THE ACCUMULATION OF CHLOROPLAST SMALL RNAs IN *Chlamydomonas reinhardtii* DURING NITROGEN DEPRIVATION

MOHD SUHAIMI BIN CHE ANI

UNIVERSITI SAINS MALAYSIA

2022

THE ACCUMULATION OF CHLOROPLAST SMALL RNAs IN *Chlamydomonas reinhardtii* DURING NITROGEN DEPRIVATION

by

MOHD SUHAIMI BIN CHE ANI

Thesis submitted in fulfilment of the requirements
for the degree of
Doctor of Philosophy

January 2022

ACKNOWLEDGEMENT

A note of gratitude to all who guided me to complete this study and thesis. Thank you, Allah SWT, for the blessings throughout these years.

Guidance from the supervisor and co-supervisor is the most crucial aspect of the research. To Professor Mohd Nazalan Mohd Najimudin and Dr. Mohd Ghows Mohd Azzam, you are always the best. Your thoughtfulness on the emphasis of the fundamental science and nurturing creative thinking in solving the problem is a gift I will always treasure. Also, thank you for your patience, enthusiasm, motivation, and knowledge. My gratitude also goes to the support I have received from the School of Biological Sciences (SBS). Thank you to all the lecturers who permitted me to use their chemical and lab apparatus. Thank you, too, for all the SBS staff who helped me directly and indirectly. I would like to thank the Ministry of Higher Education for the Fundamental Research Grant Scheme and the Ministry of Education for MyPhD scholarship in terms of financial support. I am also grateful that I was surrounded by supportive colleagues and friends, especially Lab 414 members, who helped me to survive this challenging path. A big thank you to Faisal, Mardani, Haida, Fui Ling, Aidil and Sabrina. Not forgetting, members from SBS Level 4. This journey could not be memorable without all of you.

Finally, I would like to dedicate this thesis to my family, especially my beloved parents, Puan Rodiah Binti Othman and Encik Che Ani Bin Kamis. Their belief in me since the beginning was precious, which I could not find anywhere else.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

%	Percentage
~	approximately
°C	Degree Celsius
aa	Amino acid
asRNA	Antisense RNA
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
BAM	Binary Alignment Map
BLASTn	Nucleotide Basic Local Alignment Search Tool
bp	Base pair
CC	Chlamydomonas Resource Centre
cDNA	Complementary DNA
CDS	Coding Sequence
cosRNA	Clustered organellar small RNAs
cpDNA	Chloroplast genome
Cq	quantitation cycle
cRT-PCR	Circular reverse transcriptase Polymerase Chain Reaction
csRNA	Chloroplast small RNA or Capped-small RNA-seq
CTP	chloroplast-transit peptide
cw	Cell wall
DNA	Deoxyribonucleic Acid
et al.	*et alia* (and others)
EtBr	Ethidium Bromide

h	hour
IGV	Integrated Genomic Browser
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
IR	Inverted repeat
kDa	Kilo Dalton
LA	Luria agar
LB	Luria broth
lncRNA	Long non-coding RNA
M	molar
m	meter
M factor	Maturation/Stability factor
Mb	Megabase pair
MFE	Minimum free energy
min	minute
miRNA	MicroRNA
ml	millilitre
mM	millimolar
mRNA	Messenger RNA
mt-	Mating type (-)
mt+	Mating type (+)
N	Nitrogen
NADH	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
NCBI	National Center for Biotechnology Information
ncRNA	Non-coding RNA
NEP	Nucleus-encoded polymerase

nm	nanometer
nt	nucleotide
NTC	No Template Control
OD	Optical density
OPR	Octo-trico Peptide Repeat
ORF	Open reading frame
OTAF	Organelle Trans-Acting Factors
PCR	Polymerase chain reaction
PEP	plastid-encoded polymerase
PPR	Penta-trico Peptide Repeat
PSI	Photosystem I
PSII	Photosystem II
qRT-PCR	real-time quantitative Polymerase Chain Reaction
R^2	coefficient of determination
RBP	RNA-binding proteins
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNase	Ribonuclease
rpm	Rotation per minute
rRNA	Ribosomal RNA
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
s	second
SAG	The Culture Collection of Algae at the University of Göttingen, Germany
siRNA	Small interfering RNA

snRNA	Small nuclear RNA
srcDNA	Small RNA complementary DNA
sRNA	Bacterial small RNA
srSP	small-RNA specific primer
T factor	Translation factor
TAP	Tris-Acetate-Phosphate
TAP-N	Tris-Acetate-Phosphate lacking nitrogen source
TM	Trademark
TPR	Tetra-trico Peptide Repeat
tRNA	Transfer RNA
UTEX	University of Texas Collection
UTR	Untranslated region
V	Volt
X-gal	5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside
µg	microgram
µl	microliter
µmol	micromolar

LIST OF APPENDICES

PENGUMPULAN RNA KECIL KLOROPLAS DALAM

***Chlamydomonas reinhardtii* SEMASA PENGURANGAN NITROGEN**

ABSTRAK

Kesan ketersediaan nitrogen kepada bioenergetik kloroplas sangat bermanfaat dan sangat signifikan yang perlu diterokai. RNA-RNA kecil telah dijumpai dari genom kloroplas, namun kewujudan maklumat adalah terhad berkenaan tindak balas mereka semasa pengurangan nitrogen. Dalam penyelidikan ini, penglibatan RNA-RNA kecil diterokai untuk mendapatkan pemahaman yang lebih baik mengenai fungsi mereka dalam mekanisme pengawalaturan alga semasa penyesuaian terhadap tekanan. Tinjauan transkriptomik perbandingan telah dijalankan untuk mencari RNA-RNA kecil kloroplas yang bertindak balas kepada perlucutan nitrogen di dalam mikroalga *Chlamydomonas reinhardtii*. Pertumbuhan di dalam media yang dikurangkan nitrogen menunjukkan 22 RNA-RNA kecil kloroplas terkumpul tinggi dan 18 terkumpul rendah secara signifikan. Lokasi genomik RNA-RNA kecil ini adalah pelbagai, yang merangkumi kawasan tidak tertranslasi, arah pengekodan erti atau antierti, atau hujung gen-gen tRNA dan rRNA. RNA-RNA kecil dari hujung gen-gen tRNA dan rRNA semuanya terkumpul rendah, sementara mereka yang berasal dari kawasan erti gen pengekodan protein kesemuanya terkumpul tinggi. Analisis kuantitatif masa nyata PCR (qRT-PCR) mengesahkan pengekspresan empat RNA-RNA kecil kloroplas responsif nitrogen terkumpul tinggi dan empat lagi terkumpul rendah. Pembezaan pengumpulan cosRNAs menunjukkan kepentingannya di dalam kloroplas semasa perubahan tahap nitrogen. Jujukan tiga RNA kecil kloroplas yang dipilih, cosRNA143, cosRNA148 and cosRNA206, dipelihara di dalam genom-genom kloroplas beberapa spesies mikroalga. Struktur gegelung

batang pin rambut mereka yang stabil juga adalah seperti yang terdapat di dalam RNA kecil pengawalatur yang selalu dikaji. Analisis pengekspresan relatif transkrip pengekodan protein dan protein pengikat RNA yang berkaitan dengan mereka juga telah dilakukan. RNA kecil cosRNA143 yang terkumpul tinggi menunjukkan hubungan songsang dengan transkrip pengekodan protein *petG* yang berkaitan. Tambahan pula, ia diikuti dengan tahap pengkespresan tinggi protein pengikat MCG1 yang berkaitan. Protein ini melindungi kawasan kecil transkrip semasa kemerosotan transkrip pengekodan protein dan akhirnya mendorong kepada pengumpulan cosRNA143. Kesimpulannya, penemuan ini menunjukkan pengumpulan cosRNA disebabkan oleh pemprosesan RNA ketika pengurangan nitrogen. Ini menunjukkan potensi fungsi kawalatur mereka dalam penyesuaian terhadap tekanan.

THE ACCUMULATION OF CHLOROPLAST SMALL RNAS IN *Chlamydomonas reinhardtii* DURING NITROGEN DEPRIVATION

ABSTRACT

The effects of the availability of nitrogen to chloroplast bioenergetics are beneficial and very significant that need to be explored. Small RNAs have been identified from the genome of chloroplasts, yet limited information exists on the response of small RNAs to chloroplasts during nitrogen deprivation. In this study, the participation of small RNAs is explored in order to gain a better understanding of their roles in algae's regulatory mechanisms during stress adaptation. A comparative transcriptomic survey was conducted to find chloroplast small RNAs that respond to nitrogen deprivation in the microalgae *Chlamydomonas reinhardtii*. Growth in nitrogen-deprived medium revealed 22 significantly over-accumulated and 18 less-accumulated chloroplast small RNAs. The genomic locations of these small RNAs varied covering regions that were either untranslated, or coded in the sense or antisense direction, or the ends of the tRNA and rRNA genes. Small RNAs from ends of tRNA and rRNA genes were all less-accumulated while those derived from the sense region of protein-coding genes were all over-accumulated. Quantitative real-time PCR (qRT-PCR) analysis validated the expressions of four over-accumulated and four less-accumulated nitrogen-responsive chloroplast small RNAs. The differential accumulation of cosRNAs indicated their significance in chloroplasts during changes in nitrogen levels. The sequences of three selected chloroplast small RNAs, namely cosRNA143, cosRNA148 and cosRNA206, were conserved in the chloroplast genomes of a few species of microalgae. Their stable stem-loop hairpin structures were also similar to that of the commonly studied regulatory small RNAs.

Relative expression analysis of small RNAs associated with protein-coding transcripts and RNA binding proteins were also performed. The over-accumulated cosRNA143 showed inverse correlation with its associated protein-coding transcripts of *petG*. Moreover, it was accompanied by a relatively high expression level of the associated RNA-binding protein MCG1. This protein protects short regions of transcripts during the degradation of protein-coding transcripts and eventually generates an accumulation of cosRNA143. In conclusion, these findings demonstrated the accumulation of cosRNAs due to RNA processing during nitrogen deprivation. This indicates their potential regulatory role in mediating stress adaptation.

CHAPTER 1

INTRODUCTION

1.1 Research Background

The unicellular green alga *Chlamydomonas reinhardtii* has been extensively used as a model system for studying the chloroplast's molecular biology, photosynthesis, nutrient homeostasis, and cell motility (as reviewed in Weeks, 1992; Rochaix, 1995; Harris, 2001). *Chlamydomonas reinhardtii* has a single large cup-shaped chloroplast, an organelle believed to have evolved from an endosymbiotic process. The chloroplast of *C. reinhardtii* is semi-autonomous with 83 copies of a circular genome of 205,535 base pairs (Maul *et al*., 2002; Gallaher *et al*., 2018). This chloroplast genome encodes about 72 protein-coding genes, a full complement of 29 transfer RNAs (tRNA), and five ribosomal RNAs (rRNAs). Most of these genes are involved in the formation of the photosynthetic apparatus and the transcriptional and translational factors of the chloroplast genome (Smith and Lee, 2009). Chloroplasts also require many proteins imported from the nucleus to carry out functions such as photosynthesis and synthesis of chlorophyll, fatty acid and amino acids (as reviewed in Terashima, 2011; Salomé and Merchant, 2019). In addition, these nuclear-encoded proteins are also important for chloroplast genome processing, RNA processing and stability, protein synthesis, and splicing (Rochaix, 2001, Herrin and Nickelsen, 2004).

Small RNAs are short non-coding RNA molecules that are not translated into proteins. They show multifunctional roles such as regulation of genes at the transcriptional and post-transcriptional level, DNA replication, DNA splicing,

mRNA translation, and genome protection (as reviewed Mattick and Makunin, 2006; Cech and Steitz, 2014). Small RNA molecules located in the nuclear and cytosolic compartments, especially microRNA, small interfering RNA and piwi-RNA, have been extensively studied. Small RNA molecules are present in the chloroplasts of microalgae and higher plants (Goldshmidt-Clermont *et al*., 1991; Lung *et al*., 2006; Georg *et al*., 2010). Observations on small RNA libraries extracted from purified chloroplasts have confirmed chloroplast-encoded small RNAs, and these were later named clustered organellar small RNAs (cosRNAs) (Lung *et al*., 2006; Ruwe *et al*., 2016). Many investigations are underway to understand the significance of these molecules in the chloroplast.

Accumulation of cosRNAs may be due to the action of RNA-binding proteins (RBPs) during RNA processing in the chloroplast (Lung *et al*., 2006; Ruwe and Schmitz-Linneweber 2012; Ruwe *et al*., 2016; Cavaiulo *et al*., 2017). Relaxed initiation and inefficient termination of transcription in the chloroplast genome result in full transcripts generated from both chloroplast DNA strands (Hotto *et al*., 2012; Shi *et al*., 2016). RBPs may play a significant role in protecting the termini of transcripts from subsequent nucleolytic degradation by binding to a specific short sequence. Eventually, this act of protection by RBP induces the accumulation of small RNA footprints, which can be detected by small RNA sequencing. These RNA footprints are the cosRNA molecules. Consequently, cosRNA molecules are remnants of transcript processing by RBPs.

In chloroplast RNA processing, the nuclear-encoded ‘Organelle Trans-Acting Factors’ (OTAFs) that control maturation/stability (M factors) and translation (T

factors) are among the widely characterized RBPs. These proteins carry tandem repeats of degenerated motifs of 35, 34, and 38 amino acids (aa) length and are abbreviated as helical repeat protein families of Penta-, Tetra- and Octo-tricoPeptide Repeats (PPR, TPR and OPR), respectively (reviewed by Small and Peeters, 2000; Schmitz-Linneweber, 2008; Tourasse *et al*., 2013). In *C. reinhardtii*, the dominant RBPs come from the OPR family and are involved in the processing and stabilisation of specific chloroplast RNAs (Eberhardt *et al*., 2011; Rahire *et al*., 2012; Loizeau *et al*., 2014; Wang *et al*., 2015; Viola *et al*., 2019). Their presence demonstrates that OPRs play a vital role in chloroplast metabolism of this unicellular microalgae.

Differential accumulations of cosRNAs have been demonstrated in different plant tissues in tomatoes, barley and *Arabidopsis* (Itaya *et al*., 2008; Hackenberg *et al*., 2013; Cognat *et al*., 2017), during heat stress in *Arabidopsis* and Chinese Cabbage (Georg *et al*., 2010; Wang *et al*., 2011) and phosphorus deficiency in barley (Hackenberg *et al*., 2015). These observations suggested that the accumulation of cosRNAs is affected by either developmental stages of growth or environmental conditions.

Nitrogen (N) deprivation has been shown to generate changed physiological reactions in *C. reinhardtii*, and these include variations in chloroplast expression, photosynthetic apparatus synthesis, transcription and metabolic profiling (Peltier and Schmidt 1991; Miller *et al*., 2010; Boyle *et al*., 2012; Msanne *et al*., 2012; Juergens *et al*., 2015). It was suggested that photosynthetic fluxes were sufficiently down-regulated during N deprivation to prevent stress from potential oxidative damage. Several transcriptomic studies have shown that N deprivation has led to the down-

regulation of transcripts involving structural and functional photosynthetic machineries (Juergens *et al*., 2015; Park *et al*., 2015).

Observing N deprivation as a condition affecting the expression of the chloroplast genes might explain the functionality of cosRNAs further. Since N starvation is known to have a major effect on chloroplasts' bioenergetics, these cosRNAs datasets are of important interest to the community. There is little knowledge on the response of cosRNAs and their involvement to N deprivation. The roles of small RNAs should be explored to gain a better understanding of their roles in algae's regulatory mechanisms during stress adaptation. It motivates a search for cosRNAs associated with N deprivation in *C. reinhardtii*. They might give hints on the significance of RNA processing in chloroplast to endure and adapt to stress.

The fundamental questions of this research can be summarized as below:

1. What is the optimum condition of N deprivation in *C. reinhartii*?
2. Does N deprivation affect the accumulation of cosRNAs? If yes, which cosRNAs are responsive toward N deprivation stress?
3. What are the structural characteristics of these N-responsive cosRNAs?
4. What are the potential significant of these N-responsive cosRNAs in chloroplast gene expression during this stress condition?

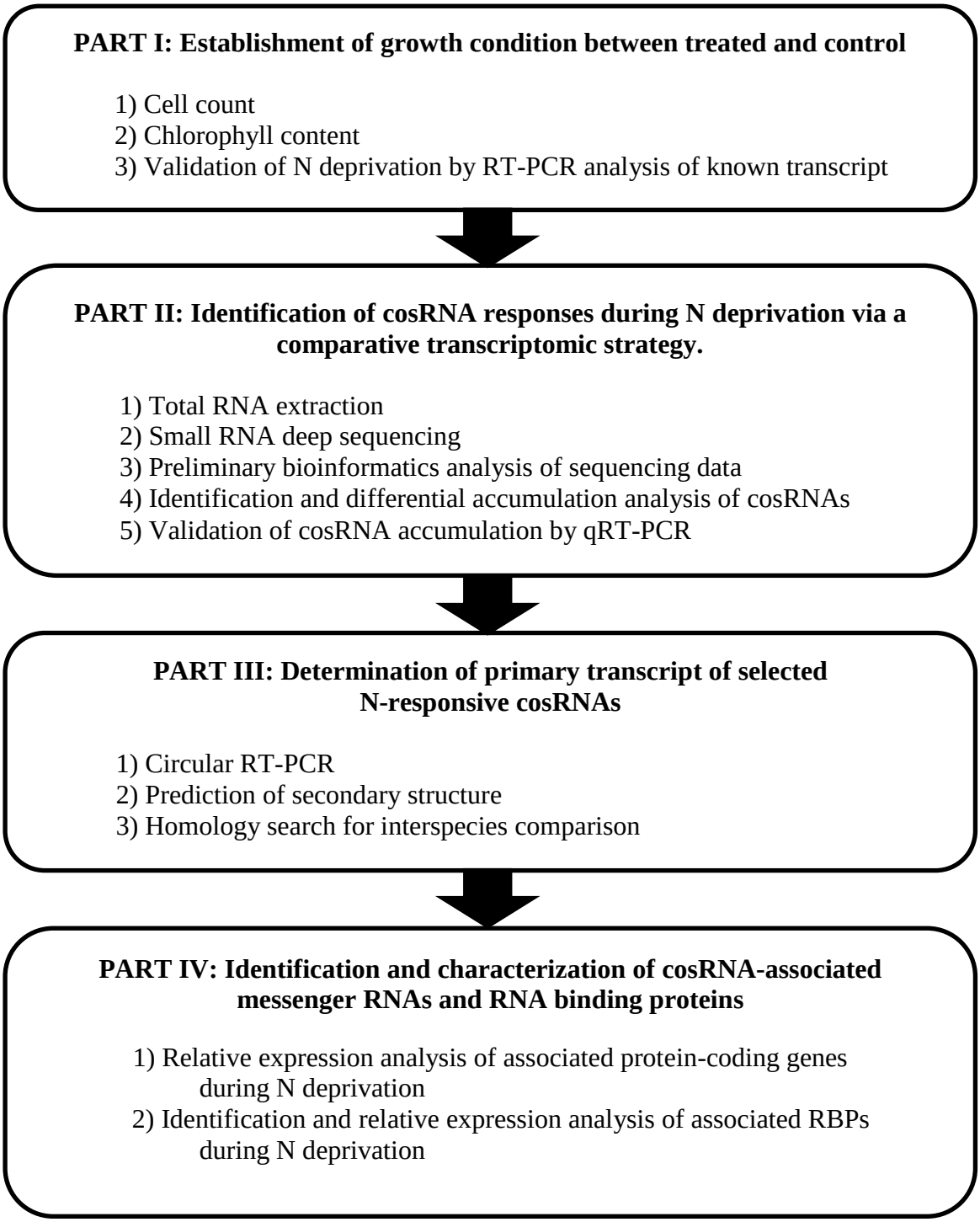
1.2 Research Objective

The aim of this thesis is to observe the accumulation of cosRNAs and identify their response during N deprivation to get a better knowledge of RNA processing in chloroplast through this physiological stress. This study might lead to the discovery of new algal regulatory systems that regulate stress adaption.

The aim of this study was achieved by performing the following specific objectives:

1) Establishment of growth condition between treated (algae grown in N-depleted medium) and control (algae grown in N-based medium).
2) Identification of cosRNA responses during N deprivation via a comparative transcriptomic strategy.
3) Structural characterization of the full-length transcript of selected N-responsive cosRNAs.
4) Identification and characterization of cosRNA-associated messenger RNAs and RNA binding proteins.

1.3 Research Flowchart



CHAPTER 2

LITERATURE REVIEW

2.1 Biology of *Chlamydomonas reinhardtii*

Chlamydomonas reinhardtii is a eukaryotic photosynthetic unicellular microalga under the family Chlamydomonadaceae. It has two flagella that are used to move. The genus *Chlamydomonas*, named by Ehrenberg in 1833, comes from two Greek words *"chlamys"* and *"monas"*, which means "a cloak" and "solitary", respectively (Harris, 2001). The names were based on these unicellular flagellates' visible physical characteristics as initially described by (Gerloff, 1940 as cited in Harris, 2001) (Figure 2.1A). The species name of *C. reinhardtii* was named after Ludwig Reinhardt, a Ukrainian botanist after his discovery on a new isolate while studying the sexual reproduction in different *Chlamydomonas* species (Dangeard, 1900 as cited in Harris, 2001).

2.1.1 The cell structure of *C. reinhardtii*

A vegetative *C. reinhardtii* cell's varies from five to ten micrometers in diameter (Gallaher *et al*., 2015). The cellular architecture of all species in *Chlamydomonas* shares a basic plan. They have an ellipsoid shape with a clear polar structure. These chlorophytes have two key features: two anterior flagella of equal length and a single sizeable cup-shaped chloroplast. The two flagella are used for cell motility and mating (Harris, 2001). The single basal chloroplast accommodates a massive extent of the volume of the cell. This oxygenic photosynthetic organelle contains one or more starch-containing pyrenoids for carbon-concentrating mechanisms (Mackinder *et al*., 2016). Like plant cells, *C. reinhardtii* cells also have

a distinct cell wall with a variable thickness within the strain (Imam *et al*., 1985). Besides, the cellular structures of *C. reinhardtii* also include the nucleus, eyespot, Golgi vesicles, starch grains, and vacuoles (Figure 2.1B). The details of the organelles can be easily spotted by transmission electron microscopy (Figure 2.1C).

2.1.2 The cultivation of *C. reinhardtii* in laboratory

The most widely studied laboratory species of *C. reinhardtii* cells take about 8 hours (h) to double in numbers. They can grow well in the following culture conditions: temperature of 20-25 ^{o}C, light intensities of 200-400 μmol m^{-2} s^{-1}, and continuous agitation for aeration and avoidance of self-shading (as reviewed in Harris, 2009). They can be grown phototrophically (synthesizing food by photosynthesis), heterotrophically (addition of carbon supply in media under total darkness), or mixotrophically (culture in light with added carbon source). *C. reinhardtii* can be grown in either liquid or agar media in the laboratory (Figure 2.2).

2.1.3 Life cycle of *C. reinhardtii*

Both asexual and sexual reproduction can occur in the life cycle of *C. reinhardtii*. Under normal circumstances, the asexual reproduction takes place by the formation of the zoospore. In this zoospore formation, the parent cells undergo simple mitotic division to form to daughter cells (Figure 2.3A). The sexual cycle involves gametogenesis (formation of gametes by meiosis) and mating (gamete copulation) as shown in detail in Figure 2.3B. After pair formation with opposite mating types (mt+ and mt-), the cytoplasmic, nuclear, and chloroplast fuse to form a diploid vegetative zygote. Eventually, the vegetative zygote undergoes mitosis to form a stable diploid cell (as reviewed in Bishop, 2003). In the laboratory, the sexual

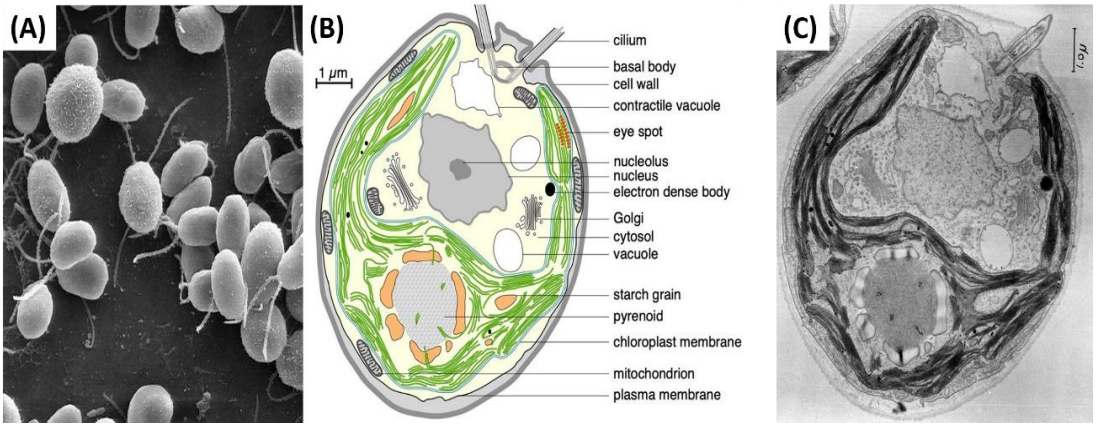


Figure 2.1 Cell structure of *C. reinhardtii*.
(A) Scanning Electroscope Image of *C. reinhardtii* cells (Ramakrishan *et al*., 2014). (B) A labelled diagram of *C. reinhardtii* cell, showing the position of each organelle (Salome and Merchant, 2019). (C) Transmission Electroscope Image of *C. reinhardtii* cell used for drawing of (B). The image is available on Cell Image Library website under ID CIL: 37252 (initially published by Ohad *et al*., 1967).

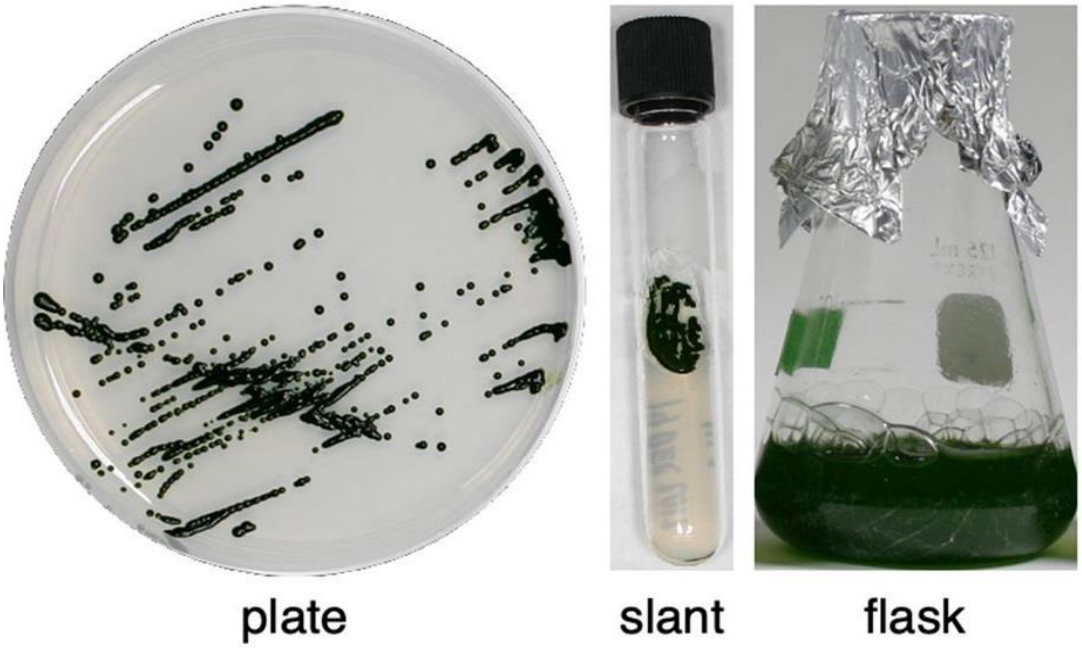


Figure 2.2 The cultivation of *C. reinhardtii* in the laboratory. (The image was adapted from Salomé and Merchant, 2019).

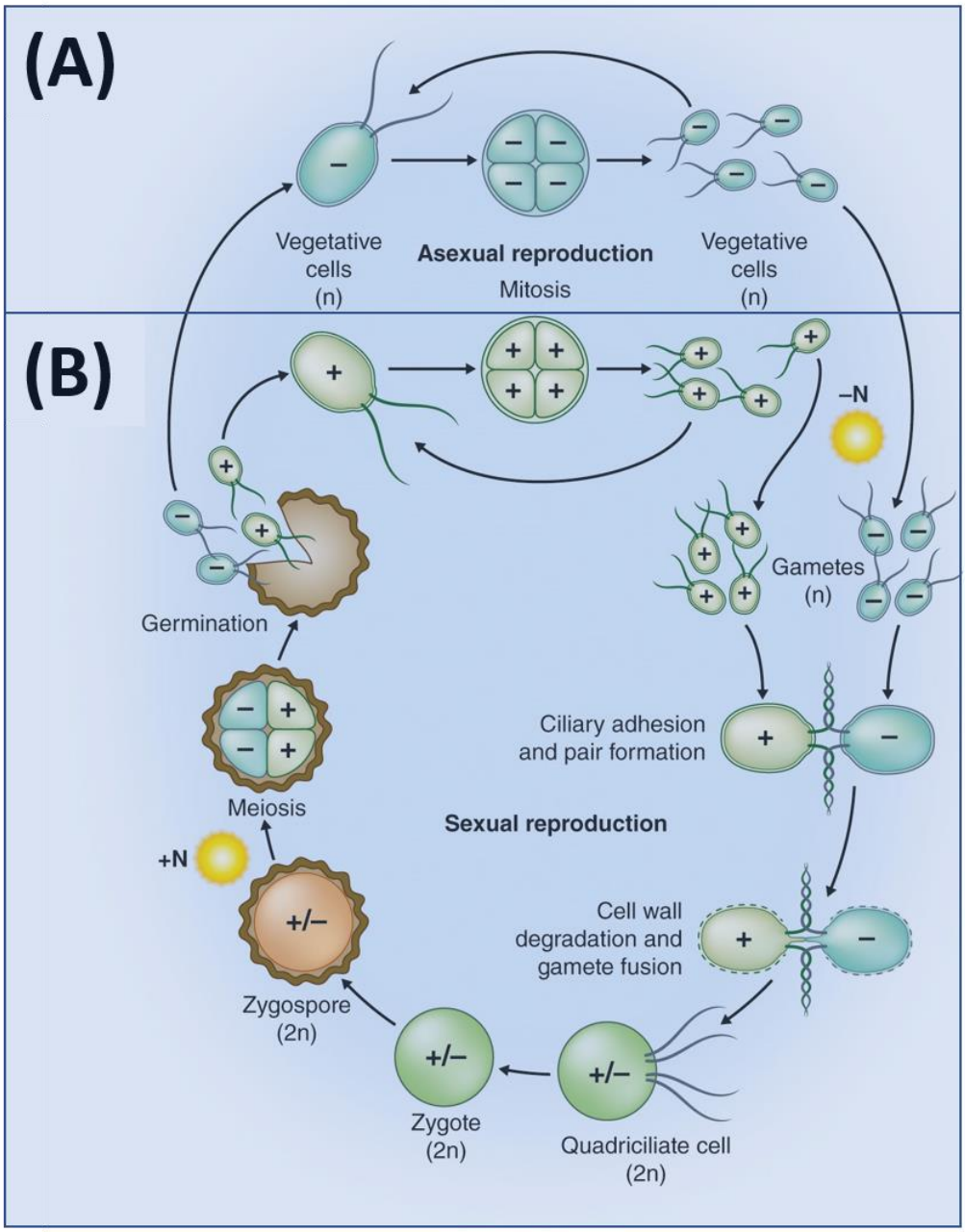


Figure 2.3 Cell cycle of *C. reinhardtii*.
(A) Asexual reproduction involves a mitosis process of haploid vegetative cells to produce four haploid daughter cells.
(B) Sexual reproduction involves the subsequent processes of gametogenesis by inducing nitrogen deprivation (-N), fusion gamete of opposing types (mt+ and mt-), and a meiosis process of the zygote to produce four haploid cells. (The image was adapted from Sasso *et al*., 2018).

reproduction cycle of *C. reinhardtii* is initiated when an appropriate condition such as nitrogen deprivation is specified (Ferris and Goodenough 1994). However, it remains uncertain how the sexual reproduction of this organism is induced in nature.

2.1.4 *Chlamydomonas reinhardtii* genome

As with other photosynthetic eukaryotes, three independent genetic systems exist in *C. reinhardtii*. These are located autonomously in the nucleus, the chloroplast, and the mitochondria. The nuclear genome was first sequenced by (Merchant *et al*., 2007) from strain CC-503 cw92 mt+, a clean wall-deficient mutant. The main nuclear genome was updated in the website *Phytozome 13* as *Chlamydomonas reinhardtii v5.6* (https://phytozome-next.jgi.doe.gov/info /Creinhardtii_v5_6; August 8, 2020) and released to the NCBI website on 2017 BioProject: PRJNA12260 and accession ID: ABCN00000000. The main assembly of the draft nuclear genome was approximately 111.1 Mb arranged on 17 chromosomes and 37 minor scaffolds. It contained 17,741 total loci of protein-coding transcripts. The chloroplast and mitochondrial genomes were smaller, with 205,535 bp and 15,790 bp, respectively (updated by Gallaher *et al*., 2018). These organellar genomes comprise around 83 copies (for chloroplast) and 130 copies (for mitochondria) per cell as deduced by high coverage shotgun sequencing analysis of four Chlamydomonas strains (Gallaher *et al*., 2018). The format for gene symbols in *C. reinhardtii* is a 3–5 letter root, followed by a number denoting isoform, or occasionally subunits. The letter roots were italicized, written in uppercase for nuclear genes or lower case for organelle genes. For example, AMT4 and *atpA* are the format of gene symbols for nuclear gene and organelle gene, respectively.

2.1.5 *Chlamydomonas reinhardtii* as a model organism

Practice of *C. reinhardtii* as laboratory strains started in 1945. Three wild-type strains have been widely used for research in genetic and biochemistry. These strains (137+, 137-, 137c) are supposedly derived from a single isolate from a soil sample taken from a potato field near Amherst, Massachusetts (Proschold, 2005). Since then, dozens of laboratory strains of *C. reinhardtii* was derived from this 1945 isolate of G.M. Smith. Figure 2.5 summarized all dominant laboratory strains of *C. reinhardtii*, most of which are publicly available from the University of Texas Algal Collection Chlamydomonas Resource Centre (with strains designated as CC), The Culture Collection of Algae at the University of Göttingen, Germany (designated as SAG strains) and American Type Culture Collection (ATCC strains). Besides, the libraries of mutants also have been collected and characterized by these organizations for research purposes.

Chlamydomonas reinhardtii serves as an outstanding model for investigating of many biological aspects such as photosynthesis, chloroplast biology, cell motility, nutrient homeostasis, and cell cycle control (as reviewed in Noordally and Millar, 2015). It has been dubbed as the "green yeast" since it shares many experimental references with the budding yeast *Saccharomyces cerevisiae* (Goodenough, 1992; Rochaix, 1995) For years, it has been used to elucidate the fundamental cellular processes. It has currently become a promising model for studying the sensory of cilia and as a bio-based factory for high-value products (as reviewed in Salomé and Merchant, 2019).

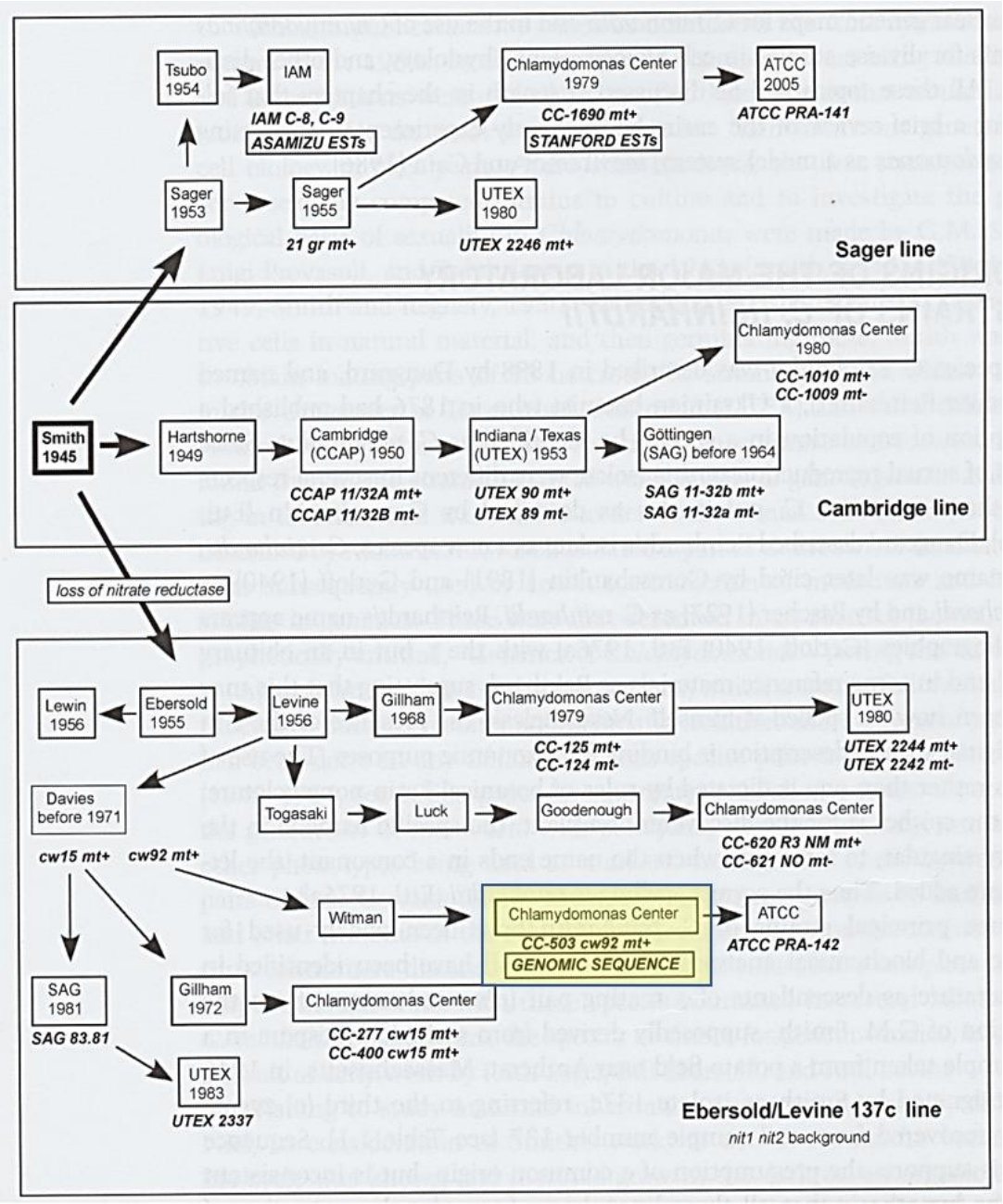


Figure 2.4 The origin of significant laboratory strains of *C. reinhardtii*.
The strains were represented by the first author and year of the published journal. The strain CC-503 cw92 mt+, used for genome sequencing and in the experimental procedure of this thesis, was highlighted in the yellow box. (The image was adapted from Harris, 2009). Originally published in (Proschold *et al*., 2005), which provides information on molecular analysis.

Due to the presence of only a single chloroplast per cell, *Chlamydomonas reinhardtii* serves as the right candidate for a study of this organelle. It is easier to detect any change that happens to the chloroplast because of this singularity. Besides, it is easy to grow this organism. It requires very little space for growth and has a short generation time compared with higher plants. Its whole genome, including that of the chloroplast, had been sequenced, yielding an immense level of information on the genes and their physical locations (Jiao *et al*., 2004; Merchant *et al*., 2007; Gallaher *et al*., 2018).

2.2 Chloroplast

A chloroplast is an organelle found in plant cells and eukaryotic algae that carry photosynthesis. According to the commonly acknowledged theory of endosymbiosis, chloroplasts were once prokaryotic microbes of cyanobacteria. The rationale was that some DNA in the pro-chloroplast genome had been transferred to the nucleus. Genome sequencing projects of plants have indisputably shown that the extant chloroplast DNA fragments are present in both the nuclear and chloroplast genomes (Farrelly and Butow, 1983; Scott and Timmis, 1984; Rousseau-Gueutin *et al*., 2011).

2.2.1 The function and structure of chloroplasts

Chloroplasts have a central role within the metabolic systems of plant cells. This organelle is where photosynthesis takes place and is responsible for much of the world's primary productivity. This reason makes chloroplasts essential to the lives of plants and animals alike. Other key metabolic activities occur in chloroplasts such as starch production and biosynthesis of amino acids, lipids and several colourful

pigments in flowers and fruits (as reviewed in Roston *et al*., 2018). Chloroplast, usually green in colour, is a type of plastid, a group physiologically related organelles found in all types of plants and alga (Wise and Hoober, 2007). The plastids group also includes chromoplast (plastid that accumulates non-green coloured pigments in flowers and fruits) and leucoplast (colourless plastids usually found in rapidly growing tissues). Chloroplasts have a unique structure to carry specialised function. The piles of thylakoids, called grana, are located inside the chloroplast. These structures place the chlorophylls, which are important for photosynthesis (as reviewed in Jensen and Leister, 2014).

2.2.2 The genome of chloroplast

In all plants and algae, the chloroplast contains a genome that encodes an essential subset of the necessary components for photosynthesis. It provides a genome that is a relic of the endosymbiont that gave rise to the organelle. Every chloroplast genome takes a single circular molecule of DNA of a unique sequence, usually designated as cpDNA or called plastome. It only contains about 50–200 genes, mostly for chloroplast proteins and the protein-synthesising machinery. These proteins of known functions include subunits of ribosomes, RNA polymerase (all core subunits), chlorophyll synthesis, Photosystem I (PSI), Photosystem II (PSII), ATP synthase, cytochrome b_6f complex, nicotinamide adenine dinucleotide (NADH) dehydrogenase (Woodson and Chory, 2008).

In *C. reinhardtii*, the circular genomic cpDNA exists in 83 copies per cell. As shown in Figure 2.5, the current updated size of cpDNA in *C. reinhardtii* (was named CPv4) is about 205,535 bp after Gallaher *et al*. (2018) sequenced and revised

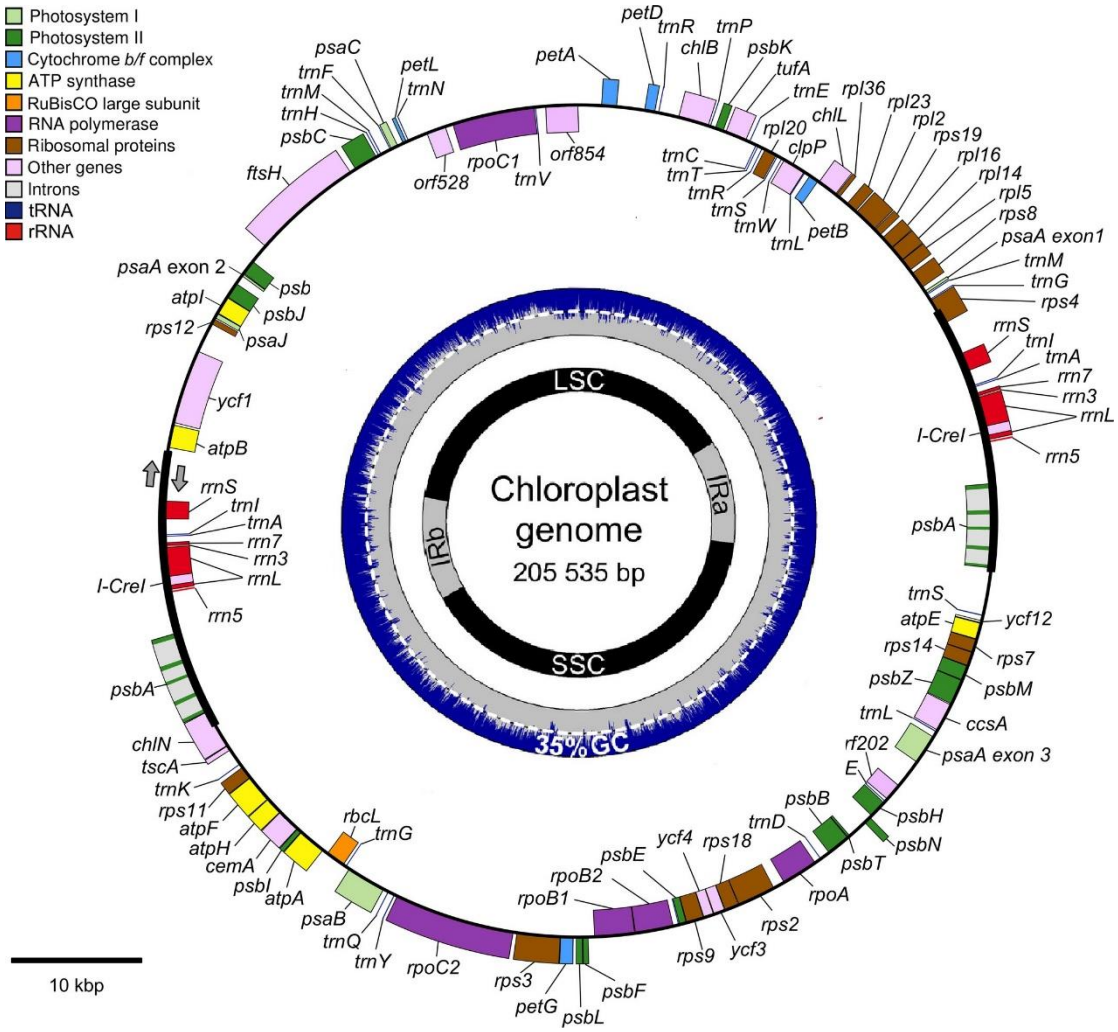


Figure 2.5 The chloroplast chromosome of *C. reinhardtii* (CPv4).
The circular genome is presented as a series of concentric rings. From outer to the inner ring, they are as follows:
1) Gene models are presented as a series of coloured boxes. Genes transcribed in a clockwise fashion are placed outside of the line, and genes transcribed counter clockwise are placed inside of the line. Categories of genes are colour-coded according to the accompanying legend,
2) A plot of the GC content in non-overlapping 25 bp windows with blue for %GC and grey for %AT.
3) Next, a ring to indicate the position of the two inverted repeats (IRa and IRb), the long single copy region (LSC), and the short single copy region (SSC).
A scale bar in the lower right indicates ten kbp.
(The image was adapted from Gallaher *et al.*, 2018).

dozens of errors in the previous genome sequences and annotations (Maul *et al*., 2002; Smith and Lee, 2009, Gallaher *et al*., 2018). The genome structure follows the typical structure in most chloroplast genomes of photosynthetic eukaryotes, which is quadripartite with two regions of single-copy and two large inverted repeats (IRs) (Palmer, 1991; Pombert *et al*., 2006). This chloroplast genome encodes about 75 protein-coding genes, a full complement of 29 transfer RNAs (tRNAs), and five ribosomal RNAs (rRNAs). Most of these genes are involved in the formation of the photosynthetic apparatus as well as transcriptional and translational factors of the chloroplast genome. Figure 2.5 summarises all protein-coding genes and functional RNA genes found in the cpDNA of *C. reinhardtii.*

2.2.3 Full Transcription of the chloroplast genome

The information in the cpDNA is copied into a new molecule of messenger RNA (mRNA) during transcription. This process is facilitated by two types of polymerase: the prokaryotic-type plastid-encoded polymerase (PEP) and nucleus-encoded polymerase (NEP) (summarised in Hess and Borner, 1999). The features similar to the prokaryotic transcriptional process involving polycistronic operons were initially believed to control the transcription of the chloroplast genome. Since various RNA outputs were discovered from the plastome, it is understood that this genome is transcribed either singly or as part of multi-gene transcription units (Shinozaki *et al*., 1986; Sugita and Sugiura, 1996).

However, an indication demonstrating numerous heterogeneous and overlapping transcripts was later revealed (Barkan, 2011; Hotto *et al*., 2011; Ruwe and Schmitz-Linneweber, 2012). This phenomenon was first observed during RNA

sequencing in barley which enabled identifications of promoters at high probability at any given genome location. It included transcripts of coding, non-coding, and antisense regions of known genes (Zhelyazkova *et al*., 2012). Inefficient transcription termination of chloroplast gene expression also contributes to the phenomenon. Mutants defective for 3' to 5' exoribonucleases revealed that many transcripts dominate variable 3' extensions (Bollenbach *et al*., 2005; Germain *et al*., 2011). Consistent with the widespread transcription initiation and inefficient termination, full transcriptions of plastome were observed in strand-specific deep sequencing of *Arabidopsis* transcriptome (Hotto *et al*., 2011). Later, a systematic transcriptomic analysis from the chloroplasts of algae, higher plants, and cyanobacteria verified that the entire plastome was transcribed (displayed in Figure 2.6) (Shi *et al*., 2016). The same phenomenon was observed in the chloroplast of *C. reinhardtii*, and many genes were found to be co-transcribed (as summarised in Harris, 2009).

2.2.4 RNA processing and stability in the chloroplast

The chloroplast genome can be transcribed in various ways, and the long primary RNAs need to be prepared into shorter RNAs. Subsequent RNA processing is necessary for the multiple primary and polycistronic transcripts to ensure stability in the chloroplast. Both exo- and endo-nucleolytic processing occur on most chloroplast mRNAs. These reactions are usually performed by RNases either at the end or internally, respectively, as shown in Figure 2.6B (Herrin and Nickelsen, 2004; Meyer *et al*., 2008). The pathway of nucleolytic processing will cause either the maturation or degradation of mRNAs (Rymarquis *et al*., 2006). RNA metabolism in the chloroplast depends on nuclear-encoded RNA binding proteins (RBPs), which are involved in RNA maturation, degradation, splicing, and post-transcriptional

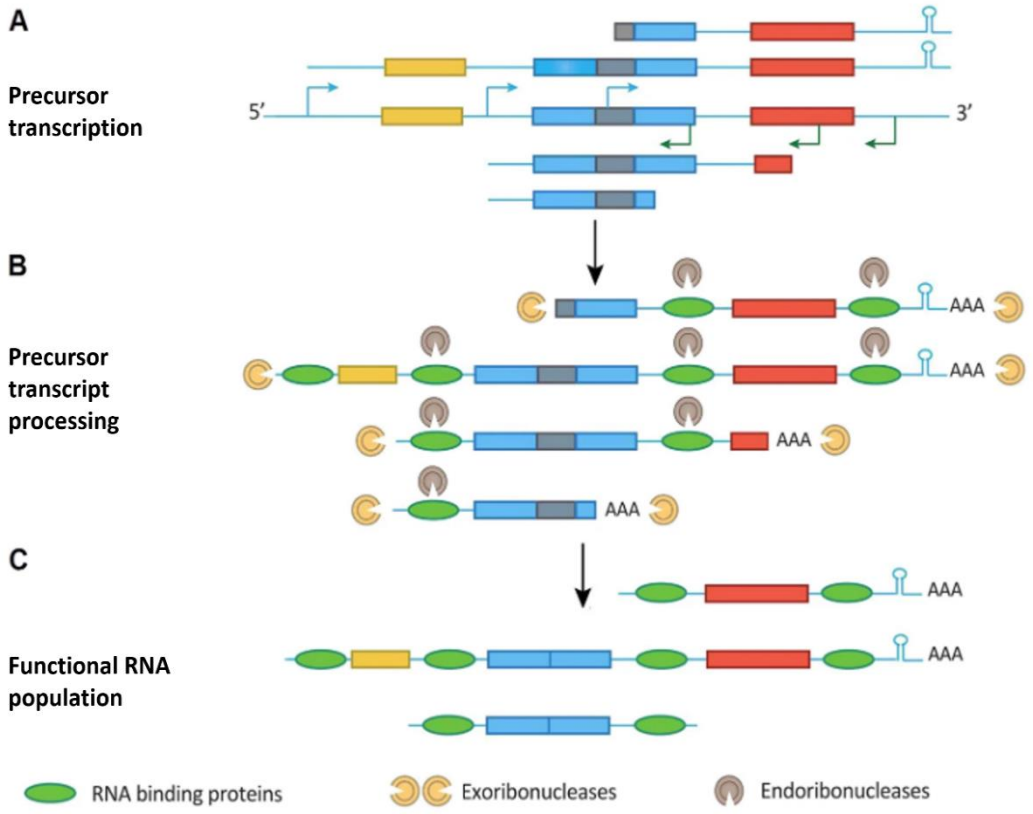


Figure 2.6: Model for the full chloroplast transcription and processing.
(A) Transcription initiation of a gene cluster occurs from multiple promoters (bent arrow) upstream of open reading frames (ORFs) or within ORFs. Together with inefficient transcription termination, this setup generates numerous precursor transcripts that can include complete or incomplete ORFs. Introns and RNA stem-loop structures are depicted as light black rectangles and hairpins, respectively.
(B) Precursor transcripts are processed by a combination of exo- and endo-ribonucleases. The precursor transcripts also can be polyadenylated by the addition of a Poly(A)-tail at the 3′-end of the transcripts. The sequence-specific RNA-binding proteins define functional RNAs followed by ribonuclease digestion. Introns and incomplete ORFs without sequence-specific RNA-binding proteins protection were digested by exo- or endo-ribonucleases. (C) RNA processing produces a pool of functional RNAs. (The image was adapted from Shi *et al*., 2016).

regulation (as reviewed in Barkans, 2011). Dozens of nuclear genes have also been identified in *C. reinhardtii* to be involved in the processing, stability, and degradation of chloroplast mRNAs (as summarised in Harris, 2009). The examples of nuclear-encoded RBPs related to RNA processing is discussed in *Section 2.3.4*. Interestingly, these identified RBPs are specific towards particular genomic regions with exclusive modes of action.

2.2.5 Interaction between nucleus and chloroplast

Due to the transfer of chloroplast genes to the nuclear chromosome, interactions between the two compartments become significant. Signals are transferred in both directions, with the chloroplast and nucleus signalling each other in response to developmental or external effects. These interactions result in changes in gene expression in both cellular compartments. Proteins encoded in the nuclear genome could be synthesised within the cytoplasm and then imported into the chloroplast (Smeekens *et al*., 1986; Friedman and Keegstra 1989; de Boer *et al*., 1991). These proteins could be transported to the chloroplast protein via the chloroplast-transit peptide (CTP) domain at their N-termini. These domains act as "zip codes" to direct the proteins towards the organelle's internal compartment. For example, those identified RBPs in *Section 2.2.4* were shown to possess the CTP sequence at their N-termini (Rahire *et al*., 2012; Ozawa *et al*., 2020). Mutant analysis of those RBPs-encoded genes revealed the defect of chloroplast mRNA processing and stability, suggesting the proteins were localised into the chloroplast. This observation also emphasises the control of mRNA stability by the interaction between the nucleus and chloroplast. There are also three other pathways to export the nuclear-encoded proteins towards the chloroplast. They may share some

components: secA-dependent, pH gradient-dependent (or Tat pathway), and Signal Receptor Protein-dependent (as reviewed in Jarvis and Robinson, 2004).

2.2.6 The availability of chloroplast transformation

The development of the chloroplast transformation method has aided research on chloroplast biology, and it is also applicable to chloroplasts of higher plants (Swiatek *et al*., 2001). In *C. reinhardtii,* the chloroplast transformation offers advantageous over nuclear transformation because of their specific localisation by homologous recombination and higher recombinant gene expression due to multiple copies of cpDNA (as reviewed in Doron *et al*., 2016). However, the current method of DNA deliveries only offers a very low transformation efficiency. The first chloroplast of *C. reinhardtii* was successfully transformed in 1987, and this success added this organism as a model organism for chloroplast biology (Sanford *et al*., 1987; Boynton *et al*., 1998). John Sanford group (from Cornell University) developed a biolistic gun method to deliver the DNA-coated tungsten microparticles into the chloroplast. They observed a relatively low cell recovery at a transformation efficiency of 1-5 per 10^6 bombarded cells (Sanford, 1990). Later, it was shown that the glass bead agitation method, usually used for nuclear transformation, also can be applied to chloroplast transformation (Kindle *et al*., 1991; Economou *et al*., 2014). However, the transformation frequency was even lower than the biolistic method since both cell wall and membrane need to be disrupted by glass beads to ensure successful delivery of DNA into chloroplast region. In *Phaeodactylum tricornutum*, the electroporation method yielded transformation efficiency of approximately around one colony per 1000 cells (Xie *et al*., 2014). The electroporation method has never been reported to be utilised for chloroplast transformation in *C. reinhardtii*.

Although only a handful of positive colonies are required, the challenge in getting even a single transformant needs to be addressed.

2.3 Small RNAs in chloroplast

There are two distinct types of RNA molecules, namely protein-coding RNAs (also known as mRNAs) and non-coding RNAs (ncRNAs). The latter is not translated into polypeptides or proteins but is functional in other biological roles. These ncRNAs include tRNA, rRNA, bacterial small RNA (sRNA), microRNAs (miRNAs), small interfering RNAs (siRNA), antisense RNAs (asRNAs), and long non-coding RNA (lncRNAs). Most of them have been reported to become either transcriptional or post-transcriptional factors in gene regulation (Kawaji and Hayashizaki, 2008; Ghildiyal and Zamore, 2009; Cech and Steitz, 2014). They also participate in DNA replication, RNA processing, splicing, and genome defence (as reviewed in Mattick and Makunin, 2006). Their expressions are affected at different development stages or in response to various external stimuli, suggesting that they are functionally involved in numerous physiological processes (as reviewed in Gottesman *et al.*, 2006).

It has been shown that these functional RNA molecules are not only present in the nuclear-cytosolic compartment. They are later discovered in the organelles of endosymbiont origin, such as chloroplast and mitochondria (Lung *et al*., 2006). In this organelle, the small ncRNAs possibly play important biological roles just as in the presumptive cyanobacterial ancestors (as reviewed in Hotto *et al*., 2012). These RNA molecules are also known as **chloroplast small RNA, csRNA,** and **clustered organellar RNA, cosRNAs** (as named in Wang *et al*., 2011 and Ruwe *et al*., 2016).

Throughout this thesis, the abbreviation of **‘cosRNAs’** is used as it directly defined the small RNA molecules from the chloroplast, while the term ‘csRNAs’ recently was used to define capped-small RNA-seq (csRNA-seq) for detection of transcription start sites at single-nucleotide resolution (Duttke *et al*., 2019). Current research focuses on the detection of cosRNAs and understanding their biogenesis and possible functions in transcriptional and post-transcriptional regulation in different photosynthetic organisms.

2.3.1 Discovery of cosRNAs in photosynthetic organisms

The cosRNAs were initially identified by cDNA cloning from total cellular RNA of Arabidopsis and tomato plants (Marker et al., 2002; Billoud et al., 2005). While cataloguing small RNA libraries from total cellular RNA, some of the reads were mapped to the chloroplast genome. In a tobacco study, cDNA cloning was done from RNA purified from chloroplast, and this small RNA library was confirmed to originate from the chloroplast genome (Lung et al., 2006). However, the scope of the cosRNA population was minimal until the advent of high throughput sequencing. Much cosRNA identification was derived from projects initially intended to profile 21-24 nt nuclear-encoded miRNAs and siRNAs (as reviewed in Hotto et al., 2012). RNA molecules of 18 to 40nt were separated from total cellular RNA extraction and used for library preparation for deep sequencing. These data sets included small RNAs derived from the photosynthetic organism's nuclear, chloroplast, and mitochondrial genomes. Since then, many researchers have identified the presence of cosRNAs in various photosynthetic organisms (Table 2.1).