogen starvation in the microalga *Chlorella zofingiensis*. *Bioresource Technology,* 152**,** 292-8.

APPENDICES

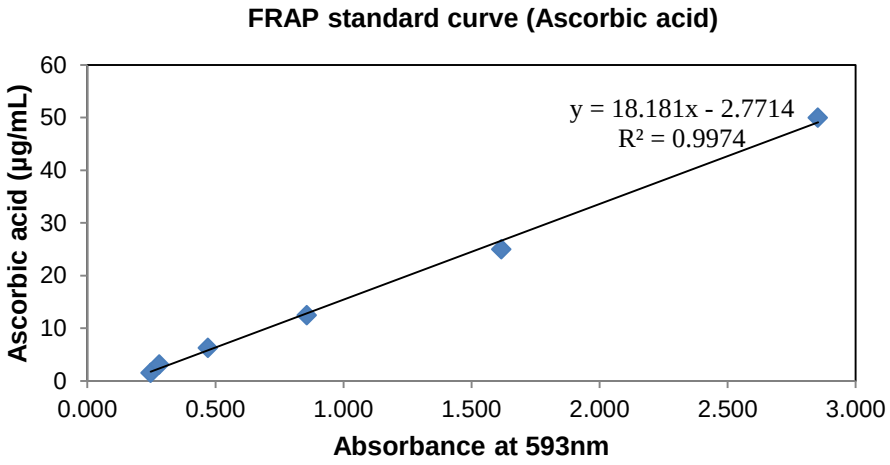
Appendix A

Total phenolic content standard curve



Appendix B

Ferric reducing antioxidant power assay (FRAP) standard curve

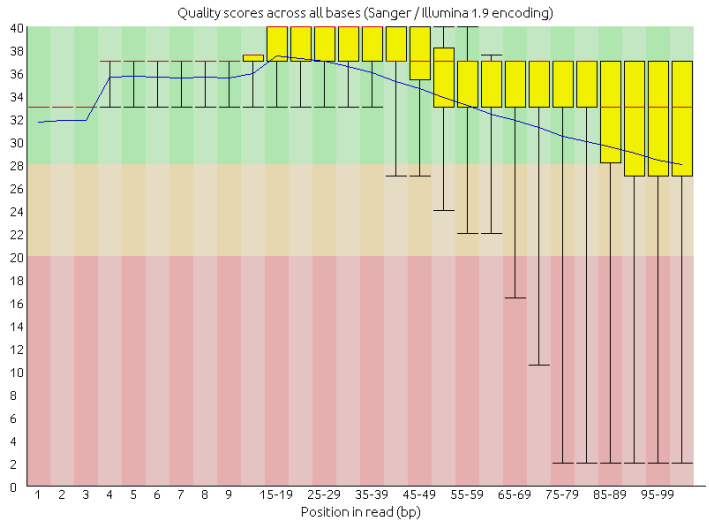


Appendix C

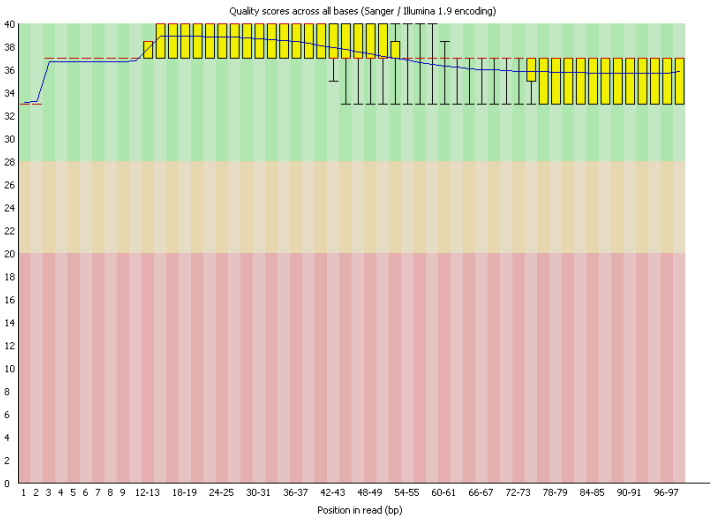
Example of FASTQC report

1) Per base quality = Q30

Before:

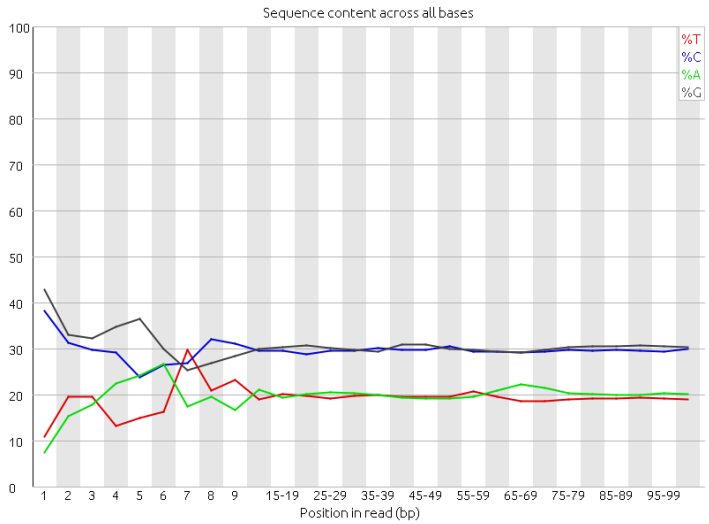


After:

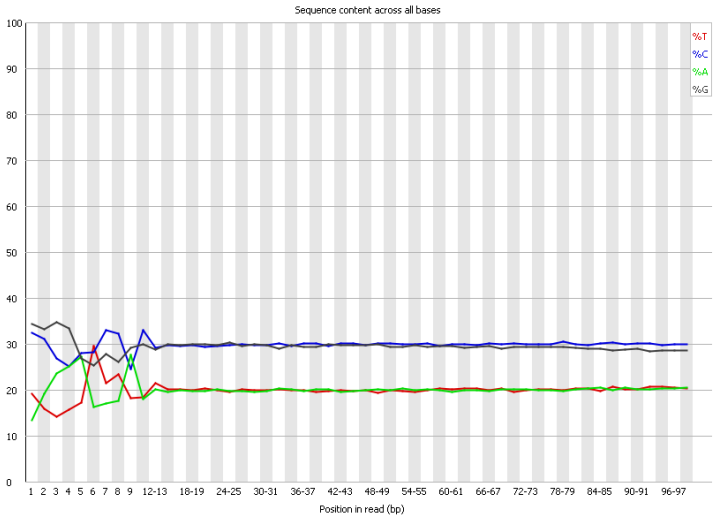


2) Per base sequence content

Before:

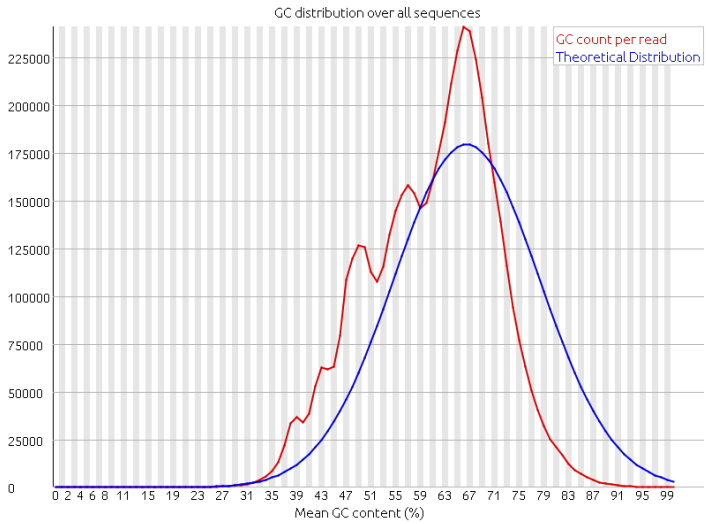


After:

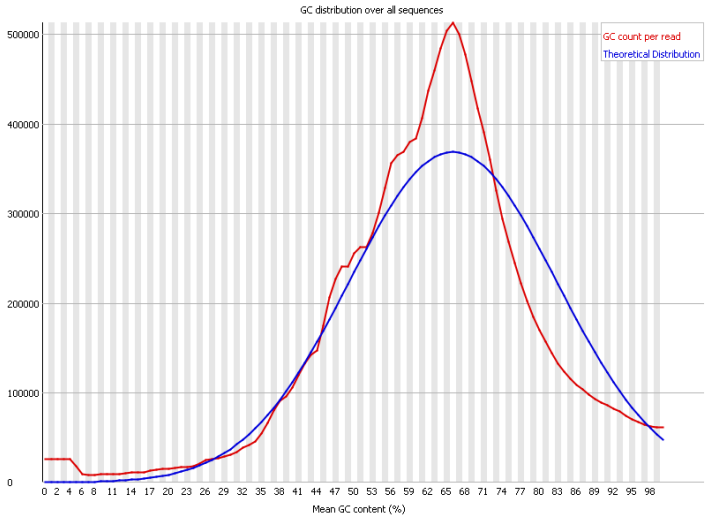


3) Per sequence gc content

Before:

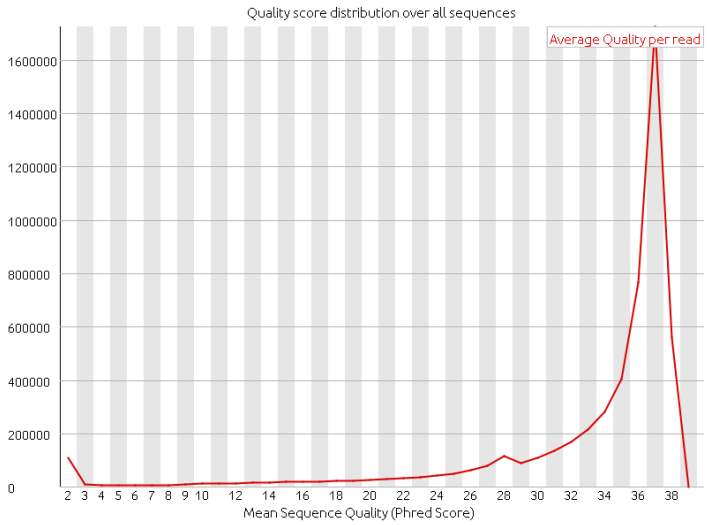


After:

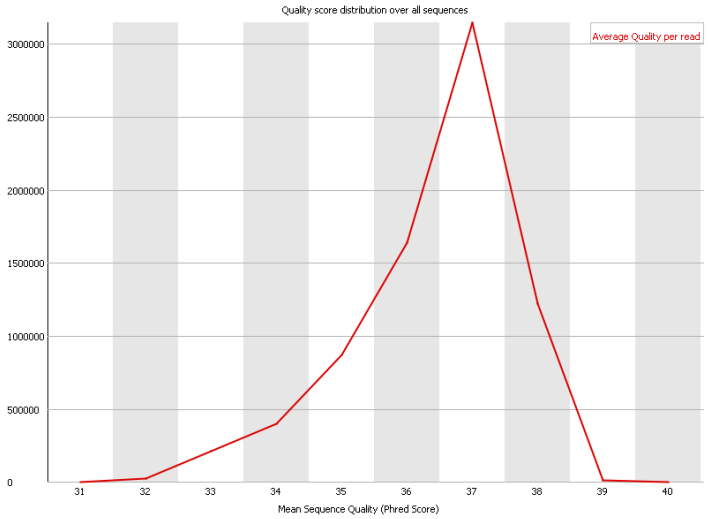


4) Per sequence quality

Before:

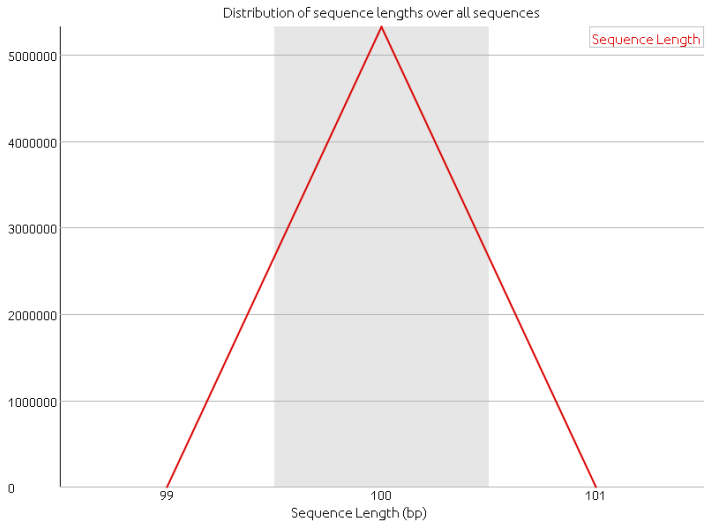


After:

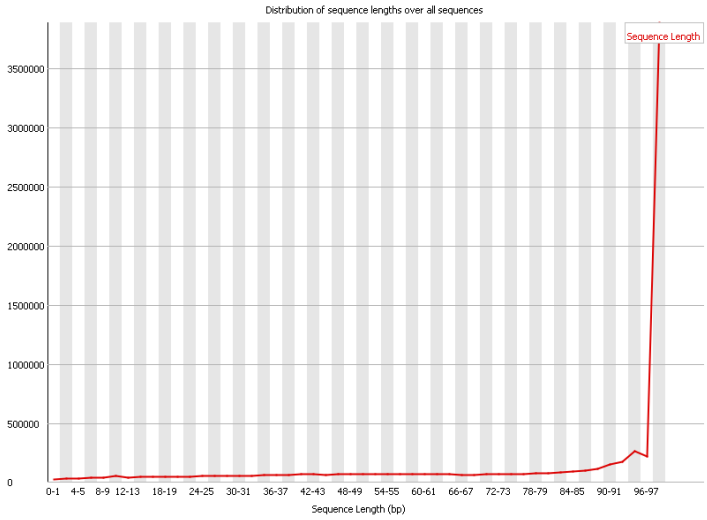


5) Sequence length distribution

Before:



After:



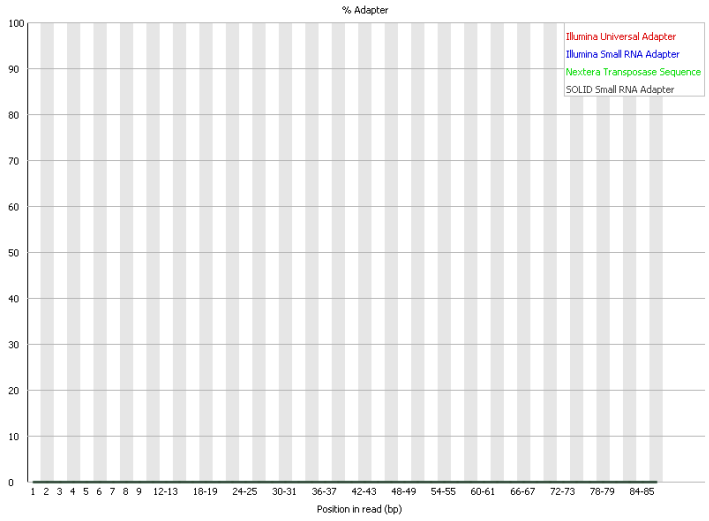
6) Adapter content

Before:

Overrepresented sequences

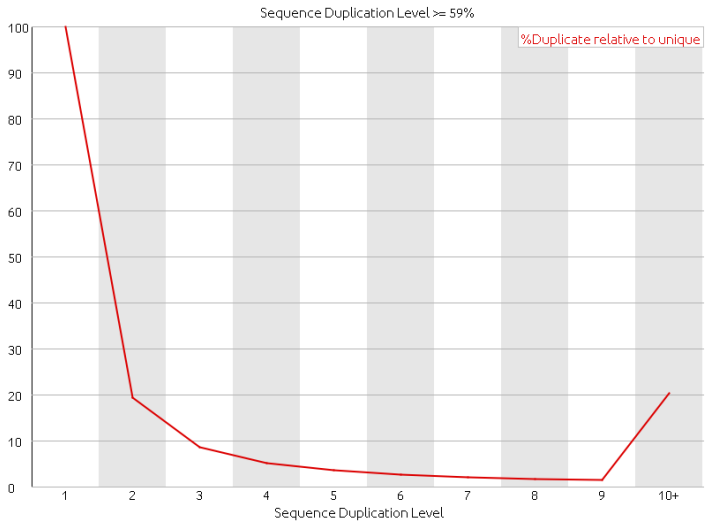
Sequence	Count	Percentage	Possible Source
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGC	88708	1.666434977325754	TruSeq Adapter, Index 1 (100% over 50bp)
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATATCGTATGC	7384	0.13871303459184478	TruSeq Adapter, Index 1 (98% over 50bp)

After

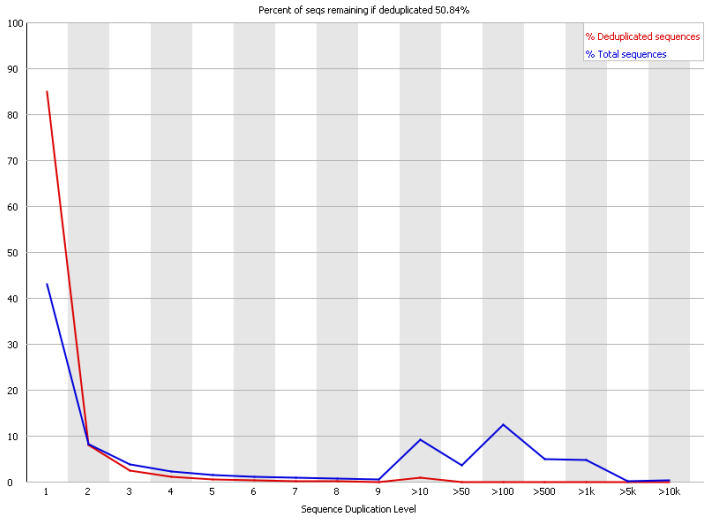


7) Duplication level

Before:

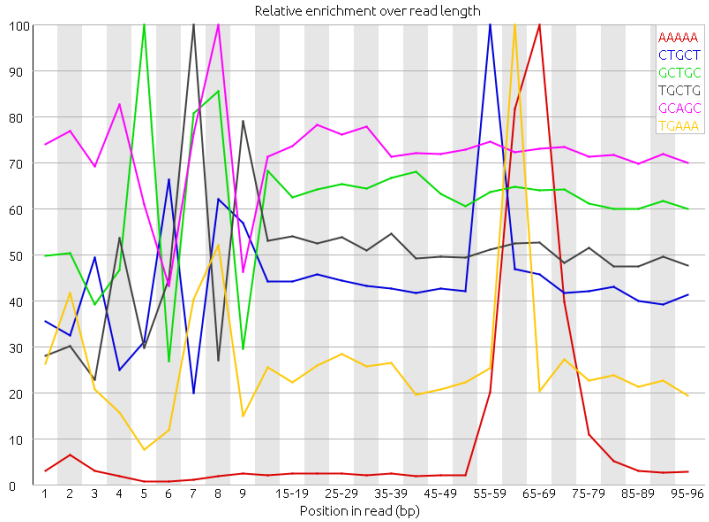


After:

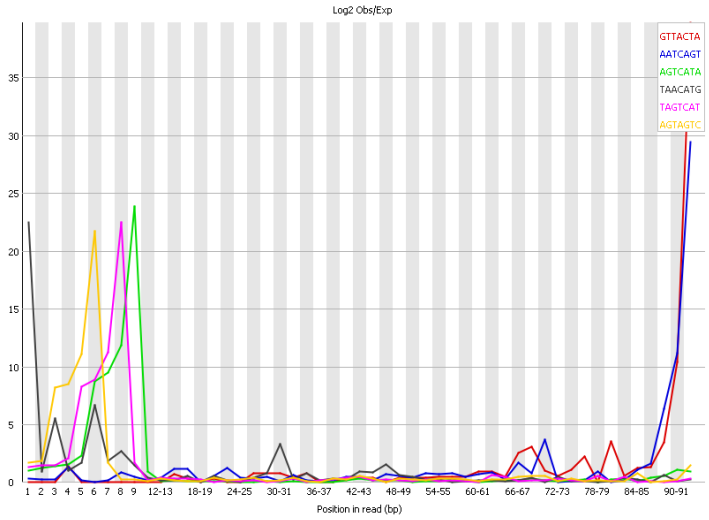


8) Kmer profile

Before:

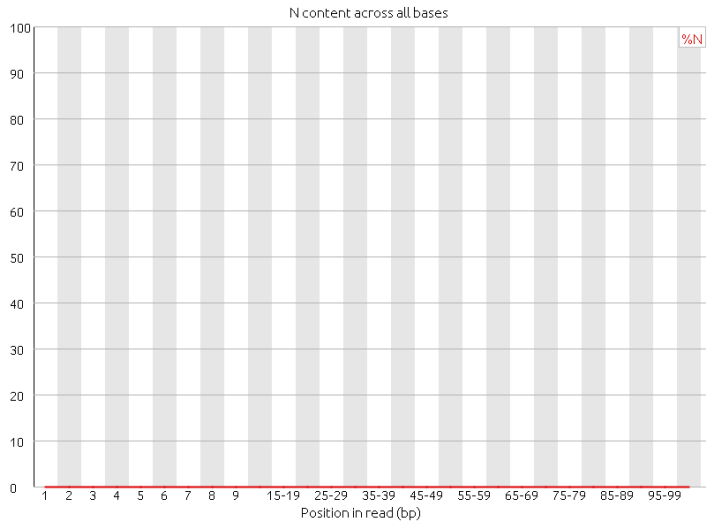


After:

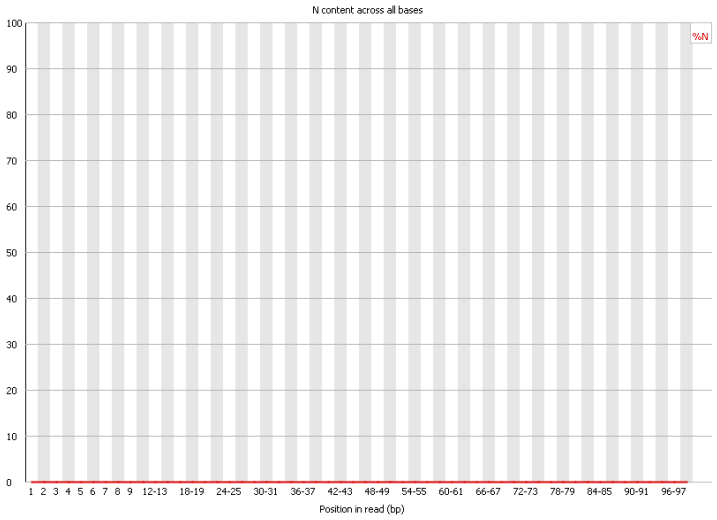


9) Per base N content

Before:



After:



Appendix D

Calculation of N50 and N90 (example)

N50 and N90 are most common statistics for comparison of assembled genomes/transcriptome. N50 is the shortest sequence in a set of sequences that together cover 50% of the total length of all assembled sequences. The calculation starts by ranking sequences according to their length in descendant order. If the cumulative length of sequences covers 50% of the total length of the transcriptome, then the order and the length of the last contig sequence in that group shows the result of the N50 statistic. The N90 calculation is the same but for contigs that cover 90% of total length of all sequences. Table 15 shows an example of N50 and N90 calculation.

Length	Sorted	Cumulative length
4	9	9
8	8	17
6	8	25
5	7	32
5	6	38
1	5	43
1	5	48
3	5	53
4	5	58
3	4	62
3	4	66
5	4	70
9	4	74
5	3	77
4	3	80
7	3	83
4	2	85
1	1	86
8	1	87
2	1	88
Total	**88**	
50%	**44**	
90%	**79**	
N50	**5**	
N90	**3**	

Appendix E

Previously sequenced green algae (Gao et al. 2014)

Table 2 Genomic features of sequenced chlorophyte green algae

	CPRO	CVAR	CSUB	CREI	VCAR	MPUS	OTAR
Taxonomic class	T	T	T	C	C	M	M
Assembly length (Mb)	22.9	46.2	48.8	121	138	21.9	12.6
GC content (%)	63	67	53	64	56	65	58
Repeat sequences (%)	6.1	8.9	7.2	16.7	23.8	8.8	5.1
Number of gene	7,039	9,791	9,851	15,143	14,520	10,575	8,166
Average gene length (bp)	2,863	2,928	3,503	4,312	5,269	1,557	nd
Average number of exons per gene	5.72	7.3	8.2	8.33	7.78	1.9	1.57
Average exon length (bp)	207	170	182	190	194	731	750
Mean length of introns (bp)	246	209	240	373	491	187	103
Coding sequence ratio (%)	3.2	4.7	5.0	8.0	9.5	2.1	1.6

CPRO: *Chlorella protothecoides* sp.0710; *CVAR*: *Chlorella variabilis* NC64A; *CSUB*: *Coccomyxa subellipsoidea* C-169; *CREI*: *Chlamydomonas reinhardtii*; *VCAR*: *Volvox carteri*; *MPUS*: *Micromonas pusilla* CCMP1545; *OTAR*: *Ostreococcus tauri*.
T: Trebouxiophyceae; *C*: Chlorophyceae; *M*: Mamiellophyceae.

Appendix F

Differential expressed gene based on pathways

Calvin-Bensen cycle

Enzyme	EC number	Gene name	Transcript found	DEseq2	lfcMLE
Phosphoribulokinase	EC:2.7.1.19	PRK	1	UP	0.47902
3-phospho-D-glycerate carboxy-lyase	EC 4.1.1.39	rbcS	5	UP	0.811944
3-phospho-D-glycerate carboxy-lyase	EC 4.1.1.39	rbcL	ND	ND	ND
phosphoglycerate kinase	EC 2.7.2.3	PGK	6	UP	0.9602
GAPDH	EC 1.2.1.12 + EC 1.2.1.59	GAPDHS	4	UP	1.29015
Glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating	EC 1.2.1.13 + EC 1.2.1.59	G3P1 GAPDH	1	UP	0.689615
Fructose-bisphosphate aldolase	EC 4.1.2.13	ALDO	8	UP	1.097251
Fructose-1,6-bisphosphatase I	EC 3.1.3.11	FBP	4	UP	0.9871
Transketolase	EC 2.2.1.1	tkt	2	X	X
Fructose-bisphosphate aldolase, class I	EC 4.2.1.13	ALDO	8	UP	1.097251
Sedoheptulose-bisphosphatase	EC 3.1.3.37		2	X	X
Transketolase	EC 2.2.1.1	tkt	2	X	X
Ribose 5-phosphate isomerase A	EC 5.3.1.6	rpiA	7	UP	1.719245

Citric acid (TCA) cycle

Enzyme	EC number	Gene name	Transcript found	DEseq2	lfcMLE
Citrate synthase	EC 2.3.3.1	CS	7	UP	2.909657
Aconitase	EC 4.2.1.3	ACO2	0	ND	ND
Aconitase	EC 4.2.1.3	ACO	1	UP	1.311402
Isocitrate dehydrogenase	EC 1.1.1.42	IDH1	1	UP	0.612413
Isocitrate dehydrogenase	EC 1.1.1.41	IDH3	2	UP	1.264615
α-Ketoglutarate dehydrogenase	EC 1.2.4.2, EC 2.3.1.61, EC 1.8.1.4	(OGDH, DLST,DLD) =OGDC)	1, 3, 3	UP	1.306047
Succinyl-CoA synthetase	EC 6.2.1.4		0, 3	UP	2.204495
Succinate dehydrogenase	EC 1.3.5.1		1,5,1,0	UP	2.280094
Fumarase	EC 4.2.1.2	fumC	2	UP	1.423942
Malate dehydrogenase	EC 1.1.1.37	MDH1, MDH2	9	UP	2.835736

Isoprenoid biosynthesis (non Mevalonate pathway or MEP pathway)

Enzyme	EC number	Gene name	Transcript found	DEseq2	lfcMLE
DOXP synthase	EC 2.2.1.7	Dxs	5	X	X
DOXP reductase	EC 1.1.1.267	IspC	2	X	X
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	EC 2.7.7.60	IspD	1	UP	1.86231
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	EC 2.7.1.148	IspE	1	X	X
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	EC 4.6.1.12	IspF	2	X	X
HMB-PP synthase	EC 1.17.7.1 , EC 1.17.7.3	IspG	1	UP	0.408189
HMB-PP reductase	EC 1.17.1.2	IspH	1	X	X
isopentenyl-diphosphate delta-isomerase	EC:5.3.3.2	idi, IDI	1	X	X

Fatty acid biosynthesis

Enzyme	EC number	Gene name	Transcript found	DEseq2	log2FC
Acetyl Coa carboxylase	EC 6.4.1.2	ACACA	3	DOWN	-2.10657
Acetyl CoA:ACP transacylase	EC 2.3.1.38, EC 2.3.1.39	ACAT, KASIII	1	UP	1.13192
Malonyl CoA:ACP transacylase	EC 2.3.1.39	FAS	2	X	X
3-ketoacyl-ACP synthase	EC 2.3.1.41 @ EC 2.3.1.179	KASI, KASII	4	X	X
3-ketoacyl-ACP reductase	EC 1.1.1.100	FAS2	3	UP	0.928745
3-Hydroxyacyl ACP dehydrase	EC 4.2.1.17	ECH	1	X	X
Enoyl-ACP reductase	EC 1.3.1.9	EAR	2	X	X
Fatty acyl-ACP thioesterase A	EC 3.1.2.14	FATA	1	DOWN	-0.69787

Fatty acid biosynthesis

Enzyme	EC number	Gene name	Transcript found	DEseq2	log2FC
Acetyl Coa carboxylase	EC 6.4.1.2	ACACA	3	DOWN	-2.10657
Acetyl CoA:ACP transacylase	EC 2.3.1.38,EC 2.3.1.39	ACAT, KASIII	1	UP	1.13192
Malonyl CoA:ACP transacylase	EC 2.3.1.39	FAS	2	X	X
3-ketoacyl-ACP synthase	EC 2.3.1.41 @ EC 2.3.1.179	KASI @ KASII	4	X	X
3-ketoacyl-ACP reductase	EC 1.1.1.100	FAS2	3	UP	0.928745
3-Hydroxyacyl ACP dehydrase	EC 4.2.1.17	ECH	1	X	X
Enoyl-ACP reductase	EC 1.3.1.9	EAR	2	X	X
Fatty acyl-ACP thioesterase A	EC 3.1.2.14	FATA	1	DOWN	-0.69787

Elongation in mitochondria (2 block missing)

Enzyme	EC number	Gene name	Transcript found	DEseq2	log2FC
Mitochondrial trans-2-enoyl-CoA reductase	EC 1.3.1.38	MECR, NRBF1	6	DOWN	-1.44502

Elongation in ER

Enzyme	EC number	Gene name	Transcript found	DEseq2	log2FC
3-ketoacyl-CoA synthase	EC:2.3.1.199	KCS	8	UP	0.499532
17beta-estradiol 17-dehydrogenase / very-long-chain 3-oxoacyl-CoA reductase	EC:1.1.1.62 EC 1.1.1.330	KAR, IFA38 (unsaturation)	1	X	X
Very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase	EC:4.2.1.134	PHS1, PAS2 (unsaturation)	4	UP	1.695083
Very-long-chain enoyl-CoA reductase	EC:1.3.1.93	TER, TSC13, CER10 (unsaturation)	3	UP	1.611245
3-hydroxyacyl-CoA dehydrogenase / enoyl-CoA hydratase / 3-hydroxybutyryl-CoA epimerase	EC:1.1.1.35, EC 4.2.1.17, EC 5.1.2.3	fadJ	1	X	X

Unsaturated fatty acid biosynthesis

Enzyme	EC number	Gene name	Transcript found	DE	log2FC
Stearyl-ACP desaturase, Δ9-D	EC 1.14.19.2	DESA1	6	UP	4.211031
Stearoyl-CoA desaturase (delta-9 desaturase)	EC:1.14.19.1	SCD, desC	2	UP	0.83664
Omega-6 fatty acid desaturase (delta-12 desaturase)	EC:1.14.19.-	FAD6, desA	1	X	X
Omega-6 fatty acid desaturase (delta-12 desaturase)	EC:1.14.19.-	FAD2	4	UP	0.897129
Omega-3 fatty acid desaturase (delta-15 desaturase)	EC:1.14.19.-	FAD8, desB	1	x	x

Fatty acid degradation or catabolism

Enzyme	EC number	gene name	Transcript found	DE	log2FC
Long-chain acyl-CoA synthetase	EC:6.2.1.3	ACSL, fadD	18	DOWN	-0.60388
Alcohol dehydrogenase	EC 1.1.1.1	frmA, ADH5, adhC	2	UP	1.08549
Aldehyde dehydrogenase (NAD+)	EC:1.2.1.3	E1.2.1.3	7	DOWN	-2.237
Alkane 1-monooxygenase	EC:1.14.15.3	alkB1_2	2	X	X
Acyl-CoA oxidase	EC:1.3.3.6	ACOX1, ACOX3	7	X	X
Acyl-CoA dehydrogenase	EC:1.3.8.7	ACADM, acd	1	X	X
Acetyl-CoA C-acetyltransferase	EC:2.3.1.9	atoB	1	X	X
Enoyl-CoA hydratase	EC:4.2.1.17	paaF, echA	1	UP	0.427783
3-hydroxybutyryl-CoA epimerase	EC 5.1.2.3	fadJ	1	X	X
Acetaldehyde dehydrogenase / alcohol dehydrogenase	EC:1.2.1.10, EC 1.1.1.1	adhE	4	UP	1.304281
Acetyl-CoA acyltransferase 1	EC:2.3.1.16	ACAA1	2	X	X
3-hydroxyacyl-CoA dehydrogenase	EC 1.1.1.35, EC 1.1.1.211	MFP2	1	X	X
3,2-trans-enoyl-CoA isomerase, mitochondrial	EC:5.3.3.8	DCI	5	X	X
Alcohol dehydrogenase	EC:1.1.1.1	yiaY	2	DOWN	-0.62254

Appendix G

DE analysis of *C. sorokiniana* small RNA-seq data using CLC Genomic Workbench

Feature ID	Small RNA - Length	EDGE test			T-test			Baggerley's test		
		Normal vs Stress, tagwise dispersions			Normal vs Stress; normalized value			Stress vs Normal normalized values		
		p-value	FC	FDR	FC	P-value	FDR	Weighted proportions FC	P-value	FDR
AAAGACAACTCTCAACA	17	1.8E-03	4024.54	0.02	inf	2.1E-03	0.03	inf	0.00	0.00
AAAGACAACTCTCAACAA	18	3.0E-05	5301.89	0.00	inf	1.4E-03	0.02	inf	0.00	0.00
AAAGACAACTCTCAACAACGA	21	9.7E-03	2528.78	0.05	inf	2.3E-03	0.03	inf	0.00	0.00
AAAGACAACTCTCAACAACGGATAA	25	5.4E-03	3041.02	0.03	inf	3.0E-03	0.03	inf	0.00	0.00
AACAGACCGAATGATTTTGACA	22	4.6E-10	8046.55	0.00	inf	5.1E-04	0.02	inf	0.00	0.00
AACGACTGAACGCCTCTA	18	3.9E-01	1108.87	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
AACGACTGAACGCCTCTAA	19	4.6E-01	893.23	0.69	inf	3.7E-01	0.47	inf	0.32	0.36
AACTCTCAACAACGAA	16	5.7E-01	-670.23	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
AACTCTCAACAACGGA	16	8.5E-02	-1.60	0.29	-1.50	8.3E-02	0.35	-1.13	0.00	0.00
AACTCTCAACAACGGAA	17	5.3E-01	-1.19	0.70	-1.43	2.5E-01	0.47	-1.07	0.00	0.00
AACTCTCAACAACGGATA	18	9.6E-01	-1.01	0.97	-1.25	3.6E-01	0.47	1.07	0.00	0.00
AACTCTCAACAACGGATAA	19	5.8E-03	-1.64	0.03	-1.94	2.4E-01	0.47	-1.41	0.00	0.00
AACTCTCAACAACGGATATA	20	4.9E-03	3114.38	0.03	inf	5.8E-04	0.02	inf	0.00	0.00
AACTCTCAACAACGGATATCA	21	5.6E-01	-1.17	0.70	-1.58	2.4E-01	0.47	-1.19	0.00	0.00
AACTCTCAACAACGGATATCTA	22	3.8E-01	1.53	0.66	1.13	5.7E-01	0.63	1.50	0.00	0.00
AACTCTCAACAACGGATATCTTA	23	4.8E-01	2.74	0.70	2.22	1.7E-01	0.42	2.31	0.02	0.04
AACTCTCAACAACGGATATCTTGA	24	5.7E-01	2.34	0.70	1.93	2.2E-01	0.47	2.00	0.06	0.12
AACTCTCAACAACGGATATCTTGGCA	26	8.4E-01	1.43	0.90	1.33	6.2E-01	0.67	1.37	0.49	0.54
AAGACAACTCTCAACA	16	9.1E-03	-3.47	0.05	-4.10	1.2E-02	0.09	-3.08	0.00	0.00
AAGACAACTCTCAACAA	17	8.7E-04	-3.79	0.01	-5.00	7.9E-03	0.06	-3.76	0.00	0.00
AAGACAACTCTCAACAACGA	20	7.6E-05	-5376.25	0.00	-inf	6.4E-03	0.06	-inf	0.00	0.00
AAGACAACTCTCAACAACGGA	21	2.1E-03	-4172.52	0.02	-inf	3.8E-03	0.04	-inf	0.00	0.00
AAGACAACTCTCAACAACGGATA	23	5.7E-01	-727.97	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
AAGACAACTCTCAACAACGGATAA	24	5.2E-08	-7233.88	0.00	-inf	1.6E-02	0.12	-inf	0.00	0.00
AAGCTGTGGGATATTA	16	7.6E-05	-2.04	0.00	-2.45	1.9E-01	0.44	-1.82	0.03	0.07
AATACCTCTGCAACATCCA	19	3.5E-02	-2184.26	0.13	-inf	2.0E-02	0.13	-inf	0.00	0.00
AATCCCCAGTGCCGTAGTGTA	21	5.8E-01	3.02	0.71	1.69	5.5E-01	0.62	3.38	0.02	0.05
ACAGACCGAATGATTTTGACA	21	7.4E-01	1.09	0.84	-1.10	7.2E-01	0.75	1.21	0.00	0.00

ACCGAATGATTTTGACA	17	7.2E-01	1.12	0.82	-1.11	2.5E-01	0.47	1.20	0.00	0.00
ACCGACCATGATCTTCTA	18	2.0E-01	-1333.82	0.52	-inf	1.2E-01	0.41	-inf	0.06	0.11
ACCGACCATGATCTTCTGTGA	21	2.9E-01	-869.56	0.64	-inf	3.7E-01	0.47	-inf	0.32	0.36
ACCTGGTTGATCCTGA	16	8.3E-01	2.23	0.90	1.36	7.9E-01	0.81	1.75	0.57	0.60
ACCTGGTTGATCCTGCA	17	8.8E-01	1.09	0.92	-1.25	4.2E-01	0.51	1.07	0.02	0.06
ACCTGGTTGATCCTGCCA	18	6.6E-02	-1.65	0.23	-1.85	1.3E-01	0.41	-1.40	0.00	0.00
ACCTGGTTGATCCTGCCAA	19	8.4E-01	-2.06	0.90	-3.21	3.8E-01	0.47	-1.99	0.51	0.55
ACGGGAATCCCCAGTGCCGTAGTGTA	26	4.7E-01	885.14	0.69	inf	3.7E-01	0.47	inf	0.32	0.36
ACTAACAGACCGAATGATTTTGACA	25	9.0E-06	5525.52	0.00	inf	1.0E-03	0.02	inf	0.00	0.00
ACTCTCAACAACGGATAA	18	2.5E-01	-6.84	0.59	-6.44	2.5E-02	0.13	-5.69	0.00	0.00
AGACCGAATGATTTTGACA	19	2.1E-03	4.18	0.02	2.81	6.2E-03	0.06	3.75	0.00	0.00
AGCCGTTGTTCTTACGATTTGTA	23	1.7E-02	2.60	0.08	2.05	1.8E-02	0.12	2.73	0.00	0.00
AGCGGTGAAATGCGTAGA	18	4.1E-01	1033.40	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
AGGCTGAAACTTAAAGGAA	19	4.0E-01	1054.96	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
AGGGCATGTCTGCCTCAGCGTCGGTA	26	1.2E-01	10.19	0.36	5.70	2.4E-02	0.13	6.70	0.00	0.00
ATCCCCAGTGCCGTAGTGTA	20	4.3E-01	1.78	0.66	1.11	8.4E-01	0.86	1.63	0.00	0.01
ATGTCTGCCTCAGCGTCGGCA	21	9.1E-01	-1.57	0.94	-3.78	2.7E-01	0.47	-1.67	0.63	0.65
ATGTCTGCCTCAGCGTCGGGA	21	4.2E-01	-652.08	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
ATGTCTGCCTCAGCGTCGGTA	21	5.6E-01	-1.11	0.70	-1.55	3.8E-01	0.47	-1.12	0.23	0.36
CAAATGAAGTAAGGCGTA	18	7.9E-01	-2.08	0.88	-2.25	5.2E-01	0.60	-1.82	0.55	0.58
CAAATGAAGTAAGGCGTAA	19	6.1E-01	-1.89	0.73	-2.61	1.3E-01	0.41	-1.78	0.12	0.20
CAAATGAAGTAAGGCGTAGA	20	8.5E-01	-1.31	0.90	-2.01	3.0E-01	0.47	-1.33	0.52	0.56
CAAATGAAGTAAGGCGTAGGA	21	3.6E-01	-1.51	0.66	-2.40	3.4E-01	0.47	-1.58	0.20	0.33
CAAATGAAGTAAGGCGTAGGGA	22	1.2E-01	-1357.61	0.38	-inf	2.0E-01	0.44	-inf	0.05	0.10
CAAATGAAGTAAGGCGTAGGGTA	23	2.8E-02	-2183.48	0.11	-inf	2.2E-02	0.13	-inf	0.00	0.00
CAACTCTCAACAACGA	16	1.3E-01	-2096.13	0.39	-inf	1.5E-01	0.41	-inf	0.05	0.11
CAACTCTCAACAACGGA	17	1.6E-01	-1831.80	0.44	-inf	1.5E-01	0.41	-inf	0.05	0.10
CAACTCTCAACAACGGATAA	20	4.0E-01	-2.86	0.66	-3.51	5.7E-02	0.27	-2.39	0.01	0.02
CAAGGCTGAAACTTAAAGGAA	21	4.1E-01	1025.31	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
CAGACCGAATGATTTTGACA	20	5.1E-01	1.19	0.70	-1.03	9.3E-01	0.95	1.27	0.04	0.08
CAGCCGTTGTTCTTACGATTTGTA	24	6.6E-01	1.21	0.78	-1.09	6.3E-01	0.68	1.22	0.00	0.00
CAGGTCTGTGATGCCCTTA	19	2.2E-01	-1347.91	0.55	-inf	1.9E-01	0.44	-inf	0.07	0.13
CAGGTCTGTGATGCCCTTAGATGA	24	5.7E-02	-2907.57	0.21	-inf	2.2E-01	0.47	-inf	0.10	0.17
CATGTCTGCCTCAGCGTCGGCA	22	5.6E-01	-681.76	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
CATGTCTGCCTCAGCGTCGGTA	22	4.9E-03	-1.66	0.03	-1.96	8.5E-02	0.35	-1.47	0.00	0.00
CCAGTGCCGTAGTGTA	16	2.4E-01	6.86	0.57	3.56	1.0E-01	0.40	7.11	0.00	0.00
CCCAGTGCCGTAGTGTA	17	5.4E-01	2.29	0.70	1.61	4.8E-01	0.56	2.43	0.01	0.02
CCCCAGTGCCGTAGTGTA	18	4.1E-01	2.32	0.66	1.39	6.6E-01	0.70	2.31	0.04	0.08
CCGAAGCTGTGGGATATTA	19	8.6E-01	-1.16	0.90	-1.68	4.5E-01	0.53	-1.16	0.53	0.57
CCGAATGATTTTGACA	16	2.6E-02	2.59	0.11	1.91	3.0E-02	0.15	2.55	0.00	0.00

CCGTGAACCATCGAATCA	18	4.3E-01	-590.40	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
CCGTTGTTCTTACGATTTGTA	21	5.7E-01	856.06	0.70	inf	3.7E-01	0.47	inf	0.32	0.36
CCTCAGTAGCTCAGCA	16	2.5E-02	-2496.47	0.11	-inf	1.6E-01	0.41	-inf	0.05	0.10
CCTGTAGCTCAGTGGACTAGAGCACA	26	1.1E-01	-1458.14	0.36	-inf	1.5E-01	0.41	-inf	0.05	0.11
CCTTGTATTCTTGCTTGA	18	2.8E-15	10.54	0.00	6.98	1.1E-03	0.02	8.16	0.00	0.00
CGAAGCTGTGGGATATTA	18	3.5E-01	1.80	0.66	1.19	6.0E-01	0.65	1.60	0.00	0.00
CGAATACAAACTGTGA	16	1.3E-01	-1349.70	0.39	-inf	2.2E-01	0.47	-inf	0.05	0.11
CGACGTTGCTTTTTGATCA	19	2.3E-02	-2.64	0.10	-3.89	2.5E-01	0.47	-2.75	0.06	0.11
CGACGTTGCTTTTTGATCCA	20	8.2E-01	-2.17	0.90	-2.87	3.7E-01	0.47	-2.14	0.44	0.49
CGGCTGGAACACCTCCTA	18	4.2E-01	1036.41	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
CGGGAATCCCCAGTGCCGTAGTGTA	25	4.1E-01	1014.53	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
CGGTGAAATGCGTAGA	16	2.2E-02	3331.47	0.10	inf	1.5E-01	0.41	inf	0.08	0.13
CGTGAGCTGGGTTTAGACCA	20	2.2E-01	-1306.61	0.55	-inf	1.6E-01	0.41	-inf	0.05	0.11
CGTTGTTCTTACGATTTGTA	20	1.7E-07	6443.18	0.00	inf	1.2E-05	0.00	inf	0.00	0.00
CTAACAGACCGAATGATTTTGACA	24	2.7E-01	1.64	0.62	1.23	1.6E-01	0.42	1.64	0.00	0.00
CTGCCGAAGCTGTGGGATATTA	22	5.9E-01	-1.37	0.71	-2.04	4.0E-01	0.48	-1.34	0.37	0.41
CTGCCTCAGCGTCGGCA	17	9.7E-01	-1.16	0.97	-2.75	3.7E-01	0.47	-1.24	0.85	0.86
CTGCCTCAGCGTCGGGA	17	4.2E-01	-585.74	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
CTGCCTCAGCGTCGGTA	17	5.4E-01	1.12	0.70	-1.15	7.3E-01	0.76	1.19	0.02	0.04
CTGCGGAAGGATCATTGA	18	2.5E-08	7175.77	0.00	inf	5.3E-04	0.02	inf	0.00	0.00
GAAGCTGTGGGATATTA	17	6.0E-01	1.95	0.72	1.44	5.6E-01	0.63	2.12	0.03	0.07
GACCGAATGATTTTGACA	18	6.8E-02	1.38	0.24	1.02	9.6E-01	0.96	1.40	0.01	0.01
GCCGAAGCTGTGGGATATTA	20	8.6E-01	-1.10	0.90	-1.62	4.4E-01	0.52	-1.14	0.50	0.55
GCCGTTGTTCTTACGATTTGTA	22	2.4E-06	5823.76	0.00	inf	6.0E-06	0.00	inf	0.00	0.00
GCGGAAGGATCATTGA	16	2.7E-07	4.28	0.00	3.59	3.0E-02	0.15	4.82	0.00	0.00
GCGGCTGGAACACCTCCTA	19	1.5E-01	1709.66	0.42	inf	1.4E-01	0.41	inf	0.05	0.10
GCGGGGTAGAGCAGTCTGGTA	21	4.2E-01	963.32	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
GCGGTGAAATGCGTAGA	17	1.6E-02	3955.55	0.08	inf	1.7E-01	0.42	inf	0.10	0.17
GGAGCGGTGAAATGCGTAGA	20	3.3E-02	2730.82	0.13	inf	1.3E-01	0.41	inf	0.06	0.11
GGCCTGTAGCTCAGTGGACTA	21	3.8E-01	-851.16	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
GGCCTTTAGCTCAGCA	16	3.2E-01	-951.00	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
GGCTGAAACTTAAAGGAA	18	3.9E-01	992.97	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
GGCTGGAACACCTCCTA	17	3.5E-03	3240.12	0.02	inf	2.3E-03	0.03	inf	0.00	0.00
GGCTTGTAACTCAGTA	16	1.8E-01	-1887.88	0.47	-inf	3.7E-01	0.47	-inf	0.32	0.36
GGGAATCCCCAGTGCCGTAGTGTA	24	4.9E-01	874.36	0.70	inf	3.7E-01	0.47	inf	0.32	0.36
GGGCATGTCTGCCTCAGCGTCGGTA	25	2.0E-09	7714.14	0.00	inf	3.0E-03	0.03	inf	0.00	0.00
GGGCTGTAGCTCAAATGGTAGAA	23	5.5E-01	859.40	0.70	inf	3.7E-01	0.47	inf	0.32	0.36
GGGCTGTAGCTCAAATGGTAGAGCA	25	5.6E-01	757.66	0.70	inf	3.7E-01	0.47	inf	0.32	0.36
GGGTTGTAGCTCAGTGGTAGAGCA	24	1.3E-01	1821.81	0.39	inf	1.5E-01	0.41	inf	0.05	0.10
GGTGCGGCTGGAACACCTCCTA	22	8.6E-02	2107.40	0.29	inf	1.2E-01	0.41	inf	0.06	0.11

GTACGAATACAAACTGTGA	19	4.2E-01	-663.78	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
GTCTGCCTCAGCGTCGGAA	19	5.5E-01	-1236.15	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
GTCTGCCTCAGCGTCGGCA	19	4.9E-01	-2.11	0.70	-3.12	1.4E-01	0.41	-1.87	0.29	0.36
GTCTGCCTCAGCGTCGGGA	19	6.4E-01	-3.76	0.76	-8.30	2.0E-01	0.44	-3.34	0.23	0.36
GTCTGCCTCAGCGTCGGTA	19	2.5E-03	-1.72	0.02	-1.25	5.8E-01	0.64	1.00	0.98	0.98
GTGGAGCCTGCGGCTTAA	18	2.5E-03	3757.15	0.02	inf	1.5E-03	0.02	inf	0.00	0.00
GTTCTTACGATTTGTA	16	4.6E-01	2.84	0.68	2.40	8.5E-02	0.35	2.42	0.01	0.01
GTTGTTCTTACGATTTGTA	19	2.3E-01	4.47	0.56	3.34	1.8E-02	0.12	3.44	0.00	0.00
TAACAGACCGAATGATTTTGACA	23	2.1E-01	1.41	0.54	1.10	6.8E-01	0.71	1.46	0.00	0.00
TAACGACTGAACGCCTCTA	19	4.9E-01	879.75	0.70	inf	3.7E-01	0.47	inf	0.32	0.36
TACTAACAGACCGAATGATTTTGACA	26	2.6E-01	1.79	0.59	1.41	6.2E-02	0.28	1.87	0.00	0.00
TATCGGTAGGGGAGCA	16	7.0E-01	-2.49	0.80	-3.03	1.8E-01	0.42	-2.68	0.10	0.16
TATCGGTAGGGGAGCGTTCTA	21	4.3E-01	-551.92	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TATCGGTAGGGGAGCGTTCTGCTA	24	4.2E-01	-579.24	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TATGACAGCAATTTTAGAAA	20	3.6E-01	-880.15	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TCAGCCGTTGTTCTTACGATTTGCA	25	5.6E-01	-635.56	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
TCAGCCGTTGTTCTTACGATTTGTA	25	6.8E-01	1.12	0.78	1.01	9.6E-01	0.96	1.35	0.00	0.00
TCCCCAGTGCCGTAGTGTA	19	6.8E-01	-1.21	0.78	-1.91	5.0E-01	0.58	-1.23	0.64	0.67
TCGACCTGAGCTCAGA	16	7.8E-01	-1.08	0.87	-1.57	4.1E-01	0.50	-1.10	0.50	0.55
TCGACCTGAGCTCAGGA	17	4.9E-01	-1.21	0.70	-1.75	1.5E-01	0.41	-1.31	0.00	0.00
TCGACCTGAGCTCAGGAA	18	5.6E-01	-612.45	0.70	-inf	3.7E-01	0.47	-inf	0.32	0.36
TCGACCTGAGCTCAGGCA	18	4.8E-03	-1.66	0.03	-1.84	3.1E-01	0.47	-1.35	0.10	0.17
TCGACCTGAGCTCAGGCAA	19	8.7E-04	-1.81	0.01	-1.96	3.3E-01	0.47	-1.43	0.15	0.24
TCGACCTGAGCTCAGGCAAA	20	5.3E-01	-3.79	0.70	-4.20	9.8E-02	0.39	-3.71	0.01	0.02
TCGACCTGAGCTCAGGCAAGA	21	2.4E-02	-1.85	0.10	-2.54	1.5E-01	0.41	-1.85	0.00	0.00
TCGACCTGAGCTCAGGCAAGAA	22	7.7E-01	-1.17	0.86	-1.60	1.8E-01	0.42	-1.20	0.00	0.00
TCGACCTGAGCTCAGGCAAGATTA	24	3.4E-01	-1.30	0.66	-1.90	2.0E-01	0.44	-1.40	0.00	0.00
TCGACCTGAGCTCAGGCGA	19	2.7E-02	-3300.73	0.11	-inf	1.9E-01	0.44	-inf	0.06	0.12
TCGACCTGAGCTCAGGTA	18	4.3E-01	-2.25	0.66	-3.48	1.4E-01	0.41	-2.03	0.14	0.23
TCTAGAAAGTGTCGAACATTCA	22	4.3E-01	-5.31	0.66	-4.68	4.1E-02	0.20	-4.13	0.00	0.01
TCTGCCGAAGCTGTGGGATATTA	23	1.6E-01	-1.46	0.44	-1.99	2.7E-01	0.47	-1.44	0.07	0.13
TCTGCCTCAGCGTCGGTA	18	1.1E-01	1.54	0.36	1.18	5.7E-01	0.63	1.61	0.00	0.01
TCTGTGAAAGGTTCGAGTA	19	3.6E-01	-770.56	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TGCCGAAGCTGTGGGATATTA	21	9.4E-01	1.39	0.96	-1.18	9.1E-01	0.93	1.45	0.80	0.81
TGCCTCAGCGTCGGTA	16	9.7E-04	2.48	0.01	1.78	1.1E-01	0.41	2.42	0.00	0.00
TGCGGAAGGATCATTGA	17	6.6E-03	10.94	0.04	5.98	4.6E-03	0.05	9.04	0.00	0.00
TGCTACGGTCATACCATCCA	20	4.4E-01	-545.40	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TGCTACGGTCATACCATCCAGAA	23	9.1E-01	-1.07	0.94	-1.60	4.8E-01	0.56	-1.13	0.59	0.62
TGCTACGGTCATACCATCCAGAACA	25	6.8E-01	1.21	0.78	-1.22	7.1E-01	0.75	1.16	0.42	0.48
TGCTACGGTCATACCATCCAGAACAA	26	4.1E-01	-629.98	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36

TGTCTGCCTCAGCGTCGGA	19	3.7E-01	-726.31	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TGTCTGCCTCAGCGTCGGAA	20	9.1E-01	-1.79	0.94	-2.88	5.7E-01	0.63	-1.58	0.76	0.78
TGTCTGCCTCAGCGTCGGCA	20	1.8E-01	-1.92	0.47	-2.73	1.3E-01	0.41	-1.99	0.04	0.09
TGTCTGCCTCAGCGTCGGGA	20	2.8E-01	-4.64	0.63	-9.77	1.6E-01	0.41	-4.10	0.08	0.15
TGTCTGCCTCAGCGTCGGTA	20	1.3E-04	-1.99	0.00	-1.32	5.4E-01	0.62	-1.06	0.77	0.78
TGTTCTTACGATTTGTA	17	2.0E-06	5841.15	0.00	inf	1.3E-03	0.02	inf	0.00	0.00
TGTTGGATAGGTTAAAGCATTCA	23	3.1E-01	1672.24	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
TTAAGTTGCTGCAGTA	16	3.7E-02	-2159.39	0.14	-inf	2.4E-02	0.13	-inf	0.00	0.00
TTCACGGAGAGTTTGA	16	9.8E-01	-1.02	0.98	-1.33	2.7E-01	0.47	1.00	0.93	0.93
TTCACGGAGAGTTTGATA	18	8.6E-01	1.16	0.90	-1.14	5.9E-01	0.65	1.17	0.00	0.00
TTCACGGAGAGTTTGATCA	19	4.2E-01	3.24	0.66	2.59	8.8E-02	0.36	3.54	0.00	0.00
TTCACTTGGGGTCGTGGGAAAA	22	4.1E-01	-722.38	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36
TTCAGCCGTTGTTCTTACGATTTGTA	26	1.5E-02	8.57	0.08	7.16	6.9E-03	0.06	7.25	0.00	0.00
TTCGATGTCGGCTCTTCCTATCATTA	26	2.8E-04	-4848.91	0.00	-inf	2.1E-02	0.13	-inf	0.00	0.00
TTGAATCGATCGAATCCA	18	4.2E-01	922.06	0.66	inf	3.7E-01	0.47	inf	0.32	0.36
TTGTTCTTACGATTTGTA	18	1.6E-01	2.04	0.44	1.55	2.6E-02	0.14	2.07	0.00	0.00
TTTCACTTGGGGTCGTGGGAAAA	23	2.2E-01	-1339.88	0.55	-inf	1.7E-01	0.42	-inf	0.06	0.11
TTTCGGGCAGAGTGCAA	17	4.3E-01	-632.24	0.66	-inf	3.7E-01	0.47	-inf	0.32	0.36

Appendix H

DE analysis of *C. zofingiensis* small RNA-seq data using CLC Genomic Workbench

Feature ID	Small RNA - Length	EDGE test			T-test			Baggerley's test		
		Normal vs Stress, tagwise dispersions			Normal vs Stress; normalized value			Stress vs Normal normalized values -		
		p-value	FC	FDR	FC	P-value	FDR	FC	P-value	FDR
AAAGAACTTTGAAAAGAGAGTA	22	0.31	-2.69	0.51	-2.82	0.06	0.15	-2.82	0.01	0.02
AAATCGGTAGGGGAGCA	17	0.03	-3969.18	0.07	- inf	0.14	0.25	- inf	0.06	0.12
AAATGAAGGCTTACGGTA	18	0.53	-1405.98	0.66	- inf	0.37	0.44	- inf	0.32	0.37
AACTCTCAACAACGGA	16	0.62	1.12	0.71	1.33	0.08	0.17	1.33	0.02	0.04
AACTCTCAACAACGGAA	17	0.77	2.47	0.85	2.42	0.42	0.49	2.42	0.37	0.43
AACTCTCAACAACGGATA	18	0.96	-1.01	0.97	1.11	0.26	0.41	1.11	0.19	0.30
AACTCTCAACAACGGATAA	19	0.42	1.13	0.59	1.45	0.08	0.18	1.45	0.02	0.04
AACTCTCAACAACGGATATCA	21	0.83	-1.05	0.91	1.12	0.08	0.18	1.12	0.02	0.05
AACTCTCAACAACGGATATCTA	22	0.88	-1.05	0.94	1.00	0.98	0.99	1.00	0.97	0.99
AAGCTGTGAGATTTTCA	17	0.46	-1566.90	0.61	- inf	0.37	0.44	- inf	0.32	0.37
AAGTAGGACATCGTCAGGCTA	21	0.00	6432.25	0.00	inf	0.00	0.00	inf	0.00	0.00
AATCTTTGAACGCATAA	17	0.36	1.44	0.56	1.60	0.04	0.11	1.60	0.00	0.01
AATCTTTGAACGCATATTA	19	0.25	1533.00	0.43	inf	0.37	0.44	inf	0.32	0.37
AATGACAACTCTCAACA	17	0.91	1.05	0.96	1.19	0.24	0.39	1.19	0.17	0.28
AATGACAACTCTCAACAA	18	0.65	1.31	0.74	1.31	0.20	0.33	1.31	0.12	0.21
AATTGACGTATAGATGAACTA	21	0.54	-1402.80	0.66	- inf	0.37	0.44	- inf	0.32	0.37
ACAGATTGAGAGCTCTTTCA	20	0.03	3393.62	0.08	inf	0.12	0.22	inf	0.05	0.09
ACCCCGAGTGCCGCAGTGTA	20	0.00	-4870.60	0.00	- inf	0.00	0.01	- inf	0.00	0.00
ACCTGGTTGATCCTGA	16	0.56	1.84	0.67	1.58	0.31	0.44	1.58	0.25	0.37
ACCTGGTTGATCCTGCA	17	0.95	1.04	0.97	1.05	0.52	0.59	1.05	0.49	0.55
ACCTGGTTGATCCTGCCA	18	0.00	-2.06	0.00	-1.67	0.00	0.00	-1.67	0.00	0.00
ACCTGGTTGATCCTGCCAA	19	0.76	1.18	0.85	1.18	0.31	0.44	1.18	0.24	0.37
ACCTGGTTGATCCTGCTA	18	0.48	-1462.24	0.64	- inf	0.37	0.44	- inf	0.32	0.37
ACCTTCGACCTGAGCTCAGGCAA	23	0.53	-1224.78	0.66	- inf	0.37	0.44	- inf	0.32	0.37
ACGACTTGGCCGAGCA	16	0.49	-2.01	0.64	-1.98	0.17	0.29	-1.98	0.10	0.16
ACTAAAGACCAATTGACA	18	0.00	-12966.29	0.00	- inf	0.00	0.00	- inf	0.00	0.00
ACTCTCAACAACGGATA	17	0.91	-1.07	0.96	-1.07	0.63	0.69	-1.07	0.60	0.66
ACTCTCAACAACGGATAA	18	0.82	1.06	0.91	1.08	0.64	0.70	1.08	0.62	0.68
ACTCTCAACAACGGATATCA	20	0.87	-1.07	0.94	1.02	0.90	0.93	1.02	0.89	0.93
ACTCTCAACAACGGATATCTA	21	0.38	-3.38	0.58	-3.59	0.06	0.15	-3.59	0.01	0.03

AGAGCATGTCTGCCTCAGCGTCGGTA	26	0.44	-1595.08	0.60	- inf	0.37	0.44	- inf	0.32	0.37
AGCCAGAACTACTGACCCTCA	21	0.09	-2864.26	0.20	- inf	0.12	0.22	- inf	0.05	0.09
AGCCCGTGTTCGGAAGATTTGTA	23	0.41	1.21	0.59	1.42	0.17	0.29	1.42	0.09	0.16
AGCGCAAGTAACCTGACTA	19	0.00	-4809.85	0.00	- inf	0.00	0.01	- inf	0.00	0.00
AGCTCATCGTCTAATGGATAGGACAA	26	0.00	-8665.39	0.00	- inf	0.01	0.03	- inf	0.00	0.00
AGGACATCGTCAGGCTA	17	0.00	5472.21	0.00	inf	0.01	0.02	inf	0.00	0.00
AGGTCATCGCCAGGCTA	17	0.00	9150.86	0.00	inf	0.00	0.00	inf	0.00	0.00
ATACTAAAGACCAATTGACA	20	0.00	-5404.10	0.00	- inf	0.00	0.00	- inf	0.00	0.00
ATGACAACTCTCAACA	16	0.15	-2571.09	0.30	- inf	0.12	0.22	- inf	0.05	0.09
ATGCAATTCCGGCGGGCA	18	0.25	1583.75	0.43	inf	0.37	0.44	inf	0.32	0.37
ATGCAATTCCGGCGGGCCA	19	0.00	5308.41	0.00	inf	0.00	0.01	inf	0.00	0.00
ATGTCTGCCTCAGCGTCGGTA	21	0.53	1.16	0.66	1.42	0.05	0.11	1.42	0.00	0.01
CAGCCCGTGTTCGGAAGATTTGTA	24	0.35	-1.25	0.54	-1.06	0.70	0.75	-1.06	0.68	0.73
CATGTCTGCCTCAGCGTCGGCA	22	0.89	-1.09	0.95	-1.15	0.66	0.72	-1.15	0.64	0.69
CATGTCTGCCTCAGCGTCGGGA	22	0.52	-1341.48	0.66	- inf	0.37	0.44	- inf	0.32	0.37
CATGTCTGCCTCAGCGTCGGTA	22	0.16	-1.39	0.30	-1.07	0.52	0.59	-1.07	0.48	0.55
CCAGGTCGGGGAGCCAA	17	0.39	1600.86	0.58	inf	0.37	0.44	inf	0.32	0.37
CCCAGGTGGCGGAATTGGTAGACGA	25	0.03	3410.93	0.07	inf	0.12	0.22	inf	0.05	0.09
CCCCGAGTGCCGCAGTGTA	19	0.00	-9728.21	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CCCGTGTTCGGAAGATTTGTA	21	0.00	3.10	0.00	4.01	0.00	0.00	4.01	0.00	0.00
CCGGTATGGTGTAGGCA	17	0.34	-2183.65	0.54	- inf	0.37	0.44	- inf	0.32	0.37
CCGGTATGGTGTAGGCGA	18	0.52	-1349.48	0.66	- inf	0.37	0.44	- inf	0.32	0.37
CCGTGTTCGGAAGATTTGTA	20	0.00	7.66	0.00	8.64	0.00	0.00	8.64	0.00	0.00
CCTTCGACCTGAGCTCAGGCA	21	0.42	1.42	0.59	1.56	0.00	0.01	1.56	0.00	0.00
CCTTCGACCTGAGCTCAGGCAA	22	0.53	1.95	0.66	1.77	0.23	0.38	1.77	0.16	0.26
CCTTCGTGATGGAACTA	17	0.00	-9758.60	0.00	- inf	0.03	0.06	- inf	0.00	0.00
CGACGTTGCTTTTTGA	16	0.96	1.22	0.97	1.06	0.97	0.99	1.06	0.97	0.99
CGACGTTGCTTTTTGATCA	19	0.59	1.14	0.68	1.21	0.44	0.51	1.21	0.39	0.45
CGACGTTGCTTTTTGATCCA	20	0.95	1.27	0.97	1.07	0.96	0.99	1.07	0.96	0.99
CGAGTGCCGCAGTGTA	16	0.21	1959.90	0.39	inf	0.37	0.44	inf	0.32	0.37
CGCCTACCGTCGTGCTTA	18	0.48	-1484.38	0.63	- inf	0.37	0.44	- inf	0.32	0.37
CGCCTACCGTCGTGCTTCA	19	0.00	-6368.25	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGCCTACCGTCGTGCTTCGA	20	0.00	-24123.08	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGCCTACCGTCGTGCTTCGAA	21	0.23	-2758.44	0.40	- inf	0.37	0.44	- inf	0.32	0.37
CGCCTACCGTCGTGCTTCGGA	21	0.00	-8.14	0.00	-7.14	0.00	0.00	-7.14	0.00	0.00
CGCCTACCGTCGTGCTTCGGCA	22	0.05	-1.59	0.12	-1.03	0.74	0.79	-1.03	0.73	0.77
CGCCTACCGTCGTGCTTCGGCAA	23	0.03	-3841.20	0.08	- inf	0.16	0.28	- inf	0.09	0.15
CGCCTACCGTCGTGCTTCGGCATTA	25	0.00	-25463.62	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGCCTACCGTCGTGCTTCGGCCA	23	0.30	-2555.15	0.49	- inf	0.37	0.44	- inf	0.32	0.37
CGCCTACCGTCGTGCTTCGGCGA	23	0.50	-1369.66	0.64	- inf	0.37	0.44	- inf	0.32	0.37

CGCCTACCGTCGTGCTTCGGCTA	23	0.00	-10.85	0.00	-10.43	0.00	0.01	-10.43	0.00	0.00
CGCCTACCGTCGTGCTTCGGCTTA	24	0.00	-15219.53	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGCCTACCGTCGTGCTTCGGCTTTA	25	0.00	-16574.14	0.00	- inf	0.00	0.01	- inf	0.00	0.00
CGCCTACCGTCGTGCTTCGGTA	22	0.14	-3.87	0.27	-3.78	0.14	0.26	-3.78	0.07	0.12
CGCCTACCGTCGTGCTTCGTA	21	0.33	-2301.54	0.53	- inf	0.37	0.44	- inf	0.32	0.37
CGCCTACCGTCGTGCTTCTA	20	0.00	-4355.68	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGGCTGTAGCTCAGTA	16	0.00	5076.24	0.00	inf	0.00	0.00	inf	0.00	0.00
CGGGAATAGCTCAGTA	16	0.23	1671.85	0.40	inf	0.37	0.44	inf	0.32	0.37
CGTGTTCGGAAGATTTGTA	19	0.00	5.32	0.00	6.22	0.00	0.00	6.22	0.00	0.00
CGTTCGTAGTCTAGGA	16	0.00	-5926.34	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGTTCGTAGTCTAGGGA	17	0.14	-1.67	0.27	-1.60	0.00	0.01	-1.60	0.00	0.00
CGTTCGTAGTCTAGGGGA	18	0.00	-4938.26	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGTTCGTAGTCTAGGGGTA	19	0.00	-4614.32	0.00	- inf	0.00	0.00	- inf	0.00	0.00
CGTTCGTAGTCTAGGGGTATGA	22	0.13	-2680.55	0.27	- inf	0.12	0.22	- inf	0.05	0.09
CTGCCTCAGCGTCGGCA	17	0.44	3.19	0.59	3.62	0.08	0.18	3.62	0.02	0.05
CTGCCTCAGCGTCGGTA	17	0.33	1.26	0.53	1.63	0.00	0.00	1.63	0.00	0.00
CTGCCTCAGCGTCGGTAA	18	0.00	4861.06	0.00	inf	0.00	0.00	inf	0.00	0.00
CTGCGGAAGGATCATTGA	18	0.00	7927.66	0.00	inf	0.00	0.00	inf	0.00	0.00
CTGCTGTAGCTCAGTGGTAGAGCA	24	0.00	5057.18	0.00	inf	0.00	0.00	inf	0.00	0.00
CTTCGACCTGAGCTCAGGCA	20	0.95	1.03	0.97	1.21	0.16	0.28	1.21	0.09	0.15
CTTCGACCTGAGCTCAGGCAA	21	1.00	-1.00	1.00	1.15	0.09	0.19	1.15	0.03	0.05
GAAAGTAGGACATCGTCAGGCTA	23	0.00	6452.21	0.00	inf	0.00	0.00	inf	0.00	0.00
GAAAGTAGGTCATCGCCAGGCTA	23	0.43	1370.29	0.59	inf	0.37	0.44	inf	0.32	0.37
GAGCATGTCTGCCTCAGCGTCGGTA	25	0.58	1.75	0.68	2.02	0.12	0.22	2.02	0.05	0.09
GAGGTAAGGGCAATTCGTGA	20	0.00	7650.22	0.00	inf	0.10	0.21	inf	0.03	0.07
GCATGTCTGCCTCAGCGTCGGTA	23	0.27	1.58	0.45	1.68	0.01	0.02	1.68	0.00	0.00
GCCCGTGTTCGGAAGATTTGTA	22	0.39	1687.53	0.58	inf	0.37	0.44	inf	0.32	0.37
GCCTCAGCGTCGGTAA	16	0.00	5337.33	0.00	inf	0.00	0.00	inf	0.00	0.00
GCGGAAGGATCATTGA	16	0.00	2.25	0.00	2.62	0.00	0.00	2.62	0.00	0.00
GGACATCGTCAGGCTA	16	0.24	1615.11	0.42	inf	0.37	0.44	inf	0.32	0.37
GGCAACACGGGGCTCCA	17	0.00	5505.03	0.00	inf	0.00	0.01	inf	0.00	0.00
GGCTCGTAGCTCAGCA	16	0.08	4053.99	0.18	inf	0.37	0.44	inf	0.32	0.37
GGCTCGTAGCTCAGCTA	17	0.34	1954.18	0.54	inf	0.37	0.44	inf	0.32	0.37
GGGAACCCCGAGTGCCGCAGTGTA	24	0.00	-8312.29	0.00	- inf	0.01	0.02	- inf	0.00	0.00
GGGCTATAGCTCAGCTGGGAA	21	0.00	5348.04	0.00	inf	0.00	0.00	inf	0.00	0.00
GGGCTATAGCTCAGCTGGGAGAA	23	0.00	6135.29	0.00	inf	0.00	0.00	inf	0.00	0.00
GGGCTATAGCTCAGCTGGGAGAGCA	25	0.06	3420.85	0.14	inf	0.12	0.22	inf	0.05	0.09
GGGGATTAGCTCAGCA	16	0.00	4960.65	0.00	inf	0.00	0.01	inf	0.00	0.00
GGGGCGTAGCGCAGCCTGGTAGCA	24	0.40	1420.85	0.58	inf	0.37	0.44	inf	0.32	0.37
GGGGTGTAGCTTAGCCA	17	0.06	3053.57	0.13	inf	0.12	0.22	inf	0.05	0.09

GGGGTGTAGCTTAGCTTA	18	0.37	1833.92	0.57	inf	0.37	0.44	inf	0.32	0.37
GGTCATCGCCAGGCTA	16	0.00	6082.72	0.00	inf	0.00	0.01	inf	0.00	0.00
GGTCTGTAGCTCAGGA	16	0.00	5293.82	0.00	inf	0.00	0.01	inf	0.00	0.00
GGTCTGTAGCTCAGGTGGTTAGAGCA	26	0.06	3442.53	0.14	inf	0.13	0.23	inf	0.05	0.10
GTACGAATACAAACTGTGA	19	0.55	-1310.62	0.66	- inf	0.37	0.44	- inf	0.32	0.37
GTCTGCCTCAGCGTCGGCA	19	0.59	-1.19	0.68	-1.06	0.81	0.85	-1.06	0.79	0.84
GTCTGCCTCAGCGTCGGGA	19	0.00	-5146.50	0.00	- inf	0.16	0.28	- inf	0.08	0.15
GTCTGCCTCAGCGTCGGTA	19	0.00	-1.63	0.00	-1.00	ï¿½	ï¿½	-1.00	1.00	1.00
GTCTGCCTCAGCGTCGGTAA	20	0.00	10445.33	0.00	inf	0.00	0.00	inf	0.00	0.00
GTGGGCGCTTAATACTGGGTGCA	23	0.00	4655.41	0.00	inf	0.00	0.00	inf	0.00	0.00
GTGTTCGGAAGATTTGTA	18	0.00	9.95	0.01	10.43	0.01	0.02	10.43	0.00	0.00
GTTCGGAAGATTTGTA	16	0.00	4.61	0.00	4.99	0.00	0.00	4.99	0.00	0.00
TAAAGACCAATTGACA	16	0.00	-5.65	0.00	-5.46	0.00	0.00	-5.46	0.00	0.00
TAAAGACCAATTGACTA	17	0.00	-21314.34	0.00	- inf	0.00	0.00	- inf	0.00	0.00
TAATTGCTCGTGCGACTTGACCA	23	0.43	1427.36	0.59	inf	0.37	0.44	inf	0.32	0.37
TACCGCCGGAATTGCATTCA	20	0.10	3202.51	0.23	inf	0.12	0.22	inf	0.05	0.09
TACCTTCGACCTGAGCA	17	0.54	-1224.79	0.66	- inf	0.37	0.44	- inf	0.32	0.37
TACCTTCGACCTGAGCTCAGGCA	23	0.87	-1.07	0.94	1.05	0.65	0.71	1.05	0.63	0.68
TACCTTCGACCTGAGCTCAGGCAA	24	0.69	-1.19	0.79	-1.00	0.99	0.99	-1.00	0.99	1.00
TACTAAAGACCAATTGACA	19	0.00	-7665.80	0.00	- inf	0.00	0.00	- inf	0.00	0.00
TAGGTCATCGCCAGGCTA	18	0.00	6672.13	0.00	inf	0.00	0.00	inf	0.00	0.00
TATCGGATTGGTCAGGCTTCGAAACA	26	0.00	7339.63	0.00	inf	0.00	0.00	inf	0.00	0.00
TCAGCCCGTGTTCGGAAGATTTGTA	25	0.03	-1.66	0.08	-1.30	0.26	0.41	-1.30	0.18	0.30
TCGACCTGAGCTCAGA	16	0.35	-2.45	0.54	-2.44	0.08	0.18	-2.44	0.02	0.05
TCGACCTGAGCTCAGGA	17	0.11	-1.94	0.23	-2.00	0.02	0.04	-2.00	0.00	0.00
TCGACCTGAGCTCAGGCA	18	0.00	-2.76	0.00	-2.50	0.01	0.03	-2.50	0.00	0.00
TCGACCTGAGCTCAGGCAA	19	0.00	-3.00	0.00	-2.61	0.00	0.01	-2.61	0.00	0.00
TCGACCTGAGCTCAGGCAAGA	21	0.00	-9404.56	0.00	- inf	0.00	0.01	- inf	0.00	0.00
TCGACCTGAGCTCAGGCAAGAA	22	0.00	-3905.07	0.01	- inf	0.00	0.00	- inf	0.00	0.00
TCGACCTGAGCTCAGGCAAGATTA	24	0.17	-3.41	0.33	-3.77	0.01	0.03	-3.77	0.00	0.00
TCGACCTGAGCTCAGGTA	18	0.37	-2047.94	0.56	- inf	0.37	0.44	- inf	0.32	0.37
TCGCGGTGAGGTTGGCACTA	20	0.12	-2762.55	0.25	- inf	0.12	0.22	- inf	0.05	0.09
TCTGCCTCAGCGTCGGTA	18	0.17	1.38	0.32	1.73	0.02	0.05	1.73	0.00	0.00
TGCAGATCTTGGTGGTAGTAGCA	23	0.55	-1377.37	0.66	- inf	0.37	0.44	- inf	0.32	0.37
TGCCTCAGCGTCGGCA	16	0.41	1387.52	0.59	inf	0.37	0.44	inf	0.32	0.37
TGCCTCAGCGTCGGTA	16	0.09	1.48	0.21	1.85	0.00	0.01	1.85	0.00	0.00
TGCGGAAGGATCATTGA	17	0.00	14380.42	0.00	inf	0.00	0.00	inf	0.00	0.00
TGCTGCAGTCATACCACCA	19	0.00	-10158.29	0.00	- inf	0.07	0.16	- inf	0.01	0.03
TGCTGCAGTCATACCACCACGA	22	0.26	-1.65	0.44	-1.58	0.39	0.46	-1.58	0.34	0.39
TGGTCAGGCTTCGAAACA	18	0.21	1782.31	0.39	inf	0.37	0.44	inf	0.32	0.37

TGTCTGCCTCAGCGTCGGAA	20	0.74	1.42	0.84	1.46	0.44	0.51	1.46	0.39	0.45
TGTCTGCCTCAGCGTCGGCA	20	0.18	-1.37	0.33	-1.26	0.46	0.52	-1.26	0.42	0.47
TGTCTGCCTCAGCGTCGGGA	20	0.11	-2.01	0.23	-1.89	0.14	0.26	-1.89	0.07	0.12
TGTCTGCCTCAGCGTCGGTA	20	0.00	-1.83	0.00	-1.00	0.00	0.00	-1.00	1.00	1.00
TGTCTGCCTCAGCGTCGGTAA	21	0.00	3.35	0.00	3.62	0.00	0.01	3.62	0.00	0.00
TGTTCGGAAGATTTGTA	17	0.00	4.11	0.00	4.43	0.02	0.05	4.43	0.00	0.00
TTCAGCCCGTGTTCGGAAGATTTGTA	26	0.00	12874.19	0.00	inf	0.00	0.01	inf	0.00	0.00
TTCGACCTGAGCTCAA	16	0.98	-1.02	0.98	1.03	0.86	0.90	1.03	0.85	0.89
TTCGACCTGAGCTCAGA	17	0.95	1.02	0.97	1.06	0.81	0.86	1.06	0.80	0.84
TTCGACCTGAGCTCAGGA	18	0.83	1.05	0.91	1.14	0.21	0.34	1.14	0.13	0.22
TTCGACCTGAGCTCAGGCA	19	0.15	-1.40	0.29	-1.07	0.62	0.69	-1.07	0.60	0.66
TTCGACCTGAGCTCAGGCAA	20	0.12	-1.44	0.25	-1.07	0.63	0.69	-1.07	0.60	0.66
TTCGACCTGAGCTCAGGTA	19	0.42	-1703.77	0.59	- inf	0.37	0.44	- inf	0.32	0.37
TTTGCTCGCCGATCCA	16	0.49	-1395.47	0.64	- inf	0.37	0.44	- inf	0.32	0.37
TTTGTTTGGAGAGTTA	16	0.00	5404.17	0.00	inf	0.00	0.00	inf	0.00	0.00

Appendix I

Summary of miRNA-seq data analysis using OmiRas for *C. sorokiniana* and *C. zofingiensis*

***C. sorokiniana*:**

Sample	Nsa	NSb	NSc	Ssa	SSb	SSc
Mature miRNA	7	5	4	13	11	12
Pre-miRNA	32	20	11	43	42	38
rRNA	12	11	4	18	14	14
snoRNA	61	43	24	93	75	73
tRNA	243	182	70	340	341	280

***C. zofingiensis*:**

Sample	Nza	NZb	NZc	SZa	SZb	SZc
Mature miRNA	10	9	12	17	12	12
Pre-miRNA	37	38	35	61	35	38
rRNA	22	18	17	23	13	17
snoRNA	87	94	104	116	86	82
tRNA	213	354	410	409	416	384

Appendix J

Details of annotation result of small RNA sequencing from *C. sorokiniana* and *C. Zofingiensis* to miRBase using CLC-Genomic workbench

C. sorokiniana	Small RNA	Expression values	Length	Count	Name	Resource
Nsa	CCTGTTACTTTGAGTAA	1	17	1	MIR169m	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAACTG	1	21	1	MIR5645c//MIR5645d	*Arabidopsis thaliana*
	AAGATCAAAAGACAAGAA	1	18	1	MIR169g	*Arabidopsis thaliana*
	GCAAAAGCTCTTAACCA	1	17	1	MIR5638b	*Arabidopsis thaliana*
	AGTAAAAAAGGATTTGGA	1	18	1	MIR159a	*Arabidopsis thaliana*
	TTGAAGAACATGAAGAA	1	17	1	MIR5017	*Arabidopsis thaliana*
	TGTTACTTTGAGTAAAA	7	17	7	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAA	2	17	2	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAAA	3	18	3	MIR169m	*Arabidopsis thaliana*
NSb	CTCCTGCTGCTCCGGCA	1	17	1	MIR909	*Chlamydomonas reinhardtii*
	TTTTTTTTTTTTTTTTA	1	18	1	MIR169a	*Arabidopsis thaliana*
	GCAAAAGCTCTTAACCA	1	17	1	MIR5638b	*Arabidopsis thaliana*
	TGTTACTTTGAGTAAAA	2	17	2	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAAA	2	18	2	MIR169m	*Arabidopsis thaliana*
NSc	CTGTTACTTTGAGTAAAA	2	18	2	MIR169m	*Arabidopsis thaliana*
Ssa	TGTTACTTTGAGTAAAA	1	17	1	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAA	1	17	1	MIR169m	*Arabidopsis thaliana*
	AAAAAAAAGAAAAAAA	1	16	1	MIR5645d	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAAA	1	18	1	MIR169m	*Arabidopsis thaliana*
	CAAAAAAAAAAAAAAAAAAAA	1	21	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAA	1	18	1	MIR5645d	*Arabidopsis thaliana*
	TACAAAAAAAAAAAAAAAAAAA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	GCAAAAGCTCTTAACCA	1	17	1	MIR5638b	*Arabidopsis thaliana*
	TTTAATGAATCTTCTACA	1	18	1	MIR844	*Arabidopsis thaliana*
	GTCTTTTTTAAAAAAAAAGAA	1	21	1	MIR419//MIR5645d	*Arabidopsis thaliana*
	CGGTTGCGGCAGCGGCA	1	17	1	MIR404	*Arabidopsis thaliana*
	TTTTAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
SSb	CAAGAACAAAAAAAAAAAAAAAAA	1	24	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAA	1	18	1	MIR5645d	*Arabidopsis thaliana*
	GAGAAGTGAAAGATGAA	1	17	1	MIR169m	*Arabidopsis thaliana*

	AAGATCAAAAGACAAGAA	1	18	1	MIR169g	*Arabidopsis thaliana*
	TGTTACTTTGAGTAAAA	1	17	1	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAA	1	17	1	MIR169m	*Arabidopsis thaliana*
	TTCGGACTCTTCTTGCGA	1	18	1	MIR831	*Arabidopsis thaliana*
	GCCTGTTACTTTGAGTA	1	17	1	MIR169m	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAA	1	19	1	MIR5645d	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAAA	1	18	1	MIR169m	*Arabidopsis thaliana*
SSc	TCAGATTGGTATTGCAA	1	17	1	MIR408	*Arabidopsis thaliana*
	AAAAATTGAACTCATTA	1	17	1	MIR5653	*Arabidopsis thaliana*
	AAGATCAAAAGACAAGAA	1	18	1	MIR169g	*Arabidopsis thaliana*
	TGATGGACTCAAATTAA	1	17	1	MIR829	*Arabidopsis thaliana*
	AGGAAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	CAGCTGCAGCTTCAGGCA	1	18	1	MIR1156	*Chlamydomonas reinhardtii*
	AAAAAAAAAAAAAAAAAAAA	1	20	1	MIR5645d	*Arabidopsis thaliana*
	AGTCTCTAAATCTGTAA	2	17	2	MIR824	*Arabidopsis thaliana*
	TGTTACTTTGAGTAAAA	6	17	6	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAA	3	17	3	MIR169m	*Arabidopsis thaliana*
	CTGTTACTTTGAGTAAAA	5	18	5	MIR169m	*Arabidopsis thaliana*

C. zofingiensis	**Small RNA**	**Expression values**	**Length**	**Count**	**Name**	**Resource**
Nza	GATGATGATGATGACGA	1	17	1	MIR5658	*Arabidopsis thaliana*
	TGAAAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	GATGATGATGATGATGA	1	17	1	MIR5658	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAA	1	17	1	MIR5645d	*Arabidopsis thaliana*
	GTGGAGGAGGAGGAGA	1	16	1	MIR5655	*Arabidopsis thaliana*
	TTTGAAAAAAAAAAAAAAAAAAAAA	1	26	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAAAAA	1	26	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAATTA	1	21	1	MIR5645d	*Arabidopsis thaliana*
NZb	AAGGAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	TTCGAAAAAAAAAAGA	1	17	1	MIR5645c	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAA	1	17	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAATTA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	AGAGAAGAAAAAAAAAAAAAAAA	1	24	1	MIR5645d	*Arabidopsis thaliana*
NZc	TGATGATGATGATGATGACGA	1	21	1	MIR5658	*Arabidopsis thaliana*
	CAAAAAAAAAAAAAAAAAAA	1	20	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAACAA	1	17	1	MIR5645d	*Arabidopsis thaliana*
	TGATGATGATGATGATGA	1	18	1	MIR5658	*Arabidopsis thaliana*
	CAAAAAAAAAAAAAAAA	1	18	1	MIR5645d	*Arabidopsis thaliana*

	GAAAAAAAAAAAAGAAAAAAA	1	21	1	MIR5645d	*Arabidopsis thaliana*
	ACAAAAAAAAAAAAAAAAAA	1	20	1	MIR5645d	*Arabidopsis thaliana*
	AACTGGCCCCGTAACCA	1	17	1	MIR908	*Chlamydomonas reinhardtii*
	AGGAGGAGGAGGAAGAA	1	17	1	MIR5658	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAA	1	19	1	MIR5645d	*Arabidopsis thaliana*
	AAAAACAAAAAAAAAAAAA	1	20	1	MIR5632//MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAAAAAA	1	26	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAA	1	17	1	MIR5645d	*Arabidopsis thaliana*
	GAAAAAAAAAAAAAAAAAAATA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	CAAGAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAA	2	23	2	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAAA	2	24	2	MIR5645d	*Arabidopsis thaliana*
Sza	CAAAAAAAAAAAAAAAAAA	1	20	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAA	1	19	1	MIR5645d	*Arabidopsis thaliana*
	TGATGATGATGATGATGAA	1	19	1	MIR5658	*Arabidopsis thaliana*
	GTTCTTCTTCTTCTGAA	1	17	1	MIR8183	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAA	1	18	1	MIR5645d	*Arabidopsis thaliana*
	ATATAAAACCAAACAGA	1	17	1	MIR169a	*Arabidopsis thaliana*
	CGACCTTGGTCTCGATCTA	1	19	1	MIR858a	*Arabidopsis thaliana*
	GCATGGAATAATAGAAA	1	17	1	MIR169e	*Arabidopsis thaliana*
	TGACAGACGTGATGACAA	1	18	1	MIR8170	*Arabidopsis thaliana*
	GGTTGCGAGAGAAGATCACTA	1	21	1	MIR835	*Arabidopsis thaliana*
	TGATGATGATGATGATGA	1	18	1	MIR5658	*Arabidopsis thaliana*
	AAACATGAAGGTAGGGA	1	17	1	MIR417	*Arabidopsis thaliana*
	ACGTGCTCGATCTGCTCA	1	18	1	MIR3932b	*Arabidopsis thaliana*
	TGACGATGATGATGATGATGA	1	21	1	MIR5658	*Arabidopsis thaliana*
	TAAAAAAAGGAAGAAAGA	1	18	1	MIR157d	*Arabidopsis thaliana*
	ATGATGATGATGATGATGA	1	19	1	MIR5658	*Arabidopsis thaliana*
	GATGATGATGATGATGATGA	1	20	1	MIR5658	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAAAAA	1	25	1	MIR5645d	*Arabidopsis thaliana*
	TGCCGCTGCTGCTGACCGCCA	1	21	1	MIR1158	*Chlamydomonas reinhardtii*
	TTAAAAAAAAAAAAAA	1	17	1	MIR5645d	*Arabidopsis thaliana*
	GTGATGATGATGATGAA	1	17	1	MIR5658	*Arabidopsis thaliana*
	GAAAAAAAAAAAAAAAAA	1	18	1	MIR5645d	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	GTCGTGTGTGCCGCTCA	1	17	1	MIR1151b	*Chlamydomonas reinhardtii*
	GATCTAGCCTTGAAATA	1	17	1	MIR5640	*Arabidopsis thaliana*
	AAGATCAGTCGCACCAA	1	17	1	MIR841b	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	AGCCGGACGCGCCGACA	1	17	1	MIR1166	*Chlamydomonas reinhardtii*

	CAAGGAGTCTAGCATGA	1	17	1	MIR160c	*Arabidopsis thaliana*
	AAAAAAAGAAAAAAAAAAAAAAAAAA	1	26	1	MIR5645d	*Arabidopsis thaliana*
	CGAGTACGACAAGATCA	1	17	1	MIR169g	*Arabidopsis thaliana*
	GCTGCTGGTGCTGGTGGA	1	18	1	MIR1144a//MIR1144b	*Chlamydomonas reinhardtii*
	CGGAAAAAAAAAAGAAA	1	17	1	MIR5645c	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAA	1	21	1	MIR5645d	*Arabidopsis thaliana*
	TTTGTTAGAACAAAACAA	1	18	1	MIR5645c//MIR5645d//MIR5645f	*Arabidopsis thaliana*
	TCAGGTGGTTAGAGCAA	5	17	5	MIR5660	*Arabidopsis thaliana*
SZb	AAAAAAAAAAAAAAAAAAA	1	19	1	MIR5645d	*Arabidopsis thaliana*
	TCAGGTGGTTAGAGCAA	1	17	1	MIR5660	*Arabidopsis thaliana*
	GGAAAAAAAAAAAAAAAAAA	1	21	1	MIR5645d	*Arabidopsis thaliana*
	GGATGATGATGATGGTGATGA	1	21	1	MIR5658	*Arabidopsis thaliana*
SZc	CTGAAAAAAAAAAAAAAAAAAA	1	23	1	MIR5645d	*Arabidopsis thaliana*
	ATGAAAAAAAAAAAAAAAA	1	19	1	MIR5645d	*Arabidopsis thaliana*
	TCTGACTTCTTAGAGGCA	1	18	1	MIR5998a	*Arabidopsis thaliana*
	AAAAAGAAAGAAAAAAAAAATA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	GATGATGATGATGATGACGA	1	20	1	MIR5658	*Arabidopsis thaliana*
	AAAAAAAAAAAAAAAAAAAAAA	1	22	1	MIR5645d	*Arabidopsis thaliana*
	TCAGGTGGTTAGAGCAA	2	17	2	MIR5660	*Arabidopsis thaliana*

BIODATA OF STUDENT

Siti Nor Ani Azaman was born on 31 August 1985 in Banting, Selangor. She received her early primary education in Sekolah Kebangsaan Sg Kelambu, Banting, Selangor (1992-1997) and continued her secondary education at Sekolah Agama Menengah Unwanus Saadah, Kanchong Darat, Banting, Selangor until 2000. In 2001, she was offered to continue her study at Mara Junior Science College Kuala Berang (MRSM Kuala Berang), Terengganu until 2002. She enrolled foundation studies at College of Matriculation in Kuala Pilah, Negeri Sembilan before being offered to pursue her studies at Universiti Putra Malaysia (UPM). In UPM, she took Biotechnology courses in Faculty of Biotechnology and Biomolecular Science, UPM and graduated on 2007.

Following her graduation, she worked as a research assistant in Laboratory Immunotherapeutic and Vaccines (LIVES) under Prof Dr Arbakariya Ariff and continued her Master's degree in the same lab with thesis title "Induction Strategies for Interferon-α2b Production in the Periplasmic Space of *Escherichia Coli*". During her candidature, she had attended several workshops, seminars, symposium and conferences. In 2009, she was awarded as the best poster presenter in 2nd National Symposium on Fermentation Technology. Besides, she has published several journal papers from the research conducted during her Master's studies from her own works and collaboration with other group members.

PUBLICATIONS

Azaman, S.N.A., Nagao, N., Yusoff, F.M., Tan, S.W., & Yeap, S.K. (2017). A comparison of the morphological and biochemical characteristics of *Chlorella sorokiniana* and *Chlorella zofingiensis* cultured under photoautotrophic and mixotrophic conditions. PeerJ 5, e3473.

Azaman, S.N.A., Satharasinghe, D.A., Nancy, L.W.C., Boo, L., Nagao, N., Yusoff, F.M., Tan, S.W., & Yeap, S.K. (2017). Small RNA transcriptomic profiling of *C. sorokinina* and *C. zofingiensis* under normal and stress conditions. Submitted to GigaScience.

Azaman, S.N.A., Yusoff, F.M., Tan, S.W., Nagao, N., & Yeap, S.K. (2018). Advance in cultivation, application and omic studies on Chlorophyta microalgae. Submitted to PeerJ.