COPYRIGHT AND USE OF THIS THESIS

This thesis must be used in accordance with the provisions of the Copyright Act 1968.

Reproduction of material protected by copyright may be an infringement of copyright and copyright owners may be entitled to take legal action against persons who infringe their copyright.

Section 51 (2) of the Copyright Act permits an authorized officer of a university library or archives to provide a copy (by communication or otherwise) of an unpublished thesis kept in the library or archives, to a person who satisfies the authorized officer that he or she requires the reproduction for the purposes of research or study.

The Copyright Act grants the creator of a work a number of moral rights, specifically the right of attribution, the right against false attribution and the right of integrity.

You may infringe the author’s moral rights if you:

- fail to acknowledge the author of this thesis if you quote sections from the work
- attribute this thesis to another author
- subject this thesis to derogatory treatment which may prejudice the author’s reputation

For further information contact the University’s Copyright Service.

sydney.edu.au/copyright
Phylogenetic analysis of photosynthesis related proteins in
Chromera velia and study of its photosynthetic response to iron
limitation

By
Hao Pan

A thesis submitted in fulfillment of the requirements for the Degree of

Doctor of Philosophy

in

Biological Sciences

Faculty of Science
University of Sydney
2015
All work presented in this thesis is from my own investigations, unless stated otherwise in the text.

Hao Pan
Acknowledgements

This work would not have been possible without the assistance of a number of people at the University of Sydney. I would like to express my appreciation to all whose who helped me during my doctoral studies. First of all, I would like to thank my supervisor, Professor Min Chen for giving me this opportunity of working on a PhD project. She has been patient, understanding and most supportive during my candidature. I have learned many things from her, not only the knowledge presented in this thesis, but also the way of academically thinking.

I would like to thank my associate supervisor, Associate Professor Dee Carter (School of Molecular Biosciences, University of Sydney) for her guidance, suggestions and discussions during my study. I would also thank Associate Professor Jan Šlapeta (School of Veterinary Sciences, University of Sydney) for his assistance on the phylogenetic analysis. I am thankful to Professor Robert Willows, for his assistance on the molecular experiments. I would also express my thanks to Dr. Patrick Loughlin (Chen’s lab, University of Sydney), for his guidance on the molecular experiments and advices with my academic writing. I would like to thank Australia Society of Plant Scientists (ASPS) and the postgraduate research support scheme (PRSS), for the travel support for attending ComBio conference (2012).

I would also express my gratitude to all members of Chen’s lab, including: Dr. Raymond Ritchie, Dr. Tendo Thomas Mukasa Mugerwa, Dr Partrick Loughlin, Dr. Rau Quirio, Dr. Moacir Torres, Mr. Zane Duxbury, Ms. Kathy Donohoe, Ms. Yaqiong Li, Mr. Yuankui Lin. It is an honor to work with you during my time in the lab. I would like to thank friends who I have met or worked with during my study in University of Sydney, Ms. Kate Weatherby, Ms. Christie Forster, and Ms.
Sonia Brazao. I thank them for the support and discussions during my research, and also many fun moments we have made together.

Last but not least, I would like to thank my parents, for their love, patience, generous encouragement and understanding during my study far away from home. This work would be impossible without their support.
Abstract

*Chromera velia* (*C. velia*) is a newly discovered algal species in Australia (Moore *et al.*, 2008). It possesses photosynthetic characteristics similar to photosynthetic dinoflagellates, but has physiological and molecular features of non-photosynthetic apicomplexan parasites. Hence, it has been proposed that *C. velia* may be the missing link between photo-autotrophic dinoflagellates and heterotrophic apicomplexans. This project aimed to: (1.) analyse the light harvesting complexes (LHC) and enzymes involved in the Calvin-Benson cycle in *C. velia* using a phylogenetic method to obtain a better understanding of the evolutionary development of light and dark reactions in photosynthesis; and (2.) characterize the photosynthetic apparatus in *C. velia* under normal and iron-stress conditions using a set of biochemical analysis methods as well as bioinformatics. LHC comprise proteins that absorb light energy and transfer it to photosynthetic reaction centres. Sequencing of the LHC in *C. velia* identified three typical membrane spanning regions (MSR). Phylogenetic analysis indicates that one group is closely related to diatoms, another to red algae, and a third one containing a LHC peptide closely related to the LI818/LHCSR group in charge of photosynthesis regulation. This is the first time LHC in *C. velia* have been analyzed using phylogenetic methods. Our results clearly support the hypothesis that there is a connection between *C. velia* and diatoms (Moore *et al.*, 2008). This relationship was also confirmed by protein sequencing of isolated LHC from *C. velia*.

The Calvin-Benson cycle is an important part of the dark reaction of photosynthesis. It fixes CO₂ and converts inorganic carbon into organic chemicals using adenosine triphosphate synthesized from light reactions. Phylogenetic analysis of the enzymes involved in the Calvin-Benson cycle in *C. velia* revealed a complicated evolutionary pathway. The majorirty of enzymes in *C. velia*
originate from proteobacteria, with two exceptions (Fructose-bisphosphate aldolase and Phosphoribulokinase). Different enzymes in *C.velia* share close relationships with green algae, red algae, diatoms, photosynthetic dinoflagellates, and Apicomplexa. We carried out our phylogenetic analysis of the enzymes involved in the Calvin-Benson cycle as a whole for the first time. Our results clearly showed that these enzymes have a mosaic pattern of evolutionary relationships with other groups, supporting the “shopping bag” theory proposed by Larkum *et al.* (2007).

Iron is an essential element for photosynthesis, vital for assembling photosystem I and other photosynthetic proteins. Given that the oxidized form of iron does not dissolve in water, iron is often limited in the marine environment. Therefore, oxygenic photosynthetic organisms have developed different strategies to cope with iron limitation during their evolutionary process. A study of iron-stress response in the primitive *C.velia* can help improve our understanding of its photosynthetic system. Our study revealed that iron-stress conditions led to decreased growth rate (with a doubling time of 7.54 days in the iron-stress culture, compared with 2.82 days in a normal culture), decreased oxygen evolution rate, decreased chlorophyll concentration per cell, shifted carotenoid composition, and shifted protein expression pattern.

As a potential evolutionary intermediate between photoautotrophs and heterotrophs, *C.velia* offers an excellent opportunity to explore the link the evolutionary development of photosynthesis in apicomplexan parasites. Thus, phylogenetic analysis and study of its photosynthetic apparatus contributes to a better understanding of its evolutionary development as well as its inherent molecular mechanism.
## Table of contents

Abbreviation list: .......................................................... 13

Publications that have arisen from the work presented in this thesis: ........................................ 15

Chapter 1. Introduction .................................................................................................................. 16

1.1. Tree of Eukaryotes ................................................................................................................. 16
    1.1.1. General introduction ............................................................................................... 16
    1.1.2. Photosynthetic Eukaryotic organisms ........................................................................ 18

1.2. Algae ...................................................................................................................................... 20
    1.2.1. Red algae ............................................................................................................. 20
    1.2.2. Green algae ........................................................................................................... 21
    1.2.3. Glaucophytes ......................................................................................................... 21
    1.2.4. Chromalveolates .................................................................................................... 22
    1.2.5. Diatoms .................................................................................................................. 22
    1.2.6. Dinoflagellates ........................................................................................................ 23
    1.2.7. Apicomplexa .......................................................................................................... 24

1.3. Endosymbiosis ......................................................................................................................... 26
    1.3.1. Primary endosymbiosis .......................................................................................... 26
    1.3.2. Secondary endosymbiosis ...................................................................................... 28

1.4. Oxygenic Photosynthesis ......................................................................................................... 30
    1.4.1. Light harvesting ..................................................................................................... 31
    1.4.2. Photosystem II ....................................................................................................... 32
    1.4.3. Photosystem I ......................................................................................................... 32
    1.4.4. Electron transport chain ........................................................................................ 33
    1.4.5. Carbon fixation ...................................................................................................... 34

1.5. Iron and Photosynthesis .......................................................................................................... 35
    1.5.1. The role of iron ....................................................................................................... 35
    1.5.2. Iron nutrient stresses in nature ................................................................................ 36
    1.5.3. The physiological changes under iron stressed condition in algae ................................ 36

1.6. Chromera velia ........................................................................................................................ 37
1.6.1. The discovery of *Chromera velia* ................................................................. 37
1.6.2. Current research status of *C. velia* ............................................................... 38
1.7. Aim ......................................................................................................................... 39
Chapter 2. Material and Methods ............................................................................. 40
2.1. Cell culture conditions ......................................................................................... 40
2.1.1. *C. velia* strain ............................................................................................... 40
2.1.2. Culture medium .............................................................................................. 40
2.1.3. Algal culture condition .................................................................................... 41
2.1.4. Counting cells and the density of culture ....................................................... 41
2.1.5. Cell harvesting ............................................................................................... 42
2.2. Pigment composition and oxygen evolution rate ............................................... 42
2.2.1. Pigment extraction and spectral analysis ....................................................... 42
2.2.2. *In vivo* absorbance reading ......................................................................... 43
2.2.3. High Pressure Liquid Chromatography (HPLC) ........................................... 44
2.2.4. Room temperature fluorescence and low temperature fluorescence spectra 45
2.2.5. Oxygen evolution rate measurement ............................................................. 46
2.3. Photosynthesis protein study of *C. velia* ......................................................... 47
2.3.1. Cell breaking .................................................................................................. 47
2.3.2. Sucrose density gradient ultracentrifugation .................................................. 47
2.3.3. Protein concentration ..................................................................................... 48
2.3.4. Sample treatment ........................................................................................... 49
2.3.5. Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE) 50
2.3.6. Western Blotting ............................................................................................ 50
2.4. Genomic study of *C. velia* ................................................................................ 52
2.4.1. DNA extraction from *C. velia* ..................................................................... 52
2.4.2. Polymerase chain reaction (PCR) ................................................................ 55
2.5. Analysis of genes in *C. velia* ............................................................................ 58
2.5.1. Retrieval of putative LHC sequences ............................................................ 58
2.5.2. LHC sequence alignment .............................................................................. 58
2.5.3. Phylogenetic tree construction of LHC sequences ....................................... 59
2.5.4. Retrieval of Calvin-Benson cycle-related sequences ........................................... 59
2.5.5. Alignment of Calvin-Benson cycle related sequences ......................................... 60
2.5.6. Phylogenetic tree construction of Calvin-Benson cycle related sequences .......... 60

Chapter 3. Phylogenetic study of Light harvesting complexes in *Chromera velia* .......... 62

3.1. Introduction .................................................................................................................. 62
3.1.1. Antenna protein complexes in oxygenic photosynthetic organism ................... 62
3.1.3. Core antenna complex system ............................................................................. 64
3.1.4. Accessory antennas ............................................................................................ 65
3.1.5. Aims .................................................................................................................... 67

3.2. Methods ..................................................................................................................... 68
3.2.1. Putative LHC sequences retrieving ................................................................. 68
3.2.2. LHC sequences alignment ................................................................................. 69
3.2.3. Phylogenetic tree construction of LHC sequences ............................................. 69

3.3. Results .................................................................................................................... 71
3.3.1. LHC sequences in *C. velia* ............................................................................. 71
3.3.2. Phylogenetic analysis of the LHC homologs from *C. velia* ......................... 73
3.3.3. Phylogenetic relationship of the *C. velia* homologs ....................................... 73

3.4. Discussion .............................................................................................................. 82
3.4.1. The relationship between *C. velia* and other algae groups indicated by phylogenetic analysis 82
3.4.2. The LHC motifs presented in *C. velia* .............................................................. 83
3.4.3. The evolutionary relationship of LHC in *C. velia* ............................................ 84

3.5. Conclusion ............................................................................................................... 86

Chapter 4. Phylogenetic analysis of Calvin-Benson cycle related enzymes in *C. velia* . 87

4.1. Introduction ............................................................................................................. 87
4.1.1. Evolutionary development of enzymes in Calvin-Benson cycle ..................... 88

4.2. Aim .......................................................................................................................... 89

4.3. Enzymes in the Calvin-Benson cycle .................................................................... 90
4.3.1. Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) ....................... 90
4.3.2. 3-phosphoglycerate kinase (PGK) ................................................................... 92
4.3.3. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)................................. 93
4.3.4. Triosephosphate isomerase (TPI).................................................................. 94
4.3.5. Fructose-bisphosphate aldolase (FBA) ............................................................ 94
4.3.6. Fructose-1,6-bisphosphatase (FBP) and sedoheptulose -1,7 -bisphosphatase (SBP) 96
4.3.7. Transketolase (TKL) .................................................................................. 96
4.3.8. Ribulose-5-phosphate 3-epimerase (RPE) ....................................................... 97
4.3.9. Ribose-5-Phosphate Isomerase (RPI) ............................................................. 98
4.3.10. Phosphoribulokinase (PRK) ......................................................................... 98

4.4. Method ............................................................................................................. 99

4.5. Results ............................................................................................................. 101
4.5.1. Phylogenetic analysis of genes involved in the Calvin-Benson cycle .......... 101
4.5.2. Phylogenetic analysis of 3-phosphoglycerate kinase .................................... 101
4.5.3. Phylogenetic analysis of cytosolic glyceraldehyde 3-phosphate dehydrogenase. 101
4.5.4. Phylogenetic analysis of triosephosphate isomerase ..................................... 103
4.5.5. Phylogenetic analysis of fructose-bisphosphate aldolase ......................... 105
4.5.6. Phylogenetic analysis of fructose-1,6-bisphosphatase .............................. 106
4.5.7. Phylogenetic analysis of transketolase ......................................................... 107
4.5.8. Phylogenetic analysis of ribulose-5-phosphate 3-epimerase ..................... 108
4.5.9. Phylogenetic analysis of ribose-5-phosphate isomerase ............................. 110
4.5.10. Phylogenetic analysis of phosphoribulokinase .......................................... 111

4.6. Discussion ....................................................................................................... 112
4.6.1. Phylogenetic analysis of ribulose-1,5-bisphosphate carboxylase oxygenase.... 112
4.6.2. Phylogenetic analysis of chloroplastic glyceraldehyde 3-phosphate dehydrogenase 113
4.6.3. The origination of the Calvin-Benson cycle related genes in C.velia .......... 115
4.6.4. Cyanobacteria derived genes (chloroplastic type) ...................................... 115
4.6.5. Proteobacteria derived genes in the Calvin-Benson Cycle ....................... 116
4.6.6. Comparison of the Calvin-Benson Cycle coding genes in C.velia with other organisms ........................................................................................................... 116
4.6.7. Comparison of C.velia with other organisms ........................................117
4.6.8. The evolution of the Calvin-Benson cycle ......................................118
4.6.9. Enzyme co-localization in the Calvin-Benson cycle .....................119
4.7. Conclusions .......................................................................................120

Chapter 5. Iron-stress and photosynthetic apparatus in C.velia ................121
5.1. Introduction ......................................................................................121
5.1.1. Iron limitation for biological system .............................................121
5.1.2. Importance of iron .......................................................................121
5.1.3. Iron-stress and photosynthesis ....................................................122
5.1.4. Current study of iron-stress in C.velia .........................................124
5.1.5. Aim ..............................................................................................124
5.2. Material and methods ......................................................................125
5.2.1. Culture conditions ........................................................................125
5.2.2. Growth rate monitoring and oxygen evolution rate measurement ...126
5.2.3. Isolation of pigment-binding protein complexes .........................127
5.3. Results .............................................................................................127
5.3.1. Cell growth rate ...........................................................................127
5.3.2. Oxygen evolution rate measurement ...........................................129
5.3.3. Pigment composition of C.velia cells from normal culture conditions131
5.3.4. Pigment composition of C.velia cells from iron-stress culture conditions132
5.3.5. Photosynthetic pigment-binding protein complexes ....................135
5.3.6. Absorption spectra .......................................................................136
5.3.7. Fluorescence reading for the isolated fractions ..............................137
5.3.8. HPLC for the isolated pigment-binding protein complexes ...........137
5.3.9. Polypeptide composition .............................................................141
5.3.10. Western blotting ..........................................................................142
5.4. Discussion .......................................................................................143

Chapter 6. General Discussion ................................................................146
6.1. Light harvesting protein complexes ................................................147
6.2. Enzymes involved in the Calvin-Benson cycle .................................148

11
Abbreviation list:

ATP: Adenosine triphosphate
BSA: bovine serum albumin
Chl: chlorophyll
CoA: Coenzyme A
*Ch. velia*: *Chromera velia*
DNA: Deoxyribonucleic acid
DoM: n-Dodecyl β-D-maltoside
EDTA: Ethylenediaminetetraacetic acid
EST: expressed sequence tag
FBA: Fructose-bisphosphate aldolase
FBP: Fructose-1,6-bisphosphatase
FCP: fucoxanthin chlorophyll binding protein
GAPDH: Glyceraldehyde 3-phosphate dehydrogenase
HLIP: High light induced protein
HPLC: high pressure liquid chromatography
isiA: iron-stress induced protein A
JGI: Joint Genome Institute
LHC: light harvesting complexes
NADPH: Nicotinamide adenine dinucleotide phosphate
NCBI: National Center for Biotechnology Information
NJ: neighbor-joining
NPQ: non-photochemical quenching
PCP: peridinin chlorophyll a/c-binding proteins
PCR: Polymerase chain reaction
PGK: 3-phosphoglycerate kinase
PRK: Phosphoribulokinase
PSI: photosystem I
PSII: photosystem II
PVDF: Polyvinylidene fluoride
RNA: Ribonucleic acid
RPE: Ribulose-5-phosphate 3-epimerase
RPI: Ribose-5-Phosphate Isomerase
RuBisCO: Ribulose-1,5-bisphosphate carboxylase oxygenase
SBP: sedoheptulose -1,7 -bisphosphatase
SRA: Sequence Read Archive
TBS: Tris-Buffered Saline
TCA: trichloroacetic acid
TKL: Transketolase
TPI: Triosephosphate isomerase
Publications that have arisen from the work presented in this thesis:


Chapter 1. Introduction

1.1. Tree of Eukaryotes

1.1.1. General introduction

Prokaryotes and eukaryotes are the two main living organisms on earth. Prokaryotes are organisms lacking membrane-bounded organelles; while eukaryotes contain cellular organelles for performing and organising their advanced cellular functions. Cellular organelles have specific functions, e.g., endoplasmic reticulum hosts the protein-folding processes, Golgi bodies are involved in protein biogenesis, lysosomes are in charge of digesting biomolecules, mitochondria is the site for respiration and energy supply, while the chloroplast is the site for photosynthesis. Some organelles are absent in certain species. For example, chloroplast does not exist in animals, while lysosome does not exist in plant or algae.

Prokaryotes consist of two domains of life: Archaea and Bacteria, while eukaryotes have only one domain called Eukaryota (Whitman et al., 1998). However, eukaryotes cover a great range of life forms, from algae and plants to animals. The total biomass of eukaryotes is equal to prokaryotes because of their larger body size (Whitman et al., 1998).

In general, species are classified based on their physical features and morphology. This classification method is reliable, if the organisms have visible differences in their appearances, or have some diagnostic features. However, microorganisms, like unicellular algae, cannot be classified easily using physical features or morphological properties. As an outcome of improved DNA sampling, sequencing and the sequence phylogenetic analysis, the classification for algae has been improved dramatically. Yet, there are still many unsolved questions, regarding the classifications of specific species, and the evolutionary relationships among different groups of
species. Eukaryotes can be divided into six supergroups: Opisthokonta, Amoebozoa, Rhizaria, Excavata, Chromalveolata, and Archaeplastida (Yoon et al. 2008; Reeb et al. 2009). Based on this hypothesis, the majority of algae are classified into two groups: Chromalveolata and Archaeplastida; while animals and fungi are mainly classified into the Opisthokonta.

Genetic sequence information and phylogenetic analysis have become powerful approaches for identifying and classifying the taxon position of unknown species. Phylogenetic analysis has improved the classification for unicellular eukaryotes dramatically. However, the outcome of phylogenetic analysis is affected greatly by the choice of gene, taxon sampling size, and phylogenetic analysis methods. The framework of the six super-groups still needs to be refined.
1.1.2. Photosynthetic Eukaryotic organisms

1.1.2.1. Plants

Plants, all red algae (Rhodophytes), Glaucophytes and green algae, all belong to Archaeplastida, also known as Plantae. Plants have inherited a remarkable strategy for harvesting and converting light energy into chemical energy from their algal ancestors. However, they have evolved to become advanced form as multi-cellular organism, with specific organs and tissues. The photosynthesis system in plants is believed to have originated from green algae.

1.1.2.2. Algae

Algae are a group of oxygenic photosynthetic organisms, with different morphologic and physiologic features developed during the long history of evolution. Algae span the phylum Chromalveolata and Archaeplastida. It is commonly recognized that prokaryotic cyanobacteria are
the ancestors of modern eukaryotic algae, since they have had photosynthesis ability for more than 2.9 billion years (Olson, 2006). Algae play an important role in deciphering the evolution of organelles and photosynthesis. In addition, there are many non-photosynthetic groups, which have closely relationship with various algal groups. Most of them may have lost their plastids and photoautotrophic life style during evolution; such organisms are not classified as algae, but algae relatives.

1.1.2.3. Plastid

The plastid is a major organelle in the cells of plants and algae. Plastids are the sites for manufacture and storage of important chemical compounds, e.g., starch, adenosine triphosphate (ATP), used by the cell. Plastids possess a double-stranded circular DNA molecule. One of the most common plastids is chloroplast, which contains the photosynthetic apparatus. All chloroplasts have at least three membrane systems—the outer membrane, the inner membrane, and the thylakoid system. The outer and inner membranes together build the chloroplast envelope. Chloroplasts that are the product of secondary endosymbiosis may have additional membranes surrounding these three membranes. Inside the outer and inner chloroplast membranes is the chloroplast stroma, in which the thylakoid system floats. In algae, photosynthetic plastids contain different types of pigments, which can change or determine the cell's colour. In different algae lineages, plastids have been given different specific names, i.e. chloroplasts (in the green lineage), rhodoplasts (in the red lineage), leucoplasts (colourless plastids), apicoplast (special plastid in Apicomplexa).
There also are other types of plastids, with specific functions other than photosynthesis. In plants, three types of plastids exist: Leucoplasts are colourless plastids that function in monoterpen synthesis; Gerontoplasts control the dismantling of the photosynthetic apparatus, and chromoplasts are in charge of the pigment synthesis and storage. Leucoplasts in algae have different functions and contain pyrenoids.

Apicoplasts in Apicomplexa are a special type of plastid that has lost its photosynthesis ability, but it is still essential for the biosynthesis of amino acids, fatty acids and isoprenoids (Harwood, 1996; Herrmann and Weaver, 1999; Rohdich et al., 2001). Similar to apicoplasts, plastids function have been radically reduced or transformed in many lineages, mostly with the loss of photosynthesis (e.g., the relict plastids of many parasitic algae and plants; Gould et al., 2008; Ralph et al., 2004; Wilson, 2002).

1.2. Algae

1.2.1. Red algae

The red algae, or Rhodophyta, are one of the oldest groups of eukaryotic algae. Plastids in red algae have a two-membranes envelop structure. Red algae use phycobilisome as their main antenna system for photosystem II (PSII), giving the “red” colour of red algae. Phycobilisomes are built with phycocyanin and phycoerythrin, located on the surface of unstacked thylakoids membranes (Woelkerling, 1991). For photosystem I (PSI), red algae use chlorophyll (Chl) a binding-light harvesting complex (LHC) as their antenna system.

Red algae lack both flagella and centrioles during their entire life cycle (Gabrielson et al., 1990; Graham and Wilcox, 2000). The plastids in red algae lack external endoplasmic reticulum and
contain unstacked thylakoids membranes. Although red algae are eukaryotic photosynthetic organisms, they share many similarities in photosynthetic apparatus with cyanobacteria (the oxygenic photosynthetic prokaryotes). Given their close connection to cyanobacteria, red algae may be the direct descendants of primary endosymbiosis (see section 1.3)

1.2.2. Green algae
Green algae are a large group of algae, from which higher plants emerged. Green algae include unicellular and colonial flagellates, most with two flagella per cell. Green algae have similar photosynthetic systems to high plants. The plastids in green algae contain Chl \(a\) and Chl \(b\), as well as two layers of chloroplast membranes. Unlike red algae, the thylakoid membranes in green algae are stacked in a similar arrangement to those in high plants (Hoek et al., 1995). Green algae use Chl-binding light harvesting complexes to capture light energy, similar to plants. With their single envelope membrane structure to chloroplast, green algae are also considered direct descents of primary endosymbiosis.

1.2.3. Glaucophytes
Glaucophytes are a small group of freshwater microscopic algae. Like red algae, glaucophytes use phycobilisomes to harvest light energy, and contain only Chl \(a\). Like green algae, they possess two flagella. Glaucophytes have a unique chloroplast, with a peptidoglycan layer. This is believed to be a relic of the endosymbiotic origin of plastids from cyanobacteria (Keeling, 2004). Thus, Glaucophytes also are considered direct descents of primary endosymbiosis, along with red algae and green algae.
1.2.4. Chromalveolates

Chromalveolates represent a diverse eukaryotic group with complicated plastids. The plastids in chromalveolates have three or four layers of envelope membranes, which is thought to be the result of secondary (possible tertiary) endosymbiosis (See Figure 1.2; section 1.4). Chromalveolates were first recognized as a group by Cavalier-Smith (Cavalier-Smith, 2002). It consists of seven subgroups: Dinozoa, Ciliophora, Stramenophile/ Heterokonta, Chomerida, Apicomplexa, Hacrobia and Cryptophyceae. Members of the chromalveolates include photosynthetic algal groups (i.e. diatoms (Heterokonta group)), as well as non-photosynthetic groups (Apicomplexa). Based on phylogenetic analyses, stramenophiles, Alveolates and Rhizaria are more closely related, than other members of the chromalveolate group. Hence, they are recognised as a clade, named as SAR.

The classification of chromalveolates is supported by several phylogenetic analyses (Janouskovec et al., 2010; Takarshita et al., 2009). However, it has also been challenged (Baurain et al., 2010; Burki et al., 2012). The problem with chromalveolate hypothesis is that their evolution has been described in a simplified way, assuming only one primary endosymbiosis event and several secondary endosymbiosis are evolved. More recently, it has been argued that tertiary and possible quaternary endosymbiosis event, as well as horizontal gene transfer, contributes to their evolution (Keeling, 2013).

1.2.5. Diatoms

Heterokonts include both photosynthetic (diatoms) and non-photosynthetic organisms. Diatoms are abundant in nearly every habitat – oceans, lakes, streams, mosses, soils, and even the bark of
trees. It is estimated that 40% of the earth’s oxygen (O\textsubscript{2}) is produced through the photosynthetic activities of diatoms in ocean. A unique feature of diatom cells is that they are enclosed within a silica shell (hydrated silicon dioxide) called a frustule. The plastids in diatoms have four chloroplast membranes, which is the result of secondary endosymbiosis. Plastids in diatoms contain Chl \textit{a} and Chl \textit{c}. They also possess a special pigment-binding protein called fucoxanthin Chl \textit{a/c}-binding protein (FCP), which functions as accessory pigment-bound protein complexes for adaption and regulation of photosynthesis (Grouneva \textit{et al.}, 2008).

1.2.6. Dinoflagellates
Dinoflagellates are common organisms in all types of aquatic ecosystems. Roughly half of the species in the group are photosynthetic (Gaines and Elbrächter, 1987); the other half is exclusively heterotrophic. Given their high diverse life styles, they are prominent members of both the phytoplankton and the zooplankton in marine and freshwater ecosystems. They can even survive in harsh benthic environments and even in sea ice. As photosynthetic species, dinoflagellates are second largest group of marine primary producers. However, they are also known to cause harmful algal blooms. About 75–80\% of toxic phytoplankton species are dinoflagellates (Cembella, 2003), they cause “red tides” that often kill fish and/or shellfish either directly with toxins, or because their large numbers that clog animal gills and deplete oxygen (Smayda, 1997). Dinoflagellates are also important components of the microbial loop in oceans, channelling significant amounts of energy into planktonic food webs. Some syndinians, notably \textit{Hematodinium}, are parasites of economically significant crustacean species. Dinoflagellates also have a pivotal role in the biology of reef-building corals, as symbionts (zooxanthellae).
Photosynthetic dinoflagellates have many unique features. Their antenna systems (peridinin Chl \(a/c\)-binding proteins; PCP) are different from other groups in both structure and composition of their binding pigments. Dinoflagellates use peridinin, as the main carotenoid for light harvesting; and Chl \(a/c\) as accessory pigments for photosynthesis adaption. The nucleus of a large majority of dinoflagellates, known as dinokaryon, is very different from other eukaryotic nuclei. Dinokaryon lack nucleosomes and its DNA content is huge, much larger than other eukaryotic cells, and even comparable to that of humans. Dinokaryon divide via a unique form of mitosis. In fact, gene products of all dinoflagellate nuclei (not only dinokaryon) are processed in a unique way in which a spliced leader is trans-spliced to all mRNA molecules. The genomes of plastids and mitochondria of this group are also unique; they are broken up into many minicircles, typically about 2–3 kbp in size. The minicircles encode fewer genes compared to other organisms. Usually, there are only one or two genes or even just fragments of genes in a single minicircle. Some minicircles include no “annotatable” coding regions at all (Zhang et al., 2002; Howe et al., 2008).

1.2.7. Apicomplexa

Apicomplexa contains a large assortment of unicellular eukaryotic organisms. While a large portion of the apicomplexans is parasites of vertebrates, including humans, they are equally diverse in both marine and terrestrial invertebrates. It is estimated that probably 1.2–10 million apicomplexan species exist on earth, while only about 0.1% have been named and described to date (Adl et al., 2007). As a group, apicomplexans are complex organisms that exhibit a wide variety of morphological shapes, depending on their genus and lifecycle stage. They are typically
host specific, inhabiting a wide array of terrestrial and marine environments depending on their host niche environments.

The most eminent of the apicomplexan parasites in humans is *Plasmodium sp.* from the Hematozoa clade, which cause malaria, and is responsible for >1 million deaths annually (Manguin *et al.*, 2010; Mackintosh *et al.*, 2004). Another parasite, *Toxoplasma gondii* is present in approximately 30% of humans worldwide as brain cysts, causing no apparent disease; however, in immunonaive pregnant women, it may seriously affect her foetus. Apicomplexan diseases of domestic animals are associated mainly with farm animals and are renowned for their large economic costs incurred by the agricultural industry. *Eimeria sp.* is known to be responsible for an annual $1.5 billion loss to the poultry broiler industry worldwide (Sharman *et al.*, 2010). Similarly, *Neospora caninum* is linked to losses in the dairy as well as beef industry (Trees *et al.*, 1999).

Apicomplexa possess common features, including four membrane surrounding an apicoplast, cortical alveoli and apical complex. Apicoplast is believed to be a relic of chloroplast. It does not have photosynthesis ability, but plays an important role in the multiple metabolism pathways. Several genetic analyses have supported the close relationship between apicomplexans and dinoflagellates (Fast *et al.*, 2001; Janouskovec *et al.*, 2010). However, as the closest relative of apicomplexans, dinoflagellates have huge gene loss in their chloroplast genome (see section 1.2.5.2). Therefore, it is difficult to compare their plastid genes directly. The newly discovered species in Chromerida phylum is expected to be able to fulfill this role (see section 1.3).
1.3. Endosymbiosis

It is generally agreed that certain organelles of the eukaryotic cell, especially mitochondria and plastids, such as chloroplasts, originated from bacterial endosymbionts. This theory is called the endosymbiotic theory, and was first articulated by the Russian botanist Konstantin Mereschkowski in 1910. This theory proposes that chloroplasts and mitochondria evolved from certain types of bacteria that eukaryotic cells engulfed through endophagocytosis. The bacteria trapped inside these cells entered a symbiotic relationship, a close association between different types of organisms over an extended time. However, to be specific, the relationship was endosymbiotic, meaning that one of the organisms (the bacteria) lived within the other (the eukaryotic cells).

1.3.1. Primary endosymbiosis

In endosymbiosis theory, the ancestral plastid is derived from a specific endosymbiosis event involving a cyanobacteria and an ancestor of Archaeplastida, which already contains mitochondria and a nucleus (Reyes-Prieto et al., 2007; Rodriguez-Ezpeleta et al., 2005). To distinguish it from other endosymbiosis events, this event is called primary endosymbiosis. The key evidence for primary endosymbiosis is that primary plastids are bound by two membranes; which are homologous to the inner and outer membranes of cyanobacteria (Keeling, 2004; Gould et al., 2008; Archibald, 2009; Keeling, 2010). The genome of this cyanobacterial endosymbiont underwent massive gene loss during evolution, accompanied by transfer of its remnant genes to the eukaryotic nucleus. Thus, the nucleus is in charge of coding proteins, which are necessary for plastid formation. A special signal peptide was added to the N-terminus of the protein product for plastid targeting (Bruce, 2001; Jarvis and Soll, 2001; Gould et al., 2008) (See Figure 1.2A-B).
Primary and secondary endosymbiosis. A–B. Primary endosymbiosis. A heterotrophic eukaryote eats a Gram-negative cyanobacterium (A), which is retained rather than being digested (B). The cyanobacterial endosymbiont is substantially reduced, and a large number of genes are transferred to the nuclear genome of the host. The protein products of these genes are targeted to the plastid by way of a transit peptide. The primary plastid is bounded by two membranes derived from the inner and outer membranes of the cyanobacterium. The presumed phagosomal membrane is lost, as is the peptidoglycan wall (except in glaucophyte algae). C–D. Secondary endosymbiosis. A primary alga (either a red or green alga) is eaten but not digested by a second eukaryote (C). This eukaryotic endosymbiont degenerates and genes encoding plastid-targeted proteins are moved from its nucleus to the secondary host nuclear genome. Some genes may also move from the plastid genome to the secondary host nucleus. These plastids would originally be bounded by four membranes derived as indicated. Plastid-targeted proteins encoded in the secondary host nucleus make their way to the plastid using a bipartite leader consisting of a signal peptide followed by a transit peptide. OM: outer membrane; IM: inner membrane. Figure credited to Keeling (Keeling, 2004).

Primary endosymbiosis is thought to have occurred once during the evolutionary processes. There are three lineages with primary plastids known to date: glaucophytes, red algae; and green algae (Keeling, 2004; Keeling, 2010). However, these groups only make up a small fraction of the plastid-bearing eukaryotes on Earth. During their long evolution, the ancestors of green algae
evolved into high plants, while ancestral red algae went through further endosymbiosis events, and evolved into chromalveolate group, with a vast biodiversity (Figure 1.3).

1.3.2. Secondary endosymbiosis
In secondary endosymbiosis, a eukaryotic cell took up another photosynthetic eukaryote that already contained a plastid (an alga); and this engulfed endosymbiotic eukaryote was then reduced and integrated into the new host. In most cases, only the plastid is remained for the benefit of photosynthesis. The typical plastid is surrounded by the relic of the previous host’s plasma membrane. Combined with its own membranes, there will be up to four membranes surrounding the plastid. In some cases, one membrane will deteriorate so that three membranes surround plastid. Compared with the two membranes surrounding the primary plastid, the outermost membrane of the secondary plastid is homologous to the phagotrophic membrane of the host, forming part of an endo-membrane system in the secondary host cell. The second outermost membrane is homologous to the plasma membrane of the engulfed eukaryotic alga. These two inner membranes correspond to the two membranes of the primary plastid (Archibald, 2009; Keeling, 2010). This complex plastid is further shuffled, undergoing huge endosymbiont genome remodelling; hence, most genes are simply lost in the process of secondary endosymbiosis, while some genes of both eukaryotic and prokaryotic origins are transferred to the nucleus of the secondary host. The primary alga and its nucleus are mostly degenerated. Plastid-targeted proteins encoded in the secondary host nucleus are directed to the required location by a bipartite targeting presequence, composed of an endoplasmic reticulum (ER) signal-transit peptide followed by a transit peptide (Gould 2008) (See Figure 1.2C/D). Secondary endosymbiosis theory is supported
by the discovery that in cryptomonads and chlorarachniophytes, a tiny relict of the algal nucleus, called a “nucleomorph” is retained. It is believed to be the relic of the secondary host, after secondary endosymbiosis (Archibald, 2005; Douglas et al., 2001; Gilson et al., 2006; McFadden et al., 1997). The two nucleomorph genomes of cryptomonad (Guillardia theta) and chlorarachniophyte (Bigelowiella natans) indicate that they are similar in size (551 and 373 kb, respectively) and in their basic genetic architecture (Lane and Archibald, 2006).

Secondary (and possible tertiary) endosymbiotic events are believed to be the main method by which plastids spread to other eukaryotic lineages (Archibald, 2005; Gould et al., 2008; Keeling,
2004; McFadden, 1999). Such events have led to the broad diversity of photosynthetic and non-photosynthetic eukaryotes on earth. Thus, endosymbiotic events have resulted in a much more sophisticated map for the evolution of plastids.

1.4. Oxygenic Photosynthesis

Photosynthesis is a biological process, whereby solar energy is captured and stored by a series of events that convert this pure light energy into the biochemical energy needed to power life. It is one of the most significant achievements of life on Earth. This remarkable process provides the foundation for essential life on earth and profoundly alters the Earth’s environment. The yield from photosynthesis provides all our food and most of our energy resources today.

Photosynthesis originated in an anaerobic atmosphere, probably 2.5 billion years ago (Summons et al., 1999). At first, photosynthesis was carried out in bacteria, which used bacteria-Chl to harvest light energy, and did not involve oxygen synthesis. This evolutionary benchmark was the advent of oxygen-producing photosynthesis (Bekker et al., 2004). With the coupling of photosynthesis and oxygen evolution, photosynthetic organisms were responsible for the transformation of this anaerobic environment into an aerobic atmosphere.

Oxygenic photosynthesis occurs in plants (angiosperms, gymnosperms, pteridophytes, and bryophytes), in green algae, and other multipigmented algae (e.g., red algae, brown algae, yellow algae, diatoms), as well as in prokaryotes (cyanobacteria, and prochlorophytes) (Whitmarsh and Govindjee, 1999). In oxygenic photosynthesis, organisms split water and produce oxygen driven by absorbed light energy. The overall equation for oxygenic photosynthesis is:

\[
\text{CO}_2 + \text{H}_2\text{O} + \sim10-12 \text{ quanta of light} \rightarrow \text{O}_2 + \{\text{CH}_2\text{O}\}_n + \text{energy loss as heat and fluorescence.}
\]
1.4.1. Light harvesting

In oxygenic photosynthesis, light reaction is catalyzed by four large photosynthetic membrane protein complexes: PSII, the cytochrome B6f complex, PSI and ATP synthase. Photosynthesis includes many steps, separated in both time and space. The first step involves light energy absorption by antenna system in PSII and PSI. Antenna systems provide an efficient way to harvest light. All antenna systems have the same basic blueprint, and are made up of proteins and binding pigments. Light energy is absorbed by pigments. Chls are the common pigments for light harvesting and energy transfer, including Chl $a$, $b$, $c$, $d$ and $f$. Another common type of pigments built into antenna systems are carotenoids, which usually function as accessory pigments in energy harvesting and quenching. The diversity of carotenoids is enormous, with over a thousand types existing in nature. Harvested energy can easily be transferred from one pigment to another. The direction of energy transfer depends on the orientation of the pigments as well as the energy at which the respective pigments absorb. The net energy transfer usually follows a gradient established by pigments, from higher to lower absorbed light energy, with the reaction centre being the ultimate recipient.

Antenna systems vary in different species (see section 3.1). Cyanobacteria use phycobilisome as their main antenna systems, while green algae/plants use LHC. Red algae use phycobilisomes for PSII, but LHC for PSI. The antenna systems in chromalveolate group are much more complex, including both red algae and green algae-derived LHC, phycobilisomes and even species-specific types of antenna system (such as FCP complexes in diatoms; PCP complexes in dinoflagellates).
1.4.2. Photosystem II

After the light energy have been absorbed by the antenna system, the second step of photosynthesis happens in the reaction centers of PSII, where a special set of Chls are excited by light energy transferred from antenna systems. An electron is “stripped away” from this set of Chls (P680, the primary donor), and transferred to a primary electron acceptor (pheophytin). At the same time this excitation occurs in the reaction centers (P680+), the manganese center (Mn4OxCa) has been activated, splitting water into four protons (H+), 4 electrons (e-) and one O₂ molecule.

PSII absorbs light, and uses this energy to extract electrons from water, releasing protons and oxygen. Several structures of algal PSII have been identified. The core of PSII consists of a dimer of membrane-spanning proteins D1 (psbA) and D2 (psbD), each possesses five transmembrane α-helices, with a pseudo C2 symmetry axis in the centre of each monomer. D1 proteins provide ligands to the catalytic site of water oxidation; therefore, it has a short life and has to be replaced every 30 min in plants under bright light. The flanking transmembrane proteins surrounding D1 and D2 are named CP43 (PsbC) and CP47 (PsbB). These proteins have six transmembrane helices, and they bind Chl pigments to help capture light energy and transfer this energy to the cores. Several small subunits are attached to CP43 and CP47. They are low molecular mass (LMN) subunits. Most of them consist of one single transmembrane spanning α-helix and function in stabilisation, assembly or dimerisation of the PSII complex.

1.4.3. Photosystem I

Photosystem I (PSI) catalyses the electron transport from plastocyanin/cytochrome C₆ to ferredoxin/flavodoxin. It also absorbs light and produces protons for transport, similar to PSII.
The electrons excited by PSI and received from electron transport chain have gone through another electron transport chain involving phylloquinone, ferredoxin and other proteins. The final receiver of electrons is nicotinamide adenine dinucleotide phosphate (NADP+).

PSI is one of the largest and most complex membrane proteins with known structure. The core of PSI is highly conserved in cyanobacteria, algae and higher plants, except that cyanobacteria have a trimer structure, where algae and higher plants possess a monomer structure. There are four LHC proteins surrounding the core of PSI (Lhca1, Lhca2, Lhca3 and Lhca4). Its external antenna is asymmetric and heterogeneous, and rapidly responds to the environmental change, such as light intensity or environmental stress.

1.4.4. Electron transport chain

Followed this light reaction, electrons generated by the special set of Chls are passed to active reaction centres. These reaction centres return to the basic state, after passing these electrons to pheophytin. Pheophytin continues the electron transport by passing it to membrane bound plastoquinone molecules, and onto the cytochrome $b_{6}f$ complex. Electrons finally arrive at reaction centers of PSI (see details in Figure 1.3). The cytochrome $b_{6}f$ complex is an enzyme found in the thylakoid membrane of chloroplasts in plants, cyanobacteria, and green algae. It transfers electrons between the two reaction complexes from PSII to PSI, thereby introducing protons into the thylakoid space. It also transfers protons generated by water splitting in the chloroplast stroma across the thylakoid membrane into the lumen. Coupled with the electron transport, a proton gradient is formed, which activates the ATP synthesis complex located on the thylakoid membrane.
1.4.5. Carbon fixation

The end product of light reaction is \( \text{O}_2 \), the reduced NADP\(^+\) (NADPH) and ATP. ATP and NADPH are then transferred to the stroma matrix, converting CO\(_2\) to carbohydrate \( \{\text{CH}_2\text{O}\}\_n \). The conversion of CO\(_2\) to \( \{\text{CH}_2\text{O}\}\) is also called carbon fixation, which involves a series of reactions catalysed by many water-soluble enzymes.
Cyanobacteria (Tabita, 1988) and photosynthetic eukaryotes utilize the reductive pentose phosphate cycle, also known as the Calvin-Benson cycle (Calvin and Benson, 1948). In this cycle, CO$_2$ is incorporated through a reaction with ribulose-1,5-bisphosphate, catalysed by the enzyme ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO).

In addition, in plants capable of C4 photosynthesis (Hatch, 1987) and Crassulacean acid metabolism (CAM) (Cushman, 2001; Black and Osmond, 2003), photosynthesis performs a pre-fixation of CO$_2$ that is catalysed by the enzyme phosphoenol pyruvate carboxylase. The resulting malate is then decarboxylated and incorporated into organic molecules by the reductive pentose phosphate cycle.

The first step in these two carbon fixation cycles relies on the carboxylation of acetyl-Coenzyme A (CoA), followed by a series of enzymes convention powered by ATP. Three molecules of carbon dioxide are converted into 3-phosphoglycerate which as an organic chemical can be transported outside of starch and used for central metabolism. At the completion of the reaction, RuBisCO is regenerated back to its original form, and used in another round of CO$_2$ fixation. This process is the most important organic biosynthesis process on the planet. Therefore, enzymes involved in this cycle are more likely to be conserved among different species, making them an ideal research subject for investigating evolution and photosynthesis (see details in section 4.1).

1.5. Iron and Photosynthesis

1.5.1. The role of iron

Iron is the fourth most abundant element in the Earth’s crust and an essential micronutrient for virtually all living organisms. Iron can be found in proteins in the form of iron-sulfur clusters, in the porphyrin ring of heme or siroheme, or as non-heme iron. Given that iron can exist in two
stable oxidation states, Fe^{2+} (ferrous) or Fe^{3+} (ferric), it is often used as a cofactor in proteins that catalyse reduction-oxidation (redox) reactions, such as respiration and photosynthesis. Therefore, bioenergetic membranes tend to be rich in iron. Thylakoid membranes of chloroplasts are also iron-rich because they use iron-containing protein complexes in the photosynthetic electron transport chain. Half of the iron requirement of photosynthesis can be attributed to PSI, which contains three iron-sulfur clusters (Ben-Shem et al., 2003).

1.5.2. Iron nutrient stresses in nature

Although iron is abundant in the Earth’s crust, its low bioavailability makes it a limiting nutrient for life. Iron seeding experiments have shown to stimulate algal blooms in the North Atlantic, Southern and equatorial Pacific oceans, demonstrating that iron is the limiting nutrient for primary productivity in oceans (Martin et al., 1994; Behrenfeld and Kolber, 1999; Boyd et al., 2000; Moore et al., 2006). With the shift in our atmosphere to aerobic conditions approximately 2.2 billion years ago (Falkowski and Isozaki, 2008; Kump, 2008), iron is now mostly found as stable Fe^{3+}-oxides, such as goethite (FeOOH) and hematite (Fe_{2}O_{3}), which are insoluble at biological pH (Guerinot and Yi, 1994). Therefore, organisms have evolved different strategies for surviving in iron-limiting conditions.

1.5.3. The physiological changes under iron stressed condition in algae

Photosynthetic organisms have an additional requirement for iron because of the abundance of iron-bound proteins in the photosynthetic apparatus and other metabolic pathways in the chloroplast. As a result of iron-stress, the abundance of iron-containing proteins is reduced. Cells become chlorotic, owing to the involvement of a di-iron enzyme in Chl biosynthesis (Tottey et al.,
PSI seems to be a prime target of such stress, probably because of its high iron content (12 Fe per PSI) (Sandmann and Malkin, 1983).

In general, photosynthetic algae remodel their photosynthetic apparatus in response to iron deficiency to minimize photo-oxidative damage, which is caused by over-excitation of the photosynthetic apparatus resulting from a reduced number of PSI reaction centres. Under iron-stress conditions, before the onset of iron-limitation and chlorosis, the LHCI antenna disconnects from PSI (Moseley et al., 2002). Following the disconnection of the antenna from the reaction centre, certain LHCI proteins are degraded or processed by proteases, and expression of new LHC proteins is induced (Moseley et al., 2002).

1.6. Chromera velia

1.6.1. The discovery of Chromera velia

Chromera velia (C.velia; Chromerida; Alveolata; Eukaryota) is a photosynthetic eukaryotic unicellular alga discovered on the eastern coast of Australia (Moore et al. 2008). During the isolation of Symbiodinium, Dr. Moore isolated a new species of brown-colored alga from the stony coral Plesiastrea versipora (Metazoa: Cnidaria: Faviidae). During its life cycle, C.velia is mostly immotile, with multiplying spherical cells, which under optimal conditions transform into motile flagellated cells (Weatherby et al. 2011; Obornik et al. 2011). It contains four membrane plastids and uses Chl a as its major pigment for light harvesting. It also contains a new type of carotenoid, which is similar to fucoxanthin in diatoms (Moore et al., 2008). This new alga has full photosynthetic function, but is genetically related to dinoflagellates and non-photosynthetic apicomplexans. Given its unique evolutionary position, C.velia has been designated to a newly
phylum, named Chromerida, which also includes the newly discovered photosynthetic species *Vitrella brassicaformis*.

1.6.2. Current research status of *C. velia*

The discovery of *C. velia* has generated great interest in protist research for many reasons. Research on *C. velia* will benefit our knowledge of photosynthesis, evolution, pharmacy and many other areas.

Based on rDNA phylogenetic analysis, the functional chloroplasts of *C. velia* are predicted to be the closest photosynthetic relatives to the relic non-photosynthetic plastids of apicomplexan parasites (Moore *et al.* 2008). Likewise, sequencing of the plastid genome suggests that the plastid in both *C. velia* and members of Apicomplexa comes from a red algae origin (Janouskovec *et al.* 2011). Therefore, the functional chloroplasts of *C. velia* can be used as a connection between apicoplast and chloroplast in dinoflagellates (Moore *et al.*, 2008). Thus, research on the chloroplast of *C. velia* is specifically helpful for understanding evolution in dinoflagellates and apicomplexans.

With its very simple, yet efficient photosystem, *C. velia* has proven similarity to diatoms, in terms of its photosynthesis adaption (Kotabová *et al.*, 2011) and light harvesting ability (Pan *et al.*, 2011; Tichy *et al.*, 2013). *C. velia* expresses a novel type of red-shifted Chl a antenna complex, when adapted to red/far-red light (Kotabová *et al.*, 2014; Bina *et al.*, 2014). It is also reported to possess a rare type of RuBisCO protein in its dark cycle (Janouskovec *et al.*, 2011; Quigg *et al.*, 2012). Thus, research on *C. velia* will also assist us to understand photosynthesis in both light and dark cycles.
Furthermore, *C. velia* is also reported to have a novel iron uptake mechanism (Sutak *et al.*, 2010). Research on this aspect of its biology will benefit the study of iron stress response in photosynthetic organisms.

Finally, people have been battling with diseases caused by apicomplexan parasites (such as Malaria) for a long time. The apicoplast in apicomplexan parasites is important in many biogenesis pathways involved in apicomplexa metabolism. Research on the closely related plastid in *C. velia* may shed light on its architecture, paving the way for the design of a new type of drugs (Lim and McFadden, 2010).

1.7. Aim

*C. velia* is a new species of algae having a unique evolutionary position. Given our interest in the photosynthetic processes, we focussed this investigation on the evolutionary relationships of proteins involved in the LHC and enzymes in the Calvin-Benson cycle of *C. velia*. Such analyses will provide clues for understanding the evolutionary position of *C. velia* with respect to its closest relatives (diatoms, photosynthetic dinoflagellates) as well as to more distantly related eukaryotes (Apicomplexa).

Previous studies of iron uptake system in *C. velia* formed the basis for my interest in understanding changes to its photosynthetic apparatus in responses to iron stress. Given its novel iron uptake system, *C. velia* has the potential to uniquely counter iron stress. In this thesis, biochemical experiments were used to investigate such photosystem changes in response to iron stress, while bioinformatics was used to analyse these results.
Chapter 2. Material and Methods

2.1. Cell culture conditions

2.1.1. *C.velia* strain

*C. velia* is a marina brown-coloured unicellular alga. *C. velia* (strain RM12) was first isolated from the stony coral *Plesiastrea versipora* (Cnidaria: Faviidae), from Sydney Harbour, New South Wales, Australia (Moore *et al.* 2008) during the isolation of intracellular symbionts of the genus *Symbiodinium*. The same *C. velia* species was also later isolated from the stony coral *Leptastrea purpurea* (Cnidaria: Faviidae) from One Tree Island, Great Barrier Reef, Queensland, Australia (Moore *et al.*, 2008). *C. velia* strain CCAP 1602/1, originating from One Tree Island, was used in this study, obtained as a kind gift from Professor Tony Larkum.

2.1.2. Culture medium

*C. velia* (strain CCAP 1602/1) was cultured in artificial seawater supplemented with K+ES medium (Keller *et al.*, 1987; Harrison *et al.*, 1980). The sea salt used for artificial seawater was purchased from local commercial brand (Ocean nature premium sea salt: Aqua Sonic, Australia). Sea salt was dissolved in distilled water to a final concentration of 3.3% (w/v) and autoclaved, followed by the addition of nutrient solutions listed in K+ES medium formula. (see Appendix-1). In order to test the influence of iron on *C. velia*, the iron was omitted from K+ES seawater medium, referred to herein as iron-stress medium (see Appendix-1). The culture flasks were soaked with 0.1 M HCL for at least 24 hours to remove any remaining iron, then rinsed completely using milliQ water. The rest of culture conditions were same as mentioned in section 2.1.3.
2.1.3. Algal culture condition

*C. velia* was cultured at 26 – 27°C under 24 h illuminations with a light intensity of 20–30 µmol photons s\(^{-1}\) m\(^{-2}\) (units also referred to as µE), using cool white fluorescent tubes (Philips, Australia).

Cultures were incubated in pre-autoclaved Erlenmeyer glass flasks with sterilised cotton stoppers (400 ml medium in a 1 L flask). The flasks were placed on an orbital flat-bed shaker (Model OP3422: Paton Scientific, Australia) with continuous shaking at 100 rpm to aerate the cultures. Cultures were inoculated in a ratio of 1:15 initiating culture: fresh medium. Cells were harvested from cultures at their late log phase (~ 3 weeks). Culture density was measured using the optical density (OD) at 730 nm (UV-2550 spectrophotometer: Shimadzu, Japan).

2.1.4. Counting cells and the density of culture

The culture growth rate was monitored by cell counting. 10 µl of the homogeneous cell culture was taken, loaded onto a hemocytometer and covered with slips, then observed under a light microscope at 40X magnification (Larorlux 12, Leitz, Germany). Cell numbers were counted manually with three replicates. There were 25 grids (5x5) observed under the microscope, which consisted of an area of approximately 0.1 mm\(^3\). The number of cells located within the grids was recorded. In the late log phase, due to the fact that the cell numbers were too high to measure (usually over 1000 cells per 25 grids), random sampling of five grids was used to indicate the total cell number. Cell numbers were recorded every three days. This measurement was applied three times to different biological samples to confirm the significance.

Another way of monitoring growth rate was by using a spectrophotometer (UV-2550: Shimadzu, Japan) with Taylor-sphere attachment (ISR-240A: Shimadzu, Japan). The culture (1 ml) was pipetted into the cuvette (1 cm light path) and the readings at 730 nm were recorded ten times,
which were used to indicate cell density. The target wavelength is far away from the photosynthesis light absorption region and only relevant to the cell number and reflectance.

Both methods were used to measure cell growth rate in normal and iron stress cultures. The results show that they are similar to each other and can be used as a proof of culture growth rate.

2.1.5. Cell harvesting
Cells were harvested at the late log phase of culture growth (3 weeks, with cell density between 500–3,000 cells per 1 mm$^3$). According to the different volume, C.veila cells were harvested by centrifugation using either 50 ml falcon tubes in a swing bucket centrifuge at 4,500 rpm, 20 °C for 5 min (Labofuge 400: Heraeus, Germany); or 400 ml capacity centrifuge bottles in an Avanti J25i Centrifuge with a JLA-10.500 rotor (Beckman Coulter, USA) at 10,000 x$g$ at 20 °C for 15 min. After centrifugation, the supernatant was discarded and the collected pellet was rinsed once with buffer A (50 mM MES pH6, 5 mM MgCl$_2$, 5 mM CaCl$_2$), and then stored in a -80 °C freezer for further use.

2.2. Pigment composition and oxygen evolution rate
2.2.1. Pigment extraction and spectral analysis
Pigment composition was recorded using methanolic cell extraction. 2 ml of C.velia culture was centrifuged at 16,000 x$g$, 20 °C for 3 min (centrifuge 5415D: Eppendorf, Germany). The harvested cell pellet was resuspended in 1 ml 100% methanol and vortexed for 3 min at room temperature; and then re-centrifuged at 16,000 x$g$ for 3 min; after which 1 ml of the supernatant was loaded into the cuvette for absorbance measurement. Absorbance was measured using a spectrophotometer (UV-2550, Shimadzu, Japan) from 400 to 800 nm. Chlorophyll $a$ (Chl $a$)
concentration was calculated according to the absorbance (A) readings at 665 and 750 nm based on the published formula: Chl \( a \) (µg/ml) = 12.945*(\( A_{665\text{nm}} \)-\( A_{750\text{nm}} \)). This formula is applicable to pure Chl \( a \) calculation in 100% methanol (Ritchie, 2006). In order to reach high accuracy of calculation, the absorbance reading at 665 nm should be located within the value range of \( A=0.1-2.0 \). In cases where the pigment concentration was too high, the pigment extract was diluted using 100% methanol until its \( A_{665\text{nm}} \) reading fits the range mentioned above.

2.2.2. \textit{In vivo} absorbance reading

Measurements of the optical characteristics of algae are the key point for estimating the efficiency of photosynthetic light utilization, interpreting the \textit{in vivo} significance of quantitative and qualitative changes in light harvesting pigments, and for determining the role of algae in modifying the transmission and reflection of light by natural waters. Despite this importance, optical measurements are made infrequently, largely because of difficulties associated with partitioning attenuation into components of absorption and scattering. In particular, measuring light absorption is complicated when scattering (reflection, refraction, and diffraction) is the dominant process that contributes to light attenuation. However, accompanied with the development of technology, the measurement of \textit{in vivo} absorbance in unicellular algae can be achieved nowadays.

An \textit{in vivo} absorption spectrum reading includes measuring the diffuse, total reflectance, and transmittance without pigment extraction. Measurements were performed on a UV–Vis spectrophotometer (UV-2550: Shimadzu, Japan) with a Taylor-sphere attachment (ISR-240A, Shimadzu, Japan). Due to the self-rotation of living cells in liquid, the readings of absorption...
would not be stable; thus, recordings at multiple times were required. A set of 10 readings was used for smoothing the absorption spectrum curves.

2.2.3. High Pressure Liquid Chromatography (HPLC)

Each sample for HPLC analysis used 2 ml of culture. Cells were harvested by centrifugation at 16,000 xg at room temperature for 3 min, and rinsed with 50% methanol to remove any remaining water and growth medium. The rinsed cell pellet was resuspended in pre-chilled 100% methanol, mixed well by vortexing, and centrifugation at 16,000 xg, at room temperature for 3 min (Eppendorf, Germany). In order to remove possible cell debris, the supernatant was centrifuged again at 16,000 xg, at room temperature for 3 min (Eppendorf, Germany). The supernatant from the second centrifugation was injected into compatible HPLC vials for further use. HPLC analysis were conducted using a C8 (4.6 X 150 mm) reverse phase column (Zorbax, Australia) attached to a Shimadzu HPLC (model 10A series) which was equipped with a diode array absorption detector (370–800 nm, SPDM10Avp, Shimazu, Japan). The mobile phase of HPLC included two solutions: solution A was 85% (v/v) methanol with 0.05 M ammonium acetate, and solution B was 100% methanol. HPLC was performed at a constant flow rate of 0.8 ml/min. The detailed program of HPLC is listed in Table 2.1 below. Generated data was recorded and plotted using Class-VP software (Shimadzu, Japan), and pigment analyses were performed using Microsoft Excel software.
Table 2.1. Details of HPLC program.

<table>
<thead>
<tr>
<th>Time (in min)</th>
<th>0–3</th>
<th>3–6</th>
<th>6–16</th>
<th>16–20</th>
<th>20–40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol concentration</td>
<td>85% constant</td>
<td>Increased to 90%</td>
<td>Increased to 95%</td>
<td>Increased to 100% constant</td>
<td>100%</td>
</tr>
<tr>
<td>Solution A (percentage)</td>
<td>100% constant</td>
<td>Decreased to 67%</td>
<td>Decreased to 33%</td>
<td>Decreased to 0%</td>
<td>0</td>
</tr>
<tr>
<td>Solution B (percentage)</td>
<td>0</td>
<td>Increased to 33%</td>
<td>Increased to 67%</td>
<td>Increased to 100% constant</td>
<td>100%</td>
</tr>
</tbody>
</table>

2.2.4. Room temperature fluorescence and low temperature fluorescence spectra

The fluorescence spectral properties of the isolated pigment-protein complexes were studied using a Varian Cary Eclipse Fluorescence spectrophotometer (Agilent, USA). All samples used for fluorescence investigation were diluted with buffer A to the final $A_{665nm}$≈0.1. The diluted solution was loaded into a glass cuvette, and then measured at different light stimulation settings. Two excitation wavelengths, 435 and 480 nm, were used to study the energy transfer profiles of the isolated complexes. For excitation of Chl $a$ at 435 nm, emission fluorescence readings were recorded between 600–800 nm; for excitation of carotenoids at 480 nm, emission fluorescence readings were recorded between 600–800 nm. One emission wavelength, 720 nm corresponding to PSI absorbance, was also used to study the energy transfer. Fluorescence excitation readings were recorded between 400 and 710 nm, and the results were recorded with at least 15 repeats to maximize accuracy.

Low temperature (77 K) fluorescence spectra were measured on the same system with additional cryostat attachment (Oxford Instruments, UK). The bands were resuspended in buffer A and
diluted with 65% (v/v) glycerol to the final $A_{605nm} \approx 0.1$. The fluorescence spectra were recorded with a slit width of 1 nm using the same light stimulation settings. To increase the signal-to-noise ratio, the fluorescence spectra were obtained as an average of 50 repeated readings with 1 min time intervals.

2.2.5. Oxygen evolution rate measurement

PSII activities can be measured by the oxygen evolution rate. This was measured using a Clark type electrode (Hansa-tech, UK). The electrode was stabilized at 25°C and used as described by Ritchie et al. (1997). The electrode was calibrated with aerated water for 100% air saturation and with N$_2$ bubbling for 15 min to adjust as zero O$_2$ condition. The oxygen concentrations in air-saturated medium were calculated according to the oxygen solubility algorithms of Carpenter (1966) and Colt (1983). 1 ml culture was used for oxygen evolution rate measurement. The culture was pipetted into the chamber after electrode calibration and acclimated in the dark condition for at least 30 min. Oxygen concentration in the chamber was recorded under illumination with different light intensities (each for 5 min) in a random order. Between measurements, samples were dark-adapted for at least 15 min. The change of oxygen concentration during the dark period was also recorded and referred to as the respiration rate in later calculations. The total oxygen evolution rate was obtained by recorded oxygen change rate minus respiration rate.
2.3. Photosynthesis protein study of *C. velia*

2.3.1. Cell breaking

*C. velia* cells were harvested by centrifugation at 10,000 xg for 10 min (JA-17: Beckman Coulter, USA) at 20 °C. The cell pellet was washed with buffer A and re-centrifuged using a swing bucket centrifuge at 4,500 rpm for 5 min (Labofuge 400, Heraeus, Germany). In order to break the cells, 0.5 ml cell pellet was mixed with 0.5 ml beads (0.1 mm diameter) and 1 ml buffer A (with freshly added 1 mM PMSF and 0.1 μM DNase). The cells were broken and lysed using a FastPrep-24 instrument (MP Biomedicals, USA) at a speed of 6.0 m/s for 20 cycles (45 s each) with 5 min interval cooling time. During the cooling time, samples were placed on ice to minimize high temperature damage. The cell lysates were centrifuged at 1,000 xg for 5 min to pellet the beads and cell debris (centrifuge 5415D, Eppendorf, Germany). The supernatant was transferred to ultracentrifuge tubes and centrifuged at 40,000 rpm for 30 min at 4 °C (SW-55 Ti: Beckman coulter, USA). After ultracentrifugation, thylakoid membranes were centrifuged down and washed once with buffer A. The pellet was dissolved in buffer A with the aid of a plastic homogenizer and stored overnight at 4 °C for the next experiment.

2.3.2. Sucrose density gradient ultracentrifugation

To prepare sucrose gradient, different concentrations of sucrose solutions were prepared (40–5% (w/v) with 5% interval separation) by dissolving sucrose in buffer A (50 mM MES pH 6.5, 5 mM MgCl₂, 5 mM CaCl₂) with additional detergent DoDM (0.03%, (w/v)). During the casting of stacked sucrose layers, different sucrose solutions were loaded into centrifuge tubes in the order from heaviest to the lightest.
Linear sucrose density gradients were also used in experiments. To prepare a linear sucrose density gradient, a sucrose solution was made (18% (w/v) sucrose, 0.03% (w/v) n-dodecyl β-D-maltopyranoside (DoDM) dissolved in buffer A) and loaded equally into ultracentrifuge tubes, then stored at -80 °C overnight. Using slow frozen method, the ultracentrifuge tubes were placed straight up gently in the cold room (6–8 °C), and then slowly thawed for building a linear gradient. This process took at least 8 h before ultracentrifugation.

In order to isolate different protein complexes from the harvested sample, detergent solubilisation was necessary. The Chl a concentration of the isolated sample was determined by methanolic extraction as described above (Section 2.1). Samples were solubilised with DoDM in a 1:10 ratio Chl a/ detergent (w/w). After 60 min rotation in the dark at room temperature, the supernatant was centrifuged at 16,000 xg at 4 °C for 15 min, and then loaded onto the sucrose density gradient. These sucrose density gradients were ultra centrifuged at 35,000 xg (SW-40 Ti: Beckman Coulter, USA) at 4°C overnight for at least 16 h. Several bands were resolved after ultracentrifugation. The isolated pigment-protein complexes were carefully collected with a syringe.

2.3.3. Protein concentration

In order to compare different bands with protein concentrations at the same level, the Bradford assay was used to monitor protein concentrations in isolated samples generated from sucrose density gradients. Bovine serum albumin (BSA) was dissolved in buffer A. Five known concentrations of BSA were used to make stand curve. First, 5 µl was taken from each BSA solution, mixed with 300 µl Bradford assay reagent (Quick Start™ Bradford 1x Dye Reagent: Bio-Rad, USA) and incubated at room temperature for 5 min. At the same time, the spectrophotometer was calibrated using pure reagent as reference. The reading at 595 nm was
adjusted to zero and then used as a baseline. The 595 nm absorbance values of mixed BSA samples were recorded and plotted against the protein concentrations using Microsoft Excel software. The linear formula generated was regarded as the standard curve and was used for future protein concentration calculation. In this study, the formula obtained was: protein concentration (µg/ml) =0.5747*\text{A}_{595\text{nm}}+0.0923; R^2= 0.9612.

Protein concentrations indifferent bands isolated from the sucrose density gradient were measured by the same way as described above, with the A_{595\text{nm}} readings used to calculate the total protein concentrations.

2.3.4. Sample treatment
Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse the protein composition in different bands isolated by sucrose density gradient. According to the protein concentration calculated, sucrose bands with same amount of proteins were mixed with trichloroacetic acid (TCA) to a final concentration of 10% (v/v). The mixture was placed on ice for at least 1 h. Mixed samples were rinsed with ice-cold acetone, and centrifuged at 16,000 xg, at 4 °C for 10 min. The pellet was inverse-dried in a fume hood. Next step, samples were rinsed with cold acetone twice. After the final dry, the pellet was mixed with 10 µl 0.5 M Tris-HCl (pH 8.0) and 10 µl sample buffer (60 mM Tris-HCl (pH 6.8), 10% (v/v) glycerol, 4% (w/v) SDS, 0.01% (w/v) bromphenol, 5% (w/v) β-Merc blue). In some cases the mixture presented a yellow colour, which implied the pH was too acidic. In order to restore the pH, 5 µl of 1 M NaOH was added until the colour changed to blue. The final step of sample treatment was boiling samples in water for 5 min, followed by cooling in ice for 5 min. The samples were then centrifuged at 16,000 xg at
20 °C for 3 min (centrifuge 5415D: Eppendorf, Germany). The supernatant was collected for SDS-PAGE.

2.3.5. Sodium dodecyl sulphate-Polyacrylamide gel electrophoresis (SDS-PAGE)
SDS-PAGE was performed on purchased 4–12% precast gels (Bolt Bis-Tris plus gels; Novex, USA), in a Bolt mini gel tank (Novex, USA) with a Mini vertical Gel system as power supply (250-4; Gradipore, Australia). Collected supernatant was loaded onto the gel, along with molecular weight markers. Electrophoresis was performed for approximately 2 h with a constant voltage setting starting from 40 V, with 20 V increases every 20 min, and ending up with 120 V for the rest of the run. After electrophoresis, the gel was fixed in 40% methanol/10% acetic acid (v/v) for at least 30 min, and then stained with R-250 Coomassie blue solution overnight (0.01% (v/v) CBB R-250, 20% (v/v) methanol, 0.5% (v/v) acetic acid). The gel was washed with 40% methanol/10% acetic acid (v/v) several times the next day to eliminate the blue colour in the background. The clear gel was scanned using Quantity One software on a GS-710 Calibrated Imaging Densitometer (Bio-Rad, USA) to obtain a high resolution image.

2.3.6. Western Blotting
2.3.6.1 Transfer of protein from the gel to the membrane
Western blotting was used to detect the specific proteins in different samples. It was performed after SDS-PAGE. Instead of fixation, the gel was soaked in 1X transfer buffer (25mM Tris pH8.5, 192 mM glycine, 10% methanol). At the same time, four pieces of filter paper slightly bigger than gel were cut and soaked in transfer buffer. A piece of PVDF membrane was also cut and wetted in 100% methanol for 5 min and then soaked in transfer buffer for 5 min. Several pieces of sponges
were also soaked in the transfer buffer. A sandwich was made on the blotting box in the following order from cathode to anode: sponges → two pieces of paper → gel → membrane → two pieces of paper → sponges. In order to remove bubbles between layers, a rod was used to press along the surface from side to side. The western blotting box was tightly connected, and filled up with transfer buffer. The western blotting transfer was performed at a constant current of 350 mA for approximately 2 h using a vertical Gel system as the power supply (250-4: Gradipore, Australia). Transfer to the membrane was checked using Ponceau Red staining before the blocking step.

2.3.6.2 Antibody staining and image scanning

After western blotting, the membrane was washed with Tris-buffered saline (TBS) buffer, and then incubated with blocking buffer (5% (w/v) Skim milker powder in TBS buffer) at 4 °C overnight. The membrane was washed with Tris-buffered saline-Tween 20 (TBS-T) buffer (1X TBS buffer with 0.1% (v/v) Tween 20) twice for at least 30 min each time. The next step was primary antibody incubation at 4 °C overnight (1:5000 dilutions in blocking buffer, as instructed on the product manual). The primary antibody solution was recycled and stored in -80 °C freezer after use. The membrane was washed with TBS-T twice for 20 min, and then incubated with secondary antibody (1: 15000 dilution in blocking buffer, as instructed on the product manual) for 2 h at room temperature with continuous shaking. The secondary antibody was discarded after use. After secondary antibody staining, the membrane was washed with TBS-T buffer twice for 20 min. The last step was incubating the membrane with immuno-detection reagents (Clarity™ Western ECL Substrate: Bio-Rad, USA). This step was performed by mixing 1 ml of each detector and
enhancer reagent in the dark, rinsing the surface of the membrane by pipetting, and incubating in
the dark for 5 minutes. The membrane was then scanned with a G:BOX (SynGene, UK) using
Gene-Snap software.

2.3.6.3 Antibody stripping and re-staining

After imaging, the membrane can be reused for another round of antibody labelling, however, the
previous stained antibody needs to be stripped. To do this, the membrane was washed with TBS-T
buffer, then incubated with stripping buffer (15 g glycine, 1 g SDS, and 10 ml Tween 20 dissolved
in 1 L; pH 2.2) for 5–10 minutes. This step was repeated once after changing the stripping buffer.
Next, the membrane was washed with TBS buffer for 10 min. And, this step was repeated with a
change of the TBS buffer. The last step was washing the membrane with TBS-T buffer twice for 5
min each. After these steps, the membrane was ready for another round of blocking and staining.

2.4. Genomic study of *C. velia*

2.4.1. DNA extraction from *C. velia*

2.4.1.1 Culture preparation

In order to extract DNA from *C. velia*, 2 ml culture was taken from the culture at the later log
phase. Cells were harvested by centrifugation at 16,000 xg, at 20 °C for 3 min (centrifuge 5415D:
Eppendorf, Germany).

2.4.1.2 Cell lyses

Cells pellets were resuspended in 250 µl freshly made Tris-EDTA (TE) containing lysozyme (20
mg/ml). The volume of TE was adjusted to accommodate the size of the cell pellet. All additional
volumes were adjusted accordingly. Cells were incubated at 37 °C for 20 min. After the incubation, 2 µl protease K (20 mg/ml) was added to the mixture and mixed by gentle shaking, and 100 µl of 20% SDS (final concentration 2%, (w/v)) was also added and mixed by gentle shaking. The mixture was incubated at 50 °C for 2 h, preferably until the cell suspension changed colour. 2 µl of 20 mg/ml RNase A was added and mixed by gentle shaking. The mixture was incubated further at 65 °C for 60 min, after which the suspension was transferred to a screw-capped tube with an O-ring seal.

2.4.1.3 DNA Purification

In the fume hood, the following steps were performed for DNA purification. All solvent contaminated items were disposed of in the hazardous waste container in the fume hood, including tips, tubes and gloves.

An equal volume of 25: 24 phenol: chloroform reagent, (Bio-Rad, USA) was added and mixed well by gentle shaking. The mixture was centrifuged for 4 min at 16,000 xg, at 20 °C (centrifuge 5415D: Eppendorf, Germany). The aqueous phase (top) was transferred to a new screw-capped tube. The previous steps were repeated one or two times if desired, after which an equal volume of chloroform was added and mixed well by gentle shaking. The mixture was centrifuged for 4 minutes at 16,000 xg (centrifuge 5415D, Eppendorf, Germany). The aqueous phase was transferred to a 1.5 ml Eppendorf tube, after which an equal volume of iso-propanol was added to the mixture and precipitated overnight at -20 °C. The mixture was centrifuged for 30 min at 4 °C, 14,000 xg (centrifuge 5415D: Eppendorf, Germany). The supernatant was removed and the pellet
was washed with 70% (v/v) ethanol, after which the pellet was air-dried and resuspended in 50–100 µl DEPC-treated water.

2.4.1.4 Assessment of DNA quality and quantity

DNA concentration was checked using a NanoDrop Spectrophotometer (ND-1000, NanoDrop). The spectrophotometer was calibrated using DEPC-treated water as the base line, then wiped clean. 5 µl of DNA solution was added to the reading plate, and the DNA concentration was calculated and recorded. The quality of the DNA was checked using two different absorption ratios: 260 nm/ 280 nm, and 260 nm/ 230 nm. DNA with a good quality should have both ratios over 1.8.

2.4.1.5 Agarose gel-electrophoresis

Agarose gel electrophoresis can also be used to check the DNA sample quality and to estimate the DNA concentration based on signal strength comparison with a DNA ladder.

Firstly, 1% (w/v) agarose was added to 50 ml running buffer (1X Tris-Acetic acid-EDTA (TAE)) in a glass bottle with a screw-top. The agarose was fully dissolved in a microwave oven (20 seconds per cycle, repeated 3–4 times) in a bottle with a loosened cap. Approximately 3 µl Syber Safe DNA gel stain (Invitrogen, Australia) was added to molten agarose, avoiding light as SYBR Safe is light-sensitive.

Next, the gel was poured into a casting tray. The well combs were inserted into the gel and stabilised for approximately 10–15 min. The gel was casted in the fridge or in the cold room for a quick setting. The gel can be stored under this condition overnight to be used in the next day. The
gel was removed from the casting tray and placed in a gel tank (Minnie gel unit; GE Healthcare, Australia), which was pre-chilled at 4°C.

Sufficient running buffer (1X Tris-Acetic acid-EDTA (TAE)) was poured into the tank to completely cover the gel and wells. After removing well combs, Samples were mixed with 6x loading buffer (New England Biolabs, USA). In the case of using a 16-wells comb, 2 µl of the mixture was loaded per well. Approximately 500 ng of DNA ladder (New England Biolabs, USA) was loaded as an appropriate size marker. Electrophoresis was carried out at 100–200 V for 1 h, with the power supplied by a Mini vertical Gel system (250-4: Gradipore, Australia). The bands were visualised on a G:BOX transilluminator (SynGene, UK).

### 2.4.2. Polymerase chain reaction (PCR)

In order to retrieve full sequences for genes of interest in *C.velia* (details in Section 5.5), PCR was performed. Different temperature settings were used for the annealing step in PCR to reach maximum productivity. The primers were designed using Primer3 software based on the conserved regions, inferred by comparing incomplete sequences with their homologous relatives.

#### 2.4.2.1 Preparation of PCR

The components of the PCR reaction mixture are listed below in Table 2.2:
Table 2.2. Components of PCR reaction mixture.

<table>
<thead>
<tr>
<th>Variable</th>
<th>20 µl reaction system</th>
<th>50 µl reaction system</th>
</tr>
</thead>
<tbody>
<tr>
<td>50mM MgCl₂</td>
<td>1 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>dNTP (10 mM)</td>
<td>0.4 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>5X MangoTaq reaction buffer</td>
<td>4 µl</td>
<td>10 µl</td>
</tr>
<tr>
<td>Water</td>
<td>11.4 µl</td>
<td>27 µl</td>
</tr>
<tr>
<td>Primers (Forward and Backward)</td>
<td>1 µl each</td>
<td>2.5 µl each</td>
</tr>
<tr>
<td>DNA template (&gt; 30 ng/µl)</td>
<td>1 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>MangoTag DNA Polymerase enzyme (added individually at the end)</td>
<td>0.25 µl</td>
<td>0.6 µl</td>
</tr>
</tbody>
</table>

2.4.2.2 Details of PCR

PCR was performed using an Eppendorf Mastercycler EP Gradient Thermal Cycler (Eppendorf, Germany). The detailed settings of PCR are listed below in Table 2.3:
Table 2.3. Settings of PCR.

<table>
<thead>
<tr>
<th>Step</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1</td>
<td>2 min at 95 °C</td>
</tr>
<tr>
<td>Step 2</td>
<td>30 s at 95 °C for denature</td>
</tr>
<tr>
<td>Step 3</td>
<td>30 s at 55 °C, 57 °C, 59 °C, and 61 °C for annealing</td>
</tr>
<tr>
<td>Step 4</td>
<td>4 min at 72 °C for extension</td>
</tr>
<tr>
<td>Step 5</td>
<td>8 min at 72 °C</td>
</tr>
<tr>
<td></td>
<td>Back to 4 °C for product storage.</td>
</tr>
</tbody>
</table>

2.4.2.3 Purification of PCR product

PCR products were purified before sequencing using the following purification protocol.

To the product tubes, 1/10 volume of 3 M ammonium acetate (pH 5.3) and 2–2.5X ethanol (100%, ice-cold) were added. Tubes were placed on ice for 10–20 min, and then centrifuged for 10 min at 16,000 xg, at 4 °C (centrifuge 5415D; Eppendorf, Germany). After centrifugation, the supernatant was carefully removed. The pellet was then washed with 500 µl ethanol (70% (v/v), ice-cold) by centrifugation for 10 min at 16,000 xg, at 4 °C (centrifuge 5415D; Eppendorf, Germany). The supernatant was carefully removed, and the pellet was inverse-dried, and then resuspended in pure water.
2.4.2.4 PCR product checking using agarose gel electrophoresis

After purification, 2 µl samples from each PCR product tube were taken, and then mixed with 0.5 µl of 6X loading buffer (New England Biolabs, USA). The mixture was loaded into a prepared agarose gel to perform electrophoresis (see details of electrophoresis in section 4.1.5). After electrophoresis, the gel was scanned using Gene-Snap software on a G:BOX transilluminator (SynGene, UK).

2.5. Analysis of genes in C. velia

2.5.1. Retrieval of putative LHC sequences

Putative LHC sequences were retrieved from an expressed sequence tag (EST) database produced from a non-motile C. velia culture (Šlapeta and Carter, unpublished data). Sequences in the database were BLAST-searched against the GenBank non-redundant protein sequence database of the National Centre for Biotechnology Information (NCBI) and the Joint Genome Institute (JGI). The candidate LHC genes retrieved from the C. velia ESTs were translated into protein sequences using standard codon and then used for subsequent phylogenetic analysis. The sequences of LHC-related peptides were aligned using ClustalW software and refined manually.

2.5.2. LHC sequence alignment

Homologous sequences of LHC candidates were retrieved by BLAST-searching candidate genes against the GenBank non-redundant protein sequence database (NCBI). Based on the conserved regions indicated by a previous publication (Durnford et al., 1999), an alignment including all putative LHCs and their homologs was constructed using BioEdit software.
2.5.3. Phylogenetic tree construction of LHC sequences

Based on previous research (Durnford et al., 1999), conserved sites within the alignment were extracted using MEGA5 (Tamura et al., 2011). A neighbour-joining (NJ) tree based on distance was constructed using the Dayhoff model in MEGA5 (Tamura et al., 2011). The NJ tree was verified with 1000 replicates. Bootstrap values that supported a node in more than 50% of the replicate trees were retained.

Based on the preliminary results of the NJ tree, representative LHCs from groups close to the LHC homologs identified in C.velia were selected for further phylogenetic analysis. Phylogenetic analysis of the subset of LHCs from C.velia and other organisms selected above was performed using Bayesian and Maximum Likelihood (ML). The ML tree was constructed using PhyML 3.0 (Guindon et al., 2010) using the LG+Γ+I model. The Bayesian phylogenetic analysis was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) using the WAG+Γ+I model. Preliminary trees, tree model selection analyses and tree annotations were conducted using MEGA5 (Tamura et al., 2011).

2.5.4. Retrieval of Calvin-Benson cycle-related sequences

Two C.velia cDNA databases were used to identify possible enzyme coding genes involved in the Calvin-Benson cycle. One was a C.velia EST database (kindly provided by Dr. Jan Šlapeta), which can be accessed from GenBank non-human, non-mouse EST database (est_other) on NCBI. Sequences in this database have been manually cured and can be directly used for sequence searches. Another source was an online dataset of C.velia transcriptomic sequences (Accession Number: SRX090189) from the Sequence Read Archive (SRA) database.
Further signal checks based on homology comparison were performed to validate these sequences.

BlastX was used to search the GenBank non-redundant (nr) protein sequences database and the Joint Genome Institute (JGI) for fishing out candidate genes homologous to known genes in the Calvin-Benson cycle. In order to obtain all closely related sequences, GenBank non-human, non-mouse EST database (est_other) and SRA databases were also searched using the tblastN method. Retrieved nucleotide sequences were translated into protein sequences based on standard codons for further analysis.

2.5.5. Alignment of Calvin-Benson cycle related sequences
Translated protein sequences were aligned with homologous sequences using ClustalW and the alignment was corrected manually. Detailed information of alignment length, involved species and percentage of similarity for each gene is listed in Appendix-2. Conserved regions of each alignment were picked using the program Gblocks (Dereeper et al., 2010)
(http://www.phylogeny.fr/version2_cgi/one_task.cgi?task_type=gblocks).

2.5.6. Phylogenetic tree construction of Calvin-Benson cycle related sequences
Based on the conserved region produced, a preliminary phylogenetic analysis using the Neighbour Joining (NJ) method was performed in MEGA5 software (Tamura et al., 2011). Representatives from groups close to the sequences identified in C.velia were picked to build a subgroup alignment for further analysis. Phylogenetic analysis of the sub-alignment was performed using the Bayesian method.
Sub-alignments were used for calculation in the MEGA 5 software to search for the most suitable model, and WAG+G with 5 discrete gamma categories was inferred as the best choice for all sub-alignments. The Bayesian phylogenetic analysis was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) using the WAG+G model.
Chapter 3. Phylogenetic study of Light harvesting complexes in *Chromera velia*

3.1. Introduction

3.1.1. Antenna protein complexes in oxygenic photosynthetic organism

Antenna systems are accessory pigment-binding protein complexes in photosynthetic organisms. They are associated with reaction centres located in the thylakoid membranes (Kühbrandt, 1994; Green, 2003). With the help of antenna proteins, the reaction centres have an enlarged absorbing area and an increased excitation energy intake. In response to different growth conditions, oxygenic photosynthetic organisms have developed different strategies to adapt to various environments, including the development of many different types of antenna protein complexes (Ferreira and Straus, 1994; Riethman and Sherman, 1988; Green and Durnford, 1996), the changed expression pattern of antenna protein encoding genes (Heddad and Adamska, 2000).

Various antenna protein complexes are presented in a broad range of photosynthetic organisms, including prokaryotic cyanobacteria, eukaryotic algae and the higher plants. They can be clearly classified into several groups with no apparent relationships in term of structure and pigments. Almost all antenna complexes are pigment–proteins, in which the Chl or other pigment is specifically associated with proteins in a unique structure. The only known exception to this rule is the chlorosome antenna complex found in the green photosynthetic bacteria, in which pigment–pigment interactions are of primary importance (Blankenship, 2002).

Antenna complexes can be broadly divided into integral membrane antenna complexes and peripheral membrane antenna complexes. Integral membrane antennas contain proteins that cross the lipid bi-layer. The pigments are often deeply buried in the membrane. In peripheral membrane
antennas, the antenna complex is associated with components buried in the membrane, but itself does not span the membrane.

The integral membrane antenna complexes are still quite diverse in terms of structure. The complexes can be divided into two groups: core antenna complexes and accessory antenna complexes. Core antenna complexes are the antenna systems located within the photosystems. Accessory antenna complexes are antenna proteins surrounding photosystems, which can be regulated directly in response to environmental change.

3.1.2. Peripheral membrane antenna complexes.

Peripheral membrane antenna complexes indirectly bind to the reaction centres. Peripheral membrane antennas include phycobilisomes in red algae, glaucophytes and cyanobacteria; chlorosomes in the green photosynthetic bacteria and peridinin-Chl $a$ proteins in dinoflagellates. Phycobilisome connect to the reaction cores of PSII through a linker protein (Chang et al., 2015). The major pigment proteins in phycobilisome are phycoerythrin, phocyerythrocyanin, phycocyanin and allophycocyanin.

The peridinin-Chl $a$ protein (PCP) is found only in dinoflagellates. PCP contains Chl $a$, Chl $c$, peridinin and diadinoxanthins. In this antenna system, carotenoids are the light harvesting chromophore. Chl $a$’s are also presented in the system, however, they are out-numbered by carotenoids (6~8 carotenoids: 2 Chl $a$). The Chls mainly function as accessory pigments to transfer energy from the carotenoids to membrane intrinsic antennas or reaction centres.
3.1.3. Core antenna complex system

Core antenna complexes are the antenna systems located within the photosystem and bound directly to the reaction centres. As defined by Green and Durnford (1996), this group includes six helices at the N-terminus of PsaA and PsaB; CP43 and CP47; IsiA and PCBs (Prochlorophyte Chl $a/b$-binding proteins). PsaA and PsaB are reaction centers in PSI with a predicted secondary structure containing 11 membrane spanning helices. The inner five helices are structurally close to the reaction centres (D1/D2) in PSII (Zouni et al., 2001), the outer six are close to CP43 and CP47, which are the core antenna connect to reaction centres in PSII. IsiA and PCBs have been discovered only in cyanobacteria so far. IsiA is found in cyanobacteria in response to Fe deprivation (Reithman and Sherman, 1988; Ferreira and Straus, 1994). It was proved to be an accessory light harvesting complex (LHC) for PSI in cyanobacteria (Bibby et al., 2001, 2003). PCBs are the Chl $a/b$ binding proteins in cyanobacteria Prochloron, Prochlorothrix, Prochlorococcus, which are distantly separated from other members of this family (La Roche et al., 1996).

The major differences among CP43, CP47, IsiA and PCBs are the length of the luminal loop between the fifth and sixth helices. CP43 and CP47 have much longer loop compared to the others. Previous research has indicated the connection between loop sequences and the acquisition of water-splitting ability in PSII (Umena et al., 2011).

Unlike other LHC, coding genes of core antenna complex (CP43, CP47) are still located in the chloroplast genome. The secondary structure of core antenna complex consists of 6 helices instead of three. This indicates a closed relationship between core antenna complex and LHC in cyanobacteria.
3.1.4. Accessory antennas

Accessory antennas include LHC surrounding PSI and PSII, as well as LH2 complexes in purple bacteria. They are membrane bound proteins. The major antenna systems in eukaryotic oxygenic photosynthesis organisms include the membrane intrinsic Chl \( a/b \), Chl \( a/c \) and Chl \( a \)-binding protein complexes with three helices in the secondary structure (Figure 3.1), the early light induced proteins (ELIPS) of green algae and plants (Green and Durnford, 1996), and the one helix proteins whose sequences are related to the first and third helices of the three-helix antennas and ELIPS. LHC homologies with two helices and four helices have also been reported recently. Two-helix stress-expressed proteins (Heddad and Adamska, 2000) and four-helix PSBS proteins (Kim et al., 1992; Wedel et al., 1992) are discovered only in higher plants.

According to the binding pigments, the three-helix LHC can be classified to specific species. An example is the fucoxanthin Chl binding proteins (FCP) in diatoms. FCP use fucoxanthin as the major carotenoids and also binds Chl \( a \) and \( c \). As diatoms are the major photosynthetic plankton in the ocean, the FCP must contribute the major role for achieving high photosynthetic efficiency.

Recently, new members in the LHC super-family were reported, and their function and relationship to known photosystems are unclear. These new LHC, designated as LHCsr/LI818 and LHCz, were discovered from the green alga *Chlamydomonas reinhardtii* (Richard et al., 2000) and other algae (Tan et al., 1997; Dittami et al., 2010). These newly discovered LHC sequences form a distinct phylogenetic group within the LHC superfamily (Pan et al., 2011). The LI818 group are stress-induced-proteins that respond directly to various environmental stresses including sulphur starvation (Zhang et al., 2004), phosphorus deprivation (Moseley et al., 2006), iron...
deficiency (Naumann et al., 2007), and high light stress (Zhu and Green, 2010). Peers et al., (2009) reported that LHCsr/LI818 plays an essential role during the LHC assembly processes.

Although the structure of LHC can be varied in different eukaryotic photosynthetic organisms with different numbers of helices, the level of amino acid conservation among these LHC indicates that they may be derived from a common ancestral protein (Neilson and Durnford, 2010). Also, these LHC polypeptides have been found to have a similar membrane topology, with three helices as standard (Green and Durnford, 1996). The crystal structure of the major LHC (LHCII) of pea revealed that the LHC peptides have three helices, which bind both Chls and carotenoids (Kühlbrandt et al., 1994). The LHC with three membrane spanning regions are found only in eukaryotic photosynthetic organisms and share homologous sequences in their membrane domains (Kühlbrandt, 1994; Liu et al., 2004). In a typical three-helix LHC, the first and third helix form an X shape through ionic bridges to bind carotenoids and Chls co-ordinately. The second helix co-operates with helix 1 and helix 3 to bind the extra Chls according to the pigmentation within the organism. Chlorophyll-binding sites are more conserved in helix 1 and 3, with helix 2 showing less homology across different LHC species.

Phylogenetic analysis of eukaryotic LHC polypeptides, along with predicted protein structure analysis and pigment binding sites, has revealed the diversity of LHC (Green, 2003; Neilson and Durnford, 2010). Genes encoding LHC proteins are located in the nucleus, which are different from pigment-binding protein complexes in reaction centers. Most essential proteins in reaction centres are encoded by genes located in the chloroplast (Archibald, 2009). However, genes encoding one-helix proteins are also found in the chloroplast genomes of red algae, glaucophytes, also in the nucleomorph genome in cryoptophytes (Douglas et al., 2001).
3.1.5. Aims

Although the C.velia plastid has been reported to be related to red algal chloroplasts (Janouskovec et al., 2010), it is not known whether genes encoding LHC share a similar evolutionary pathway. In this study, LHC genes are retrieved from expressed sequence tag (EST) sequences of C.velia and their phylogenetic relationship with other photosynthetic organisms are investigated using sequence alignment, structural modelling and further study on phylogenetic relationship.
3.2. Methods

3.2.1. Putative LHC sequences retrieving

Candidate LHC sequences were retrieved from an expressed sequence tag (EST) database produced from a non-motile *C. velia* culture (Šlapeta and Carter, unpublished data). The EST library contains 2856 sequences which can be accessed from the following link (http://www.ncbi.nlm.nih.gov/nucest/?term=chromera%20velia). Sequences in the EST database were blasted against Genbank non-redundant protein sequences database on the National Centre for Biotechnology Information (NCBI) and Joint Genome Institute (JGI). Possible sequences with an expect value threshold less than 10E-15 were selected as candidates. The LHC candidate genes retrieved from the *C. velia* ESTs were translated into 6 open reading frames (ORFs) using the standard genetic codon, in order to find the right reading frame to generate deduced amino acid sequence of putative LHC. The sequences of LHC-related peptides were aligned using ClustalW software and refined manually according to conserved domains based on previous studies (Durnford *et al.*, 1999). The conserved domains represent the three transmembrane helices. The three helices structure is also confirmed using secondary structure prediction software: Jpred 3 (http://www.compbio.dundee.ac.uk/jpred) (Cole *et al.*, 2008).

Transmembrane regions in proteins usually have higher hydrophobicity due to the need to cross the lipid bi-layer membrane. Therefore, transmembrane regions in LHC can also be confirmed using a hydrophobicity calculator. Toppred is an online tool for hydrophobicity calculation (VonHeijne, 1992). Default settings were used for the hydrophobicity check when using candidate LHC sequences as queries. Toppred is available from the link provided below:

http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::toppred
3.2.2. LHC sequences alignment

Homologous sequences of LHC candidates were retrieved by blasting candidate genes against the Genbank non-redundant protein sequences database (NCBI). Sequences with an Expect value threshold less than 10E-15 were selected and used for following total sequences alignment. The total alignment contained 644 sequences, including 24 LHC-related sequences from *C.velia* and 620 LHC-related sequences from other organisms. The latter included 132 sequences from 16 species of green algae; 167 from 53 higher plants; 22 from four species of red algae; three from two species in Raphidophyte; one from *Giraudyopsis stellifer* (Chrysomerophyte); 3 from *Vaucheria litorea* (Xanthophyte); 36 from four species of Haptophyte; 24 from 2 species of cryptomonads; 94 from 7 species of diatoms; 68 from 5 species of brown algae; 14 from 2 species in *Chlorarachniophyte*; 15 from *Euglena gracilis* (Euglenozoa) and 41 from 7 species of dinoflagellates. One sequence from *C.velia* formed a long branch within the *C.velia* group. This was resulted by the different amino acids at the third conserved domains, which is most likely to be a misreading during sequencing. Therefore, it was removed from the final phylogenetic analysis (see details in section 3.3.2 and Figure 3.3).

3.2.3. Phylogenetic tree construction of LHC sequences

Based on previous research (Durnford *et al.*, 1999), conserved sites within the alignment (approximately 139 amino acids in this study) were extracted using MEGA5 (Tamura *et al.*, 2011). A Neighbour-joining (NJ) tree based on distance was constructed using the Dayhoff model in MEGA5 (Tamura *et al.*, 2011) in order to assess the phylogenetic relationship of all of the *C.velia*
LHCs. The NJ tree was verified with 1000 replicates. Bootstrap values that supported a node in more than 50% of the replicate trees were retained.

In order to save time, representative LHCs from different branches were selected for further phylogenetic analysis. The 24 LHC sequences from *C.velia* were clustered into one major branch and several small branches (Figure 3.5). Representatives of the major branches and all members of small branches were selected for further study. The further phylogenetic analysis was based on 23 *C.velia* LHC sequences and 56 LHC sequences from other photosynthetic organisms. Phylogenetic analysis of the subset of LHCs from *C.velia* and other organisms selected above was done using Bayesian and Maximum Likelihood (ML). The ML tree was constructed using PhyML 3.0 (Guindon *et al.*, 2010) using the LG+Γ+I model. The Log likelihood of the best maximum likelihood tree was -12722.4117, alpha = 1.298032, and the proportion of invariant sites = 0.030429. The robustness of the ML tree was evaluated by bootstrapping with 100 replicates using PhyML. The Bayesian phylogenetic analysis was performed using MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) using the WAG+Γ+I model. Metropolis-coupled Markov chain Monte Carlo analyses were run with one cold and three heated chains (temperature set to default 0.2) for 4,000,000 generations and sampled every 100 generations. This process was performed three times from a random starting tree and run well beyond convergence. Trees before convergence (the first 30%) were discarded for the reconstruction of the consensus Bayesian tree with posterior probabilities. The Log likelihood arithmetic mean was -12019.7, harmonic mean = -12775.54, alpha = 1.297 (variance 0.04) and proportion of invariant sites (pinvar) = 0.105 (variance 0.001).

Preliminary trees, tree model selection analyses and tree annotations were conducted using MEGA5 (Tamura *et al.*, 2011).
3.3. Results

3.3.1. LHC sequences in *C. velia*

A total of 24 peptide fragments with homology to LHC sequences were recovered from the *C. velia* ESTs. All of the 24 LHC homologs had a minimum translated peptide of 155 amino acids (aa), with coverage of conserved regions encompassing more than 63% of their full length. Twenty two of the 24 retrieved sequences represented the full length of the predicted LHC protein of 213–259 aa, and two had a truncated sequence. All retrieved LHC homologs were analysed by Jpred3 for determining secondary structure patterns (Cole *et al.*, 2008). Jpred3 calculated the hydrophobicity in query sequences, which indicates possible transmembrane regions. Based on the prediction produced by Jpred3, three likely transmembrane regions exist in all 24 sequences (Figure 3.2). These regions corresponded to the three MSR of the LHC superfamily, which were highly conserved (Figure 3.3). Even the shortest sequence retrieved from *C. velia* (Cv1_05C08; 155 aa), had these three conserved MSR regions (Figure 3.3). Two highly conserved sequence regions known as “retention motifs” (Green and Pichersky, 1994; Jansson, 1999; Hoober *et al.*, 1999) were observed in MSR 1 and 3 of the *C. velia* LHC homologs (Figure 3.3). The first motif, -EXXHXR-, was found in the MSR 1 in six LHC homologs that branched away from the main clade shown by the preliminary phylogenetic analysis (Figure 3.3), while isomer motif, –EXXNXR-, was observed in the MSR 1 in main LHC homologs belonging to the main clade. Those motifs are important for stabilising the LHC structure and for interacting with other protein complexes in the thylakoid membranes. All six of the recognised Chl *a* binding sites within MSR 1 and MSR 3 (Gantt *et al.*, 2003) were found in the LHC homologs: E167, H/N170, R172, G180,
E290, N/H293, R295 and Q307/H308 (Figure 3.4). To compare the retrieved LHC homologs and confirm their putative function, the 24 LHC genes were used to search the LHC superfamily in NCBI database.

According to the LHC EST abundance (Šlapeta and Carter, personal communication), 24 LHC homologs represent only 0.8% of the total EST sequences. Interestingly, three EST sequences, Cv1_Contig242, Cv1_Contig293 and Cv1_contig332, represent the dominant LHC homologs. These three LHC homologs are grouped as sisters in the main group of LHC homologs. The major
LHC sequences represent 0.6% of total EST sequences and 83% of total LHC homologs. (Figure. 3.5).

3.3.2. Phylogenetic analysis of the LHC homologs from *C.velia*

In order to gain a general idea of the relationship of the LHC sequences in *C.velia* and other common organisms, a preliminary tree was constructed. Figure 3.5 shows the phylogenetic tree based on 23 *C.velia* LHC homologs. One *C.velia* sequence (Cv1_14G04) was removed from analysis due to divergence in the MSR3 conserved region (Figure. 3.3), which caused a long-branch with a low bootstrap support (data not shown). As shown on the phylogenetic tree, seventeen sequences formed a main clade that includes three well supported sub-groups. The remaining six *C.velia* LHC sequences diverged from the main clade and formed individual branches.

3.3.3. Phylogenetic relationship of the *C.velia* homologs

The trial phylogenetic tree constructed using 643 LHC homologs (data not shown) indicated that LHC homologs from *C.velia* are closely related to algae containing Chl a/c, but diverged from the Chl a/b-binding LHC of higher plants and green algae. ML and Bayesian analyses were used to construct phylogenetic consensus trees for 66 selected representative LHCs (Figure. 3.5a,b). The ML and Bayesian consensus trees were similar in topology, and both had significant bootstrap support for major clades (Figure. 3.5a,b).

Four major divisions are resolved on the phylogenetic trees, including one containing the Chl a/b-binding proteins (encoded by lhcb and lhca genes), a cluster of LI818/LI818-like proteins, a
fucoxanthin Chl a/c-binding protein (FCP) group, and the red algae and crytomonad LHC group (Figure 3.5). Most of the *C.velia* LHC homologs formed a distinct cluster that grouped with diatoms LHC sequences, *Thalassiosira pseudonana* LHCf11, which is sister related to the fucoxanthin Chl a/c-binding proteins including those from *Macrocytis pyrifera* FCPs (chrysophyceae), *Giraudyopsis stellifer* CAC (diatoms), *Phaeodactylum tricornutum* FCPs and *Heterosigma carterae* FCPs (raphidophytes). Two sequences positioned closed at the low branch of this FCP clade. Figure 3.5 shows that *C.velia* LHC are the only known branch within the FCP group that do not have a fucoxanthin pigment, although *C.velia* does possess a novel carotenoid that has been suggested to be a fucoxanthin isomer (Moore *et al.*, 2008). The high bootstrap and posterior probability values (60–100) suggest that the LHC in these taxa, which include the main LHC clade of *C.velia* identified in Figure 3.4, share a single common ancestor.

Three of 23 LHC homologs found in *C.velia*, Cv1_02D11, Cv1_02E07 and Cv1_contig274, are grouped with a clade containing LHCr from diatoms, which is closely related to the red alga LHCa group. Two LHC homologues, Cv1_05C08 and CV1_contig4, were in a sister relationship to LHC1 from the yellow-green alga, *Vaucheria litorea*, in the early branching group of Chl a/Chl c-binding LHC protein complexes. One LHC homolog, Cv1_contig56, was grouped with LI818/LI818-like protein complexes from various different photosynthetic organisms, strongly suggesting that it potentially functions as an LI818 protein, a unique member in the LHC superfamily. In order to make a direct comparison, all LHC sequences from *C.velia* were aligned with FCP and LI818 proteins, as shown in Figure 3.6a,b.
Figure 3.3 *C. velia* LHC candidate sequences aligned with potential transmembrane regions marked. Three conserved regions were indicated with boxes. *C. velia* 14G04 (arrow indicated) was removed from further analysis due to the fact that differences in MSR 3 lead to an unsupported long branch on the phylogenetic tree. Conserved Chl binding sites and the retention motif are labelled as stars.
Figure 3.4 First dataset alignment with three transmembrane conserved regions (helices). LHC sequences of Pea and Spinach were used as the references to highlight key features in the alignment (Liu et al. 2004). Black star, Chl a binding sites; grey star, Chl b binding sites; grey diamond, Lutein binding sites.
Figure 3.5. Phylogenetic placement of the 23 C.velia LHC homologues with LHC from other algae. a, Neighbour-joining tree; b, Baysian tree. The final alignment included 139 informative amino acid positions.
The numbers above the branches indicate bootstrap support based on 1,000 replicates in the NJ tree and 100 replicates in the ML tree (a), and posterior probabilities 50% (b) on the Baysian tree. The Baysian tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The 23 C. velia LHC sequences are indicated with black dots. Sequences included in this Figure: Raphidophyte: Heterosigma carterae (Hc) Fucoxanthin Chl binding protein (FCP) 1, (Accession NO. Q39969); Brown alga: Macrocystis pyrifera (Mp) FCPa (Accession NO. Q40297) and FCPb (Accession NO. Q40296); Diatoms: Phaeodactylum tricornutum (Pt) FCPc (Accession NO. Q08586) and FCPa (Accession NO. Q08584); Thalassiosira pseudonana (Tp) FCPf1 (Accession NO. B8CFW3) and FCPf2 (Accession NO. B8CEV8); Haptophyte: Isochrysis galbana (Ig) FCP (Accession NO. Q39709) and Euglena gracilis (Eg) LHca3 (Accession NO. A4QPI1); Green algae: Chlamydomonas reinhardtii (Cr) LHCl (Accession NO. Q9FVE3), LHCh2 (Accession NO. Q9ZS15); Chlorarachnion: Bigelowiella natans (Bn) LHCbm3 (Accession NO. Q9LE97) and LHCy1 (Accession NO. A4QPI6); Red alga: Porphyridium cruentum (Pe) LHCa1 (Accession NO. P93449) and LHCa2 (Accession NO. P93450); Guillardia theta (Gt) LHC1 (Accession NO. Q5K288) and LHC2 (Accession NO. Q5K286); Higher plant: Lycopersicon esculentum, (Tomato, Le) LHC II (Lhcb1, Accession NO. P07369) and LHC I (Lhca1, Accession NO. P12360); Arabidopsis thailiana (At) LHC I (Lhca1, Accession NO. P27521) and LHC II (Lhcb4, Accession NO. P04778); Spinacia oleracea (Spinach, So) LHCp (Accession NO. P12333); Pisum sativum (Pea, Ps) LHCI (lhcA-P4, Accession NO. Q9SQL2), LHCII (LHcb, Accession NO. P27490 and LHCa3, Accession NO. Q32904).
Figure 3.6a. The potential LHC sequences from C.velia (Cv) aligned with FCP sequences. Sequences included in this analysis: Haptophyte: Isochrysis galbana (Ig) FCP (Accession NO. Q39709) and an isomer FCP iso14 (Accession NO. Q2IA65); Raphidophyte: Heterosigma carterae (Hc) FCP1 (Accession NO. Q39969); Diatoms: Phaeodactylum tricornutum (Pt) FCPc (Accession NO. Q08586), FCPa (Accession NO. Q08584) and LHCr13 (Accession NO. B7G871); Thalassiosira pseudonana (Tp) FCP1 (Accession NO. B8CFW3), FCP2 (Accession NO. B8CEV8), LHCr, LHCa and LHCr3 (Accession NO. B8C0K4, B8BUU4 and B8C2K6, respectively); Skeletonema costatum (Sc) FCP (Accession NO. P93857); Cyclotella cryptic (Cc) FCP4 (Accession NO. O81932); Brown algae: Macrocystis pyrifera (Mp) FCPa (Accession NO. Q40297), FCPb (Accession NO. Q40296) and FCPE (Accession NO. Q40301); Ectocarpus siliculosus (Es) LHCp1 (Accession NO. D8LB8) and LHCp3 (Accession NO. D8L4R4); Saccharina latissima (Sl) LHCf3 (Accession NO. Q9DFZ6) ; Chroococytic alga: Vaucheria litorea (Vl) LHC1 (Accession NO.Q9ATC7), Red algae: Rhodomonas sp. (Rs) LHCc5 and LHCc6 (Accession NO. Q4GWU4 and Q4GWU3); Porphyridium cruentum (Pe) LHCa1 (Accession NO. P93449); Guillardia theta (Gt) LHCr1 (Accession NO. Q5K288); Griffithsia japonica (Gj) LHC1 (Accession NO. Q7XYZ2) and Galdieria sulphuraria (Gs) LHC4 (Accession NO. Q9FDZ6); Dinoflagellate: Heterocapsa triquetra (Ht) LHC (Accession NO. Q5ENL9) and Pyrocystis lumula (Pl) Chl a/c binding protein (Accession NO. Q8GZE4).
Figure 3.6b. The potential LHC sequences from *C.velia* (Cv) aligned with identified and indicated (by blastX result) LI818 sequences. Sequences included in this analysis: *Micromonas pusilla* (Mp) (strain CCMP1545) LI818 (Accession NO. C1MIN8); *Micromonas sp.* (Ms) (strain RCC299 / NOUM17) LI818 (Accession NO.C1E3Q9); *Micromonas sp.* (Ms) CCMP490 LI818-1 and LI818-2 (Accession NO. A4QPN0 and A4QPN4, respectively); *Bigelowiella natans* (Bn) LI818-1, LI818-2.2 and LI818-3 (Accession NO. Q7XYJ2, A4QPI1 and A4QPI7, respectively); *Mesostigma viride* (Mv) LI818-1 and LI818-2 (Accession NO. A2SY28 and A4QPM49); *Thalassiostra pseudonana* (Tp) LI818-1,2,3,4 (Accession NO. B8CGG0, B8BSG2, B8CGG1 and B8C364, respectively); *Phaeodactylum tricornutum* (Pt) LI818 (Accession NO. B8CGG2); *Dinophysis acuminata* (Dc) LI818 (Accession NO. D918K4); *Gymnochlora stellata* (Gs) (Accession NO. B5A417); *Scenedesmus obliquus* (So) LI818 (Accession NO. A3QQ0); *Chlamydomonas reinhardtii* (Cr) LI818r-1 (Accession NO. P93663); *Cyclotella cryptic* (Cc) LI818 like protein (Accession NO. O81933); *Ostreococcus tauri* (Ot) LI818r (Accession NO.O3B9U1); *Isochrysis galbana* (Ig) LI818 like protein (Accession NO. Q21A70) and *Karlodinium micrum* (Km, Dinoflagellate) LI818 like protein (Accession NO. A7YX8).
3.4. Discussion

3.4.1. The relationship between C.velia and other algae groups indicated by phylogenetic analysis

The existence of multiple copies of LHC homologs and their phylogenetic relationships indicates that the majority of C.velia LHC is related to diatoms (Lepetit et al., 2010). Three different LHC groups could be identified in C.velia using phylogenetic analysis: (1) a major LHC group within the branch containing characterised FCP in diatoms; (2) a group related to the Lhcr proteins; and (3) a single sequence within the LI818 protein group, which contains the ancient Lhcx protein (Figure 3.5).

The light-harvesting strategy used by red algae is different to that used by C.velia, according to their pigment profiles. Extrinsic soluble pigment-binding protein complexes, the phycobilisomes, are the major LHC in red algae, especially for PS II (Gantt et al., 2003). The intrinsic Lhcr mainly services PS I as an accessory LHC. The presence of three distinct LHC groups in C.velia, and their predicted functions based on homology with other LHC sequences, may indicate that the relationship between C.velia and red algae lineage is not as close as what was suggested by Janouskovec et al. (2010). The different evolutionary relationship among LHC in C.velia, diatoms and red algae may reflect the different light-harvesting strategies developed and used by photosynthetic organisms to adapt into their ecological niches.

The two phylogenetic consensus trees indicated that most LHC homologs in C.velia group with Fucoxanthin Chl a/c-binding LHC from Chl c-containing algae, including diatoms, brown algae, chrysophyceae and dinoflagellates, although they occur on a strongly supported branch that is separate from the main group (Figure. 3.5). Fucoxanthin is the major carotenoid in Chl c-containing organisms and is the main carotenoid in FCPs, which is encoded by lhcf genes in diatoms. One of the main carotenoids found in C.velia was a previously unknown carotenoid that was suggested to be an isomer of fucoxanthin and to have similar chemical properties. In this study, the phylogenetic trees indicate that the major clade of LHC in C.velia share similar features to the FCPs in diatoms, which suggests that the new carotenoid found in C.velia may function in a
similar way to fucoxanthin in diatoms, although *C.velia* has no Chl c and only a trace amounts of MgDVP, a chlorophyll c-like pigment. Considering that a putative isomer of fucoxanthin is a major carotenoid and occurs together with violaxanthin in *C.velia* (Moore et al., 2008), it is possible that the LHC in *C.velia* was inherited from early species of Xanthophyceae by several horizontal gene transfers, which is a common feature during LHC evolution (Neilson and Durnford, 2010). Two *C.velia* LHC homologs, CV1_contig4 and CV1_05C08, are positioned as two separated low branches together with *Vaucheria litorea* LHC in the clade of FCPS from Chl c-containing algae.

In the second LHC group identified in *C.velia*, three homologous peptides (Cv1_02D11, Cv1_02E07 and Cv1_contig274), lay within the LHCr group that is found in red algae and diatoms. The name LHCr originally represented LHC from red algae, and has since been extended to include other highly homologous LHC sequences from other algae (Neilson and Durnford, 2010). LHCr binds no other bound accessory Chls except Chl a, which is typically associated with red algae. Pigment analysis in *C.velia* suggested that Chl a is the major photosynthetic pigment.

One LHC sequence, which was grouped in a LI818/LI818-like cluster with a bootstrap support of 85–95%, is suggested to be an LI818-like protein (Figure. 3.5). Phylogenetic analysis suggested it may have a similar role to other known LI818 proteins, functioning mainly as photoprotection under stress conditions (Zhu and Green, 2010; Naumann et al., 2007; Zhang et al., 2004; Moseley et al., 2006). Further biochemical and photophysiological investigations are required to characterise the function of this LI818 homolog.

### 3.4.2. The LHC motifs presented in *C.velia*

Three conserved MSR regions were identified in the *C.velia* LHC alignment, which is consistent with the current LHC model (Kühlbrandt 1994). MSR 1 and 3 interact with each other in the thylakoid membrane, forming a scaffold that helps to stabilize the LHC in the membranes. The “retention motifs”, -EXXH(N)XR- in MSR 1 and MSR 3 are common to all functional LHC. The
MSR1 motif -EXXHXR-, which was found in the six *C. velia* LHC homologs that lay outside the major clade in Figure 3.4, is a common feature of LHCb proteins, while the isomer –EXXNXR-, found in LHC in the major clade, is common to all LHCa proteins (Green and Pichersky, 1994, Jansson, 1999, Hoober *et al.*, 1999). The isomer motif –EXXNXR- is also found in Chl a/c binding LHC in photosynthetic dinoflagellates (Green and Kühlbrandt, 1995; Hiller *et al.*, 1999). There is no phycobiliprotein system in *C. velia*, and there is no known extrinsic antenna system associated with PS II (Gantt *et al.*, 2003). EST sequences represent genes being expressed at the moment that the mRNA was isolated, hence the 23 retrieved LHC can be considered to be functional proteins expressed by actively photosynthesising, immotile *C. velia* cells. Based on their MSR 1 motifs and their relationship to Lhcf of diatoms (Figure 3.5), LHC peptides in the major clade seen on Figure 3.3 may function as LHCII associated to PS II. This proposed function is consistent with the function of Lhcf in diatoms (Lepetit *et al.*, 2010). The other LHC that contain the -EXXHXXR- motif in MSR1 and group with the Lhcr of red algae (Figure. 3.5) may function as LHCI, which mainly associates with PS I. This assumption is in agreement with the Lhcr function in diatoms, as it is only found in isolated PS I complexes (Veith *et al.*, 2009).

Motif –EXXNXR- in MSR 3 is another retention motif that is recognised to be important for LHCII assembly (Flachmann and Kühlbrandt, 1996). The *C. velia* sequence Cv1_14G04 was removed from further phylogenetic analysis because it lacked this important motif in MSR3, which resulted in it residing on an unsupported long branch away from all other LHC homologs. All 22 other homologs had this motif in MSR 3. The Cv1_02D11 sequence contains a related isomer motif: -EXXHXR-. It is uncertain whether this motif has the same functions as the usual MSR3 motif.

3.4.3. The evolutionary relationship of LHC in *C. velia*

Eukaryotic photosynthetic organisms are believed to originate from an accidental endosymbiosis event which involves a eukaryotic host and a prokaryotic photosynthetic organism. During the long process of endosymbiosis, part of the photosynthetic prokaryotes genes were moved to the
nucleus, and then integrated into the host genomes. Some genes, including a major part of the LHC coding genes, have migrated to the nucleus but still function within the prokaryotes (Archibald, 2009). Their sequences have been modified to attach a signal peptide for transport across additional membranes. Endosymbiosis has separated the location of LHC transcription and function, and completely changed the LHC structure in eukaryotes. Combined with other genetic modification methods, it has led to a great diversity of LHC presented in eukaryotic photosynthetic organisms.

With so many types of LHC in existence, the evolutionary relationships between them are debatable. The discovery of four-helix PSBS is of great help. In PSBS, the first pairs of helices are highly related to the second pair of helices, suggesting that a two-helix ancestor might undergo tandem gene duplication, and then fused the duplicates into one sequences with four helices. The four-helix product may lose one helix later, which would lead to the ancestor of three-helix LHC (Green and Kühlbrandt, 1995). In consideration of the wide-spread of one-helix LHC, the two-helix LHC might also originate from a fusion of a single-helix High light induced protein (HLIP) with another sequence encoding a potential membrane spanning helix. This model is later supported by the fact that one-helix and two-helix proteins of LHC were found in the Arabidopsis genome (Jansson et al., 2000; Heddad and Adamska, 2000).

Also, three-helix LHC have not been discovered in glaucophytes up to date. Glaucophytes are considered as the possible first photosynthetic eukaryotes branch from the ancestor of all photosynthetic eukaryotes (Baldau et al., 2000; Moreira et al., 2000). Glaucophytes plastid genome carries only one-helix HILP. There are reports of the discovery of one-helix HLIP in the nucleus genome of higher plants (Jansson et al., 2000; Heddad and Adamska, 2000). Therefore, one-helix LHC most likely have migrated to the nucleus after glaophytes, in the common ancestor of red and green algae. After that, the one-helix LHC underwent a series of fusion and duplication events to become the ancestral three-helix LHC (Bryant, 1992; Tomitani et al., 1999). Therefore, as the latest form of LHC evolution, three-helix LHC has only been discovered in the nuclear genome up to date.
3.5. Conclusion

In this study, 23 LHC homologs were successfully retrieved from EST sequences. The structure prediction tools indicate that LHC in *C.velia* is likely to have three helices with typical binding motifs for Chl *a*.

Based on the phylogenetic analysis, LHC candidates in *C.velia* can be separated into four different clades. The major clade included 17 LHC sequences that were closely related to FCP and may function as LHCII for PS II. Two LHC sequences were grouped in a clade with *Vaucheria litoria*, with a close relationship to FCP. Three LHC sequences were grouped into the Lhcr clade and may function as LHCI for PS I. One LHC showed high homology to the LI818/LHCz group. The phylogenetic analysis in this study suggests that in terms of LHC evolution, *C.velia* has a close relationship with diatoms and the heterokonts group, but it is also connected to the red algae.
Chapter 4. Phylogenetic analysis of Calvin-Benson cycle related enzymes in C. velia

4.1. Introduction

The Calvin-Benson cycle includes a series of biochemical reactions that take place in the stroma of chloroplasts in photosynthetic organisms. It is central to carbohydrate synthesis in photoautotrophs. The Calvin-Benson cycle uses the energy generated from light-dependent reactions to convert CO$_2$ into organic compounds, but it is essentially light independent. There are 13 reaction steps in the Calvin-Benson cycle, involving ten enzymes. The Calvin-Benson cycle can be divided into three phases: carbon fixation, reduction, and Ru1,5P$_2$ (ribulose-1,5-bisphosphate; Ru1,5P$_2$) regeneration (Figure 4.1). In phase I, a CO$_2$ molecule is incorporated into the 5-carbon compound Ru1,5P$_2$, producing the unstable intermediate 2-carboxy-3-keto-arabinitol-1,5-bisphosphate. This intermediate has a very short life and immediately breaks down into two stable 3-carbon compounds (3-phosphoglycerate; 3PGA) (Farazdaghi, 2009). Thus, three CO$_2$ are fixed at the beginning of the Calvin-Benson cycle, resulting in glyceraldehyde-3-phosphate (GA3P) being exported to the cytosol (Figure 4.1). Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) is the enzyme in charge of CO$_2$ incorporation (Campbell and Reece, 2007).

Phase II of the Calvin-Benson cycle involves reduction of 3PGA to 1,3-bisphosphoglycerate (1,3-BPGA). The final product of phase II also is GA3P. Thus, each CO$_2$ incorporation into Ru1,5P$_2$ produces two 3PGA (phase I), while six 3PGA molecules (3 CO$_2$ + 3 Ru1,5P$_2$) are used for the production of six 1,3-BPGA (phase II). This reaction is catalysed by phosphoglycerate kinase (PGK), followed by further reduction of 1,3-BPGA to GA3P. This phase of the Calvin-Benson cycle uses the energy generated from the light-dependent reactions of photosynthesis (Campbell and Reece, 2007).

Phase III of the Calvin-Benson cycle involves regeneration of Ru1,5P$_2$. There are ten steps that convert GA3P to Ru1,5P$_2$, involving seven enzymes in this complicated regeneration process. Of the Six GA3P that enter the third phase of the Calvin-Benson cycle, five are recycled back to
regenerate three Ru1,5P. Only one GA3P is exported to the cytosol for cellular metabolism. At the completion phase III, Ru1,5P is recycled for the next round of the Calvin-Benson cycle (Russell et al., 2010). Many of the enzymes involved in the third phase have shared functions in the glycolysis and gluconeogenesis metabolic pathways.

4.1.1. Evolutionary development of enzymes in Calvin-Benson cycle
In algae, most of genes encoding enzymes taking part in the Calvin-Benson cycle have been migrated into the nucleus, with the exception of large subunit RuBisCO. These genes have merged with the genome and function well in the nucleus, while still regulating and governing the Calvin-Benson cycle in the chloroplast. All enzymes involved the Calvin-Benson cycle are targeted at chloroplasts. According to the theory of chloroplast evolution (see section 1.3), these enzymes should be homologous to those in cyanobacteria.

Some enzymes in the Calvin-Benson cycle are functionally equivalent to the enzymes used in other metabolic pathways, such as glycolysis, gluconeogenesis, and the pentose phosphate pathway, which are central to cellular metabolism in the cytosol (Martin and Schnarrenberger, 1997). For example, PGK and glyceraldehyde 3-phosphate dehydrogenase (GAPDH) also work in the glycolysis pathway; while fructose-bisphosphate aldolase (FBA) and fructose-1,6-bisphosphatase (FBP) also catalyse reactions in gluconeogenesis pathway (Figure 4.1). Although most of these enzymes have encoded genes that have migrated to the nucleus during the evolutionary development, some enzymes still remain as copies in the chloroplast. These enzymes usually have two copies, one functions in the chloroplast, and the other functions in the cytosol. Although these gene copies catalyse the same reactions in different organelles, they have structural differences.

According to the evolutionary theory for chloroplast (see section 1.3), the enzymes targeted in the chloroplast should be homologous to those in cyanobacteria, while the enzymes involved in gluconeogenesis and glycolysis pathways should be homologous to those in proteobacteria. However, evolution of metabolic pathways is very complicated given all the possible events, involving lateral gene transfer, gene duplication, and replacement. Clearly, some genes have
inherited one copy from a single source (either chloroplast or mitochondria); some have
duplicated copies, after inheriting from a single source; some genes have kept copies from
different sources and have distinct functions; while some have inherited copies from different
sources during multiple endosymbiosis and lateral gene transfer events.

4.2. Aim

Because the Calvin-Benson cycle converts inorganic carbon into organic compounds using light
ergy, it is an important pathway for photoautotrophic organisms. In contrast, glycolysis and
gluconeogenesis are essential to cellular metabolism, vital for both photoautotrophic and
heterotrophic organisms.

Since *C.velia* is reported to favour mixotrophic growth conditions (Forster *et al.*, 2014), it may
possess the capability of two different life styles, under certain growth conditions. Thus, *C.velia*
represents an excellent opportunity for investigating the development of autotrophic *versus*
heterotrophic life styles, due to its capacity to undergo a metabolic switch under different
environments. The phylogenetic analysis of proteins in the Calvin-Benson cycle, and investigation
of their relationships with their corresponding enzymes in both glycolysis and glucogenesis
pathways, will be helpful to further understand the evolutionary position of *C.velia*. Again, study
of *C.velia* —the closest photosynthetic relative of Apicomplexa, might shed some light on the
evolutionary relationship between photoautotrophic and parasitic living styles. This study may
also advance our knowledge of the structural evolution of chloroplast and apicoplast. Determining
the origin of enzymes in the Calvin-Benson cycle may also benefit our understanding of the
Calvin-Benson cycle, as a possible replacement of the heterotrophic style metabolic pathway.

Previous studies have focused on characterising individual enzymes in *C.velia*, such as RuBisCO
(Janouškovec *et al.*, 2010) and chloroplastic GAPDH (Takarshita *et al.*, 2009). The study of
RuBisCO revealed that type II RuBisCO exists in *C.velia*, similar to that in apicomplexans, and
possibly derived from proteobacteria (Janouškovec *et al.*, 2010). The phylogenetic study of
GAPDH in *C.velia* has suggested that multiple lateral gene transfers have led to the unique type
GapC in the chromalveolate group (Takishita *et al.*, 2009).
In this study, for the first time, a systematic study of the enzymes in the Calvin-Benson cycle in *C.velia* was carried out. Enzymes encoding genes involved in the Calvin-Benson cycle were retrieved from expressed sequence tag (EST) and the transcriptome database on NCBI (http://www.ncbi.nlm.nih.gov/sra). Through comparing these genes with their homologues, a better understanding of the evolution of *C.velia* and its life style was achieved.

Figure 4.1 Enzymes of the Calvin cycle in spinach chloroplasts and their cytosolic homologues. Enzymes regulated through the thioredoxin system are indicated. Suggested evolutionary origins for the nuclear genes are color-coded. Substrate/product abbreviations are: Ru1,5P2: ribulose-1,5-bisphosphate; 3PGA: 3-phosphoglycerate; 1,3BPGA: 1,3-biphosphoglycerate; GA3P: glyceraldehyde-3-phosphate; DHAP: dihydroxyacetone phosphate; F1,6P2: fructose-1,6-bisphosphate; F6P: fructose-6-phosphate; E4P erythrose-4-phosphate; Xu5P: xylulose-5-phosphate; Su1,7P2: sedoheptulose-1,7-bisphosphate; Su7P: sedoheptulose-7-phosphate; R5P: ribose-5-phosphate; Ru5P: ribulose-5-phosphate. Open arrowheads indicate transport rather than conversion. Figure is modified from Martin and Schnarrenberger (Martin and Schnarrenberger, 1997)

4.3. Enzymes in the Calvin-Benson cycle

4.3.1. Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO):

RuBisCO catalyses the reaction described as the following equation:
2 3-phospho-D-glycerate + 2 H⁺ ↔ D-ribulose 1,5-bisphosphate + CO₂ + H₂O

RuBisCO catalyses the fixation of CO₂ in phase I of the Calvin-Benson cycle (Figure 4.1). It assists the incorporation of a CO₂ molecule into Ru1,2P₂, forming an unstable product, which immediately breaks down into two 3PGA. RuBisCO consists of two types of subunits, known as the large chain and the small chain (Yoon et al., 2001). The gene encoding the large chain protein is part of the chloroplast DNA molecule in plants and in most algae. The gene encoding the small chain protein is usually located in the nucleus. The synthesised small chain protein in the cytosol is imported to the stromal compartment of chloroplasts by crossing the outer chloroplast membrane.

RuBisCO is the speed limiting step in the Calvin-Benson Cycle and its functional efficiency is affected by many environmental factors, e.g., light intensity, CO₂ concentration, oxygen concentration, etc. (Ellis, 2010). Thus, it is a key target gene for increasing crop yield in genetic engineering (Whitney et al., 2011).

There are four known types of Rubisco found in nature: classes I, II, III, and IV. Class I RuBisCO is about 560 kDa and made up of eight large subunits and eight small subunits. Most organisms encode class I RuBisCO in their chloroplast DNA or cyanobacterial genome, although higher plants usually transfer the small subunits gene into the nucleus for regulation. Class I RuBisCO is most common in photosynthetic eukaryotes and cyanobacteria. Phylogenetic analysis has further subdivided this enzyme into two subbranches: RuBisCO in red algae is classified as red type, while the enzyme in green algae, plants and cyanobacteria is classified as green type (Tabita et al., 2007). These two types of RuBisCO do share 50–60% sequence identity; therefore, it is most likely that the ancestors of green and red algae possessed a common gene, which was duplicated and lost differentially in the two separate lines (Martin and Schnarrenberger, 1997)

Class II RuBisCO consists of 2 to 8 large subunits only, and is normally encoded in the nucleus (Morse et al., 1995). Class II RuBisCO is common in proteobacteria, dinoflagellates and C.velia (Janouskovec et al., 2011). Unlike Class I RubisCO, the gene encoding Class II RuBisCO is located in the nucleus. This infers a lateral gene transfer occurred in the ancestor of
dinoflagellates, in which the original RuBisCO was replaced by the proteobacterial type (Whitney et al., 1995).

Class III RuBisCO is only found in some archaea and is comprised of dimers of large subunits; either a single dimer or a five dimer-ring are formed (Tabita, 2007). In addition, a novel type of putative RuBisCO was discovered in the green sulphur phototrophic bacterium *Chlorobium tepidum*, having distinct features. This enzyme has proved to be incapable of catalysing Ru1,5P2-dependent CO2 fixation. Therefore, class IV RuBisCO is also known as RuBisCO-like-proteins (Tabita et al., 2007). Class IV RuBisCO consists of large subunits only, occurring in a dimer form, with a parallel structure (Li et al., 2005).

4.3.2. 3-phosphoglycerate kinase (PGK)

PGK catalyses the reaction described as the following equation:

3-phospho-D-glycerate + ATP ↔ 1,3-bisphospho-D-glycerate + ADP

PGK is an enzyme that works in both the cytosol and chloroplast. In the chloroplast, it works in the Calvin-Benson cycle, having a molecular weight of approximately 41 kDa. It catalyses the phosphorylation of 3PGA, producing 1,3-bisphosphoglycerate (1,3-BPGA) and ADP. In the cytosol, it works in the glycolysis and gluconeogenesis pathways. In the glycolysis pathway, PGK catalyses the reversible transfer of a phosphate group from 1,3-BPGA to ADP, forming the products 3PGA and ATP (Figure 4.1). In the gluconeogenesis pathway, PGK catalyses similar reactions as in the Calvin-Benson cycle. The cytosolic copy of PGK is similar to the chloroplast copy, but with slightly smaller molecular weight (40.7 kDa) (Köpke-Secundo et al., 1990).

Higher plants have two copies of PGK, corresponding to those of chloroplast and cytosol, respectively. However, in the green algae *Chlamydomonas*, there is only one chloroplast copy (Schnarrenberger et al., 1990), shared by two different metabolic pathways. Based on phylogenetic analysis, PGK in red alga and its derivates has branched separately from green algae and higher plants group (Archibald et al., 2003). PGK from cyanobacteria are close to red and green algae as well as plants (Kaneko et al., 1996), which suggests that ancestors of red and green algae possess a common PGK, most likely derived from cyanobacteria. PGK might be duplicated
during the evolution of plants. McFadden et al. revealed that PGK in apicomplexan parasites has a different type from that in cyanobacteria and higher plants, which is likely derived from proteobacteria. Is this a similar case to RuBisCO, where a lateral gene transfer happened in the ancestor of dinoflagellate and apicomplexa? The answer is still not known.

4.3.3. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH)

GAPDH catalyses the conversion of GA3P to D-glycerate 1,3-bisphosphate in phase II of the Calvin-Benson cycle. GAPDH also catalyses this reaction in the cytosol and chloroplast. In the cytosol, GAPDH catalyses a step in glycolysis, in which glucose is broken down to synthesise energy and organic molecules. In the chloroplast, GAPDH catalyses the reduction of 1,3BPGA to GA3P, a central step in phase II of the Calvin-Benson cycle (Figure 4.1).

Currently, three types of GAPDH are known that share only 15–20% identity in their protein sequences. One type of GAPDH, known as GapA, exists in cyanobacteria, green algae, red algae and plants (Petersen et al., 2006). GapA is associated with a redox-sensitive protein (CP12) from PSII for the benefit of regulation. The CP12 protein acts as a linker, essential to the assembly of a core complex of PRK/GAPDH. This complex coordinates the reversible inactivation of chloroplast enzymes GAPDH and PRK during darkness in photosynthetic tissues (Wedel et al., 1997). Higher plants and some green algae have an extra copy of GAPDH, which is different from GapA. This extra copy is known as GapB. GapB has its own regulatory mechanism, directly governed by the light-dark cycle. In the chromalveolates group, a third type of GAPDH is found, known as GapC. GapC is more likely derived from the cytosolic form, being only 48% to 61% homologous to GapA (Harper and Keeling, 2003; Liaud et al., 2000; Figge et al., 1999). GapC is widely spread within the chromalveolate group. Therefore, it is often used as evidence to support the chromalveolate evolutionary hypothesis (Harper and Keeling, 2003).
4.3.4. Triosephosphate isomerase (TPI)

TPI is an enzyme that catalyses the reversible inter-conversion of the triose phosphate isomers dihydroxyacetone phosphate (DHAP) and GA3P, in the initial step of the phase III of the Calvin-Benson cycle (Figure 4.1). In animal and other heterotrophic eukaryotes, the TPI enzyme is localized in the cytosol, and involved in glycolysis metabolism only. In photosynthetic eukaryotes, TPI also exists in the chloroplast, and functions in the step converting GA3P to DHAP in the Calvin-Benson cycle (Figure 4.1).

Based on phylogenetic analysis, cytosolic TPI in eukaryotes is likely descended from a homolog in bacteria, excluding cyanobacteria (Keeling and Doolittle, 1997). The chloroplastic TPI is more likely to be a duplicate of the cytosolic copy, rather than a copy inherited from prokaryotic cyanobacteria (Schmidt et al., 1995).

However, in some green and red algae including Chlamydomonas reinhardtii and Cyanidioschyzon merolae, only a single chloroplastic TPI enzyme was detected (Klein, 1986; Matsuzaki et al., 2004; Schnarrenberger et al., 1990). In contrast, there are three copies of TPI in diatoms, one located in the cytosol, one in the chloroplast, and an extra copy imported into the mitochondria as a fusion protein of GapC and TPI (Liaud et al., 2000). The reason for the variable copies of TPI in nature is still under investigation.

4.3.5. Fructose-bisphosphate aldolase (FBA)

FBA catalyses the reaction described as the following equation:

\[
\text{D-fructose 1,6-bisphosphate} \leftrightarrow \text{dihydroxyacetone phosphate} + \text{D-glyceraldehyde 3-phosphate}
\]

FBA catalyses a reversible reaction that splits the fructose 1,6-bisphosphate (F1,6P2) into DHAP and GA3P. The FBA enzyme is involved in the gluconeogenesis and glycolytic pathway, as well as the Calvin-Benson cycle. In the gluconeogenesis pathway, FBA catalyses the conversion of phosphoenolpyruvate (PEP) to F1,6P2. In glycolysis, FBA catalyses the conversion of F1,6P2 to PEP; this is the reverse reaction to the glucogenesis pathway. In the Calvin-Benson cycle, FBA
catalyses the conversion of 3PGA to F1,6P2, which is a different reaction from its function in glycolysis and gluconeogenesis pathway.

FBA is the enzyme with the most complicated evolutionary origin in the Calvin-Benson cycle. There are two classes of FBA, with little sequence similarities. In most oxygenic photosynthetic eukaryotes (red algae, green algae and plants), two copies of FBA exist, one is the chloroplastic copy and the other is the cytosolic copy. These copies of FBA belong to class I, but they form two distinct branches, thought to be the result of gene duplication at the stage before organelle differentiation (Gross et al., 1999; Kruger and Schnarrenberger, 1983). Class I FBA was also detected in the cytosol of apicomplexans and cyanobacteria. Class II FBA mainly exists in bacteria, including cyanobacteria, but diverged into two distinct branches, known as type A and type B. Besides bacteria, type A FBA is also found in fungi, euglena and diatoms (Marsh and Lebherz, 1992; Pelzer-Reith et al., 1994; Plaumann et al., 1997; Rutter, 1964).

Recent reports indicate that type B Class II FBA is diversely presented in the plastid of glaucophyte, cyanobacteria, bacteria, and the cytosol of green algae descendants (Gross et al., 1994; Henze et al., 1998; Nickol et al., 2000; Sanchez et al., 2002). Later, diatoms were found to contain five FBA genes, including both Class I and Class II FBA in their cytosol and chloroplast, with different functions in carbon concentration metabolism, nutrient and light response (Allen et al., 2012).

It is most likely that in the chloroplast, the original class II FBA was replaced by class I FBA in green and red algae, after glaucophytes diverged (Rogers and Keeling, 2004). However, in the chromalveolate group, type A class II FBA exists rather than the more common class I FBA in red/green algae and plants. Considering that the chromalveolate group originates from red algae, the presence of type A class II FBA in the chromalveolate group (instead of class I FBA) is very interesting. This unique presence of type A class II FBA has been used as evidence to support single endosymbiosis leading to the evolution of the chromalveolate group (Patron et al., 2004). More recently, Baurain et al. have argued for a serial endosymbiosis model, based on statistical analyses of phylogenomic data (Baurain et al., 2010; Keeling, 2013).
4.3.6. Fructose-1,6-bisphosphatase (FBP) and sedoheptulose-1,7-bisphosphatase (SBP)

FBP is an enzyme that converts F1,6P2 to fructose 6-phosphate (F6P) in the gluconeogenesis pathway and the Calvin-Benson cycle. SBP is an enzyme with a similar function to FBP, but it is uniquely present in plants and uses sedoheptulose-1,7-bisphosphate (Su1,7P2) as its substrate. SBP cleaves Su1,7P2 into sedoheptulose-7-phosphate (Su7P), releasing an inorganic phosphate ion into solution.

In bacteria, FBP is encoded by a single gene, with dual specificity for corresponding substrates (Tamoi et al., 1996; Yoo and Bowien, 1995). In higher plants and algae, there are two distinct genes encoding FBP and SBP, respectively. FBP in green algae, red algae and plants are different from those in cyanobacteria; they are more closely related to the cytosolic type, which originated from bacteria. FBP found in the apicomplexan group is clustered with bacterial FBP in phylogenetic analyses (Rogers and Keeling, 2004). The most possible scenario describing the evolution of this enzyme would be similar to class I FBA described above. In chloroplast, original cyanobacterial FBP was likely replaced by the duplication of the cytosolic copy of FBP (Martin et al., 1996).

In plants and green algae, SBP is only found in the chloroplast, without a corresponding cytosolic copy. SBP homologous have been reported in non-photosynthetic fungi. Given that cyanobacteria use a dual FBP/SBP enzyme, green algae and plants appear to have bypassed the original cyanobacteria enzyme and obtained their SBP from elsewhere. Based on its phylogenetic analysis, SBP seems to have appeared later (Rogers and Keeling, 2004).

4.3.7. Transketolase (TKL)

TKL functions in the chloroplast (Calvin-Benson cycle) as well as in the cytosol (oxidative pentose phosphate pathway in higher plant). In the Calvin-Benson cycle, TKL catalyses the conversion of Su7P and GA3P to two pentoses: R5P and Xu5P. In the cytosol, TKL catalyses the conversion of Xu5P and R5P into S7P and GA3P—the reverse reaction of the one in the chloroplast. In higher plants, TKL copies have been detected in both the cytosol and the
chloroplast, except in Spinach, which only possess a chloroplastic copy (Flechner et al. 1996). The cytosolic and chloroplastic copies share 77% sequence identity (Flechner et al., 1996). TKL has also been detected in the cyanobacterium *Synechocystis*, as well as bacteria (Kaneko et al., 1996).

TKL sequences are composed of two multiple function domains. The N terminal contains a thiamine diphosphate binding domain and a pyrimidine binding domain. The sequences of the C terminal are conserved between a group of enzymes, including the C terminal of pyruvate dehydrogenase E1 component, branched-chain alpha-keto acid dehydrogenases, as well as domain II of pyruvate-ferrodoxin oxidoreductase. Based on phylogenetic analysis, it is clear that both cytosolic and chloroplastic TKL in higher plants are branched together with bacteria (Flechner et al., 1996). Surprisingly, TKL in plants share 65–70% sequence identity with bacteria, which indicates that TKL was conserved during evolution. Thus, the most likely scenario for the TKL evolutionary pathway is that TKL in plants (both cytosolic and chloroplastic copies) were derived via duplication of an ancestor originating from bacteria.

4.3.8. Ribulose-5-phosphate 3-epimerase (RPE)

RPE, also known as phosphopentose epimerase, is an enzyme that catalyzes the inter-conversion between R5P and Xu5P. RPE functions in both cytosol and chloroplast. It has been detected in animals, bacteria, and higher plants (Karmali et al., 1983; Kusian et al., 1992; Nowitzki et al., 1995; Teige et al., 1995). Recently, RPE has been discovered in the apicomplexa *Plasmodium*. RPE is an exciting potential target for drug design (Caruthers et al., 2006). From the few RPE sequences detected to date, chloroplastic RPE in plants is closest to that of cyanobacterium *Synechosystis*, suggesting that plants inherited this gene from cyanobacteria (Wise et al., 2004). In contrast, cytosolic RPE is reported to be closer to animals and yeast, which indicates the cytosolic copy of RPE comes from proteobacteria (Kopriva et al., 1999). It is most probable that cytosolic and chloroplastic RPE are derived from proteobacteria and cyanobacteria, respectively.
4.3.9. Ribose-5-Phosphate Isomerase (RPI)

RPI is an enzyme that catalyzes the conversion between R5P and Ru5P. RPI plays a vital role in biochemical metabolism in both the pentose phosphate pathway and the Calvin-Benson cycle. Currently, only a few RPI sequences have been reported, including a chloroplastic copy from spinach (Schnarrenberger et al., 1995), and two copies from E. coli (Zhang et al., 2003a; Zhang et al., 2003b). It is assumed that one of the copies of RPI from E. coli (named as RPIA) is common to most organisms.

RPI exists as two distinct protein forms, termed RPIA and RPIB, which show no sequence or overall structural homology. RPIA has a molecule weight of 25 KDa. It is common and highly conserved in bacteria, plants, and animals. RPIB is present in bacteria, with a main function role as is to activate as a backup, or in response to specific circumstances (Zhang et al., 2003b). In the pentose phosphate pathway, RPIA converts Ru5P to R5P, which can be incorporated into the glycolysis pathway. In the Calvin-Benson cycle, RPIA catalyses the reverse reaction, generating Ru5P from R5P. Ru5P is converted into Ru1,5P₂, which is the molecule used for CO₂ fixation in the next cycle of the Calvin-Benson cycle.

Recently, RpiA is reported to be vital to the apicomplexan parasite Plasmodium. It functions in the cytosolic pentose phosphate pathway, which provides a large amount of NADPH for rapid DNA synthesis in the parasite, as well as degradation of hemoglobin. RpiA could be a potential target for drug design against malaria (Becker et al., 2003).

4.3.10. Phosphoribulokinase (PRK)

PRK catalyses the reaction described as the following equation:

\[ \text{ATP} + \text{D-ribulose 5-phosphate} \leftrightarrow \text{ADP} + \text{D-ribulose 1,5-bisphosphate} \]

PRK is an enzyme that catalyzes the reversible conversion of R5P and Ru1,5P₂. PRK and RuBisCO are the only two enzymes which work exclusively in the Calvin-Benson cycle. The PRK family can be divided into two distinct classes, with only ~20% protein sequence identity (Martin and Schnarrenberger, 1997). Class I PRK is derived from proteobacteria, while class II PRK is derived from cyanobacteria. PRK has been detected in several organisms, including green
algae, higher plants, and algae groups, which were involved after secondary/tertiary endosymbiosis (Milanez and Mural, 1988; Raines et al., 1989; Roesler and Ogren, 1990; Horsnell and Raines, 1991; Archibald et al., 2003). All of these PRK have a close relationship to cyanobacterial PRK genes, indicating that these genes have been transferred to the host cell nucleus after primary endosymbiosis and have never changed back (Martin and Schnarrenberger, 1997). However, during the long history of evolution, green and red algae differentiated their PRK sequences, which led to two subgroups of red and green type PRK. Only green type PRK has been detected in chromalveolates, rather than red type, which indicates that a possible horizontal gene transfer event happened in the ancestor of this chromalveolate group. This supports hypothesis that a single secondary endosymbiosis led to formation of the chromalveolate group (Petersen et al., 2006).

In higher plants, PRK is combined with GAPDH (see section 4.1.3) and the small chloroplast protein CP12 to form a complex in the dark, which regulates enzyme activity quickly and efficiently against light intensity change, as well as CO₂ flux (Wedel et al., 1997; Wedel and Soll, 1998; Maberly et al., 2010). However, in chromalveolates, the CP12 protein is missing, and GAPDH has evolved from an ancestral cytosolic copy (GapC, instead of Gap A/B in higher plants), which results in a high expression of PRK and less sensitivity of redox control in some chromalveolate species (Maberly et al., 2010).

4.4. Method

There are 13 enzymes listed in the Calvin-Benson cycle. Candidate gene sequences encoding for these enzymes have been retrieved from the EST library and aligned individually with their homologues (see section 2.5.5). Because the RuBisCO large subunit gene is encoded by the chloroplast genome, it does not appear in the cDNA dataset used in this study. Among all retrieved genes related to the Calvin-Benson cycle in *C. velia*, three enzymes have two copies, and 11 have only one copy.

The alignments indicated that the candidate gene sequences for TKL, PRK and PGK retrieved from *C. velia* were incomplete. TKL had two small sequence copies, with lengths of 195 and 205
aa, corresponding to N and C terminals, respectively. The PRK sequence had a length of 253 aa, but lacks the C terminal; while PGK had two sequence copies, with lengths 245 and 143 aa, corresponding to N and C terminals, respectively.

In this case, PCR primers were designed to sequence these genes, and to determine their actual lengths. To obtain their full lengths, primers were designed based on conserved terminal regions inferred by their homologous sequences. PCR amplification was performed on the complete sequence from C. velia. The primers designed for amplification are listed in Appendix-4. The details for PCR amplification are given in Section 2.4.

The PCR products were purified and sequenced using Macrogen Co service (South Korea). The sequencing results were checked manually and compared against the recently published C. velia genome database. The verified sequences were compared with homologues for phylogenetic construction. The full-length alignments were manually tailored, according to the lengths of candidate sequences for C. velia. The details of these alignments are listed in the Table 4.1 below.

<table>
<thead>
<tr>
<th>C. velia Sequence names</th>
<th>Putative encoding enzymes in the Calvin-Benson cycle</th>
<th>Length of retrieved sequences (aa)</th>
<th>Length of homolog sequences (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV23C07</td>
<td>FBA</td>
<td>258</td>
<td>300–380</td>
</tr>
<tr>
<td>CV1 18F03</td>
<td>RPI</td>
<td>193</td>
<td>180–300</td>
</tr>
<tr>
<td>CV1 contig221</td>
<td>FBP</td>
<td>153</td>
<td>120–400</td>
</tr>
<tr>
<td>CV1 contig51</td>
<td>GAPDH (cytosolic)</td>
<td>299</td>
<td>170–330</td>
</tr>
<tr>
<td>CV1 01A04</td>
<td>TPI</td>
<td>156</td>
<td>150–260</td>
</tr>
</tbody>
</table>

Given the fact that some gene copies of the same enzyme may possess little or no sequence identity, separate phylogenetic tree construction was carried out for all genes in this study, rather than on concatenated multi-gene phylogenies. The phylogenetic tree was constructed, using the same method described in section 2.5.6. The full sequences for TKL, PRK and PGK, together with other sequences retrieved from the database are listed in Appendix-5.
4.5. Results

4.5.1. Phylogenetic analysis of genes involved in the Calvin-Benson cycle

As shown in Figure 4.1, PGK, GAPDH, TPI, FBA and FBP are the essential enzymes for the Calvin-Benson cycle in chloroplast, as well as glycolysis and glucogenesis pathways in the cytosol. The PGK enzyme is involved in the carbon fixation process, catalysing the synthesis of 1,3BPGA, while in the cytosol it helps metabolize 1,3BPGA. TPI and FBP catalyse the synthesis of DHAP and F6P in chloroplast, respectively. They also catalyse the reverse actions in the cytosol. GAPDH and TPI were reported to have two copies in C.velia, while only one copy of FBA and PGK were obtained to date.

4.5.2. Phylogenetic analysis of 3-phosphoglycerate kinase

One copy of the PGK gene was detected in the gene database used in this study (Figure 4.2). The PGK gene obtained from C.velia branched closely with haptophytes, apicomplexans, and bacterial groups. This indicates a potential common origin from bacteria for both haptophytes and C.velia. In contrast, diatoms, green algae, red algae, heterokonts, and higher plants shared close relationships, indicating that their PGK enzymes probably originated from a common ancestor, most likely to be cyanobacteria.

4.5.3. Phylogenetic analysis of cytosolic glyceraldehyde 3-phosphate dehydrogenase

Two copies of GAPDH were obtained from C.velia from the gene database. The chloroplastic copy of GAPDH was previously analysed by Takishita et al. (see section 4.6.2), while the cytosolic copy was analysed in this study (Figure 4.3). Based on Figure 4.3, the cytosolic copy of GAPDH in C.velia branched closely with apicomplexans and heterokonts, indicating a common origin. In terms of phylogenetic relationships, the cytosolic copy of GAPDH is similar to the chloroplastic copy in C.velia (Figure 4.12).
Figure 4.2 Phylogenetic alignment of 3-phosphoglycerate kinase (PGK) coding enzyme in *C. velia*. The numbers at the nodes of this tree show support values derived from a ML bootstrap analysis. The branches with support value over 50% are labelled. Major taxonomic groups are labelled on the right. The accession numbers of sequences used in the phylogenetic analysis are listed in Appendix-5.
4.5.4. Phylogenetic analysis of triosephosphate isomerase

Two copies of TPI genes were detected in the gene database. Both the cytosolic (Figure 4.4a) and the chloroplastic (Figure 4.4b) copies of TPI from *C. velia* were analysed in this study.

The cytosolic copy of TPI gene from *C. velia* branched closely with cytosolic copies from apicomplexans and chlorophycea groups. The close relationship between chlorophycea and *C. velia* indicates that they may inherit their TPI genes from the same origin. Given that chlorophycea inherited the TPI gene from proteobacteria (see section 4.3.4), the cytosolic TPI gene in *C. velia* is most likely also originated from proteobacteria.

The chloroplastic copy of the TPI gene in *C. velia* branched closely with diatoms, apicomplexans, and fungi-yeast, indicating a possible horizontal gene transfer between yeast and the common ancestor for *C. velia*, apicomplexans, and diatoms.

The close relationship between chlorophycea and *C. velia* indicate that the TPI gene in *C. velia* is homologous to higher plants and algae (both cytosolic and chloroplastic copies; see Figure 4.4a,b),
which originated from proteobacteria (see section 4.3.4). Therefore, the TPI gene in *C.velia* most likely originated from proteobacteria.

![Phylogenetic tree](image)

**Figure 4.4a** Phylogenetic analysis of putative cytosolic copy of TPI coding enzymes in *C.velia* (*CV 13G02*). The numbers at the nodes of this tree shows support values derived from a ML bootstrap analysis. The branches with support values over 50% are labelled. Major taxonomic groups are labelled on the right. The accession numbers of sequences used in the phylogenetic analysis are listed in Appendix-5.
4.5.5. Phylogenetic analysis of fructose-bisphosphate aldolase

Only one copy of the FBA gene was detected in the gene database (Figure 4.5). The FBA obtained from *C.velia* is branched within the cytosolic class II FBA cluster, together with proteobacteria, dinoflagellates, and diatoms. This indicates a common origin for the cytosolic copy of FBA gene for these species. The FBA gene detected in *C.velia* is closed to other members of the chromalveolate group; therefore, it is most likely to be type A Class II FBA. The close relationship between chloroplastic FBA and cytosolic FBA indicates that they may be duplicate copies of one gene.
4.5.6. Phylogenetic analysis of fructose-1,6-bisphosphatase

Only one copy of the FBP gene was detected in the gene database (Figure 4.6). The FBP gene obtained from *C.velia* branched within the apicomplexan group. This large cluster includes higher plants (both chloroplastic and cytosolic copies), apicomplexans, fungi, heterokonts, and cyanobacteria. This cluster also branched closely with proteobacteria groups, indicating that all FBP genes from these species originated from proteobacteria. Therefore, the FBP in *C.velia* most likely originated from proteobacteria.
4.5.7. Phylogenetic analysis of transketolase

Only one copy of TKL gene was obtained from C.velia. Its phylogenetic analysis is presented in Figure 4.7. The TKL gene from C.velia branched within the proteobacteria group. This is quite interesting, as all other photosynthetic organisms clustered together, including red algae, green algae, higher plants, dinoflagellates, and diatoms. Even heterotrophic apicomplexans were clustered within this group. In contrast, the TKL gene in C.velia is clearly different from that of both photosynthetic (dinoflagellate) and heterotrophic (apicomplexans) organisms. Instead, it branched closely with proteobacteria. This indicates that the TKL gene in C.velia is inherited via horizontal gene transfer from a proteobacteria.
4.5.8. Phylogenetic analysis of ribulose-5-phosphate 3-epimerase

Two copies of RPE genes were detected in the gene database. Phylogenetic analyses were performed separately on both sequences (Figure 4.8a,b).

According to Figure 4.8a, the first RPE obtained from C.velia (CV337) branched with chromalveolate species Biogelowiella natans, Emiliania huxleyi, Aureococcus anophagefferens, Ectocarpus siliculosus, and diatom groups. The chromalveolate group branched adjacent to higher plants and cyanobacteria, forming a large cluster, next to the proteobacteria group. This indicates that the RPE gene in C.velia, diatoms, green algae, and higher plants originated from proteobacteria.
Figure 4.8a Phylogenetic analysis of putative copy of RPE coding enzymes in *C. velia* (CV337). The numbers at the nodes of this tree shows support values derived from a ML bootstrap analysis. The branches with support values over 50% are labelled. Major taxonomic groups are labelled on the right. The accession numbers of sequences used in the phylogenetic analysis are listed in Appendix-5.

Figure 4.8b Phylogenetic analysis of putative copy of RPE coding enzymes in *C. velia* (CV311). The numbers at the nodes of this tree shows support values derived from a ML bootstrap analysis. The branches
with support values over 50% are labelled. Major taxonomic groups are labelled on the right. The accession numbers of sequences used in the phylogenetic analysis are listed in Appendix-5.

According to Figure 4.8b, the RPE gene \((CV31I)\) branched close to diatoms and apicomplexans. These branches, together with many other smaller branches, including heterokont, protist, human, and animal components, form a large cluster. This cluster branches next to higher plants and green algae, located next to the proteobacteria groups. This suggests that RPE genes in \(C.\text{velia}\), diatoms, apicomplexans, heterokonts, green algae, and higher plants are from proteobacteria.

4.5.9. Phylogenetic analysis of ribose-5-phosphate isomerase

Only one copy of RPI was detected in the gene database. Results of the phylogenetic analysis are presented in Figure 4.9. Here, the RPI gene \((Cv1\ 18F03)\) from \(C.\text{velia}\) branched closely with apicomplexans, green algae, and higher plants. These groups formed a large cluster, close to the branch for the heterokont group.

It seems that all these RPI sequences share high homology with the \(E.\text{coli}\) RPI copy. Therefore, these RPI genes are likely to be of RPIA type, which originated from proteobacteria (see section 4.3.9).
Figure 4.9 Phylogenetic analysis of putative copy of RPI coding enzymes in *C.velia* (CV1 18F03). The numbers at the nodes of this tree shows support values derived from a ML bootstrap analysis. The branches with support values over 50% are labelled. Major taxonomic groups are labelled on the right. The accession numbers of sequences used in the phylogenetic analysis are listed in Appendix-5.

4.5.10. Phylogenetic analysis of phosphoribulokinase

Only one copy of PRK gene was detected in the gene database. The phylogenetic analysis of this PRK gene in *C.velia* is shown in Figure 4.10. Based on its phylogenetic analysis, PRK genes from higher plants, green algae, red algae, and most of the chromalveolate group branched together. This indicates that the PRK genes in these species likely originated from the same ancestor, probably from cyanobacteria. In contrast, the PRK gene from *C.velia* located next to this group, with a positive support value of 53%. This indicates that this PRK gene probably also originated from cyanobacteria. The PRK gene from *C.velia* also branched away from the dinoflagellate group, which does not support the hypothesis that there was a common photosynthetic ancestry between *C.velia* and dinoflagellates. It is possible that the PRK in *C.velia* was differentiated during the evolution from a common ancestor with dinoflagellates.
4.6. Discussion

4.6.1. Phylogenetic analysis of ribulose-1,5-bisphosphate carboxylase oxygenase

Only one copy of the RuBisCO gene was detected in the *C. velia* gene database. The RuBisCO gene in *C. velia* has been analysed previously by Janouškovec *et al.* (2010) (Figure 4.11). According to Figure 4.11, *C. velia* clearly branched with the dinoflagellates clade, next to the proteobacteria. Since dinoflagellates have proteobacteria type II RuBisCO (see section 4.3.1), *C. velia* is very likely to also possess a type II RuBisCO, based on this phylogeny.
Figure 4.11. Phylogenetic alignment of RuBisCO coding enzyme in *C. velia*. Figure credit to Janouškovec et al. (2010). α, β, γ represents the three subclass of proteobacteria: alphaproteobacteria, betaproteobacteria and gammaproteobacteria, respectively. The numbers at the nodes of this tree shows support values derived from a RAxML bootstrap analysis and support values based on rooting this tree using distantly related non-class II RuBisCO genes. CCMP3155 represents *Vitrella brassicaformis CCMP3155*.

4.6.2. Phylogenetic analysis of chloroplastic glyceraldehyde 3-phosphate dehydrogenase

One chloroplastic and one cytosolic copy of GAPDH were identified in this study. The chloroplastic copy of GAPDH was previously analysed by Takishita *et al.* (2009). Based on Figure 4.12, the GapC1 type phylogenetic tree is divided into two major clades: Clades A and B. The GapC1 homologues from haptophytes, dinoflagellates (belonging to the genera *Karenia, Karlodinium*, and *Lepidodinium*), the apicomplexan *Toxoplasma*, and *C. velia* form Clade B. This indicates that chloroplastic GAPDH in *C. velia* is homologous to the copy in apicomplexans. In contrast, the majority of dinoflagellates and diatoms are clustered within clade A, suggesting that a lateral gene transfer occurred during the evolution of organisms in clade B.
Figure 4.12 Phylogenetic analysis of putative plastid-targeted copy of GAPDH coding enzymes in C.velia. Figure credit to Takishita et al. (2009). The GapC1 type tree was divided into two major clades, Clades A and B, highlighted by green and orange shades, respectively. The stramenopile homologues are written with bold letters. ML bootstrap probabilities (RAxML/PhyML) over 50% are shown at the branches. The thick branches represent Bayesian posterior probability over 0.95. Major taxonomic groups are labelled on the right.
4.6.3. The origination of the Calvin-Benson cycle related genes in *C. velia*

The origin of genes involved in Calvin-Benson cycle in *C. velia* is summarized in Table 4.2. To facilitate a direct comparison, the relative information on four related groups, including diatoms, photosynthetic dinoflagellates, and apicomplexans, are also listed.

Table 4.2 Origin of genes related to the Calvin-Benson cycle in *C. velia* and related groups (based on the literature and results of this study).

<table>
<thead>
<tr>
<th>Origination of enzymes</th>
<th><em>C. velia</em></th>
<th>Diatoms</th>
<th>Dinoflagellate (photosynthetic)</th>
<th>Apicomplexa</th>
</tr>
</thead>
<tbody>
<tr>
<td>RubisCO</td>
<td>proteobacteria</td>
<td>cyanobacteria</td>
<td>proteobacteria</td>
<td>N/A</td>
</tr>
<tr>
<td>PGK</td>
<td>proteobacteria</td>
<td>cyanobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>GAPDH (cytosolic)</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>TPI (cytosolic and plastidic)</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>FBA</td>
<td>Class II type A</td>
<td>Class II type A</td>
<td>Class II type A</td>
<td>Class I</td>
</tr>
<tr>
<td>FBP</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>TKL</td>
<td>proteobacteria</td>
<td>cyanobacteria</td>
<td>cyanobacteria</td>
<td>cyanobacteria</td>
</tr>
<tr>
<td>RPE (cytosolic and plastidic)</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>RPI</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
<td>proteobacteria</td>
</tr>
<tr>
<td>PRK</td>
<td>cyanobacteria</td>
<td>cyanobacteria</td>
<td>Distant from cyanobacteria</td>
<td>N/A</td>
</tr>
</tbody>
</table>

4.6.4. Cyanobacteria derived genes (chloroplastic type)

Among all phylogenetic trees constructed, only two *C. velia* sequences out of 14 have likely originated from cyanobacteria, including one copy of FBA and one copy of PRK. However, the PRK is the only one involved in the Calvin-Benson cycle. This is probably the reason for its cyanobacterial origin in *C. velia*. The FBA is unique among the chromalveolates (class II FBA),
compared with red and green algae (class I FBA). It is reported that putative class I FBA also exists (Petersen et al., 2014). However, no clear detection of FBA was found in the combined database used for this study.

4.6.5. Proteobacteria derived genes in the Calvin-Benson Cycle

Among all the phylogenetic trees constructed, twelve C.velia sequences out of 14 are likely from proteobacteria, including two copies of GAPDH, two copies of TPI, two copies of RPE, and one copy each of FBP, TKL, PGK, FBA, PRK, and RuBisCO. Previous reports have identified RuBisCO and GAPDH genes in C.velia as proteobacterial in origin (Janouskovec et al., 2011; Tarkashita et al., 2009). GAPDH and RPE are more conserved within the chromalveolate group, but too far away to compare with red or green algae.

4.6.6. Comparison of the Calvin-Benson Cycle coding genes in C.velia with other organisms

Through comparison of coding genes in the Calvin-Benson cycle of C.velia with those from higher plants, and other members of the chromalveolates group, e.g., diatoms, dinoflagellates, and apicomplexans, a different origin to the Calvin-Benson cycle genes is established (Table 4.2). Clearly, there are four enzymes, with different origins compared with higher plants: RuBisCO, FBA, and GAPDH have been reported previously (Janouskovec et al., 2011; Obornik et al., 2008); but RPE is a new discovery. Comparing C.velia with diatoms—another member of the chromalveolates, shows that FBA and RuBisCO are different in these two organisms. Diatoms possess both types of FBA and use them for carbon concentration. Diatoms also contain a class I RuBisCO compared with the class II type found in C.velia. The TKL gene found in C.velia is clearly different from diatoms (Figure 4.7), while the PRK gene in C.velia also is distant from diatoms. These comparisons suggest there was clear differentiation during the evolution of these organisms (Figure 4.10).
Comparing *C.velia* with apicomplexans—a non-photosynthesis parasitic organism that possess a relic plastid, shows that apicomplexans possess a class I FBA, instead of the class II type common to chromalveolates (Rogers and Keeling, 2004). The TKL gene in *C.velia* is close to the proteobacterial group, while the copy in apicomplexans is closed to the cyanobacterial group (Figure 4.7). RuBisCO and PRK are absent in apicomplexans.

**Table 4.3 Summary of the origin of Calvin-Benson cycle-related enzymes in *C.velia*.**

<table>
<thead>
<tr>
<th>C.velia enzymes</th>
<th>Cyanobacterial origin</th>
<th>Proteobacterial origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>FBA, PRK,</td>
<td>RuBisCO, FBA, GAPDH (2 copies), RPE (2 copies), FBP, RPI, TPI (2 copies), PGK, TKL</td>
<td></td>
</tr>
</tbody>
</table>

### 4.6.7. Comparison of *C.velia* with other organisms

A comparison of *C.velia* with other organisms shows that a better understanding of biodiversity through evolution can be achieved. Non-photosynthetic apicomplexans being the closest relative of *C.velia*, share similar types of enzymes in the Calvin-Benson cycle, with the exception of RuBisCO, PRK, TKL and FBA. The lack of RuBisCO and PRK in apicomplexans indicates that these enzymes are essential for photosynthesis in the Calvin-Benson cycle. Their different TKL and FBA origins indicate a possible horizontal gene transfer occurred during the evolution of a common ancestor of *C.velia* and Apicomplexa.

When comparing *C.velia* with diatoms, the most significant differences arose in the types of RuBisCO and FBA. Diatoms have a cyanobacterial derived RuBisCO (class I RuBisCO), while *C.velia* has a proteobacterial copy (Class II RuBisCO). Diatoms have different types of FBA, including both FBAI and FBAII types, with multiple copies in some organisms. In *C.velia*, only FBAII was found in this study, with only one copy. RuBisCO helps to fix carbon, while FBA in diatoms carries out carbon concentration. Thus, these differences in enzyme types may lead to a difference in carbon fixation efficiency. This may be part of the reason for the high abundance of diatoms in oceans, as one of the major phytoplankton.
When comparing *C.velia* with higher plants, we noted that the enzymes GAPDH, FBA and Rubisco were different. This difference is typical for the chromalveolate groups. The unique presence of these genes in *C.velia* proves its similarity to other chromalveolates. This can be used as evidence to support a common ancestor for chromalveolates. Furthermore, a distinct origin for the RPE genes was discovered in this study. Our current knowledge of the evolution, function, and structure of RPE is limited, thus further investigation is required.

4.6.8. The evolution of the Calvin-Benson cycle

After reviewing the literature related to the enzymes in the Calvin-Benson cycle, there are four enzymes generally reported to have two different classes, sharing low similarity. These are: Rubisco (30%), GAPDH (15–20% for GapA/B and GapC), FBA (no detectable similarity) and PRK (20%). The chromalveolate group has a proteobacterial origin to these genes, rather than the cyanobacterial copies present in higher plants, except for PRK. PRK may be involved during primary endosymbiosis, since it only works in the chloroplast, while it appears that the other three distinct enzymes were introduced to the chromalveolate group much later from proteobacteria. Given that Rubisco differs between dinoflagellates/*C.velia* and others in the chromalveolate group (Janouskovec *et al*., 2011), and that multiple lateral gene transfers are responsible for its GAPDH distribution (Tarkashita *et al*., 2009), multiple endosymbiosis must have occurred at least twice during its evolution. The first one likely introduced a different type of FBA and GAPDH into the ancestor of the chromalveolate group, while following ones took place in the common ancestor of dinoflagellates, *C.velia*, and apicomplexans. This brought a different type of GAPDH into haptophytes and dinoflagellates, as well as class II Rubisco into dinoflagellates and *C.velia*. Apicomplexa possess a class I FBA, different from the class II type found in *C.velia* and other chromalveolates. This indicates that another lateral gene transfer event may have occurred during the differentiation of Apicomplexa from the chromalveolate group. The TKL in *C.velia* is also distinct from other members of the chromalveolate group, indicating that *C.velia* also underwent lateral gene transfer during its differentiation from the chromalveolate group.
Comparison of *C.velia* genes with their homologues suggests that gene evolution occurred over a long history. It appears to be very complicated, especially for the enzymes in the Calvin-Benson cycle, which are an essential to the biochemical pathway for photoautotrophs. The mosaic origin features of enzymes in the Calvin-Benson cycle clearly support the shopping bag model proposed by Larkum *et al.* (2007).

4.6.9. Enzyme co-localization in the Calvin-Benson cycle

During study of the Calvin-Benson cycle-related enzymes, it was discovered that some of the enzymes are co-localized on the stroma for the benefit of regulation. In the cytosol of higher plants, TPI and FBA, GAPDH and FBA, GAPDH and TPI, PGK and GAPDH are co-localized (Anderson and Carol, 2005). In the chloroplast of higher plants, each enzyme in the Calvin-Benson cycle is co-localized with the enzyme that forms its substrate and uses its product, with the exceptions of FBA, SBP and TKL. GAPDH and RuBisCO, GAPDH and TKL are also co-localized, despite the fact that they are not really sequentially related in the Calvin-Benson cycle (Anderson *et al.*, 2006). More recently, it was discovered that GAPDH, PRK, FBA, and photosystem protein CP12 form a complex in the chloroplast of green algae (Erales *et al.*, 2008). Given the knowledge that all enzymes in the Calvin-Benson cycle are closely related, it is possible that co-localized enzymes may also go through the same evolutionary pathway, e.g., have been replaced in the same lateral gene transfer. In this study, it was shown that in *C.velia*, RuBisCO, GAPDH and FBA are of special types compared with other organisms of the chromalveolate group. However, these three enzymes are not sequentially related to each other in the Calvin-Benson cycle. Given that enzymes in both former and later steps are similar to those in higher plants, it is possible that gene replication in the Calvin-Benson cycle did not markedly affect the formation and function of this complex.
4.7. Conclusions

In this study, we analysed the set of Calvin-Benson cycle-related genes in *C.velia*. Mostly, the result do support the chromalveolates evolutionary hypothesis, since *C.velia* genes are mainly clustered with genes from other species in the chromalveolate group, such as diatoms, dinoflagellates, haptophytes, and apicomplexans. However, some genes may support a single endosymbiosis hypothesis, while others (like GAPDH) support a multiple endosymbiosis hypothesis. One possible scenario for evolution of the Calvin-Benson cycle in *C.velia* is that a set of genes encoding enzymes in the Calvin-Benson cycle has been transferred to the nucleus in one endosymbiosis event, probably during primary endosymbiosis. However, some genes have clearly been replaced by lateral gene transfer, either accidently or intentionally during its evolution (to account for TKL and RuBisCO types), while other genes were reserved (such as PRK). The replacement genes may have arisen through secondary or tertiary endosymbiosis, or a cytosolic copy may have originated from mitochondria (and ultimately proteobacteria). It is also possible that both situations occurred.
Chapter 5. Iron-stress and photosynthetic apparatus in *C. velia*

5.1. Introduction

5.1.1. Iron limitation for biological system

Iron is one of the essential elements for photosynthetic organisms. Iron exists naturally in two stable oxidation states, Fe^{2+} (ferrous) or Fe^{3+} (ferric). It is often used as a cofactor in proteins that catalyse reduction-oxidation (redox) reactions, such as electron-transfer chains in respiration and photosynthesis. Iron is mostly found as stable Fe^{3+}-oxides such as goethite (FeOOH) and hematite (Fe_{2}O_{3}) in the ocean, which are insoluble at biological pH and not useful for biological systems (Guerinot and Yi, 1994). Iron availability is a limiting factor for phytoplankton growth in the sea, with growth impaired by low iron concentration. Previous reports indicate that about 40% of the ocean, mostly in the Southern Ocean, Pacific Ocean and North Pacific Ocean, has sufficient nutrients but low carbon fixation and organic production due to iron limitation (Moore *et al.* 2002).

5.1.2. Importance of iron

Iron is an important element in heme-proteins, such as cytochromes, myoglobin, haemoglobin and iron-sulfur clusters. Iron-containing proteins work as a cofactor in many metabolic pathways, including photosynthesis, respiration, nitrogen fixation, pigment and DNA biosynthesis.

Iron is a vital component in the photosynthesis system, mainly functioning in the PSI and electron transport chain (z-scheme) (Figure 1.4). There are three iron-sulphur [4Fe-4S] clusters that function as primary electron acceptors in PSI, named as Fx, F_{A}, and F_{B} (Bibby, 2001; Blankenship, 2001). Ferredoxin is a soluble iron-sulphur [2Fe-2S] protein that receives electrons from PSI and transfers them to an enzyme for reducing NADP into NADPH (Fukuyama, 2004). PSII consists of only two heme-binding protein complexes, which results in a much lower requirement for iron compared to the number of iron atoms in PSI. Cytochrome *b_{6}f* contains three heme binding complexes and a [2Fe-2S] cluster, and it is the main component in the electron transport chain. The iron availability is
essential for maintaining the integrity of the electron transport chain and the activities of photosystems.

5.1.3. Iron-stress and photosynthesis

Different photosynthetic organisms use different strategies to adapt and thrive in iron-stress conditions. Under iron-starved conditions, plant cells will lose their ability for synthesising Chls, which is a common phenomenon called “chlorosis” (Moseley et al., 2002). The cellular Chl content will be decreased, and the growth rate will be inhibited as the result of impaired photosynthesis.

The onset of iron-stress has been investigated broadly in oxygenic photosynthetic organisms, including plants, algae and cyanobacteria. In cyanobacteria, iron-stress conditions will not only lead to a decrease in the ratio of PSI/PSII, but will also lead to the synthesis of a new protein, the iron-stress-induced protein A (isiA) (Fraser et al., 2013). IsiA is a Chl-binding protein complex homologous to CP43 and CP47 (La Roche et al., 1996). The interaction between PSI and its light-harvesting components is modified with this specialized accessory Chl-binding protein complex (isiA), which is thought to play a photoprotective role under iron-stress conditions (Burnap et al., 1993). Under iron-stress conditions, isiA acts as an accessory light-harvesting antenna for PSI and forms an 18 isiA/3 PSI supercomplex (Boekema et al., 2001; Bibby et al., 2001). This supercomplex formation presumably protects and enhances the function of PSI under iron-stress conditions (Ryan-Keogh et al., 2012).

In terms of photosynthesis, iron has been shown to influence the assembly of the photosynthetic apparatus in green algae (Varsano et al., 2006). In the green alga *Chlamydomonas reinhardtii*, iron-stress conditions induce the disconnection of PSI from LHCI, resulting in decreased energy transfer efficiency between light harvesting antenna and reaction centres (Yadavalli et al., 2012). Similar changes in response to iron-stress conditions have also been revealed in the red alga *Rhodella violacea* (Desquilbet et al., 2003; Doan et al., 2003). Interestingly, algae also synthesise “new” LHCs, named as LI818, in response to iron-stress conditions, and the LHCs function as
accessory light-harvesting components mainly for PSI (Peers et al., 2009). These new LHC proteins have a homologous structure to the LHC protein family, but are distinguished from the IsiA in cyanobacteria. Moseley et al. (2002) indicated that new types of LHC were synthesised and assembled a 4 LHC trimers/1 PSI super-complex under iron-stress conditions. The LHC/PSI super-complex protects the function of PSI and enhances the PSI efficiency in response to iron limitation (Moseley et al., 2002). In diatoms, LI818 was significantly up-regulated under iron-stress conditions (Lommer et al., 2012).

Flavodoxin is a small electron-transfer protein that has a similar function as ferredoxin but without Fe-S clusters. The replacement of ferredoxin with flavodoxin is an important strategy for photosynthetic organisms thriving under iron-stress conditions. In Thalassiosiroid Diatomsis, flavodoxin-encoding gene was shown to be up-regulated in response to iron-stress conditions (Whitney et al., 2011). This phenomenon has also been observed in other eukaryotic algae (LaRoche et al., 1993) and prokaryotic cyanobacteria (Morrissey and Bowler, 2012). Flavodoxin is reported to be involved in an alternative electron transport pathway in photosynthetic organisms in response to oxidative stress conditions, which can also be applied as a bio-marker for iron-stress conditions (La Roche et al., 1996; Tikkanen et al., 2012).

Decreased energy conversion efficiencies in PSII and the electron transport rate between photosystems have been discussed as another potential strategy involved in iron-stress responses (Greene et al., 1992).

Recently, it has been noticed that the petF gene, encoding ferredoxin, has migrated from the plastid to the nucleus in a species of diatoms under iron-stress conditions (Lommer et al., 2010). This migration of petF may benefit cells with effective iron-regulation of photosynthesis within the plastid.

Fructose-bisphosphate aldolase (FBA) catalyses the reaction: fructose 1-phosphate ↔ DHAP + glyceraldehyde, which is an important step in glycolysis (respiration pathway) and the Calvin-Benson cycle (photosynthesis pathway). Previous studies based primarily on EST
sequencing suggested that certain FBA genes are responsive to iron availability in diatoms (Allen et al., 2006; Maheswari et al., 2010). The iron-regulation of FBA provides the possibilities of shifting cellular metabolic reactions, achieving a balance between respiration and photosynthesis.

5.1.4. Current study of iron-stress in *C.velia*

FBA represents the switching point between respiration and photosynthesis in response to the iron availability (Allen et al., 2006; Maheswari et al., 2010). Based on the result of Chapter 4, *C.velia* has a type of FBA that is homologous to diatoms and proteobacteria, rather than homologous to cyanobacteria. *C.velia* has the potential to adapt to iron-stress conditions while keeping its photosynthesis ability, which is also a common phenomenon observed in diatoms (Lommer et al., 2012). Additionally, *C.velia* is reported to have the petF gene (for ferredoxin biosynthesis) migrated to the nucleus (Burki et al., 2012). This gene migration gives *C.velia* a potential advantage when growing under iron-stress conditions, as reported in diatoms (Lommer et al., 2010).

Twenty-three LHC homologs have been retrieved from a *C.velia* EST database, including one homologous to the LI818 class, three homologous to PSI related LHC, and 17 homologous to PSII-related LHC (see Chapter 3; Pan et al., 2012). LI818 has been shown to be up-regulated under iron-stress conditions in diatoms (Lommer et al., 2012). The presence of LI818 homologs in *C.velia* might be a potential strategy for cells to cope with iron-stress conditions. Sutak et al. (2010) studied the iron-transport systems in *C.velia* and revealed that *C.velia* has a unique iron absorption ability that is different from any known iron-uptake system. Therefore, *C.velia* may have the potential to adapt to a severe iron-stress environment with an unknown mechanism.

5.1.5. Aim

*C.velia* is a newly discovered algal species with a close phylogenetic relationship to diatoms (Janouskovic et al., 2011). Previous reports showed that diatoms are able to thrive under
iron-stress conditions (Lommer et al., 2010, 2012). Thus, C.velia has the potential to survive under iron-stress conditions. Study of the photosynthesis-related responses to iron-stress conditions will improve the understanding of the biology of C.velia. By comparison with diatoms, possible information regarding the photosystem regulation and evolutionary relationship may be revealed. In Chapter 3, 23 LHC polypeptides were phylogenetically analysed in C.velia. These peptides were designated into different categories, and three sequences were shown to be homologous to LHCI. By studying the photosynthetic apparatus changes in response to iron-stress conditions, especially the functional relationship between LHC peptides and photosystems, the new regulatory mechanism by iron availabilities may be characterised and determined.

This chapter focused on the development of methods for the isolation of pigment-binding protein complexes because of the difficulties in breaking C.velia cells due to its tough cell wall structure (Weatherby et al., 2011). The isolated photosynthetic protein complexes and their functions were characterised by spectral analysis, HPLC and SDS-PAGE. The results presented here will provide hints for further study on photosynthetic apparatus rearrangements in response to iron-stress conditions. The interaction between LHC and reaction centres is discussed.

5.2. Material and methods

5.2.1. Culture conditions

Both iron-stress and control cultures were set up as described in Chapter 2, section 2.1.2. For iron-stress culture setup, it is common to use the same growth medium as the control (normal culture), but without the addition of iron (Fraser et al., 2013; Ryan-Keogh et al., 2012). In this study, we selected artificial sea water with K+ES medium for both cultures. C.velia has also been cultured in filtered natural seawater medium in our lab, which produced similar cultures as artificial seawater medium. We believe that the growth medium conditions used in this study (artificial seawater with K+ES supplements) is a good mimic of oceanic oligotrophic sites, representing the realistic natural environment.
Cells were harvested after 3 weeks of growth in the iron-stressed or control conditions (details in Chapter 2, section 2.1.5). Harvested cells were rinsed once with fresh artificial seawater and stored at -80°C until further investigation.

5.2.2. Growth rate monitoring and oxygen evolution rate measurement

In order to monitor the physiological feature of *C. velia* culture under iron-stressed conditions, three 1 ml culture suspensions from both iron-stress and control cultures were sampled every two days over a course of 18 days for cell-counting, oxygen evolution rate measurement, and pigment analysis. Cell samples were counted manually under a light microscope (40X magnification). To improve the statistical significance of counted cell numbers, the values for each sample were determined from biological samples with three technical replicates (n=12).

Oxygen evolution rate was measured using an oxygen electrode under illumination (see details in Chapter 2, section 2.5). The relative respiration rate was measured at the same time in the dark. After dark-adaptation, the control and iron-stress cultures were illuminated using white light with intensities of 35, 70, 140, 350, 420, 560 and 700 µE for 10 min. The photosynthetic O₂ evolution rate was calculated according to the Chl *a* concentration (µg/ml) or the number of *C. velia* cells. Chlorophyll *a* concentration was measured using a spectrophotometer, and calculated according to a published formula (Ritchie *et al.*, 2006; see details in Chapter 2, section 2.2.1).

The pigment compositions of normal and iron–stress *C. velia* cultures were analysed weekly using HPLC to monitor the dynamic changes of pigment during their growth phases over the course of three weeks (see details in Chapter 2, section 2.3). For each culture, three replicate samples were collected at the initial week (week 0), week 1, week 2, and week 3. Pigment extractions and HPLC were performed as described in Chapter 2, section 2.3.
5.2.3. Isolation of pigment-binding protein complexes

In order to isolate pigment-binding protein complexes from the thylakoid membranes, harvested cells were broken using a bead-beater (details in section 2.3.1). Isolated thylakoid membrane was collected by ultracentrifugation at 40,000 rpm for 30 min at 4°C (SW-55 Ti; Beckman Coulter, USA). The pigment-binding membrane protein complexes were solubilized using 1% DoDM and separated using sucrose density gradient ultracentrifugation (See Chapter 2, section 2.3.1). The fractions generated from sucrose density gradient were collected carefully using a syringe. Absorption spectral and fluorescence spectral analyses were performed for each collected fraction at room temperature (293K) and low temperature (77K). The pigment and protein compositions of each fraction were determined using HPLC, SDS-PAGE, and western blotting (see details in Chapter 2).

5.3. Results

5.3.1. Cell growth rate

The growth curves of normal and iron-stress *C. velia* cultures, each based on average cell numbers counted over 8 sampling points during 3 weeks, are presented in Figure 5.1. The average cell numbers with Standard error of the mean (SEM) are presented in Figure 5.1. To calculate the growth curve, cee numbers were re-plotted using a logarithmic scale (Ln; Figure 5.1 inset). The doubling time (Y, exponential growth rate) was calculated according to the formula: $Y = \frac{\ln(2)}{K}$, where $K$ represents the fitting slope between days 4–12 (covering 5 sampling points). The statistical data of the growth curve are listed in Table 5.1.

Cells in the normal culture appeared to have a higher growth rate than the iron-stress culture. After 3 weeks growth, both cultures stayed at their log phase, although cells in the iron-stress culture had a much slower, but still stable growth rate (Figure 5.1). Under normal culture conditions, the maximal growth rate (doubling time) of *C. velia* was reached at ~2.82 days, while the culture under
iron-stress conditions only reached a doubling time of 7.54 days, which was about 2.5 times lower than the normal culture.

Table 5.1 Growth curve measurement of *C. velia* in control and iron stress cultures

<table>
<thead>
<tr>
<th>Culture type</th>
<th>Doubling time (days)</th>
<th>R2 value of curve fitting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control (normal culture)</td>
<td>2.82 ± 0.18</td>
<td>0.98</td>
</tr>
<tr>
<td>Iron-stress culture</td>
<td>7.54 ± 0.47</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Figure 5.1. Growth rate of *C. velia* cultures. Cell numbers were counted from day 0–18, from four biological replicates, with 3 repeats for each data point (n=12). The error bars represent the Standard error of the mean (SEM). Insert, logarithm growth fitting curve of *C.velia* in control and iron-stress culture. Doubling times were calculated based on the formula: \( Y = \frac{\ln(2)}{K} \), where \( Y \) represents the doubling time (days), and \( K \) represents the slope of the fitting line. The maximum doubling time was calculated based on the fitting data covering from days 4 to 12 for both control and iron-stress culture. Round dots, cells in control culture; squares, cells in iron-stress culture.
5.3.2. Oxygen evolution rate measurement

Oxygen evolution rate was measured using an oxygen electrode (Hans-tech, UK). The gross oxygen evolution rate represents the net oxygen evolution rate monitored under light, together with the oxygen consumed rate observed at the dark period (regarded as respiration rate). To compare the photosynthesis ability in the iron-stress and control cultures, the calculated results were normalised to the cell numbers and Chl concentration, respectively. The results are plotted in Figure 5.2 a,b.
The results indicated that under the same light intensity illumination, cells from the normal culture had higher oxygen evolution ability per cell compared to the iron-stress culture. *C. velia* reached its maximal oxygen evolution rate of $1.6 \times 10^{-7}$ nmol O$_2$/h per cell at 70 µE. The O$_2$ evolution rate dropped down to $0.6 \times 10^{-7}$ nmol O$_2$/h per cell at 350 µE, only 37% of maximal O$_2$ evolution rate, indicating the optimal light intensity for *C. velia* culture is ~70 µE. The light intensity beyond 100 µE would cause photoinhibition under the culture conditions. The O$_2$ evolution rate remained stable at $0.4 \times 10^{-7}$ nmol O$_2$/h per cell when cells are exposed to light intensities of $>350$ µE. Under iron-stress conditions, the maximal O$_2$ evolution rate can only reach $0.3 \times 10^{-7}$ nmol O$_2$/h per cell at 140 µE, less than 20% of the normal culture. However, this culture showed different responsive profiles to the changed light intensities. The O$_2$ evolution rate remained similar following the increased light intensity from 70 µE–420 µE, suggesting the iron-stress culture is tolerant to relative high-light intensity.
Based on the Chl concentration, the maximal O₂ evolution rate of *C. velia* normal culture was 1.29 nmol O₂/h per µg Chl at 70 µE, while iron-stress culture showed 0.612 nmol O₂/h per µg Chl at 70 µE under moderate light condition. The [O₂] rate of iron-stress cells was approximately 50% of the value for normal cells, suggesting the amount of PSII was decreased with the increased Chl *a* per PSII. Interestingly, under high light intensity (>350 µE), cells in iron-stress culture showed a higher O₂ evolution rate of 0.72 nmol O₂/h per µg Chl, which was approximately 30% higher than normal culture (0.48 nmol O₂/h per µg Chl), indicating the iron-stress culture prefers high light conditions. Overall, *C. velia* cells in iron-stress conditions had a low but stable oxygen evolution rate, and also preferred high light conditions.

5.3.3. Pigment composition of *C. velia* cells from normal culture conditions

The HPLC results for the pigment composition analyses of the normal culture are summarized in Figure 5.3. Five major pigments were resolved from the normal culture: Chl *a*, violaxanthin, isofucoxanthin, an unknown carotenoid and β-carotene. Concentrations of three major pigments (Chl *a*, isofucoxanthin and violaxanthin) were calculated based on the area of the corresponding HPLC peaks, which were recorded at their coefficient wavelength: 665 nm for Chl *a*, 480 nm for isofucoxanthin, and 435 nm for violaxanthin. The ratio of the major carotenoids (violaxanthin/isofucoxanthin) was calculated based on the peak areas, and is presented in Figure 5.4. The third peak which comes after isofucoxanthin was an unknown carotenoid that is thought to be produced by *C. velia* in response to adhesive growth of cells in culture flasks (Foster, personal communication).

The Chl *a* concentration can represent the growth rate of cultures. In the control (normal) culture, the Chl *a* concentration increased from 4 µg/L to 12 µg/L during 3 weeks of growth. As a proportion of the total pigments, the Chl *a* increased from 63% to a final value of 84%. The isofucoxanthin/violaxanthin ratio remained similar during the three week growth period (Figure 5.4).
5.3.4. Pigment composition of *C.velia* cells from iron-stress culture conditions

The pigment composition of the iron-stress culture was analysed by HPLC, and the results are analysed and summarized in Figure 5.5. Similar to the HPLC profiles from the normal culture, five main pigments were resolved from the iron-stress culture: Chl *a*, violaxanthin, isofucoxanthin, an unknown carotenoid and β-carotene. Concentrations of three major pigments (Chl *a*, isofucoxanthin and violaxanthin) were calculated based on the areas of the corresponding HPLC peaks. The violaxanthin/isofucoxanthin ratio was calculated based on the peak areas and is presented in Figure 5.4.

Under the iron-stress conditions, the growth of *C.velia* was significantly inhibited. The content of isofucoxanthin in *C.velia* was decreased, while violaxanthin concentration was increased during growth (Figure 5.5). The total amount of Chl decreased to 5.5 µg/L under iron-stress conditions. In
contrast, the relative amount of violaxanthin (% of total pigments) increased from 5% to 27% during continuous growth under iron-stress conditions.

Figure 5.4. The ratio change of violaxanthin/ isofoxcxanthin in normal (dark grey bars) and iron-stress (light grey bars) *C. velia* cultures. The black line represents the point at which violaxanthin was present in the same amount as isofoxcxanthin. The amounts of pigments were calculated based on the corresponding HPLC peak areas presented using the HPLC software (reading at 665 nm for Chl *a*, 400–470 nm for carotenoids).
Figure 5.5 HPLC chromatograms for pigments extraction from *C. velia* in iron-stress culture conditions. Data were normalised to the intensity of the Chl *a* peak (indicated on the graph). Peaks corresponding to major pigments are indicated with arrows. Chromatograms are represented as the maximal spectral readings between 400–700 nm. Peaks are assigned based on their retention times and their online spectral properties.

After 3 weeks of growth, the pigment compositions of both iron-stress and normal cultures were compared by HPLC (Figure 5.6). *C. velia* cells in the control (normal) culture had more Chl *a* and carotenoids, which suggests a greater *in vivo* function of photosynthesis. The β-carotene in both cultures remained at a relatively similar amount.
5.3.5. Photosynthetic pigment-binding protein complexes

In order to understand the changes of photosynthetic membranes caused by iron-stress conditions, pigment-binding protein complexes were isolated from the normal and iron-stress *C. velia* cultures. The solubilised pigment-binding protein complexes were loaded onto sucrose density gradient tubes (10%–40% in buffer A; see details in Chapter 2, section 2.3). After ultracentrifugation at 150,000 xg for a minimum of 16h, four coloured fractions were resolved from both iron-stress and normal cultures (Figure 5.7).

Four green bands, named as NR1, NR2, NR3 and NR4 were resolved from the normal culture, which is consistent with the results of Tichy *et al.* (2013). The NR1 fraction was on the top with a green colour, the NR2 fraction was the main fraction with a brownish/green colour, NR3 was a distinct fraction with a light green colour, and NR4 was the heaviest fraction and was barely visible. Compared with the normal culture, the bands resolved from the iron-stress culture showed a more yellowish/brown colour, and were named as IR1, IR2, IR3 and IR4. IR1 was a yellowish/brown
coloured fraction that had a lower density compared to NR1, and may be composed of LHC and monomer of reaction centres. IR2 had a similar colour to IR1, and showed the same density as NR2, but had a much lower Chl concentration. This was consistent with the pigment composition result obtained from spectral analysis (Figure 5.8b). IR3 was the main fraction generated from the iron-stress culture, which had the same density as the corresponding band NR3. IR4 was the heaviest fraction, also a very faint band, similar to NR4 from the normal culture.

Figure 5.7 Isolation of pigment-binding protein complexes using sucrose gradient ultra-centrifuge. Fractions were taken using syringes for use in later experiments. a, sucrose density gradient after ultracentrifugation; b, the absorption spectra for thylakoid membrane samples isolated from iron-stress and normal culture. Absorption spectra were normalised to the same height for a peak at 665 nm.

5.3.6. Absorption spectra

The fractions resolved by sucrose density gradient ultracentrifugation were collected and diluted in buffer A (see details in Chapter 2, section 2.3) for spectral analysis at room temperature (Figure 5.8).

Four fractions generated from the normal culture showed similar absorption spectral profiles. The main absorption peaks at the red wavelength region were centred at ~670 nm at room temperature (RT), although a slightly blue-shifted (~1–2 nm) peak was noticed from NR1 and NR2 fractions. The 418 nm peak could be attributed to the contribution of violaxanthin. The shoulder at 480–520 nm could be attributed to the presence of isofucoxanthin and was observed from NR1 and NR2 fractions.
Interestingly, the four fractions from the iron-stress culture had absorption spectra different from those of the normal culture (Figure 5.8). In the IR2, IR3, and IR4 fractions, high content of isofucoxanthin resulted in an increased shoulder at 480–520 nm. A relatively high amount of violaxanthin was also observed in the absorption spectra. The higher reading at ~418 nm reached ~2 times that of the Chl \textit{a} reading at 670 nm in iron-stress fractions, whereas this ratio was only ~1.3 in normal culture fractions.

Interestingly, the IR2 fraction showed an extra absorption peak at ~700 nm, which was not observed from any other isolated pigment-binding protein complexes. In IR3 and IR4 fractions, there were two small peaks, centred at 720 nm and 730 nm, which suggest that a new pigment-binding protein complex may be induced under iron-stress conditions.

5.3.7. Fluorescence reading for the isolated fractions

Figures showing the room temperature and 77K low temperature fluorescence readings are presented in Appendix-7. The results for the bottom bands (NR4 and NR3) showed correspondence with PSI-Lhcr supercomplex (714 nm) and Lhcr (676nm), respectively. Under iron-stress conditions, only one peak was observed in the bottom bands (IR3 and IR4) with a wavelength reading at 685 nm, which corresponded to the recently discovered red shifted antenna complex that is produced by \textit{C.velia} under red/far-red light illumination (Bina \textit{et al.}, 2014). The NR2 band from the normal culture presented a peak around 690 nm, which corresponded to the expected readings of an FCP antenna complex (Lepetit \textit{et al.}, 2007; Veith \textit{et al.}, 2009).

5.3.8. HPLC for the isolated pigment-binding protein complexes

The pigment composition of isolated pigment-binding protein complexes was analysed using butanol pigment extraction followed by HPLC. The HPLC chromatograms, along with absorption spectra for individual peaks, are presented in Figure 5.9. Due to limited sample amount, there is no HPLC analysis data for the NR4 and IR4 samples.
Figure 5.8. In vivo absorption spectra of sucrose gradient fractions isolated from *C. velia*. a, control (normal) culture samples, with absorption spectra normalised to the height of the 435 nm peak. b, iron-stress culture samples, with absorption spectra normalised to the height of the 416 nm peak. Major peaks are indicated with arrows and absorption wavelengths.
Chl a was the main photopigment in fractions generated from the normal culture (HPLC peak around 25 min). There were only three pigments detected in NR3: violaxanthin, Chl a and pheophytin a. This supports the theory that NR3 was most likely enriched with photosystem reaction centres, while isofucoxanthin was reported to be the main pigment in LHC (FCP) (Tichy et al., 2013). The top band of normal culture, NR1, consisted of a higher amount of Chl a, as well as detectable violaxanthin and isofucoxanthin (12.6 and 14.9 min retention times, respectively). According to previous studies and our peptide analysis (see below section 5.4), NR1 was likely enriched with LHC mixed with photosystem reaction centres.

For iron-stress samples, the most abundant pigment in IR1 was pheophytin a, rather than Chl a. Pheophytin a is an important cofactor that serves as the first electron carrier intermediate in the electron transport pathway of PSII, which plays a protective role under iron-stress conditions. Higher pheophytin a may also be an indication of pigment degradation occurring during iron-stress treatment. Pheophytin a may be degraded Chl a, released as a free-agent from pigment-protein complexes. Chl a was more abundant in IR2 and IR3 samples, which indicated that these two samples contained less free pigments. IR3 was the main fraction resolved from sucrose gradient (Figure 5.7). IR3 had a very low amount of carotenoids, a relatively high amount of Chl a, and a similar carotenoid/chl a ratio as NR3. This indicated that IR3 may contain similar fragmented/preserved reaction centres like NR3 (Figure 5.9a).
Figure 5.9. HPLC analysis of sucrose gradient fractions isolated from control (normal; a) and iron-stress (b) cultures. Data are presented normalised to the same height of the Chl a peak in normal culture (25.312 min) and iron stress culture (25.867 min). Absorption spectra with retention times are presented inside the Figure.
5.3.9. Polypeptide composition

The polypeptide composition of the fractions isolated from the sucrose gradients were analysed by SDS-PAGE (Figure 5.10). The bottom bands (NR3 and NR4) in samples from the normal culture had both PSII and PSI. The twin bands around 43 kDa and 47 kDa corresponded to the CP43 and CP47 proteins in PSII, while the bands around 60 kDa correspond to PSI. In contrast, the bottom bands in iron-stress samples (IR3 and IR4) lacked the twin CP43 and CP47 bands. IR4 had a visible 60 kDa band as well as some small peptides (below 26 kDa), which suggested a similar PSI polypeptide pattern to that published by Tichy et al. (2013).

An extra 20kDa band was observed in iron-stress fractions IR2–IR4. The common LHC bands in normal samples were located between 15–20 kDa, which was most likely to be the FCP-LHC proteins (Tichy et al., 2013). The iron-stress samples had weak signals of FCP-LHC bands in this region, indicating that the LHC system was reduced under iron-stress conditions.

![Figure 5.10 SDS-PAGE of sucrose gradient fractions from normal (a) and iron-stress (b) C.velia cultures. Indications of signature proteins in PSI and PSII are labelled with boxes. The protein marker sizes are listed on the left side of each gel image.](image-url)
3.10. Western blotting

Three antibodies were applied for detecting specific proteins in the sample, targeting at LHC1, PsaC and PsbA proteins, respectively. LHC1 is a small protein inside LHCs around PSI, with a size between 20–15 kDa. PsaC is a protein located between the reaction centres of PSI and ferredoxin, with a molecular size of 9 kDa. PsbA is the centre protein in PSII, with an expected molecular weight of ~30 kDa.

The results of the western blot staining with the three antibodies of interest are presented in Figure 5.11. For the LHC1 antibody, signals were detected mainly around 20kDa and 15kDa (Figure 5.11b). This result fits the expected range of FCP-LHC protein size. NR1 and NR2 have strong signals corresponding to LHC1 antibody. In comparison, NR3 and NR4 had stable but weaker signals. The western blotting result was consistent with the SDS-PAGE result, where most of FCP-LHC is present in the top fractions (NR1 and NR2). On the other hand, only IR2 gave positive but weak signals to the LHC1 antibody. This suggested that the LHC system in the iron-stressed culture is inhibited.

For PsbA antibody detection, the strongest signal came from bands higher than 25 kDa, which is close to the expected 28–30 kDa (Figure 5.11c). There were some weak signals below the 25 kDa marker, which could be due to the degraded copies of PsbA. Signals were very strong in NR1 and NR2 samples, indicating that these fractions contain a lot of PSII fragments. The signals were weak in NR3, IR2 and IR3, suggesting that these fractions do not contain a large amount of PSII.

For PsaC antibody detection, there was only one strong signal at the 10 kDa range in the NR3 sample (Figure 5.11a). This indicated that NR3 contained most of the degraded/complete PSI complex. This corresponded to the SDS-PAGE result (Figure 5.10), where NR3 and NR4 contained most components of the PSI system. Samples from the iron-stress culture did not show any signals in response to the PsaC antibody (Figure 5.11a), which indicates a limited number of PSI or disassembled PsaC in *C. velia* under iron-stress conditions (Yadavalli *et. al.*, 2012).
5.4. Discussion

It appears that iron is an important nutrient element for *C. velia*. In order to set up an iron-stress culture, the culture flasks with acid to remove any potential iron compounds. However, the iron-stress culture is iron depletion set-up, instead of “iron-free”, due to unforeseen trace amount of iron from chemicals or other unknown resources. The iron-stress culture does contain much less amount of iron, compared to the normal culture ([Fe⁺] = 2 μg/ml; Appendix-1), which justifies the comparison between iron-stress and normal culture.

*C. velia* cultures can reach a doubling time of ~3 days under normal culture conditions. The growth rate of the normal culture (0.24 day⁻¹, indicated in Figure 5.1) is consistent with a previously reported *C. velia* culture growth rate of 0.27 day⁻¹ (Foster *et al*., 2014).
Under iron-stress condition, *C.velia* has produced less Chl *a*. Accompanied with the decrease of Chl *a*, the carotenoids composition has also shifted. As seen in figure 5.4, the ratio of violaxanthin/isofucoxanthin has increased under iron-stress condition. Quigg *et al.* (2012) reported that violaxanthin/zeaxanthin cycle is an important strategy in NPQ for *C.velia* cells under iron-stress conditions. The increased violaxanthin in *C.velia* indicated that the NPQ in *C.velia* may be enhanced, possibly due to the need for balancing light energy intake and limited PSI function. A small decrease in beta-carotene and relative inhibition of Chl composition mainly was caused by the loss of reaction centres in PSII. A typical PSI contains ~90 Chl *a* and 22 β-carotene, while PSII contains only 35 Chl *a* and 12 β-carotene (Saenger *et al.*, 2002). The loss of PSII reaction centres and increased NPQ resulted in an increased ratio of LHC components per reaction centre.

The HPLC experiments confirm the increase of violaxanthin and decrease of Chl *a* under iron-stress conditions. For the fluorescence spectra experiments, it appears that *C.velia* cells might have shifted their antenna complexes under iron-stress conditions, similar to the response known to occur under red/far-red light illumination (Bina *et al.*, 2014). However, this still needs to be investigated further.

Due to the fact that iron is more abundant in PSI and the electron transport chain, iron-stress conditions will limit the function of PSI more than PSII. Therefore, the energy absorbed by the antenna complex of PSII needs to be dissipated properly. A previous report has indicated that the ratio of PSII to PSI complexes and the non-photochemical quenching (NPQ) capacity are both elevated under iron-stress conditions (Allen *et al.*, 2008). However, from the results of sucrose density gradient ultracentrifugation, SDS-PAGE and western blotting, this is not the case for *C.velia*. The SDS-PAGE and western blotting experiments have shown that the iron-stressed sample has a lower amount of PSII proteins (indicated by PsbA antibody labeling). An interesting fact is that the bottom bands from iron-stress samples have a band at 100 kDa, which contains PsaC. While PSII in *C.velia* is decreased significantly, PSI did not change very much compared to the control culture. This is likely the reason why *C.velia* can survive under iron-stress conditions.

The features of iron-stress samples cannot be defined clearly in SDS-PAGE and western blotting results; further investigation is required by techniques such as liquid chromatography-mass
spectral analysis. The target band for PsaC antibody is about 9 kDa, which was out of the range of the protein ladder (10–250 kDa). PsaC is also a protein that is embedded in PSI, which makes it difficult to detect. In order to obtain a better result, it may be worth trying to switch PsaC to some other commercially available antibodies for western blotting. LHC1 antibody is designed from the red alga *Porphyridium cruentum* (Agisera, Lot AS08–282). It is confirmed with high reactivity to red algae and diatoms, but with low reactivity to the LHC2 in spinach. *C.velia* showed a positive signal to the LHC1 antibody; the signal strength was moderate and corresponded to the FCP pattern on the gel (Figure 5.11). Therefore, the LHC1 antibody is capable of LHC detection in *C.velia*.

The sample amounts fractioned from the sucrose gradients were limited. Therefore, the results for SDS-PAGE and western blotting were not very satisfactory. The signal strengths for the bottom samples (IR4, NR4) were too low. A repeat set of analyses from a larger amount of *C.velia* culture is recommended.
Chapter 6. General Discussion

*C.velia* is a newly discovered autotrophic alga, which is regarded as the closest relative of heterotrophic apicomplexan parasites (Moore *et al.*, 2008). However, Forster *et al.* recently reported that *C.velia* possesses the potential for a mixotrophic lifestyle (Forster *et al.*, 2014). In order to understand the biology of *C.velia*, the photosynthetic system of *C.velia* has been investigation in this study. New knowledge of the photosynthetic systems in *C.velia* was acquired, including the classification of LHC proteins, the evolutionary relationship and the origination of enzymes in Calvin-Benson cycle (Chapter 3 and 4). The photosynthetic protein organization and pigment composition changes in response to iron-stress conditions have also been investigated (Chapter 5).

The work presented here clarified the origination and classification of two groups of proteins in *C.velia*: one group is the Light harvesting proteins (Chapter 3) and the other group is enzymes involved in the Calvin-Benson cycle (Chapter 4). These proteins play important roles in the photosynthetic light reactions and dark reactions. Clarification of the evolutionary relationship of these proteins will improve current knowledge of *C.velia*, and shed light into the research of nutrient requirements for stressed lifestyle.

Calvin-Benson cycle includes a series of chemical reactions that convert carbon dioxide (inorganic carbon source) into glucose (organic carbon molecule). This cycle takes place in the stroma of chloroplasts, outside of the thylakoid membranes and is driven by the energy generated from light reactions. There are 10 enzymes involved in 13 reactions. The regeneration of RuBP process in Calvin-Benson cycle shared many common reactions with the glycolysis and gluconeogenesis pathway (Figure 4.1). Therefore, enzymes involved in the Calvin-Benson cycle affect both autotrophic and heterotrophic pathways. As an organism reported with possible mixotrophic lifestyle (Forster *et al.*, 2014), *C.velia* may have unique characteristics in carbon utilizing. By studying the enzymes involved in the Calvin-Benson cycle, a complicated origination map of Calvin-Benson cycle in *C.velia* was presented (Table 4.2). This map reflects part of the evolutionary pathway of *C.velia*. Based on the phylogenetic analysis, enzymes in
Calvin-Benson cycle in \textit{C.velia} can be classified into 2 different originations: two enzymes shared homology with cyanobacteria, closed to red algae and green algae; eight enzymes shared homology with proteobacteria, closed to apicomplexa. \textit{C.velia} is an organism close to apicomplexa parasites and has potential mixotrophic features. The study of Calvin-Benson cycle in \textit{C.velia} will certainly be inspirational for the study of other organisms.

This was the first time photosystems in \textit{C.velia} had been studied in response to the iron-stress conditions. Although only a pilot was reported in the chapter 5, the methods developed for investigating the photosynthetic apparatus in \textit{C.velia} under iron-stress condition has been proved to be applicable and will be useful for further studies. Iron-stress is a common nutrient limitation in the ocean environment, which limits the photosynthetic activity of primary producers. The study of response to iron-stress conditions in \textit{C.velia} revealed that there are similar reactions including decreased photosynthesis activities and the pigment composition changes (Figure 5.2 and 5.5). However, the PSI/PSII ratio was increased under iron-stress condition, which is different to other photosynthetic organisms (Allen \textit{et al.}, 2008). Since the whole genome of \textit{C.velia} was finally available from the late 2014, advanced technologies such as Liquid Chromatography-Mass Spectrometry (LC-MS) and Matrix-assisted laser desorption/ionization-time-of-flight mass spectrometry (MALDI-TOF) would be useful for identifying the protein organisation and characterising the regulatory function of photosynthetic complexes under iron-stress conditions.

6.1. Light harvesting protein complexes

An introduction to photosynthetic antenna protein complexes was described in chapter 3. The light harvesting protein complexes (LHC) in \textit{C.velia} have been studied specifically using the phylogenetic analysis method (Pan \textit{et al.}, 2011). LHC are the proteins in charge of light energy absorption and regulating the energy flowing pathways in photosystems. Therefore, LHC plays an important role in photosystem regulation and adaptation, which directly links to the environment changes. Recently, a red shifted antenna complex was discovered under red/far-red light
illumination in *C. velia* (Kotabova et al., 2014; Bina et al., 2014). Phylogenetic analysis revealed that there are three main groups of LHC: FCP related type, red algae like type (LHCr) and LI818 type. Out of 23 retrieved LHC peptides, 17 peptides belong to FCP related type, 3 peptides belong to LHCr group, 1 peptide belongs to LI818 group. According to published information of cDNA database of *C. velia*, FCP related type and LHCr type are the main LHC under normal conditions. The classification of LHC and the potential functions of LHC in *C. velia* need to be verified in the future.

Moore *et al.* (2008) reported that *C. velia* possesses a novel pigment which is suspected to be an isomer of fucoxanthin. Fucoxanthin is an important pigment component of LHC (Fucoxanthin-Chl binding protein complexes, FCP), especially in diatoms. FCP is a type of LHC with high efficiency in photosynthetic reactions. Based on pigment analysis, LHC in *C. velia* use iso-fucoxanthin instead of the original fucoxanthin in FCP type LHC. In this case, a close relationship between the LHC in *C. velia* and diatoms would be expected.

In this work, for the first time, the close relationship between LHC in *C. velia* and diatoms has been revealed (Pan *et al.*, 2011). The majority of LHC in *C. velia* (17 out of 23 peptides) is close to FCP group, which is the major LHC in diatoms. Therefore, for the first time, a close relationship between *C. velia* and diatoms has been demonstrated. This close relationship supports the idea that iso-fucoxanthin has replaced the function of fucoxanthin in *C. velia*. The investigation by isolating photosystems and LHC polypeptides from *C. velia* agrees well with the theory proposed before (Tichy *et al.*, 2013; Pan *et al.*, 2011). Quigg *et al.* examined the photosynthetic capabilities of *C. velia* and demonstrated that *C. velia* possesses a highly efficient photosynthetic system including a highly efficient LHC (Quigg *et al.*, 2012). The finding of multiple gene copies of LHC in *C. velia* in this study agrees well with their conclusion.

### 6.2. Enzymes involved in the Calvin-Benson cycle

In Chapter 4, sequences of enzymes involved in the Calvin-Benson cycle have been analyzed using the phylogenetical method. Calvin-Benson cycle is an important metabolism pathway that
fixes CO₂ and converts it into organic compounds. Due to the fact that most of the enzymes involved in the Calvin-Benson are coded in the nucleus, the sequences of these genes could be retrieved from DNA databases. However, at the beginning of this study, the genome sequencing project of *C.velia* was not completed and no preliminary data was available. Therefore, in order to obtain as much information as possible, *C. velia* sequences were retrieved from several different DNA databases, including unpublished expressed sequencing tag (EST) database and the transcriptome database (see details in section 2.5.5). By the end of 2014, the information of whole genome sequences of *C. velia* is available to public. Petersen et al. (2014) have noticed several phylogenetically interested genes during the sequencing, and the results of their phylogenetic analyses were consistent with the results reported here in chapter 4. We performed phylogenetic analyses on nine enzymes involved in the Calvin-Benson cycle and discussed their phylogenetic relationship, while Petersen et al. only investigated five out of ten enzymes from the Calvin-Benson cycle, which did not reflect the Calvin-Benson cycle as a whole process. The summary of the phylogenetic analysis result has revealed the complicated evolutionary relationships between *C. velia* and other related organisms, avoiding the bias from a single perspective.

Due to the limits of current knowledge on the related genes and protein functions information, the phylogenetic tree constructed for some genes may be very simple and lack of credibility. For example, the RPI and TKL trees constructed were significantly limited by the numbers of available homologous sequences. More information of homologous genes of related organisms can improve the reliability of trees constructed. The original PRK, PGK and TKL sequences obtained from *C. velia* database were not full sequences. Therefore, the three enzymes encoded genes were sequenced using PCR synthesized DNA fragments and extensive sequence searching against NCBI. The reliability of these trees can be confirmed according to updated genome sequences.
6.3. Photosynthetic apparatus response to iron-stress conditions in *C. velia*

In chapter 5, the photosynthetic apparatus of *C. velia* in response to the level of iron nutrients has been studied. The primary result indicated that under iron-stress conditions, *C. velia* demonstrated several physiological changes including pigment composition change, decreased growth rate, decreased photosynthetic activities (inhibited photosynthetic oxygen evolving rate, the re-modeled photosystems and up-regulated additional LHC-related proteins).

*C. velia* cells are reported to have a thick cyst wall that is hard to break (Weatherby *et al.*, 2011). Therefore, in order to obtain enough proteins for the subsequent analysis, I have tested several methods for breaking the cells, including enzyme lyses, French pressure cell press, fast prep cell lyses and Bead-beat cell lyses. The optimized method for breaking the cells is Bead-beater cell lyses, although there is only 60% cells can be broken. Therefore, a large scale of culture is applied to improve the photosynthetic membranes yield. In this study, 20 liters of *C. velia* culture was successfully grown under both normal and iron-stress conditions. Both cultures were inoculated and started with the same number of cells (~3000 cells per ml). However, cells under iron-stress conditions grew much slower than the normal culture, and eventually the yields of harvested cells after about 3 weeks was only 20% of normal culture. To make a comparable result, a similar mass of cells (~5 g, wet weight) from both cultures were used for protein extraction and analysis.

Considering the limited activity of photosynthesis under iron-stress conditions, the amount of PSI and PSII protein complexes obtained from the iron-stress culture was decreased dramatically compared to the normal condition culture, as shown in the sucrose gradient and SDS-PAGE result (Figure 5.7, 5.10 in section 5.3).

In this study, common antibodies against PSI (PsaC), PSII (PsbA) and LHC1 are obtained from the Agisera Company. All of these antibodies are designed for targeting at conserved protein domains across photosynthetic species. PsbA interacts against protein D1 in the reaction centres of PSII, and LHC1 interact against LHC associated to PSI instead of the LHC associated to PSII, although there are multiple LHC peptides (Figure 5.10). PsaC antibody interacts against the PsaC peptide in PSI, which is adjacent to PsaE and PsaD. PsaC, PsaE and PsaD form the docking site...
for ferredoxin and flavodoxin on the stromal side of the thylakoid membrane (Fromme et al., 2003). Under iron-stress condition, PsaC helps docking the flavodoxin instead of ferredoxin, which should be stable independent to the iron-stress conditions. However, no satisfied western blots were obtained by using PsaC antibody, which may due to the less specific binding features of PsaC antibody. A different selection of antibody with a relative higher affinity may be helpful to improve the result.

6.4. Evolutionary relationship of *C. velia* in the chromalveolate group

The chromalveolate hypothesis was firstly proposed by Cavalier-Smith (Cavalier-Smith, 1999). In the theory, four phyla of algae including cryptomonads, haptophytes, stramenopiles, diatoms and alveolates were united together under a branch in the evolution tree. Cryptomonads contain chloroplasts that are recognized as nucleus relic; haptophytes are important species in the algal bloom; stramenopiles (also named as heterokonts) are a diverse group containing more than 100,000 species, with typically unequal flagella. This large family includes diatoms which are the dominant component of marine phytoplankton, which is reported to be close to *C. velia* (Pan et al., 2011); the last phylum is alveolates which are a diverse group as well. Alveolates contain non-photosynthetic species such as the apicomplexa, the ciliates, perkinsids, colpodellids and ellobiopsids. Alveolates also include photosynthetic dinoflagellates. Newly discovered *C. velia* and *Vitrella brassicaformis* CCMP3155 are also located within this group (Moore et al., 2008).

The four phyla of algae within chromalveolate group show distinguishing differences, however, they do share some common features, but not in all groups. Morphologically, cellulose is presented in most cell walls. Phylogenetically, most cells share the same type of glycerol-3-phosphate dehydrogenase (GAPDH) and fructose bisphosphate aldolase (FBA), as well as a novel class of Rab GTPase (Elias et al., 2009).

Based on these similarities, Cavalier-Smith believed that these groups of algae have acquired their chloroplast by a single secondary endosymbiosis of a free-living photosynthetic red alga (Archibald, 2009; Green, 2011). There are several arguments according to a single or a serial of endosymbiosis events (Takishita et al., 2009; Baurain et al., 2010; Keeling 2013). In addition,
Dorrell and Smith argue that this endosymbiosis was happened via green algae rather than red algae (Dorrell and Smith, 2011). Due to the limited number of available sequenced red algae genomes, the hypothesis for lateral gene transfer or endosymbiosysis, and the possibility of lateral gene transfer rather than endosymbiosis are still under debate.

As an organism with unique features and important evolutionary position, *C. velia* has been a hotspot of research since the discovery. Based on the same EST database as this study, Woehle et al. (2011) analyzed the nuclear coded genes in *C. velia* and detected both red and green phylogenetic signals. They argue that a single secondary endosym is not likely; the most possible scenarios should include several secondary endosymbiosis event in chromalveolates. Based on the chloroplast genome analysis, Janouskovec et al. (2011) supported a red algae origination for the chloroplast in *C. velia*. Based on the recently published genome information, Petersen et al. (2014) analysed five enzymes in *C. velia* and proposed that *C. velia* originated from the red algae. Furthermore, they proposed a new theory to replace chromalveolate hypothesis, named as rhodoplex hypothesis. In this hypothesis, not only the four phylums under chromalveolates hypothesis, but also ciliates, rhizaria are united in one group with a red algae origination. These phylums are not united by a single secondary endosymbiosis. Instead, independent secondary endosymbiosis events, in combination with following tertiary and quaternary endosymbiosis events in some lineages, resulted in a complicated algae supergroup.

In this thesis, the evolutionary position of *C. velia* has been analyzed from two perspectives: structural protein (LHC) in light reaction, and enzymes in the Calvin-Benson cycle. Conclusions based on these analyses cannot fully reflect the evolutionary pathway of *C. velia*. Nevertheless, these conclusions contribute to the study of individual protein families, and are helpful for understanding the evolutionary relationship between *C. velia* and other photosynthetic organisms.

In the case of LHC in antenna complexes group, the phylogenetic analysis concluded that the majority of LHC are mostly closed to the unique group of LHC (FCP) in diatoms, although a small portion of LHC (3 out of 23 retrieved LHC) are red lineage originated. Considering the red algae origination of chloroplast in *C. velia* (Janouskovec et al., 2011), it is likely that the *C. velia*
originated from red algae. During the evolution, the common ancestor of diatoms and C. velia has involved the new pigment and developed their unique type of LHC, with a minor trace of ancestral LHC being kept.

Among enzymes involved in the Calvin-Benson cycle in C. velia, the phylogenetic trees indicated that PRK and RPI are more close to green algae. Most of the other genes are unique in chromalveolate groups compared to green and red algae, or have originated from proteobacteria which are common in all three lineages (red, green algae and chromalveolates). Therefore, the debate of origination of chromalveolate remains unsolved. Perhaps the best answer is that chromalveolates are derived from both green and red lineage (Moustafa et al., 2009).

The summary of the phylogenetic analysis (Table 4.2) indicated that C. velia has a complicated evolutionary origination. C. velia is reported to have a form II RuBisCO enzyme (Janouskovec et al., 2011), which is popular in proteobacteria and dinoflagellates. RuBisCO is the speed limiting step for carbon fixation (Ellis, 2010). PGK, GADPH, TPI, FBA and FBP are the enzymes working in the Calvin-Benson cycle, glycolysis and glyconeogenesis pathway; all of them have originated from proteobacteria (Table 4.2). The unique type of RuBisCO, and the preference of proteobacteria originated enzyme in C. velia, indicated that C. velia, like its relative dinoflagellate, has some potential advantages in the carbon-utilization environment. This conclusion fits the report that C. velia grows faster in a mixotrophic media (Forster et al., 2014).

6.5. Further research direction

As the closest photosynthetic organism to photoautotrophic apicomplexan parasites, C. velia possesses a great research potential. This investigation mainly focused on two perspectives of C. velia: the evolutionary development of photosynthetic related proteins, and the photosynthetic response to iron-stress environment in C. velia.

Based on the result, the further possible directions of C. velia research can be separated into different ways:

1. Characterizing the photosynthetic membrane protein complexes in C. velia: In chapter 5, an applicable method for extracting the photosystem complexes from C. velia has been proved
working. Possible future analysis including possible Liquid Chromatography Mass Spectra analysis, two-dimensional gel electrophoresis can be helpful for characterising the photosynthetic related proteins in *C.velia*.

2. The genome of *C.velia* has been released in 2014. Therefore, a deep digging into this database may be helpful for clarifying the evolutionary development of *C.velia*. In fact, the very first paper related to the genome database was focused on the five enzymes involved in the Calvin-Benson cycle (Petersen *et al.*, 2014). In chapter 4, a comprehensive set of analysis of the enzymes involving in the Calvin-Benson cycle has been performed. Similar studies can be performed on key proteins with high evolutionary research values, such as photosystem core proteins, oxidize/reduce functional proteins.

3. *C.velia* possesses a novel and unique carotenoid named as isofucoxanthin, which is close to fucoxanthin based on pigment analysis and phylogenetic analysis. Fucoxanthin has been proved to be helpful for weight loss (Abidov *et al.*, 2010) at certain level. The potential role of isofucoxanthin as a weight loss supplement is interesting.
Reference


156


sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3: 109-136.


Smayda TJ (1997) Harmful algal blooms: Their ecophysiology and general relevance to phytoplankton blooms in the sea. Limnol Oceanogr 42: 1137-1153


Tan S, Cunningham FX, Gantt E (1997) LhcaR1 of the red alga Porphyridium cruentum encodes a polypeptide of the LHCl complex with seven potential chlorophyll a-binding residues that are conserved in most LHCs. Plant Molecular Biology 33: 157-167. doi:Doi 10.1023/A:1005715528297


Appendix

Appendix-1. Modified K+ESM media for *C. velia*

Artificial sea-water:

A) Artificial sea water stocking (200X):

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>KCL</strong></td>
<td>59.648g</td>
</tr>
<tr>
<td><strong>CaCl_2.H_2O</strong></td>
<td>29.97g</td>
</tr>
<tr>
<td><strong>H_3BO_3</strong></td>
<td>0.06g</td>
</tr>
<tr>
<td><strong>Sr(Cl_3).6H_2O</strong></td>
<td>4.00g</td>
</tr>
<tr>
<td><strong>NaF</strong></td>
<td>0.042g</td>
</tr>
<tr>
<td><strong>NaBr</strong></td>
<td>0.103g</td>
</tr>
</tbody>
</table>

Up to 500ml with dd water

B) main salts

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NaCl</strong></td>
<td>23.9604g/L</td>
</tr>
<tr>
<td><strong>MgCl_2.6H_2O</strong></td>
<td>6.05056g/L</td>
</tr>
<tr>
<td><strong>MgSO_4.7H_2O</strong></td>
<td>10.353g/L</td>
</tr>
<tr>
<td><strong>NaHCO_3</strong></td>
<td>0.21g/L (after autoclave)</td>
</tr>
</tbody>
</table>

Add 5ml stocking in 1L water with main salts together
**K+ESM medium**

<table>
<thead>
<tr>
<th>1. Major Nutrients: (1000X stock solution)</th>
<th>Na$_2$-glycerophosphate</th>
<th>8.1g</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NaNO$_3$</td>
<td>200g</td>
</tr>
<tr>
<td></td>
<td>K$_2$HPO$_4$</td>
<td>5g</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bring up to 1000 ml with dd water</td>
<td></td>
</tr>
</tbody>
</table>

| 2. Fe-Mn EDTA: (1000X stock solution)      | FeSO$_4$·7H$_2$O          | 2g (Dissolve this first, not applied in iron stress culture stock solution) |
|                                             | Na$_2$-EDTA               | 2.98g (Dissolve separately and then add together) |
|                                             | MnSO$_4$                  | 1.24g |
|                                             |                           |      |
| (For iron stress culture, use Mn-EDTA instead of Fe-Mn-EDTA) | | |
|                                            | Bring up to 1000 ml with dd water | | |

| 3. Vitamins: (1000X stock solution)        | Thiamine-HCl              | 200mg |
|                                            | Biotin                    | 1.5mg |
|                                            | Vitamin B12               | 1.5mg |
|                                            |                           |      |
|                                            | Bring up to 1000 ml with dd water | | |

| 4. Trace minerals: (1000X stock solution)  | MnCl$_2$·4H$_2$O          | 178mg |
|                                            | ZnSO$_4$·7H$_2$O          | 2.3mg |
|                                            | CoSO$_4$·7H$_2$O          | 1.2mg |
|                                            | Na$_2$MoO$_4$·H$_2$O      | 7.2mg |
|                                            | CuSO$_4$·5H$_2$O          | 2.5mg |
|                                            | Na$_2$-EDTA               | 37g  |
Appendix-2. Detailed information of alignment length, involved species and percentage of similarity for each gene in Calvin-Benson cycle

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Alignment Length (aa)</th>
<th>Numbers of species used in the phylogenetic analysis</th>
<th>Percentage of similarity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PGK</td>
<td>500</td>
<td>53</td>
<td>65</td>
</tr>
<tr>
<td>GAPDH (cytosolic)</td>
<td>360</td>
<td>21</td>
<td>78</td>
</tr>
<tr>
<td>TPI (13G02)</td>
<td>260</td>
<td>26</td>
<td>68</td>
</tr>
<tr>
<td>TPI (01A04)</td>
<td>270</td>
<td>29</td>
<td>50</td>
</tr>
<tr>
<td>TKL</td>
<td>700</td>
<td>24</td>
<td>59</td>
</tr>
<tr>
<td>FBA</td>
<td>380</td>
<td>28</td>
<td>54</td>
</tr>
<tr>
<td>FBP</td>
<td>325</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>RPE (CV337)</td>
<td>230</td>
<td>26</td>
<td>95</td>
</tr>
<tr>
<td>RPE (CV311)</td>
<td>230</td>
<td>32</td>
<td>65</td>
</tr>
<tr>
<td>RPI</td>
<td>230</td>
<td>24</td>
<td>63</td>
</tr>
<tr>
<td>PRK</td>
<td>355</td>
<td>35</td>
<td>48</td>
</tr>
</tbody>
</table>
Appendix-3. The detailed setting of phylogenetic trees constructed in chapter 4

All phylogenetic trees constructed in chapter 4 were using Bayesian method with WAG+G+I model. Metropolis-coupled Markov chain Monte Carlo analyses were run with one cold and three heated chains (temperature set to default 0.2) for 4,000,000 generations and sampled every 250 generations. This process was performed three times from a random starting tree and ran well beyond convergence. Trees before convergence (the first 25%) were discarded for the reconstruction of the consensus Bayesian tree with posterior probabilities.

<table>
<thead>
<tr>
<th>Gene names</th>
<th>Log likelyhood</th>
<th>Harmonic mean</th>
<th>α</th>
<th>variance</th>
<th>Proportion of invariant sites (pinvar)</th>
<th>variance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>arithmetic mean</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGK</td>
<td>-7165.19</td>
<td>-7216.54</td>
<td>1.389</td>
<td>0.03</td>
<td>0.136</td>
<td>0.001</td>
</tr>
<tr>
<td>GAPDH</td>
<td>-9999.71</td>
<td>-10052.41</td>
<td>1.199</td>
<td>0.01</td>
<td>0.150</td>
<td>0.008</td>
</tr>
<tr>
<td>TPI (13G02)</td>
<td>-6618.07</td>
<td>-6651.17</td>
<td>1.055</td>
<td>0.02</td>
<td>0.135</td>
<td>0.001</td>
</tr>
<tr>
<td>TPI (01A04)</td>
<td>-4872.91</td>
<td>-4915.79</td>
<td>1.318</td>
<td>0.05</td>
<td>0.185</td>
<td>0.002</td>
</tr>
<tr>
<td>TKL</td>
<td>-2876.60</td>
<td>-2898.67</td>
<td>0.839</td>
<td>0.04</td>
<td>0.114</td>
<td>0.004</td>
</tr>
<tr>
<td>FBA</td>
<td>-7166.92</td>
<td>-7199.37</td>
<td>1.490</td>
<td>0.04</td>
<td>0.095</td>
<td>0.001</td>
</tr>
<tr>
<td>FBP</td>
<td>-2966.63</td>
<td>-2996.85</td>
<td>1.119</td>
<td>0.04</td>
<td>0.093</td>
<td>0.002</td>
</tr>
<tr>
<td>RPE (CV337)</td>
<td>-5737.52</td>
<td>-5772.12</td>
<td>1.373</td>
<td>0.01</td>
<td>0.097</td>
<td>0.001</td>
</tr>
<tr>
<td>RPE (CV311)</td>
<td>-6139.30</td>
<td>-6181.99</td>
<td>1.681</td>
<td>0.06</td>
<td>0.158</td>
<td>0.001</td>
</tr>
<tr>
<td>RPI</td>
<td>-2307.42</td>
<td>-2329.01</td>
<td>1.906</td>
<td>0.31</td>
<td>0.070</td>
<td>0.002</td>
</tr>
<tr>
<td>PRK</td>
<td>-4039.17</td>
<td>-4070.58</td>
<td>1.307</td>
<td>0.04</td>
<td>0.041</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Appendix-4. Primer constructed for PCR

<table>
<thead>
<tr>
<th>PRK primers-full length</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>CTACCATCTTCGCGAGTGCC</td>
<td>20</td>
<td>60.87</td>
<td>60</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>CGATGTAGCGAGAGATGGGC</td>
<td>20</td>
<td>60.39</td>
<td>60</td>
</tr>
<tr>
<td>Product length</td>
<td>707</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PRK primers-based on conserved region</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>CTGCCTGGATGACTACCACA</td>
<td>20</td>
<td>59.10</td>
<td>55</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>CCTCTCTCTCTATGATCTCTCTG</td>
<td>23</td>
<td>58.92</td>
<td>52.17</td>
</tr>
<tr>
<td>Product length</td>
<td>339</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TKL primers-1</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>CGAGGCTCTTTTGAGGTCTG</td>
<td>20</td>
<td>60.11</td>
<td>60</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>TGTCAGAGAGCAGCGTAG</td>
<td>20</td>
<td>60.04</td>
<td>55</td>
</tr>
<tr>
<td>Product length</td>
<td>843</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TKL primers-2</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>GTCTGGGAATTTTCGAGGC</td>
<td>20</td>
<td>60.11</td>
<td>60</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>GATGTGCAGCAGAGACCTGT</td>
<td>20</td>
<td>60.04</td>
<td>55</td>
</tr>
<tr>
<td>Product length</td>
<td>843</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PGK primers-1</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>AAGAAGTCGTGGAGGAGACCT</td>
<td>20</td>
<td>60.18</td>
<td>55</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>TGGGTCCGTTCCAGACAAATG</td>
<td>20</td>
<td>59.96</td>
<td>55</td>
</tr>
<tr>
<td>Product length</td>
<td>781</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PGK primers-1</th>
<th>Sequence (5'-&gt;3')</th>
<th>Length</th>
<th>Tm</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward primer</td>
<td>TCCAGAGGGTGCTGAAAGA</td>
<td>20</td>
<td>60.18</td>
<td>55</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>GTAACAGACCTCGAGCTGGT</td>
<td>20</td>
<td>59.96</td>
<td>55</td>
</tr>
<tr>
<td>Product length</td>
<td>781</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix-5. Protein sequences of enzymes involved in Calvin-Benson cycle in *C.velia*

**PGK:**

>`CV.PGK

MKTIDDFNFENKKALIRVDFNVPLDADFNVTDTNRIEAAPKTIKVLLEDGGSAILMSHLGRPKGQRNPSSLKHI
CSKVSEIIGVQVFKVDSSVGENVESAASELPGEVLLLENFLFYAEEEKGQGFAEKLRLGDIYVNADFHTA
HRAHASTTIVAQFDQDKCFFGLALAKEIDAEIAEKVRMTGKEPKVGLILGGAKSVSKITIENILDKVDHLIIGGGMT
YTFVKAOQGQGYVGSICDDMKELALELGQAQKKGVEVHLPDSVLAANDENryptoADTVVVEVKIPDGWQGL
DAGPKTLESFREVILKCRTILWNPVGVFEMERFAKGTIAVGNYIDEATQGAFSLVGGDSVAAVKQFGFED
KVSVYSTGGAMLESLEGKTLPGIAAIIGE

**GAPDH:**

>`CV 51

RVSLRFAAPPSTTSSFSLPKMLGKINGFGRGGLVFRAAIEKGSIVTCAINDPMPVDYMLYQLKYDSVHG
RFFPCDVKDGKLVNKGTEVSEKDEAPAAIKWGAAGADVYCESVVFSTSDKALHCXGAKKVIIASAPKD
ATPMFVMGQVHEKYSALKVSNASCTTLAPLAKVINDKFGEVGLMTTVHAMMTATQVTDGPSKGGKD
WRGGRAGANIANPSTGAAKAKGVKIPLENQGLTGMAFRVPTFDSVGDLLTCLAKPKYADIIAIAEKASAG
PMKGVLGW

**TPI:**

>`CV 13G02

NVSNGAYTGEGVSMIKDMINLSSWTLIGHSERSSSYGETDEVVADKVEACQKGGIANAVCIGEVLEERE
GGKTTEEVVKQVEAFIPKVTDSKIVIAYEPWVAITGKVATPEQAOHTACIRKLIKEKCGDAAVAAIVYVG
GSVSNDSCVGLFANEDIDGFLVGGASLKPAFIPVIDSAKAK

>`CV1 01A04

RDDVVALKTRIALDQGLEVMPICGKKEQREAGTTMIDVADQMTALVKLPEDWANLVLAYEPWVAITG
LTAPTHHVQDTQKGIRDWDDSKVSPSIAENVRLYGGSVKAAANAEELFAEPIDIGFLVGASLNKEFIAINAST
TPKKA

**TKL:**

>`cvtkl.cv

MDAVPPASQDDSSNPSMDHQMHRDMANAIARAMDAVQKANSNHGMPMAGMADVATVFLFQFLKFD
AKPDWPRODRFVLSAGHGSMLLALLHTGYEDMTMEILRNFRQLSRTAGHPYHGAAGIETTTGPGGQG
LANAVGMAERLLNARFGDDLVDHHTYVIAGDGCLMEGISHEAISLAGHLKLNKLIVLFDNACICIDGSTDLT
TVSDDQIKRFESAGWAATRIGHDPAIAAIAIAAIAAKSDRPSLIAACKTIGGYAPQNKGTAATHGAPLGDDEIA
GAREALGWPHAPFEPVDDVLKPWRAAGARGAADSAAAWQERLNDTAAEPRAFQKQMAELPGWREALSA
AFKAEVVAAEAPTAVRVSQKTDLTAVAPAMIAGSADLTGSSNTKSKSQAASVSDADDGAYIHYGVREHG
MAAAMNGIALHGGILPYGTFLVFDTYCRPAIRLSMALMKQRVIIYVMTHDSEGIEDGQPVEQIAALRAIP
NLVFRPCDVTETAECWELALDSDEAPSVLALTRQGLPTLRKDGSENHSAYGAYIMVPADSERQTVLASSGSE
VQIAVEAQKMLKEEGISAAVVIPSCEFELFQPPHYRDEVLPGLHIAEIAAPSFGERWIESGGGFVGMRSF
GASAPAKDLKYKDFNITAEAVVEAVKTRF

FBA:
>CV 23-c07
QGRALVFLLSLHTHTVCLFYSSSRHPDPTEFSKSELRKEKMPTLKEALGVGVTGDKVMLLLFEAAAKKHGAIPA
NYTTSSSANAVLEAARDGISSLQASNGGAAFFAGKSVNKEPEQAIIAGAVACALHRQVAPYGIPVVLH
SDHCAKLLPWFDGMLKADEEYFAHKHEPLFSSHMLDLSEETDPENIGTCKAYFEKMAPMKFLEMEIGITGG
EEDGVDNTHADQSKLYTQSPIWAVYETLSKIAPNFS

FBP:
>CV1 contig 221
GTRVQFISHCQSKAMACRYLGSVMADFHRNFLKGGIYIYPPTFEDPKPSISMIFOCNPLAFICEQAGGKASDGFT
RILAMEPTHILHRVCFCCSRKMVEEAEFSMALKKMADGGWCRDEAGAESPVHSNSPSLPLNESPVMYSS
KHRAKV

RPE:
>CV 337
ARGSLQFWPKRTEPSWVESWREDRKEKRNHMKLVSVSFSALAGASAESGAAAFVGSFTCAPAAARKRNLRSSL
KMEMTDGYPTHTGFIAPSALSADFKEGLGEEDVNLQAGADTVHFDVMHNYVPNLTIGPMVCNLARKHGIK
APIDVHLMVDPVDSMVEFIKAGASYITFHPEASRHRVDTLQKJQGGCKAGLVFNPAPTPLDWLRLYICDKVID
VLLMSVPNFPFGQFKQFIPYTLDKLQDKARKIIIQESGRQIRQVDFGGVANNIREVAEAGADMVAGSAIFNTPDY
KETVDAMRKELAAAASKPS
>CV311
RVTSPFCSRFQFRLSRSRFPPGSSLQFTLICPAHFSARVTSGLSRVEQNRQVVRKERRKELHSTFPFIIMPSQIE
EAICPSILSADLSNLTKCQELIKKGADWHLDVMDHGFPVNITFGAPVKNLRKINAPFFDCHLMVSNPQG
WIEDFFKAGAQFHTHELTCGVEKAIELIDRIVATGMKAGITVKPSTPVEDSVFPVLACPGKIHVTLIMTVE
PGFGGQSFMEMMANVE
RPI:

>CV1 18F03

EFPMRSLACALLVSLQCSALLSRARRGFSSLRMSVSQDEMKKNVGYKAVDDYVKSGMLVGLGTGSTVYFAVERLGEKLKSGLDIVGPTSETREQAESLGILPLITLDQRSDIDVAGADEVPKLSSLVKGRGG

ALLREKQIEECAKEFIVIDQSKIQDGLGTDGAMPVEVNKCCDYL

PRK:

>gi|588481381|gb|AHK23661.1| chloroplast phosphoribulokinase [Chromera velia]

MKFTIVATIFASANAGLANLRGRQSAFVLPVGVSHPGSHRDGPATLLEALKALKEKPVVGLAASGCGKTTFMRRIASILEGELKNPLNPKNTDENNQTQVSDLITLIDLDDYHKLDRKGRAETGSLAHPDANDFVTMEKOVVIMVKTEAPKTFDVIYHVTGIDEAEPVLTDLMIEGHLPMFAEGVRQTLDYAIIYLDVDDKFAWKIQRDHEERGHSIESIIKSEGKRPDFSAYIEPQSKSDCVDVCQILPSDFPGANKGRDLRVLRTITPNKAKGHPDFFITPDMIEFNAKDSEGIFVVKSYKESWMGKDADVIDDGDSTINMLIQVDKLGNIEKARVKEGLTKFADAPGSTNASGLLQTLGAMIVKVDNTHGKADAAIL

Appendix-6. Information of sequences used in phylogenetic analysis in Chapter 4

Sequences in PGK alignment: species (accession number in NCBI protein database):

Prochlorococcus marinus (124024984); Triticum aestivum (129916); Acaryochloris marina (158333828); Chlamydomonas reinhardtii (159482940); Clostridium leptum (160932292); Paenibacillus larvae (167461756); Physcomitrella patens (16805247); Chloroherpeton thalassium (501490882); Paulinella chromatophora (194476823); Phaeoaetum tricornutum (219124224; 219126933; 7288203); Chloroflexus aggregans (501684335); Thalassiosira pseudonana (223997336; 224005154; 224014072); Geobacillus sp. (239828276); Cyanobium sp. (254432738); Chondrus crispus (27446627); Veillonella parvula (282850379); Coprococcus catus (291522309); Ectocarpus siliculosus (299115459; 299117084; 299117171; 299470729); Phytophthora infestans (301122065); Volvox carteri (302840184); Chlorella variabilis (30711253); Peptostreptococcus stomatis (307244672); Bacillus sp. (31031865); Bigelowiella natans (32307580); Aureococcus anophagefferens (323453579); Albugo laibachii (325182865); Clostridium sp. (325265027); Prochlorococcus marinus (33864370); Leuconostoc sp. (339491316); Candidatus Arthromitus (342731876); Plasmodium Falciparum (343197194; 343197195); Dorea formicigenerans (346306413); Heterocapsa triquetra (58613475); Isochrysis galbana (58613477); Laminaria digitata (6453561); Synechococcus sp. (86605036); Pismum sativum (9230771); Emiliana huxleyi (JGI accession number: 264369; 272547); Lacinutrix sp. (336172155); Marivarca tructuosa (313677606); Galdieria sulphuraria (545713745)
Sequences in GAPDH alignment: species (accession number in NCBI protein database):

Tetrahymena thermophila (118351857); Plasmodium Falciparum (124810131); Heterosigma akashiwo (13377477); Tetrahymena thermophila (13377481); Blepharisma intermedium (13377483); Paramaecium tetraurelia (13377485); Lepidodinium chlorophorum (168279473); Perkinsus marinus (294943386; 294952693; 294955938); Amphidinium carterae (317135011); Toxoplasma Gondii (342351184); Plectospiro myriandra (34329029); Thraustotheca clavata (34329035); Heterocapsa triquetra (35210478); Scripsiella trochoidea (35210482); Odontella sinensis (52547710); Phaeodactylum tricornutum (6979050); Achlya bisexualis (7274154); Plasmodium yoelii (83317699); Chromera velia (194319782)

Sequences in TPI alignment: species (accession number in NCBI protein database):

13G02:

Arabidopsis thaliana (7076787); Thalassiosira pseudonana (220972769); Phaeodactylum tricornutum (219112703); Tetrahymena thermophila (146162132); Schistosoma haematobium (146741274); Babesia bovis (156084332); Plasmodium vivax (156100529); Leishmania major (157870099); Chlamydomonas reinhardtii (159463610); Schizosaccharomyces pombe (19075524); Zea mays (195605636); Plasmodium Falciparum (217035241); Perkinsus marinus (294938358; 294941714); Phytophthora infestans (301103109); Chlorella variabilis (307105526); Albugo laibachii (325188358); Ichthyophthirius multifiliis (340504670); Solanum chacoense (38112662); Cryptosporidium parvum(66362312); Trypanosoma brucei (71755425); Plasmodium yoelii (82540546); Theileria annulata (84996815); Toxoplasma gondii (92399547); Emiliana huxleyi (JGI accession: 115859)

01A04:

Coprinopsis cinerea (1036782); Caenorhabditis elegans (1036784); Aspergillus terreus (115399504); Scheffersomyces stipitis (126139725); Meyerozyma guilliermondii (146419545); Dugesia japonica (167643879); Homo sapiens (16877874; 226529917); Salmo salar (209734322); Paracoccidioides brasiliensis (225677471); Lepeophtheirus salmonis (225714114); Zea mays (226529672); Mus musculus (226958349); Toxoplasma gondii (237837823); Ectocarpus siliculosus (298712036); Phytophthora infestans (301103109); Chlorella variabilis (307105526); Drosophila melanogaster (3184326); Euglena gracilis (54328433); Oryza sativa
Japonica (553107); Trypanosoma brucei (71755425); Trypanosoma brucei (730975); Euglena intermedia (91075901); Toxoplasma gondii (92399545; 92399547); Emiliania huxleyi (JGI accession: 401226; 438339); Trypanosoma brucei (71755425); Thalassiosira pseudonana (B8C8U5; 224011591; B5YLS7); Spinacia oleracea (P48496)

Sequences in TKL alignment: species (accession number in NCBI protein database):

Toxoplasma gondii (672563370); Plasmodium falciparum (583227485); Chlamydomonas reinhardtii (158281100); Micromonas sp. (226518445); Arabidopsis thaliana (332646583); Galdieria sulphuraria (452823767); Chondrus crispus (507106160); Thalassiosira pseudonana (220977521); Phaeodactylum tricornutum (217410391); Heterocapsa treiquetra (58613541); Synechocystis sp. (451780030); Cyanobacterium aponinum (428682927); Bradyrhizobium elkanii (654879662); Rhodopseudomonas palustris (499759114); Rhodobacterium udaipurense (739380226); Mesorhizobium sp. (563930694); Acetobacter sp. (547265864); Aurantimonas coralicida (737725610); Phaeospiillum molischianum (488818615); Thalassospora permensis (740144319); Sneathiella glossodoripedis (739637766); Kiloniella laminariae (648570016); Rhodovibrio salinarum (653036631)

Sequences in RPE alignment: species (accession number in NCBI protein database):

CV337:

Haemophilus influenza (145631755; P44756); Phaeodactylum tricornutum (219110571); Thalassiosira pseudonana (224004198); Pseudomonas syringae (237803375); Yersinia aldoave (238758718); Acinetobacter radioresistens (255318302); Vibrio parahaemolyticus (28899515); Ectocarpus siliculosus (299469833); Escherichia coli (320197340; P0AG07); Bigelowiella natans (32307592); Aureococcus anophagefferens (323447374); Methylococcus capsulatus (53803277); Oceanospirillum sp. (89094165); Serratia marcescens (P45455); Mycoplasma genitalium (P47358; P75522); Rhodospirillum rubrum (P51013); Helicobacter pylori (P56188); Synechocystis sp. (P74061); Spinacia oleracea (Q43157); Solanum tuberosum (Q43843); Methanocaldococcus jannaschii (Q58093)

CV311

Trichomonas vaginalis (123398284); Giardia lamblia (159119502); Chlamydomonas reinhardtii (159483647); Arabidopsis thaliana (18395962); Phaeodactylum tricornutum (219117749); Homo
sapiens (219879828); Toxoplasma gondii (237835673); Mus musculus (27532955); Perkinsus marinus (294891829); Ectocarpus siliculosus (299473285); Phytophthora infestans (301118582); Volvox carteri (302848131); Micromonas pusilla (303279797); Ostreococcus tauri (308804988); Salpingoeca sp. (326428717); Hordeum vulgare (326515430); Dictyostelium fasciculatum (328866250); Saccharomyces cerevisiae (6322341); Dictyostelium discoideum (66816491); Plasmodium berghei (68006356); Trypanosoma cruzi (71421279); Emiliana huxleyi (JGI accession: 63508); Escherichia coli (P0AG07); Haemophilus influenza (P44756); Serratia marcescens (P45455); Mycoplasma genitalium (P47358); Rhodospirillum rubrum (P51013); Helicobacter pylori (P56188); Synechocystis sp. (P74061); Mycoplasma pneumonia (P75522); Spinacia oleracea (Q43157); Solanum tuberosum (Q43843); Methanocaldococcus jannaschii (Q58093)

Sequences in TPI alignment: species (accession number in NCBI protein database):

Spinacia oleracea (18654317); Arabidopsis thaliana (21592672); Chlamydomonas reinhardtii (159467673); Heterocapsa triquetra (58613535); Oxyrrhis marina (190683032); Synechocystis sp. (16331615); Acaryochloris marina (158335565); Thalassiosira pseudoaonana (223996325); Phaeodactylum tricornutum (219120917); Emiliana huxleyi (264585); Pavlova lutheri (77024233); Cyanidioschyzon merolae (54215034); Plasmodium Falciparum (85544643); Toxoplasma gondii (221503640); Aspergillus fumigatus (159124517); Schizosaccharomyces pombe (6138903)

Sequences in PRK alignment: species (accession number in NCBI protein database):

Emiliana huxleyi (JGI accession: 261032); Chromera velia (588481381); Oryza brachyantha (573920260); Triticum urartu (473968259); Oryza sativa Japonica (115448091); Spinacia oleracea (125579); Vaucheria litorea (13398515); Heterosigma akashiwo (146739106); Arabidopsis thaliana (158338026); Chlamydomonas reinhardtii (159471788); Synechococcus sp. (170079259); Zea mays (19565472); Phaeodactylum tricornutum (219113745); Chloroflexus aggregans (219847575); Thalassiosira pseuadona (223999217); Acidimicrobium ferrooxidans (256370985); Galdieria sulphuraria (27526436); Bigelowiwia natans (32307608); Lepidodinium chlorophorum (338746108); Glaucocystis nostochinearum (349585010); Gymnochlora stellata (349585014); Closterium peracerous-strigosum-littorale complex (349585018); Gamma proteobacterium (356960438); Heterocapsa triquetra (58613469); Isochrysis galbana (58613471); Pavlova lutheri (60101670); Guillardia theta (60101672); Lingulodinium polyedrum (60101674); Pyrocystis lunula (60101678); Prymnesium parvum (60101684); Karlodinium
micrum (77024133); Synechococcus elongates (8099163); Triticum aestivum (21839); Chlorella variabilis (307107355)
Appendix-7. Low temperature fluorescence spectrum of *C.velia*

Figure 1. Low temperature fluorescence spectra of *C.velia* normal culture. NR1, 2, 3, 4 represents the four fractions isolated from sucrose gradient ultracentrifuge. Samples were excited at 435nm, and the fluorescence was recorded between 600nm to 800nm region.

Figure 2. Low temperature fluorescence spectra of *C.velia* iron-stress culture. IR1, 2, 3, 4 represents the four fractions isolated from sucrose gradient ultracentrifuge. Samples were excited at 435nm, and the fluorescence was recorded between 600nm to 800nm region.
Figure 3. Room temperature fluorescence spectra of *C. velia* normal culture. NR1, 2, 3, 4 represents the four fractions isolated from sucrose gradient ultracentrifuge. Samples were excited at 435nm, and the fluorescence was recorded between 600nm to 800nm region.

Figure 4. Room temperature fluorescence spectra of *C. velia* iron-stress culture. I1, I2, I3, I4 represents the four fractions isolated from sucrose gradient ultracentrifuge. Samples were excited at 435nm, and the fluorescence was recorded between 600nm to 800nm region.