The University Of Sheffield.

Developing metabolic engineering tools for enhanced synthesis of high-value products in *Chlamydomonas reinhardtii*

Josie McQuillan

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

Faculty of Engineering

Department of Chemical and Biological Engineering

Submitted 4th March 2021

Abstract

This work aimed to improve the biotechnological potential of *Chlamydomonas reinhardtii* by increasing its capacity for lutein production using forward and reverse genetic engineering approaches, and by developing genetic tools to enhance and expand the metabolic engineering toolkit available for this model green microalga.

The ORANGE protein is a post-translational regulator of a key enzyme in the carotenoid biosynthetic pathway in higher plants; its overexpression has previously led to the increased accumulation of carotenoids in several plant species. Here, a *C. reinhardtii* homologue of ORANGE was identified, cloned, and overexpressed from the *C. reinhardtii* nuclear genome for the first time, which resulted in a 2.0-fold increase in lutein production compared to the wild-type strain.

A semi high-throughput screening platform to isolate *C. reinhardtii* chemical mutants exhibiting increased carotenoid biosynthesis was developed, generating 5 strains that produce significantly higher total carotenoids than the wild-type, the highest of which (EMS-Mut-5) synthesised 5.4-fold more lutein than the parental strain. EMS-Mut-5 was characterised using a label-free quantitative proteomics workflow, which highlighted potential metabolic engineering targets for further enhancement of lutein production.

Lastly, novel genetic devices for metabolic engineering were then developed to facilitate transgene expression from the *C. reinhardtii* nuclear genome. Promoter sequences of highly expressed genes were computationally analysed to find DNA sequences that contribute to their high expression. 14 DNA motifs identified in this study were cloned into fluorescent protein reporter vectors for *in vivo* analysis, and their expression activity measured by flow cytometry. Ten out of 14 DNA motifs tested displayed significantly higher fluorescent protein expression compared to a minimal core promoter, and promoter 12 (pCRE-12) presented instances of higher expression than the current strongest *C. reinhardtii* promoter (Hsp70A-RbcS2). The outcome is a new suite of synthetic promoters that can drive a dynamic range of recombinant protein expression levels.

Acknowledgements

I would firstly like to thank my supervisor Dr Jagroop Pandhal for his support, guidance and unwavering patience throughout my project – I know it's been a challenge! I would also like to thank my sponsors EPSRC, and the University of Sheffield for giving me the opportunity to complete this PhD.

Thank you to all of the amazing researchers in CBE that have offered me advice and support over the years, particularly Dr Caroline Evans, Dr Phil Jackson, Dr Joy Mukherjee, Dr Esther Karunakaran and Dr Adam Brown. I'd also like to thank James G, Kasia, Yash and Gloria.

It's also been a pleasure working alongside all the members of the Pandhal group: Alaa, Fikayo, Wan, Hannah, Joanna, Helen, Mengxun, Zongting, Ali, José and Charlotte. Thanks for being amazing lab mates! Massive thanks to the 'G4boys' for teaching me all things molecular biology, but also for the lab DJ sets and discussions: Dr Stephen Jaffe, Dr Greg Fowler, both the Bens, Simon, Leila and Juliano.

I'd like to thank Dr Mark Scaife for teaching me the Chlamy culturing and molecular biology basics, and Prof Alison Smith for allowing me to visit her laboratory to learn.

Thanks to the algae team at University of California, San Diego for giving me the opportunity to test my synthetic promoters and learn new and exciting techniques, particularly Dr Ashley Sproles, Dr Anthony Berndt, Francis Fields and Prof Stephen Mayfield for letting me loose in his lab for a month! Thanks also to the HVCfP network for providing the training funds that allowed this to happen.

I have to say a big thank you to (the soon to be Dr) Christina Vanhinsbergh, both professionally for being an incredible HPLC tutor, and personally for being my rock and my best friend. Thanks to all my friends for keeping me sane(ish) throughout my studies, especially John, Ryo, Mik, José, Laura, Hannah, Marco, Lewis and Damiano. Thank you Neri for getting me though this with all your support and affection. You made these years the best of my life.

I'd finally like to thank my parents Lisa and John McQuillan (even if they still can't tell me what a carotenoid is after all this time!) for all the support, encouragement and love I could ever ask for. I couldn't have done it without them!

Declaration

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not been previously been presented for an award at this, or any other, university.

Table of Contents

List of Figures

List of Tables

List of abbreviations

°C	degrees Celsius
6mA	N^6-adenine methylation
Å	angstrom
A	adenine
A.	*Arabidopsis*
aa	amino acids
ALE	adaptive laboratory evolution
AME	analysis of motif enrichment
amiRNA	artificial micro RNA
AR-1	Hsp70A-RbcS2 promoter
AU	absorbance units
BKT	β-carotene ketolase
Bp	base pair
C	cysteine/ cytosine
C.	*Chlamydomonas*
C_X	carbon, where X represents number of carbon atoms
$CaCl_2$	calcium chloride
Car	Total carotenoids
Chl.	*Chlorella*
^{3}Chl	triplet chlorophyll
Chl	chlorophyll
cm	centimetre
CO_2	carbon dioxide
CRE	*cis*-regulatory element
CRISPR	clustered regularly interspaced short palindromic repeats
crOR	*Chlamydomonas reinhardtii* ORANGE
CRTISO	carotenoid isomerase
D.	*Dunaliella*
Da	daltons
DCW	dry cell weight

dH_2O	deionised water
DHA	docosahexanoic acid
DMAPP	dimethylallyl diphosphate
DNA	deoxyribonucleic acid
DTT	dithiothreitol
DXP	1-deoxy-D-xyulose 5-phosphate
DXR	1-deoxy-D-xyulose 5-phosphate reductoisomerase
DXS	1-deoxy-D-xyulose 5-phosphate synthase
E.	*Escherichia*
EDTA	ethylenediaminetetraacetic acid
Em	emission
EMS	ethylmethanesulphonate
EPA	eicosapentanoic acid
Ex	excitation
FACS	fluorescence-activated cell sorting
FDR	false discovery rate
fg	femtogram
FI	fluorescence intensity
FMDV	foot-and-mouth-disease virus
FPP	farnesyl disphosphate
g	gram
G	guanine/ glycine
G3P	glyceraldehyde 3-phosphate
gDNA	genomic deoxyribonucleic acid
GGPP	geranylgeranyl pyrophosphate
GMO	genetically modified organism
GO	gene ontology
GOI	gene of interest
GPP	geranyl pyrophosphate
GPPS	geranyl pyrophosphate synthase
h	hour

H.	*Haematococcus*
H_2O	water
His_6-tag	6x poly-histidine tag
HL	high light
HOMER	hypergeometric Optimization of Motif EnRichment
HPLC	high performance liquid chromatography
HRP	horseradish peroxidase
Hsp70A	Heat-shock protein 70A
IDI	IPP:DMAPP isomerase
IDS	IPP:DMAPP synthase
IPP	isopentenyl diphosphate
iRbcS2	ribulose bisphosphate carboxylase small subunit 2 intron 1
kbp	kilobase pair
kDa	kilodalton
λ	wavelength
L	litre
LB	Lysogeny broth
LC	liquid chromatography
LFQ	label-free quantification
LHCSR1	light harvesting complex stress related protein 1
LHCSR3	light harvesting complex stress related protein 3
LI	light intensity
LIMMA	linear models for microarray data
LYCb	lycopene β-cyclase
LYCe	lycopene ε-cyclase
µF	microfarad
µg	microgram
µL	microlitre
µm	micrometre
µM	micromolar
µmol	micromoles

m	metre
M	molar
Mbp	megabase pairs
MEME	Multiple Expectation maximization for Motif Elicitation
MEP	methyl-D-erythritol 4-phosphate
mg	milligram
$MgCl_2$	magnesium chloride
min	minute
miRNA	micro RNA
mL	millilitre
mm	millimetre
mM	millimolar
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger ribonucleic acid
MS	mass spectrometry
MS/MS	Tandem mass spectrometry
MVA	mevalonate
n	number (of replicates/ samples)
N	nitrogen
NaCl	sodium chloride
NCBI	National Center for Biotechnology Information
NEB	New England Biolabs
NF	norflurazon
nfH_2O	nuclease-free water
ng	nanogram
NHEJ	non-homologous end joining
nm	nanometre
NPQ	non-photochemical quenching
O.	*Ostreococcus*
O_2	oxygen
OD	optical density

OR	ORANGE protein
ORF	open reading frame
P	phosphorus
P.	*Pheodactylum*
PAP	plastid lipid-associated protein
PCA	principal component analysis
PCR	polymerase chain reaction
pCRE	putative *cis*-regulatory element
PDS	phytoene desaturase
Pg	picogram
pI	isoelectric point
pmol	picomoles
PSI	photosystem I
PSII	photosystem II
PSY	phytoene synthase
PTM	post-translational modification
PUFA	polyunsaturated fatty acid
PVDF	polyvinylidene fluoride
PWM	position weight matrix
qE	energy-dependent non photochemical quenching
RbcS2	ribulose bisphosphate carboxylase small subunit 2
RNA	ribonucleic acid
RNAi	RNA interference
ROS	reactive oxygen species
rpm	rotations per minute
RSAT	regulatory sequence analysis tools
s	second
SD	standard deviation
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SGR	specific growth rate

sgRNA	single guided ribonucleic acid
SNP(s)	single nucleotide polymorphisms
SORLIP	sequences overrepresented in light-induced promoters
ssDNA	single stranded deoxyribonucleic acid
T	thymine
TAE	tris-acetate-EDTA
TAG(s)	triacylglyceride
TALEs	transcription activator-like effectors
TAP	tris-acetate-phosphate media
TF	transcription factor
TFA	trifluoroacetic acid
TFBS	transcription factor binding site
T_m	annealing temperature
TSP	total soluble protein
TSS	transcription start site
UCSD	University of California, San Diego
USD	United States dollars
UTR	untranslated region
UV	ultra violet
UVM	UV-mutagenised
VDE	violaxanthin de-epoxidase
V	volts
V.	*Volvox*
v/ v	volume/ volume
w/ v	weight/ volume
WT	wild-type
xg	times gravity
ZDS	ζ-carotene desaturase
ZEP	zeaxanthin epoxidase
Z-ISO	15-cis-ζ-carotene isomerase

Chapter 1: Literature Review

1.1. Introduction to microalgae as biotechnological hosts

Microalgae are a metabolically and physiologically diverse group of photosynthetic microorganisms, many of which can produce valuable metabolites and chemicals such as lipids, biofuel precursors, polysaccharides and pigments. Their ability to harness solar energy and convert carbon dioxide (CO_2) into useful products make microalgae attractive sustainable biotechnology hosts; this is in contrast to other microbial platforms, such as yeast and bacteria, which require external carbon sources. Additionally, many microalgal species can be safely consumed by humans. Advantages to using microalgal systems over higher plants for the production of biofuels and other products include microalgal unicellularity, rapid growth, higher productivity per land unit, water and nutrients can be recycled, and their ability to thrive in a range of environments (Chisti, 2008). The incredible metabolic diversity across microalgal species gives them great potential for industrial exploitation and for the production of commercially useful and new chemicals, emphasising the importance of developing new metabolic engineering tools in microalgae. The physiological diversity of microalgae is highlighted in **Figure 1.1**.

This literature review will cover microalgae as industrial hosts, then focus on a specific class of microalgal products: carotenoids. Furthermore, the model green microalga *Chlamydomonas* (*C.*) *reinhardtii* and its potential as a metabolic engineering host for carotenoid production will be reviewed, as well as the tools currently available with which this can be achieved.

1.2. Natural products of microalgae

1.2.1. Low-value natural products

1.2.1.1. Whole-cell biomass: Foods and feeds

Algal biomass can be used as a direct food source or as a nutritional supplement for both humans and animals due to their high protein, antioxidant and vitamin content (Spolaore *et al*., 2006). The main microalgal genera cultivated for human consumption are spirulina (*Arthrospira*; **Figure 1.1D**), *Chlorella*, *Dunaliella* (**Figure 1.1C**) and *Aphanizomenon*, which can be consumed in the forms of tablet or powder, or incorporated into other food products to increase their nutritional value (Spolaore *et al*., 2006; Caporgno and Mathys, 2018). Microalgae are also important feed for aquaculture, where they provide the pink pigmentation of salmonoids and shrimp (Yaakob *et al*.,

2014; Shah *et al*., 2018). Health benefits for using microalgae as feed for livestock have also been demonstrated (Holman and Malau-Aduli, 2013; Yaakob *et al*., 2014).

1.2.1.2. Biofuels

During stressful conditions such as nitrogen deprivation, certain microalgal species redirect their metabolism from growth to the production of triacylglycerides (TAGs), which can be extracted and converted to biodiesels via transesterification (Chisti, 2008). TAG-producing microalgal species include *Chlorella sp.*, *Phaeodactylum (P.) tricornutum* (**Figure 1.1B**) and *Nannochloropsis sp*. These microalgae have the potential to be a superior source of biodiesel compared to plants, as they reproduce more quickly, require less land mass and water, can thrive on nonarable land, and are more efficient at converting CO_2 and light into TAGs (Chisti, 2008). Furthermore, some algal species such as *Botryococcus Sp.* (**Figure 1.1E**) can accumulate up to 86 % of their dry weight as hydrocarbons, which can be used directly as fuel after purification without transesterification (Brown *et al*., 1969). However, the cost of producing fuels from algae is still too high, and further advancements in harvesting and lipid extraction technologies would be required to render algae an economically viable feedstock for biofuel production (Chisti, 2008; Lam and Lee, 2012; Batan *et al*., 2016). Diesel currently costs around 2.5 United States dollars (USD) per gallon in the United States of America (USA; as of 14-12-2020; https://www.globalpetrolprices.com/diesel_prices/#hl228). The cost of algal culturing, lipid extraction and processing, biofuel transportation and storage combined must be of a similar value to be competitive with fossil fuels. Recent techno-economic assessments have estimated algal biofuel costs to be between $2.20–$31.60 USD per gallon (Nagarajan *et al*., 2013; Richardson *et al*., 2012), meaning that significant improvements must be made to improve the algal biofuel economy.

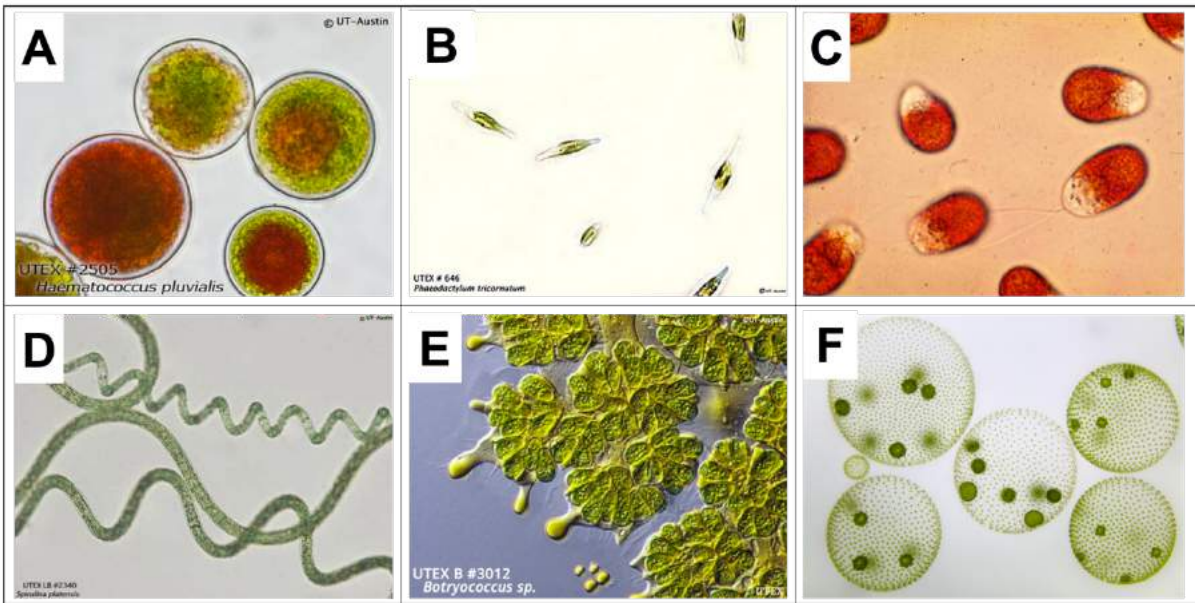


Figure 1.1: Microscope images of various microalgal species. A - *Haematococcus pluvialis* UTEX 2505 cells turning from green to red as they accumulate the high-value carotenoid astaxanthin. **B** - *Phaeodactylum tricornutum* UTEX 646 shown here is a dinoflagellate species that can produce high levels of PUFAs and pigments. Genetic engineering tools have been developed in this alga enabling production of terpenoid molecules (Fabris *et al.*,2020). **C** - *Dunaliella salina* cells hyperaccumulating the orange/ red high-value carotenoid β-carotene. Image adapted from http://cfb.unh.edu/phycokey/Choices/Chlorophyceae/unicells/flagellated/DUNALIELLA/Dunaliella_Image_page.html. **D** - *Spirulina platensis* UTEX LB 2340, also known as *Arthrospira platensis*, is a blue-green cyanobacterial strain characterised by its helical filamentous structure. This alga is cultivated as a food supplement. **E** - *Botryococcus sp.* UTEX 3012 is shown secreting hydrocarbons into its extracellular matrix. **F** - The multicellular microalga *Volvox carteri* shown here is an important model organism for investigating the origins of multicellularity (Figure adapted from https://microbewiki.kenyon.edu/index.php/File:Vc.jpg#filelinks). Figures 1A, 1B, 1D and 1E were adapted from https://utex.org/ (photo credit ©2020 UTEX Culture Collection of Algae).

1.2.2. High-value natural products

1.2.2.1. Long-chain polyunsaturated fatty acids (PUFAs)

Long-chain polyunsaturated fatty acids (PUFAs) are important nutrients required for human eye and brain development in infancy, as well as having positive effects on cardiovascular health and immune system function (Siscovick *et al.*, 1995; Ruxton *et al.*, 2005); however, humans and many other animals are unable to synthesise PUFAs, and so must acquire them from their diets. PUFAs are usually acquired through fish and fish oils, but due to overfishing and worries over toxin

accumulation in fish, finding another source of PUFAs is essential (Martins *et al*., 2013). Microalgae could provide a sustainable and vegetarian alternative to fish oils, as some species (e.g. *Nannochloropsis Sp*. and *P. tricornutum*) can accumulate high levels of PUFAs such as docosahexanoic acid (DHA) and eicosapentanoic acid (EPA) under certain physiological conditions (Martins *et al*., 2013). One of the main limitations to using microalgae for PUFA biosynthesis is that the PUFAs accumulate within cell membranes, making the extraction process relatively difficult and expensive (Martins *et al*., 2013). Using metabolic engineering to increase PUFA production is one strategy for improving economic viability of microalgae as a PUFA source; for example, DHA production was increased eight-fold in *P. tricornutum* by introduction of heterologous fatty-acid biosynthesis genes from *Ostreococcus (O.) tauri* (Hamilton *et al*., 2014, 2016).

1.2.2.2. Carotenoids

Carotenoids are structurally diverse, high-value pigment molecules synthesised by photosynthetic organisms, as well as some non-photosynthetic bacteria and funghi; their anti-oxidative and colourful characteristics make them of commercial interest to areas such as the aquaculture, cosmetics, pharmaceutical and nutraceutical industries. Microalgae are natural carotenoid producers, often generating high amounts of these pigments under certain conditions. Examples include accumulation of the commercially valuable carotenoid β-carotene under high light and salt stress in the halotolerant alga *Dunaliella (D.) salina* (Lamers *et al*. 2010), and astaxanthin production in *Haematococcus (H.) pluvialis* (**Figure 1.1A**) and *Chlorella* (*Chl.*) *zorifingiens* (Boussiba, 2000; Ip and Chen, 2005). Other valuable carotenoids include the brown algae-produced fucoxanthin, and the plant and green algae-produced carotenoids lutein and zeaxanthin. **Figure 1.2** shows the chemical structures of common carotenoids found in microalgae.

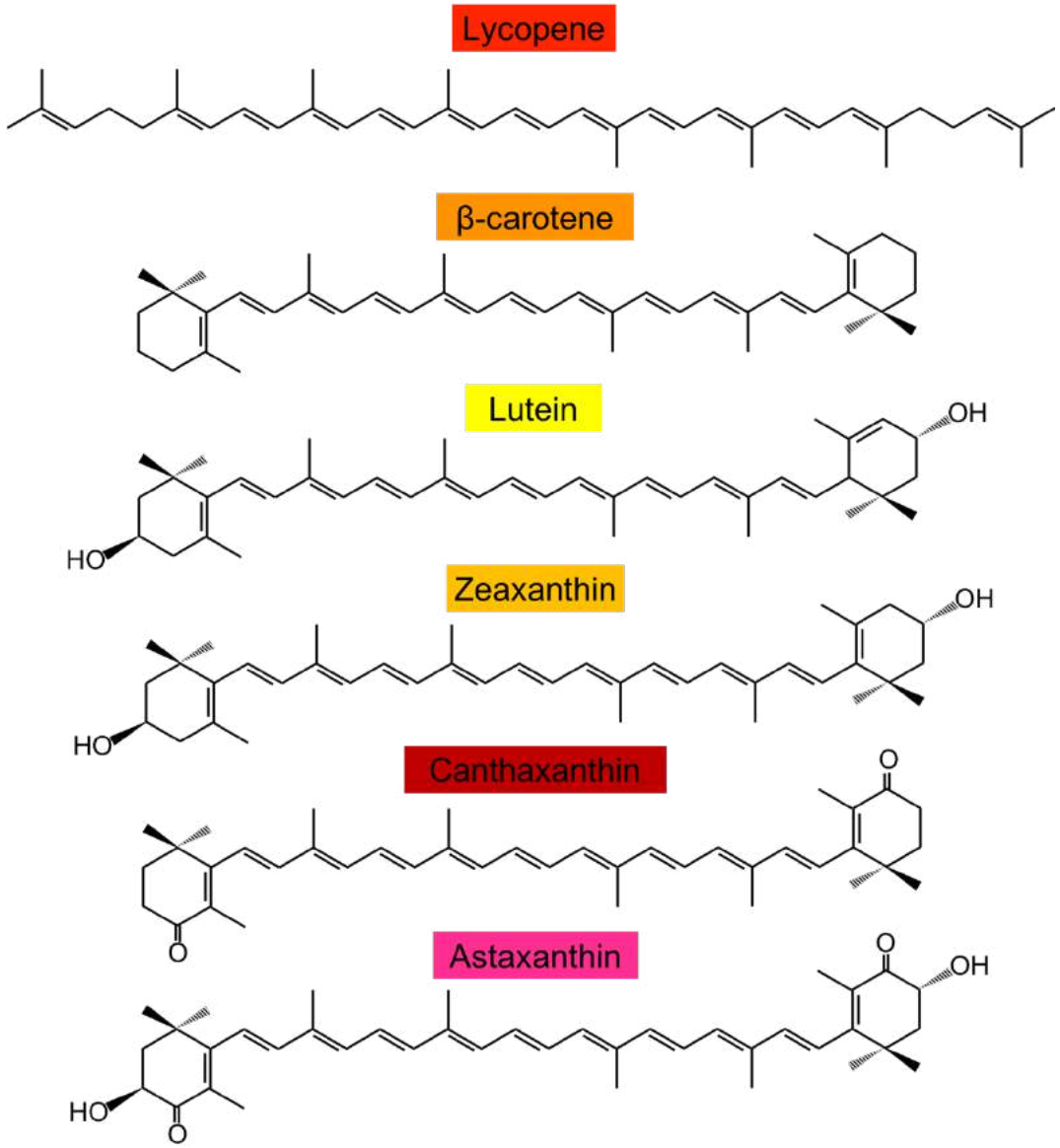


Figure 1.2: Molecular structures of common high-value carotenoids produced by microalgae. The C_{40} structures of colourful microalgal carotenoids with industrial value are shown. Lycopene and β-carotene are carotene hydrocarbon structures, whereas lutein, zeaxanthin, canthaxanthin and astaxanthin are xanthophyll structures, containing oxygen moieties. The colour of each carotenoid is highlighted; the colours are determined by the position and number of double bonds within the conjugated π-electron systems.

1.2.2.3. Other high-value natural products

Other important high-value algal biochemicals include polysaccharides, amino acids, phycobilins, sterols, polyhydroxyalkonates, and hydrogen gas. See Borowitzka (2013) for a detailed review of these products, their regulation and commercialisation.

1.3. Carotenoids

This section will examine carotenoids as a commercial product in more detail.

1.3.1. Carotenoids market

The carotenoids market is projected to grow from the estimated $1.5 billion USD in 2019 to $2.0 billion USD by 2026 (Markets and Markets, 2020). The predominant constituents of this market are β-carotene and astaxanthin, followed by lutein, lycopene, astaxanthin zeaxanthin, canthaxanthin. Lutein is projected to be the fastest growing carotenoid market segment between 2020 and 2026 (Markets and Markets, 2020).

1.3.2. Carotenoid commercial uses

The main commercial uses of carotenoids are in the areas of supplements, food, feed and cosmetics. Animal feed and nutrition was the dominant carotenoid market area in 2019 (Markets and Markets, 2020); animals tend to grow faster and healthier when fed with a carotenoid-rich diet (Madeira *et al*., 2017). Increasingly more beneficial effects of dietary and supplementary carotenoids are being discovered. This, along with increasing consumer awareness and desire for convenience, made carotenoid supplements the fastest growing segment of the carotenoids market from 2014–2019 (Markets and Markets, 2014). The antioxidant properties of carotenoids confer antiaging properties to these molecules, making them valuable in the cosmetics industry, especially in aging populations (Masaki, 2010).

1.3.3. Health benefits of carotenoid consumption in humans

The anti-oxidative properties of carotenoids make them excellent free-radical scavengers, capable of limiting toxic singular oxidative species produced as by products from cellular activity. Mainly for this reason, the consumption of carotenoids can enhance the immune system, protect against and treat certain cancers, lower the risk of heart disease and diabetes and prevent degradation of the eye, amongst other benefits (Fiedor and Burda, 2014).

1.3.4. Carotenoid sources

The majority of carotenoids on the market are synthesised chemically (Borowitzka, 2013; Gong and Bassi, 2016); however, there is increasing demand from consumers for supplements to be naturally sourced, and evidence is growing for the health benefits of consuming natural over chemically synthesised carotenoids (Borowitzka, 2013). Carotenoids that are chemically sourced are typically

derived from fossil fuels. Furthermore, synthetically produced carotenoids contain unnaturally high levels of all-trans isomers, which can dramatically reduce the biological activity. For example, the antioxidant activity was 50 times lower in synthetically produced astaxanthin compared to biologically produced astaxanthin (Capelli *et al*., 2013). The global projected market size for naturally-sourced carotenoids was $348.5 million USD by 2019 (Markets and Markets, 2014).

Carotenoids are synthesised by all chlorophyllous photosynthetic organisms, including plants, algae and photosynthetic bacteria, as well as some fungi and non-photosynthetic bacteria. Microalgae are the main natural source for carotenoids (Borowitzka, 2013). The microalga *D. salina* was the first microalga to be commercialised for carotenoid production, namely for β-carotene (Borowitzka, 2013). β-carotene currently has the second largest carotenoid market, and is grown commercially in open ponds in many locations worldwide. *H. pluvialis* is another commercially valuable organism, producing large amounts of astaxanthin. For a review of other commercially valuable carotenoid producing strains, see Borowitzka (2013) and Gong and Bassi (2016).

The quality of the carotenoid products is affected by each step in the supply chain, in particular synthesis, extraction and storage. This is less of a problem for *D. salina* and *H. pluvialis*, as the desired carotenoids make up ~90 % of their total carotenoid content, and the commercial carotenoids are stored in globules which enhances separation (Borowitzka, 2013). Other microalgal species that currently produce a lower percentage of the desired carotenoid must generally undergo expensive separation procedures to isolate the carotenoid of choice from other pigments. The purity level of carotenoids is critical when it comes to gaining approval for sale in various industries, particularly for human consumption.

Genetic modification and metabolic engineering of microalgae could overcome this issue by optimising output towards the desired carotenoid, but negative public attitudes towards genetically modified organisms (GMOs) would significantly lower the market value of carotenoids produced in this way. Applying an adaptive laboratory evolution method, exposure to small amounts of mutagenic ultra violet (UV) light or chemical mutagen, or overexpressing native genes under native promoters could potentially allow engineering of the cells without having to declare them GMOs (Schierenbeck *et al*., 2015).

1.3.5. Lutein

As mentioned above, lutein is projected to have the fastest growing value of all of the carotenoids currently on the market (Markets and Markets, 2020). Growing evidence of lutein's ability to

prevent and treat human cataracts and macuolar degeneration, as well as enhance visual processing speed when taken as a supplement to diet, is the main driver for this increase (Bernstein *et al*., 2016; Tian *et al*., 2015; Bovier and Hammond, 2015). Protection of the skin from UV radiation is also afforded by oral supplementation of lutein (Grether-Beck *et al*., 2017), as well as anti-inflammatory effects in patients with coronary heart disease (Chung *et al*., 2017). Additional to human benefits, lutein is an important additive to poultry feed where it intensifies the yellow pigmentation of poultry egg yolks, fat and skin (Leeson and Caston, 2004).

Currently, natural lutein is primarily extracted from marigold oleoresin, where lutein concentrations can very between 0.17 and 5.7 milligrams (mg) per gram $(g)^{-1}$, depending on the species (Piccaglia *et al*., 1998). This is unlike most other carotenoids, which are microalgal-sourced. High microalgal producers of lutein do exist, and include *Muriellopsis sp*. and *Scenedesmus almeriensis* and *Chlorella* species (eg *minutissima, sorokiniana, vulgaris*); between 3.4 and 9.8 mg g^{-1} dry cell weight (DCW) can be achieved in these microalgal strains, which exceeds that found in marigold petals (Fernández-Sevilla *et al*., 2010; Gong and Bassi, 2016).

As highlighted in **Section 1.1**., the production of biochemicals in microalgae has some advantages over plants, including higher productivity, faster growth, simpler growth parameters and a lack of seasonal dependence; despite this, microalgae do not currently produce lutein to high enough levels to be competitive with marigold, when the cost of extraction from algae is considered (Fernández-Sevilla *et al*., 2010). A genetic engineering strategy could be employed to enhance lutein yields in microalgae, but genetic modification tools in the aforementioned lutein-rich species are presently lacking.

The model green microalga *C. reinhardtii* produces lutein to moderate levels (2.8 mg g^{-1} DCW; Cordero *et al*., 2011b). Using the genetic tools available for this organism, lutein production has the potential to be enhanced; this possibility will be explored in **Chapter 3** and **Chapter 4**.

1.4. Chlamydomonas reinhardtii

1.4.1. History and evolution

C. reinhardtii (**Figure 1.3**) is a unicellular green microalga with a rich history of study that began in the 1950's, and it has since become a model organism for the study of basal bodies, phototaxis, photosynthesis, carbon concentrating mechanisms, chloroplast biology, cell cycle and sexual reproduction, and many other important cellular processes (Keller *et al*., 2005; Kianianmomeni and

Hallman, 2014; Rochaix *et al*., 2012; Umen, 2011). *C. reinhardtii* is also a model algal organism for the study of lipid accumulation and metabolic engineering for useful bioproducts (Chisti, 2008; Scranton *et al*., 2015).

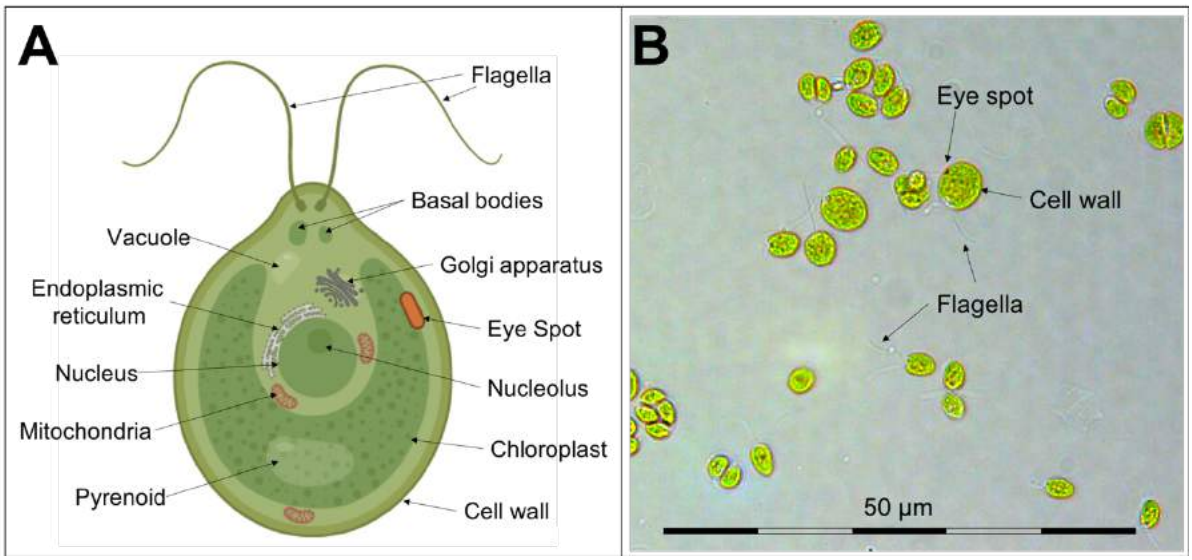


Figure 1.3: *Chlamydomonas (C.) reinhardtii*. A - Schematic diagram of *C. reinhardtii* showing the major organelles. Image generated using Biorender. **B** - Light microscope image of *C. reinhardtii* CC-125 with visible characteristics annotated. Image taken at 40x zoom, scale bar 50 µm.

C. reinhardtii belongs to the *Chlorophyceae* class and is closely related to the multicellular green alga *Volvox (V.) carteri* (**Figure 1.1F**), as well as other biotechnologically significant green microalgae such as *O. tauri*, *Botryococcus braunii*, *H. pluvialis* and *D. salina* (Merchant *et al*., 2007; Harris, 2009). Being of the green lineage, *C. reinhardtii* and land plants share a common ancestor, meaning many findings in this alga can be directly translated to plant models (Merchant *et al*. 2007). Other laboratory *Chlamydomonas* species include the snow alga *C. nivalis*, the acidophilic *C. acidophila* and the agriculturally significant *C. mexicana* (Harris, 2009).

1.4.2. Physiology

C. reinhardtii is a biflagellate organism with the following defining features: one large cup-shaped chloroplast, a pyrenoid carbon-concentrating organelle, a red-orange carotenoid-rich eyespot, a cell wall, contractile vacuoles, and the standard eukaryotic organelles including several mitochondria, the Golgi apparatus, endoplasmic reticulum, nucleus, and a prominent nucleolus (**Figure 1.3**; Harris *et al.*, 2009).

1.4.3. Genetics

The first draft genomic sequence for *C. reinhardtii* was completed by Grossman *et al*. (2003), which was later fully completed by Merchant *et al.,* (2007); this revealed the sequences for the nuclear, mitochondrial and chloroplast genomes. Re-drafts of the genome have been completed and annotated since (Dutcher *et al*., 2012; Blaby *et al*., 2014; Goold *et al*., 2016; Gallaher *et al*., 2018). The mitochondrial genome is ~16 kilobase pairs (kbp) and encodes only 13 genes, with a guanine + cytosine (GC) content of ~45% (Vahrenholtz *et al*., 1993; Remacle *et al*., 2006; Gallaher *et al*., 2018). The chloroplast genome is circular, ~205 kbp in size, and carries ~99 genes; its GC content is ~35% (Smith and Lee, 2009; Gallaher *et al*., 2018). Both the chloroplast and mitochondrial genomes are transformable by homologous recombination (Kindle *et al*., 1991; Remacle *et al*., 2006).

With a size of ~11.1 megabase pairs (Mbp), the *C. reinhardtii* nuclear genome has around ~19,500 predicted gene models, and has a relatively high GC content of ~64% (Blaby *et al*., 2014; Merchant *et al*., 2007). The nuclear genome is haploid and has two mating types (*plus* [+] and *minus* [-]), allowing for sexual and asexual propagation. The haploid nature of the cells causes all mutations to be dominant, making *C. reinhardtii* a useful organism for genetic studies. Under optimal growth conditions, *C. reinhardtii* reproduces vegetatively, although gametogenesis can be induced by stress conditions such as nitrogen (N)-deprivation (Harris, 2009). For a review on the *C. reinhardtii* cell cycle, see Cross and Uman (2015).

1.4.4. Growth and metabolism

Chlamydomonas species are versatile organisms that can be found thriving in polar, tropical, soil, fresh and marine water environments (Harris, 2009). *C. reinhardtii* has a flexible metabolism, capable of growth under photoautotrophic (light + CO_2 + water [H_2O]), heterotrophic (dark + acetate) and mixotrophic (light + external carbon source e.g. acetate) metabolic conditions (Harris, 2009). Growth is relatively fast in *C. reinhardtii*, with a doubling time of less than 7 hours (h) under optimal conditions. It can be grown under continuous or day:night cycles, and can survive under a temperature range of 15–35 degrees Celsius (°C), demonstrating its versatility as a laboratory strain (Harris, 2009). Upon nitrogen deprivation, *C. reinhardtii* is capable of producing high levels of lipids as a protective carbon storage response. *C. reinhardtii* is generally adapted to conditions with relatively low light intensities for mixotrophic growth, such as 100–200 micromoles (µmol) photons per metre $(m)^2$ per second (s), and requires slightly higher light photon flux densities for

photoautotrophic growth (200–400 µmol photons m^2 s^{-1}); *C. reinhardtii* tends to bleach and die at light intensities > 1500 µmol photons m^2 s^{-1} (Harris, 2009).

1.5. Carotenoid Biosynthesis in *Chlamydomonas reinhardtii*

1.5.1. Introduction

Carotenoids are 40-carbon (C_{40}) tetraterpenoid lipophilic molecules, synthesised by the step-wise condensation of 8 C_5 isoprene units; the resulting carotenoid skeleton can then be modified in several ways, such as desaturation, cyclisation, oxidation, hydroxylation and epoxidation. Carotenoids are generally categorised as carotenes, which are pure hydrocarbons, and xanthophylls, which contain one or more oxygen moieties. Differing levels of desaturation throughout the carotenoid backbone contribute to their conjugated π-electron system, which provides the pigments with their various colourful and antioxidant properties.

Carotenoids are bound within light-harvesting complexes in the chloroplast of *C. reinhardtii*, where their functions include light-harvesting, photoprotection, non-photochemical quenching, modulating photosynthetic complex assembly and photosystem antenna size (Frank and Cogdell, 1996; Demmig-Adams and Adams, 1996; Dall'Osto *et al*., 2015).

Under standard growth conditions, the following percentages of carotenoids are produced by *C. reinhardtii*: β-carotene (25%); lutein (25%); violaxanthin, neoxanthin and loroxanthin between 10–20% each (Krinsky and Levine, 1964; Eichenberger *et al*., 1986; Niyogi *et al*., 1997). Stressful conditions, such as high light intensity, induce the conversion of a significant proportion of the violaxanthin pool back to zeaxanthin via an antheraxanthin intermediate (**Figure 1.5**, **Section 1.5.5.**).

Figure 1.4 describes the carotenoid biosynthetic pathway in *C. reinhardtii*, which can be considered in two stages: isoprenoid biosynthesis (**Section 1.5.2.**) and committed carotenoid biosynthesis (**Section 1.5.3.**).

All enzymes in the isoprenoid and carotenoid pathways are nuclear-encoded, and guided to the chloroplast by an N-terminal transit peptide (Lohr *et al*., 2005). An exception includes farnesyl disphosphate (FPP) synthase, which possesses no targeting peptide, and accumulates in the cytosol (Lauersen *et al*., 2016b). Only one copy of each enzyme in the carotenogenic pathway is predicted to be encoded within the *C. reinhardtii* genome (Lohr *et al*., 2005).

1.5.2. Isoprenoid biosynthesis

The isoprenoid building-blocks that form the carotenoid backbone are two C_5 methylated diphosphate isomers: isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP).

Two convergent evolutionary pathways exist for the synthesis of IPP and DMAPP: the mevalonate pathway (MVA) and the methyl-D-erythritol 4-phosphate pathway (MEP; for review, see Lichtenhaler, 1999). The MVA pathway is the best characterised of the two; it is solely cytosolic and is formed from three molecules of acetyl-CoA. The MVA pathway is the primary isoprenoid anabolic pathway in archaea and eukaryotes. The MEP pathway, a relatively recent discovery, is of prokaryotic origin (Rohmer *et al*., 1993). Most organisms of the kingdom planta carry both such pathways; the cyanobacterial endosymbiotic event leading to chloroplast incorporation introduced the MEP pathway to plants, which works in tandem with the original nuclear eukaryotic MVA pathway (Schwender *et al*., 1997; Lichtenthaler, 1999). *C. reinhardtii* and other green Chlorophyta provide an exception to this rule; they have completely lost the ancestral MVA pathway, and use only the MEP pathway, which over time has been incorporated into the nuclear genome (Schwender et al, 1997; Disch *et al*., 1998; Schwender *et al*., 2001; Lohr *et al*., 2005).

The precursors to the *C. reinhardtii* MEP pathway are glyceraldehyde 3-phosphate (G3P) and pyruvate, which are both derived directly from the Calvin-Benson cycle. G3P and pyruvate are condensed via the enzyme 1-deoxy-D-xyulose 5-phosphate synthase (DXS) to form 1-deoxy-D-xyulose 5-phosphate (DXP; Lichtenthaler, 1999; **Figure 1.4**). DXP can then branch from the pathway to form pyridoxol and thiamine. The isomerisation of DXP to MEP is then catalysed by 1-deoxy-D-xyulose 5-phosphate reductoisomerase (DXR) by oxidation of a NADPH cofactor (**Figure 1.4**); this is the first committed step towards isoprenoid biosynthesis, and is specifically inhibited by fosmidomycin (Takahashi *et al*., 1998; Schwender *et al*. 2001; Jomaa *et al*., 1999). The enzymes DXS and DXR are thought to be important regulatory points within the MEP pathway. For example, the triterpene super-producing green alga *Botryococcus brauni* race B possesses three isoforms of DXS, hinting towards the enzyme's role as a 'gate keeper' (Matsushima *et al*., 2012).

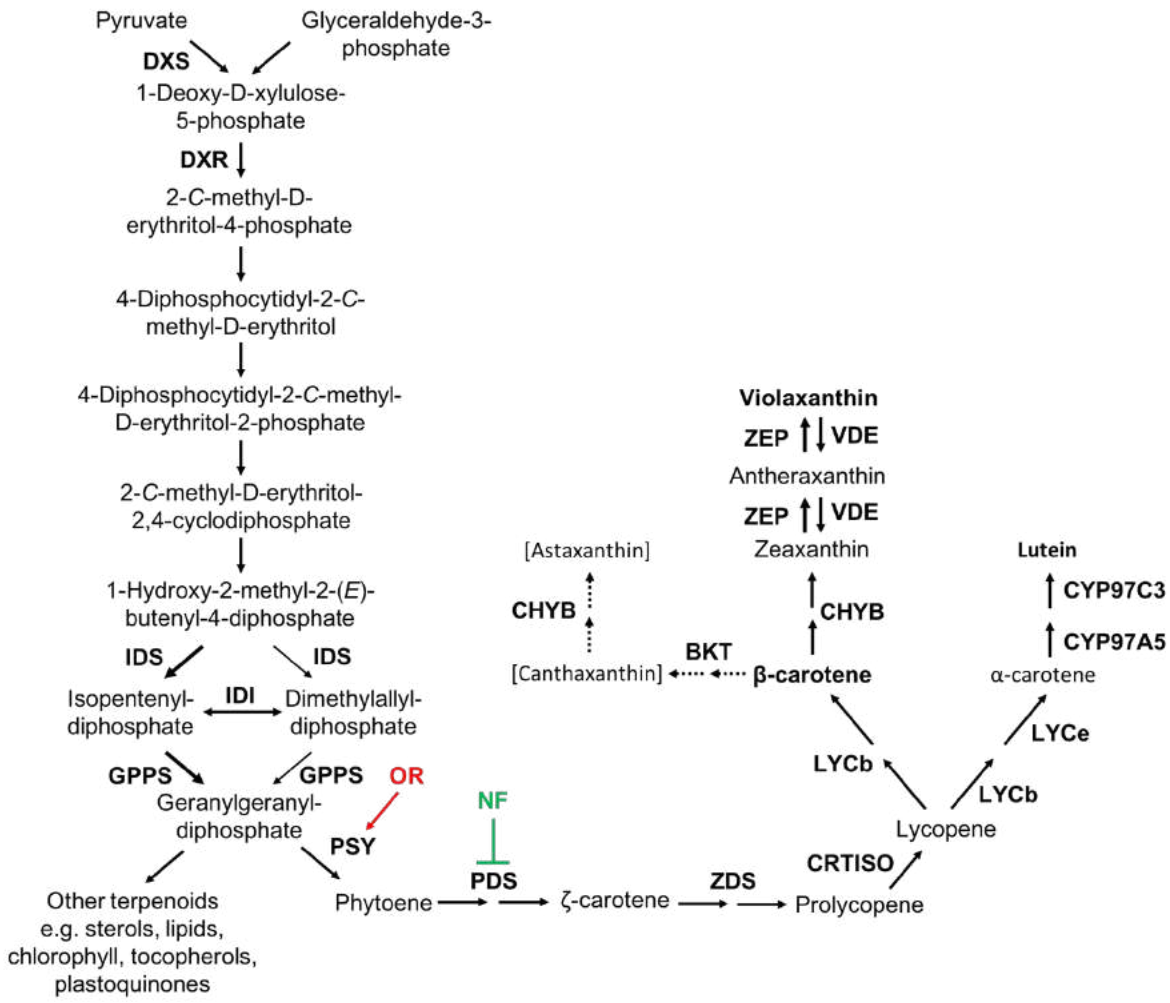


Figure 1.4: The isoprenoid and carotenoid biosynthetic pathways in *C. reinhardtii*. Pathway enzymes are depicted in bold beside the arrow representing the reaction they catalyse. Carotenoids predominantly found in *C. reinhardtii* are emboldened. Full names for key enzymes in the pathway above can be found in **List of Abbreviations**. The proposed interaction between the ORANGE protein and phytoene synthase (PSY) is displayed in red; the inhibitory relationship between the herbicide norflurazon (NF) and phytoene desaturase (PDS) is shown in green. The postulated astaxanthin biosynthetic pathway is included in the figure, denoted by striped arrows.

MEP is then converted to IPP and DMAPP via five more reactions, the last of which is catalysed by IPP:DMAPP synthase (IDS; **Figure 1.4**), which synthesises both IPP and DMAPP. IPP:DMAPP isomerase (IDI) maintains the equilibrium between the two isomers by catalysing their interconversion in response to cellular metabolic requirements. Isoprenoid synthesis can occur without the presence of IDI, however the efficiency of terpenoid production is affected by expression of IDI, suggesting that it is an important control point in terpenoid biosynthesis (Berthelot *et al*., 2012).

Units of chain-elongating IPP can then be added stepwise to one initiator DMAPP molecule to form several classes of terpenoid with chain lengths of multiples of 5, catalysed by prenyltransferase enzymes (**Figure 1.4**). The monoterpene geranyl pyrophosphate (GPP) is produced by condensation of one IPP and one DMAPP by geranyl pyrophosphate synthase (GPPS); GPP can then be successively elongated by addition of IPP units, producing for example C_{15} FPP and C_{20} geranylgeranyl pyrophosphate (GGPP). As well as forming the carotenoid backbone, isoprenoids are the precursors to several essential metabolic processes, including the biosynthesis of sterols, lipids, the chlorophyll phytyl tail, tocopherols, and plastoquinones (Ruiz-Sola *et al*., 2016). A transport system must therefore exist to export isoprenoid units from the chloroplast to the rest of the cell; a candidate transmembrane antiporter has been identified (Weber *et al*., 2006), however this has yet to be confirmed experimentally for *C. reinhardtii*. The CPSFL1 protein, which is necessary for carotenoid accumulation in *C. reinhardtii* and *Arabidopsis* (*A.*) *thaliana*, has been proposed to act as a GGPP delivery system from the stroma to the chloroplast envelope for carotenoid biosynthesis (Hertle *et al*., 2020; García-Cerdán *et al*., 2020).

1.5.3. Committed carotenoid biosynthesis

The first committed step of carotenoid biosynthesis is the head-to-head condensation of two GGPP molecules by phytoene synthase (PSY) to form the colourless carotenoid phytoene (**Figure 1.4**; McCarthy *et al*., 2004). The enzyme phytoene desaturase (PDS) then catalyses a two-step desaturation of phytoene, producing ζ-carotene (**Figure 1.4**); this step is inhibited by the bleaching herbicide norflurazon (Breitenbach *et al*., 2001; Tran *et al*., 2012). Norflurazon was applied in **Chapter 4** to select for carotenoid-rich mutants. PDS is directly linked to the electron transport chain of photosynthesis, as it uses plastoquinone as its hydrogen acceptor (Grossman *et al*., 2004).

ζ-carotene is then converted to lycopene via further desaturation and isomerisation steps by ζ-carotene desaturase (ZDS), 15-cis-ζ-carotene isomerase (Z-ISO; Chen *et al*., 2010) and carotenoid isomerase (CRTISO; Lohr *et al*., 2005). At this point, the carotenoid pathway branches; depending on the cyclisation reaction type that occurs at the end groups of lycopene, either lutein or xanthophyll cycle carotenoids are produced.

For the synthesis of xanthophyll cycle pigments, lycopene is cyclised at both ends by the enzyme lycopene β-cyclase (LYCb), which introduces two β-ionone rings to form β,β-carotene (β-carotene). β-carotene is hydroxylated at its C3 and C3' positions by the non-haem di-iron hydroxylase CHYb to form zeaxanthin (Cunningham *et al*., 1996; Lohr *et al*., 2005), which is then epoxidated by zeaxanthin

epoxidase (ZEP) to form violaxanthin via the intermediate antheraxanthin (**Figure 1.4**, **1.5**; Baroli *et al*., 2003). Under stressful conditions such as light-saturation, violaxanthin can be converted back to zeaxanthin via the recently characterised enzyme, violaxanthin de-epoxidase (VDE; **Figure 1.4**, **1.5**; Li *et al*., 2016). The interconversion of zeaxanthin and violaxanthin is known as the xanthophyll cycle, the function and regulation of which will be discussed briefly in **Section 1.5.5**.

Similar to xanthophyll cycle pigment synthesis, the production of lutein involves the incorporation of a β-ionone ring by LCYb; however, this only occurs at one end of the lycopene molecule, the other end being cyclised by lycopene ε-cyclase (LCYe) to form β,ε-carotene (α-carotene). The ratios of lutein and zeaxanthin produced within the cell are therefore, at least in part, determined by the relative activities of LCYb and LCYe. The P450 enzymes CYP97A5 and CYP97C3 catalyse the hydroxylation of the β- and ε-ionone rings of α-carotene, respectively, to form lutein (Lohr *et* al., 2005; Kim *et al*., 2009; Niu *et al*., 2020).

Both violaxanthin and lutein are hydroxylated further to produce neoxanthin and loroxanthin, respectively; however, no neoxanthin synthase or loroxanthin synthase gene orthologues have yet been identified in *C. reinhardtii* in comparative studies with the *A. thaliana* genome (Lohr *et al*., 2005). Interestingly, although keto-carotenoids have not been detected in *C. reinhardtii*, a β-carotene ketolase (BKT) orthologue is present within the *C. reinhardtii* genome, however it is poorly expressed and resides there solely as an inactive pseudogene (Lohr *et al*., 2005; Merchant *et al*., 2007; Perozeni *et al*., 2020).

1.5.4. Carotenoid biosynthesis regulation

The biosynthesis of chlorophyll and carotenoid pigments are under stringent control by the cell. Sensory factors that are involved in the regulation of pigment production include light, heat, cellular redox state, and nutrient availability (Bohne and Linden, 2002; Napaumpaiporn and Sirikhachornkit, 2016; Couso *et al*., 2012; Juergens *et al*., 2015). The availability of light harvesting complex (LHC) and photosystem (PS) apoproteins also influences pigment levels (Polle *et al*., 2003)

Isoprenoid and carotenoid biosynthesis enzymes are regulated at the transcriptional, translational and post-translational levels. Studies regarding the effect of light on carotenoid gene transcripts have revealed upregulation of mRNA encoding DXR, DXS, IDI, PSY, PDS, LYCb and BchY under high light conditions (Bohne and Linden, 2002; Im *et al*., 2006; Sun *et al*., 2010). Blue light has also been found to increase levels of PSY and PDS, detected by a blue light phototropin, PHOT (Bohne and Linden, 2002; Im *et al*., 2006).

Increases in PSY and PDS transcripts do not necessarily equate to an increase in carotenoids, suggesting higher levels of biosynthetic regulation (Bohne and Linden, 2002). An example of post-translational regulation is the discovery that GGPP synthase undergoes feedback inhibition by GGPP (Sun *et al*., 2010). In *A. thaliana*, the PSY enzyme is activated by the DnaJ-like protein ORANGE, suggesting that there are other unknown factors at play in the regulation of carotenogenesis (Li *et al.,* 2012; Zhou *et al*., 2015). An ORANGE protein homologue in *C. reinhardtii* was cloned and endogenously overexpressed to increase carotenoid production in **Chapter 3**.

1.5.5. The xanthophyll cycle

The xanthophyll cycle is an important photoprotective mechanism in plants and algae, whereby zeaxanthin is reversibly epoxidated to violaxanthin (Demmig-Adams and Adams, 1996). Violaxanthin is found under standard, low light conditions in *C. reinhardtii*; this carotenoid is efficient at harvesting light energy. Under high-light or other stress conditions such as N-deprivation, violaxanthin is de-epoxidated to zeaxanthin (**Figure 1.5**). Zeaxanthin has an increased number of double bonds compared to violaxanthin and hence a greater conjugated π-electron system, which in turn lowers the molecule's state transition energy, thus facilitating energy acceptance from damaging free radicals such as triplet chlorophyll (^{3}Chl) and singlet oxygen and dissipating it as heat (Frank and Cogdell, 1998; Baroli *et al*., 2003). Furthermore, structural changes associated with de-epoxidation of violaxanthin trigger the rearrangement of photosynthetic complexes within the thylakoid membrane to lower the antenna size of photosystem II (PSII) and decrease the probability of photon saturation and free radical generation in high light (Minagawa and Tokutsu, 2015).

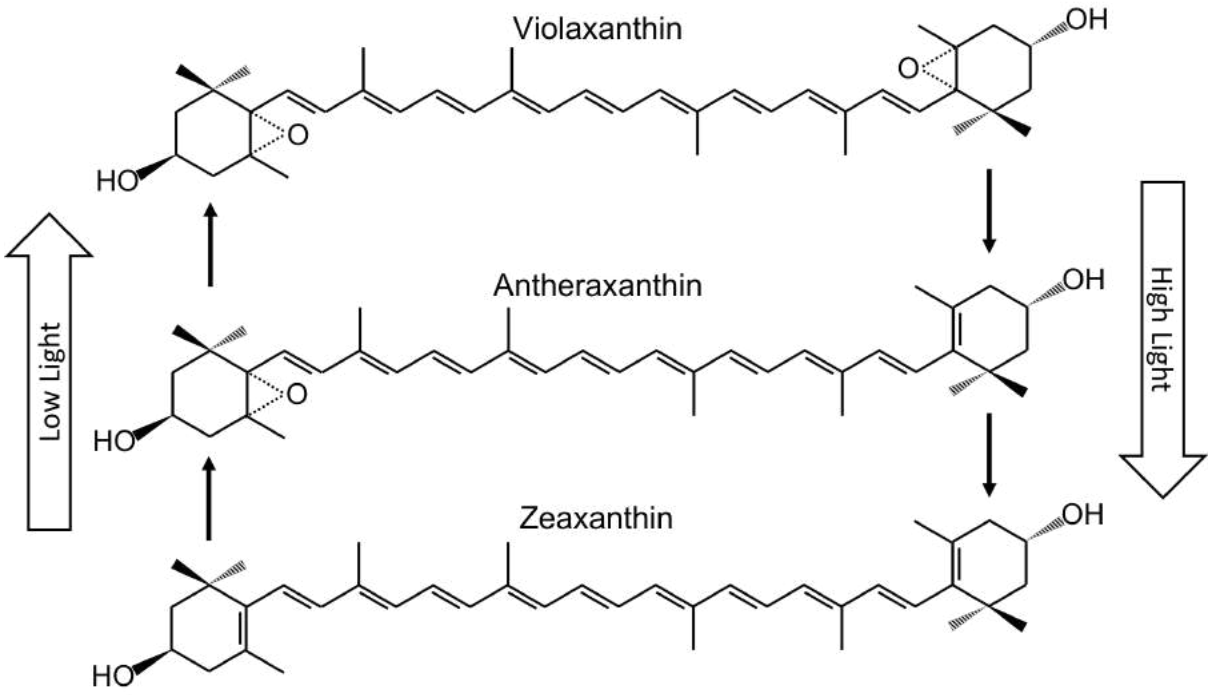


Figure 1.5: The xanthophyll epoxidation cycle. Under low light conditions, zeaxanthin is epoxidated by zeaxanthin epoxidase to generate violaxanthin via an antheraxanthin intermediate. High light triggers de-epoxidation of violaxanthin back to zeaxanthin, catalysed by violaxanthin de-epoxidase.

The xanthophyll cycle is regulated by the activities of ZEP and VDE (Baroli *et al*., 2003; Li *et al*., 2016), and de-epoxidation is activated by acidification of the thylakoid lumen (Demmig-Adams and Adams, 1996). Acidification of the thylakoid lumen occurs under high light stress, as an abundance of protons accumulates due to saturation of the electron transfer cycle. Higher levels of zeaxanthin relative to violaxanthin are detectable in *C. reinhardtii* under high light and N-deprivation stress conditions when compared to standard conditions, suggesting post-translational activation of the VDE enzyme (Couso *et al*., 2012; Juergens *et al*. 2015).

1.5.6. Lutein

Lutein is the predominant carotenoid involved in non-photochemical quenching (NPQ) under high light stress, and mutants lacking lutein are extremely light-sensitive (Niyogi *et al*., 1997). Under high light stress, greater quantities of lutein have been reported (Pineau *et al*., 2001; Couso *et al*., 2012). Lutein has also been shown to accumulate in *C. reinhardtii* when acclimated to harmful UV radiation (Korkaric *et al*., 2015).

Lutein is mostly found in the thylakoid membrane bound to PSII antenna, and can also be found in photosystem I (PSI) where it is the primary carotenoid (Pineau *et al*., 2001). The PSII proteins that bind lutein include CP29 which can bind 1 lutein cofactor, and CP26, LhcbM1, LhcbM2, and LhcbM3, each of which can each bind 2 molecules of lutein (Sheng *et al*., 2019). The associated energy dependant non-photochemical quenching (qE NPQ) protein light harvesting complex stress related protein 3 (LHCSR3) additionally contains a lutein binding site (Bonente *et al*., 2011).

1.6. The molecular toolkit for engineering the *C. reinhardtii* nuclear genome

1.6.1. Genetic engineering in the *C. reinhardtii* nucleus

Over the past few decades, the development and application of genetic engineering techniques in *C. reinhardtii* has increased exponentially (Scaife *et al*., 2015), rapidly advancing its effectiveness as a biotechnological host.

Many recent genetic manipulation endeavours in *C. reinhardtii* for recombinant protein expression have targeted the chloroplast genome, in part due to its ease of genomic manipulation by homologous recombination and high levels of protein accumulation (Rasala and Mayfield, 2015). However, in order to achieve more complex metabolic engineering goals in *C. reinhardtii*, the nuclear genome must be edited and exploited. For example, transit peptide sequences can be added to nuclear-expressed proteins, meaning that protein products can be effectively secreted, and heterologous enzymes can be directed to specific regions of the cell for more efficient substrate channelling (Rasala *et al*., 2012; Lauersen *et al*., 2013a, 2013b, 2015, 2016b). A broad range of post-translational modifications (PTMs) for protein products is also available for nuclear encoded proteins, such as glycosylation (Mathieu-Rivet *et al*., 2013); the chloroplast is generally limited to ancient prokaryotic-like protein synthesis machinery, limiting the PTMs available (Chen and Melis, 2013; Rasala and Mayfield, 2015). Furthermore, modifications that target *C. reinhardtii* nuclear gene expression, such as gene knock-outs, gene-silencing, or overexpressing a transcription factor, must be performed in the nuclear genome.

The reasons above make nuclear genomic editing an essential factor for *C. reinhardtii* to become a successful host for metabolic engineering; however, there are some serious set-backs that have slowed progression of nuclear engineering in *C. reinhardtii*. Deoxyribonucleic acid (DNA) introduced into *C. reinhardtii* generally integrates into the nuclear genome via non-homologous end joining (NHEJ) at seemingly random positions (Shimogawa *et al*., 1998; Kindle *et al*., 1990). Homologous recombination can occur, but at such low rates (1000:1 non-homologous to homologous DNA

uptake) that it is not efficient to rely on this method (Sodeinde and Kindle, 1993; Zorin *et al*., 2005). This increases the difficulty of editing endogenous genes and prevents the integration of exogenous DNA at specific loci, making standardisation of transformations nigh on impossible. Transgene silencing is also a major problem in the *C. reinhardtii* nucleus, often giving rise to undetectable exogenous protein levels, despite there being high levels of the corresponding mRNA (Cerutti *et al*., 1997). A loss of transgene expression over time has also been noted in some strains (Cerutti *et al*., 1997; Yamasaki *et al*., 2008).

Despite these issues, there has been some success in nuclear genome editing; the following sections of this review will assess the tools currently available to surmount these complications and manipulate the *C. reinhardtii* nuclear genome.

1.6.2. Strains

A diverse range of *C. reinhardtii* strains and mutants is available for use as biotechnological hosts, each with properties suitable for certain biotechnological processes. Important resources and culture collections include the Chlamydomonas Resource Centre at the University of Minnesota (http://www.chamycollection.org/) and the Culture Collection of Algae and Protozoa (http://www.ccap.ac.uk/). One example highlighting the importance of strain selection is in choosing the thickness of the cell wall. Cell wall-deficient strains are generally used for genetic engineering for several reasons; they are more readily transformable, secrete products more efficiently, and are more easily lysed thus potentially lowering downstream processing costs. Conversely, cell wall-intact strains may be more appropriate for particular tasks such as high-volume bioreactor growth, due to their robustness and resistance to shear force.

If the desired modification requires transgene expression in the nucleus, one strategy to overcome epigenetic silencing in *C. reinhardtii* is to use mutant strains with enhanced heterologous gene expression. Two UV-mutagenised (UVM) cell-wall deficient strains isolated by Neupert *et al*. (2009), UVM4 and UVM11, exhibit increased nuclear transgene messenger ribonucleic acid (mRNA) and protein expression. UVM4 has become a widely-used tool for nuclear protein expression and genetic manipulation (Lauersen *et al*., 2013b; Kong *et al*., 2014; Lauersen *et al*., 2016a, 2016b; Jarquín-Cordero *et al*., 2020; Mehrshahi *et al*., 2020). MET1 is another *C. reinhardtii* mutant strain that exhibits increased expression of transgenes; this was created by insertional mutagenesis, which induced the disruption of the maintenance methyltransferase 1 (*MET1*) gene (Kong *et al*., 2015). This strain was further mutagenised to produce a greatly enhanced transgene expresser,

demonstrating the presence of methylation and non-methylation-based transgene silencing mechanisms in *C. reinhardtii* (Kurniasih *et al*., 2016).

These overexpression strains are useful but not perfect, as they still exhibit some transgene silencing and they lack cell walls; further mutagenesis experiments, or eventually direct engineering, could be applied to *C. reinhardtii* to create stable strains capable of high levels of transgene expression.

1.6.3. Mutagenesis

A relatively simple approach for generating biotechnologically useful *C. reinhardtii* strains is via mutagenesis. Many comprehensive strain and mutant collections exist, which can then be screened for desired phenotypes (Gonzalez-Ballester *et al*., 2011; Dent *et al*., 2015; Terashima *et al*., 2015; Li *et al*., 2015; Li *et al*., 2019).

Chemical mutagenesis is one method for generating *C. reinhardtii* mutants. Popular chemical mutagens include nitrosoguanidine, methyl methanesulphonate and ethylmethanesulphonate (EMS). Another useful mutagenesis technique is using UV light; this technique has the advantage of being safer to apply than by handling chemical mutagens. Additionally, very low levels of UV can be applied to cells without having to declare the resulting strains GMO; this could be particularly useful in increasing synthesis of natural products that are of interest to nutraceutical and health food industries (Schierenbeck *et al*., 2015). EMS mutagenesis was applied to generate *C. reinhardtii* carotenoid-producing mutants in **Chapter 4**.

Insertional mutagenesis is another option for generating strains; this technique exploits the *C. reinhardtii* natural NHEJ DNA uptake mechanism by randomly inserting a selective insertion cassette within the nuclear genome (Vila *et al*., 2013; Zhang *et al*., 2014; Dent *et al*., 2015). Generating mutants by insertional mutagenesis offers the advantage of being able to pin-point the site of insertion by using the known sequence of the insertion cassette to 'pull out' its flanking sequences using complementary primers or affinity tags, allowing the genetic basis of the phenotype to be established (Gonzalez-Ballester *et al*., 2005, 2011; Zhang *et al*., 2014); this avoids the complex and expensive whole-genome sequencing required to identify point mutations generated by UV or chemical mutagenesis. The downside to insertional mutagenesis is that mutations are limited to gene disruption, ruling out the potential 'gain of function' mutations that can be achieved through generation of single nucleotide polymorphisms (SNPs) with chemical or UV mutagenesis.

1.6.4. DNA delivery: transformation techniques

For genetic engineering to take place, the DNA, protein or RNA components required to alter cellular output must pass through the cell wall and/ or lipid bilayer to become effective; this must also occur without damaging the biological tool itself.

The most widely used method for gene delivery is via agitation with glass beads, first described by Kindle (1990); this method is quick and cost-effective. Another technique for biological tool delivery is biolistic particle bombardment of cells by gene gun. This is one of the most versatile methods of DNA delivery, as DNA can be transported to mitochondria, chloroplast and nucleus using this method (Mussnung, 2015). Microparticles, generally gold, are coated in exogenous DNA and shot into the cells. The equipment, however, is relatively complex and expensive, bringing one downside to this method of transformation. *Agrobacterium tumefaciens* has also been used to deliver DNA into *C. reinhardtii* cells, although not many examples have been noted since Kumar *et al*. (2004) developed the method. This method is cheap and easy as it merely involves transformation of *Agrobacterium tumefaciens* with a binary vector containing desired DNA, then the transformant bacteria mixed with *C. reinhardtii* (Kumar *et al*., 2004).

Electroporation of *C. reinhardtii* cells with foreign DNA achieves the highest transformation efficiencies of all of the methods listed above (Brown *et al*., 1991; Shimogawara *et al*., 1998; Jinkerson and Jonikas, 2015). This method can be applied to strains with or without a cell wall (Brown *et al*., 1991). Foreign proteins and RNAs can also be introduced into *C. reinhardtii* via this technique, for example the cas9 nuclease and single guided RNA (sgRNA) was added to *C. reinhardtii* cells, allowing clustered regularly interspaced short palindromic repeats (CRISPR)/ cas9 nuclear genome editing to successfully take place (Shin *et al*., 2016; See **Section 1.6.7.3.**).

1.6.5. Tools for improving transgene expression in *C. reinhardtii*

As mentioned above (See **Section 1.6.1.**), there are some unique problems that come with nuclear expression of foreign genes in *C. reinhardtii*. A variety of regulatory components have been identified or designed to overcome these hurdles and enhance genetic transformation of *C. reinhardtii*, as discussed below.

1.6.6.1. Promoters

Strong promoters are essential components to any metabolic engineer's toolkit in order to manipulate transgene expression. Constitutive promoters, which stimulate constant expression of

the gene under their command, are extremely useful for simple overexpression applications. The dominant promoter system used for transgene expression in *C. reinhardtii* is the *Hsp70A-Rbsc2* expression cassette, which incorporates the core promoter from the constitutively expressed ribulose bisphosphate carboxylase small subunit 2 (RbcS2) and an enhancer element from the heat-shock protein 70A (Hsp70A) promoter (Schroda *et al*., 2000; Sizova *et al*., 2001). PsaD is another widely used strong constitutive promoter (Fischer and Rochaix, 2001). High expression synthetic promoters have been designed *in silico* based on common DNA sequences found within the promoters of highly expressed nuclear *C. reinhardtii* genes (Scranton *et al*., 2016). A related study was undertaken in **Chapter 5**, where DNA motifs discovered within strong constitutive promoters were tested for promoter activity individually, and their ability to drive protein expression to a range of levels was demonstrated.

For synthetic biology systems to be realised in *C. reinhardtii*, effective inducible promoters are required to allow conditional gene expression circuits to function within the cell. Inducible promoters are also valuable for fine-tuning enzymatic gene expression to maximise flux through a metabolic pathway, and to control when a recombinant protein is expressed as some products are toxic if expressed in too early a growth phase. Inducible promoters available for *C. reinhardtii* include copper-responsive CYC6 (Quinn and Merchant, 1995), Iron-responsive ATX1 (Fei and Deng, 2007), CO_2-responsive CA1 (Villand *et al*., 1997), ammonium responsive NIT1 (Ohresser *et al*., 1997), phosphorus-responsive SDQ1 (Ruecker *et al*., 2008), cobalamin b_{12} suppression (Helliwell *et al*., 2014), high light inducible protein promoter (LIP; Baek *et al*., 2016a) and sodium chloride (NaCl)-responsive Femu2 (Li *et al*., 2017).

Scrapping the use of promoter systems altogether by transforming naked genes has also resulted in high transgene expression. Díaz-Santos *et al*. (2016) co-transformed promoter-less genes into *C. reinhardtii* using glass-bead method; many transformants successfully integrated, transcribed and translated the exogenous genes, which had integrated in frame of endogenous promoters due to random NHEJ. The lack of species-specific components makes this a potential broad-use strategy for transgene expression in multiple species of microalgae. However, for targeted or more complex metabolic engineering goals, the integration of a promoter system is necessary.

1.6.5.2. Introns

Compared to other microalgal species, the *C. reinhardtii* nuclear genome has a high intron density, with an average of 7.3 introns per gene (Merchant *et al*., 2007). The incorporation of the first intron

of RbcS2 was demonstrated to enhance expression of transgenes in *C. reinhardtii* (Lumbreras *et al*., 1998). Eichler-Stahlberg *et al*. (2009) took this further by including an RbcS2 intron 2 within the gene of interest (GOI), further increasing protein production. RbcS2 introns are now regularly used for enhancing expression of exogenous genes (Lauersen *et al*., 2015, 2016b, 2018; Wichmann *et al*., 2018), and additional intron sequences capable of improving transgene expression have been identified (López-Paz *et al*., 2017). Incorporating multiple intron copies spaced throughout large transgenes can improve protein yields (Baier *et al*., 2018) and the number and positions of introns can be optimised using the web tool Intronserter (Jaeger *et al*., 2019).

1.6.5.3. Functional transport peptides

Compartmentalisation of metabolic reactions is one of the advantages to using a eukaryotic nuclear transgene expression system, as sub-cellular localisation of enzymes to their functional sites increases access to their substrate. Functional peptide sequences can be added to a protein to direct it to various cellular compartments. Recently, specific localisation signal peptides were fused to fluorescent protein reporters in *C. reinhardtii*; the fluorophores were successfully transported to such locations as the cytoplasm, nucleus, chloroplast, mitochondria and microbodies, as well as being secreted into the extracellular medium, thus increasing the tools available for metabolic engineering in *C. reinhardtii* (Rasala *et al*., 2014; Lauersen *et al*., 2015).

1.6.5.4. Codon optimisation

As the GC content of the *C. reinhardtii* nuclear genome is relatively high (Merchant *et al*. (2007), optimising the coding regions of exogenous genes to reflect this GC richness can reduce transgene silencing (Shao and Bock, 2008). This is due to in an increase in mRNA stability when genes are codon-optimised, as well as preventing heterochromatin formation induced by unfavourable GC content (Presnyak *et al*., 2015; Barahimipour *et al*., 2015). Fine-tuning codon usage surrounding the translation initiation site can further boost protein abundance (Weiner *et al*., 2018).

1.6.5.5. RNA devices

RNA devices such as ligand-sensitive riboswitches, silencing micro RNAs (miRNAs) and interchangeable RNA aptamer domains offer an alternative means of controlling gene expression to generic transcription factor/ promoter expression systems, and enable the design of complex synthetic biology workflows (Liang *et al*., 2011). There are currently very few RNA-based expression tools available for *C. reinhardtii*, but recent efforts to change this are promising (Navarro and Baulcombe, 2019); in a recent example, a riboswitch that alters protein expression levels in

response to ligands in a dose-dependent manner was developed for *C. reinhardtii*, facilitating fine-tuning of transgene expression (Mehrshahi *et al*., 2020).

1.6.5.6. Selection markers

Following transformation of cells with foreign DNA, cells which have successfully integrated the GOI need to be distinguished from those that have not. Incorporating antibiotic resistance to an expression cassette is a common method for selection of successful transformation by growing transformants on selective media.

Examples of antibiotic resistance genes that have been used for this purpose in *C. reinhardtii* include introducing resistance to hygromycin B by *aphVII* from *Streptomyces hygroscopicus* (Berthold *et al*., 2002), paromomycin by *aphVIII* from *Streptomyces rimosus* (Sizova *et al*., 2001), bleomycin-antibiotic-family (eg zeocin) by *Sh ble* from *Streptoalloteichus hindustanus* (Stevens *et al*., 1996), AadA spectinomycin resistance gene from *Escherichia coli* (*E. coli*) plasmid R538-1 (Meslet-Cladiere *et al*., 2011), a synthetic *tetX* tetracycline resistance gene (Garcia-Echauri and Cardineau, 2015) and modified *nptII* kanamycin resistance gene from *E. coli* transposon Tn5 (Barahimipour *et al*., 2016).

Complementation of auxotrophic mutants with their missing lethal metabolic enzyme is another selection method, for example transforming *NIT1* into ammonium-requiring auxotrophs (Kindle *et al*., 1989) and *ARG7* into arginine-requiring mutants (Debuchy *et al*., 1989; Nour-Eldin *et al*., 2016).

Cotransformation of *C. reinhardtii* with enhanced native genes that provide resistance from certain biocides can also be utilised as dominant selectable markers; improved acetolactate synthase provides resistance to sulfometuron methyl (Kovar *et al*., 2002) and Cry1 grants resistance to cryptopleurine and emetine (Nelson *et al*., 1994). The herbicide norflurazon is a demonstrated selection marker, where transformants are selected for by introduction of a modified norflurazon-resistant *C. reinhardtii* PDS, or a bacterial PDS that is naturally unaffected by norflurazon (Liu *et al*., 2013; Molina-Márquez *et al*., 2019).

1.6.5.7. Reporter genes

The first fluorescent reporter gene was codon-optimised green fluorescent protein expressed under the influence of the RbcS2 promoter (Fuhrmann *et al*., 1999). Since then, a many fluorescent reporter genes of different colours have been codon-optimised and expressed in *C. reinhardtii* such as cyan (mCerulean), yellow (mVenus), orange (tdTomato), red (mCherry) and blue blue (mTagBFP; Rasala *et al*., 2013, 2014; Lauersen *et al*., 2015). These fluorescent tags are visible against *C.*

reinhardtii chlorophyll autofluorescence, and can be fused to other proteins as a way to measure and locate protein expression (Rasala *et al*., 2014; Lauersen *et al*., 2016b). Bioluminescent reporters, such as luciferases, are also effective reporters in *C. reinhardtii* (Fuhrmann *et al.,* 2004; Shao and Bock, 2008). The endogenous ARS2 gene that codes for an arylsulphatase has also been adapted as a reporter gene; its extracellular secretion can be quickly detected and quantified using a colorimetric assay (Specht *et al*., 2015).

1.6.5.8. Bicistronic DNA and self-cleaving peptides

One strategy employed to reduce the effects of transgene silencing of foreign genes is to fuse the GOI with the *ble* antibiotic resistance marker via a foot-and-mouth-disease virus (FMDV) 2A self-cleaving peptide linker region (Rasala *et al*., 2012). The GOI and *ble* marker are transcribed as one mRNA transcript, which is later cleaved at the translational level by a proposed 'ribosome skipping' mechanism, thus producing two separate proteins. Xylanse1 expressed under this expression system resulted in ~100-fold increase in protein product when compared to the open reading frames (ORFs) expressed separately (Rasala *et al*., 2012). Other bicistronic expression systems have been developed by Onishi and Pringle (2016) and Dong *et al*. (2017); these tools can greatly simplify transformations by eliminating the need for secondary screening for GOI expression.

1.6.5.9. Standardised vectors and modules

Standardised customisable vectors that contain a variety of DNA modules such as untranslated regions (UTRs), promoters, reporters, selection markers, transit peptides and tags are available to facilitate transgene expression in *C. reinhardtii,* such as the pOptimised vectors that contain sites for GOI insertion by restriction enzymatic digestion (Lauersen *et al*., 2015; Wichmann *et al*., 2018), and the MoClo vectors and genetic modules that can be recombined using Golden Gate cloning (Crozet *et al*., 2018). The standardisation of transformation systems and genetic parts in *C. reinhardtii* has greatly advanced its potential to becoming a key organism for synthetic biology (Crozet *et al*., 2018).

1.6.5.10. Co-transformation and mating

In order to manipulate and build upon a metabolic pathway in *C. reinhardtii*, it is likely that multiple transgenes will need to be expressed in a cellular system. One method for co-transforming cells is by mating two transformant strains of opposite mating types and selecting for progeny that express both sets of transgenes. This was demonstrated by Rasala *et al*. (2014), who mated strains expressing mCerulean and mCherry together to produce double-mutants. The double mutant was

then mated with another fluorescent protein-expressing strain, to create a triple mutant, and so on until a rainbow algae strain was created (Rasala *et al*., 2014). Interestingly, the progeny of mated cells expressed the same or higher levels of transgene, suggesting that silencing is reduced in daughter cells (Rasala *et al*., 2014).

Alternatively, cells can be transformed with the first GOI, then transformed again with the second GOI using a different selection marker, and so on until a multi-transformant strain expressing several transgenes is achieved (Lauersen *et al*., 2015; Wichmann *et al.,* 2018).

1.6.7. Tools for targeted genome editing in *C. reinhardtii*

The direct engineering of endogenous *C. reinhardtii* genes has been limited by its lack of reliable homologous recombination machinery and random integration of transformed DNA. New technologies are however being successfully applied to target specific nuclear genomic targets, so it is only a matter of time before more direct engineering approaches can be made in this organism. Examples of these technologies are discussed below.

1.6.7.1. RNA interference-based methods for endogenous gene silencing

RNA interference (RNAi) is an effective strategy for reducing native gene expression (Rohr *et al*., 2004). It is generally accepted that *C. reinhardtii* naturally possesses machinery that recognises and degrades double stranded RNA sequences into small interfering RNAs, which can then bind complementary mRNA sequences as a signal for degradation (Cerutti, 2003); this can be exploited by transforming cells with the antisense or inverted repeat sequence of the target gene, which when transcribed will bind to the target gene mRNA, causing transcript degradation (Rohr *et al*., 2004; Schroda, 2006). Artificial micro RNA (amiRNA) based on micro RNA precursors are also effective tools for native gene knock-downs (Molnar *et al*., 2009; Zhao *et al*., 2009; Hu *et al*., 2014b). RNA-based gene silencing strategies are relatively easy to apply and generally effective; however, disadvantages include possible unwanted off-target effects, incomplete silencing, and strain-to-strain variation in silencing effectiveness (Rohr *et al*., 2004). Recently, artificial miRNA was applied to downregulate autophagy in *C. reinhardtii*, which in turn increased β-carotene output by 2.3-fold per g DCW (Tran *et al*., 2019).

1.6.7.2. Directed endonucleolytic cleavage

Due to the *C. reinhardtii* endogenous NHEJ DNA repair mechanism, cleaved nuclear DNA often results in deletions or insertions following repair, causing a gene knock-out. Fusing highly specific

DNA-binding domains, such as zinc-finger domains and transcription activator-like effectors (TALEs), to a non-specific nuclease (eg FokI) can target restriction enzyme activity to a target endogenous gene. Recently, zinc finger nucleases and TALEs have been successfully applied to knock out genes in *C. reinhardtii* (Sizova *et al*., 2013; Gao *et al*., 2014, 2015).

1.6.7.3 CRISPR

The recently discovered clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) system allows for highly specific editing of genomic DNA. The first attempt to use this tool in *C. reinhardtii* was unsuccessful, as plasmids containing genes encoding the ribonucleoproteins were toxic to the cell (Jiang *et al*., 2014). A DNA-free method was later developed, where the cas9 proteins and their sgRNAs were introduced via electroporation; *C. reinhardtii* survived, and successful gene knock-outs and knock-ins were achieved (Baek *et al*., 2016b; Shin *et al*., 2016). Recent applications of CRISPR in *C. reinhardtii* include an investigation into photoreceptor function through gene disruption (Greiner *et al*., 2017), improvement of zeaxanthin accumulation through knocking out of a zeaxanthin epoxidase (Baek *et al*., 2018), and enhancement of lipid production through expression interference of a carboxylase enzyme (Kao and Ng, 2017).

1.7. Metabolic engineering in *C. reinhardtii*

1.7.1. Introduction

Metabolic engineering can be defined as: “The directed improvement of product formation or cellular properties through the modification of specific biochemical reaction (s) or the introduction of new one (s) with the use of recombinant DNA technology.” (Stephanopoulos *et al*., 1998). This section will highlight examples of metabolic engineering strategies that have been applied to *C. reinhardtii* and related organisms, with a main focus on editing the nuclear genome, and on engineering isoprenoid-related pathways.

1.7.2. Adaptive evolution and spontaneous mutations

Adaptive laboratory evolution (ALE) could be an alternative means of identifying novel metabolic engineering targets. Adaptive evolution relies on small individual mutations that confer a selective advantage to individuals under a particular condition. Repetitive rounds of selection for the advantageous phenotype eventually result in a pool of organisms exhibiting an improved trait. The basis of the enhancement can then be determined by -omics comparisons of different time-points during the ALE procedure. Velmurugan *et al*. (2014) applied this strategy in *C. reinhardtii* to produce

strains with improved lipid accumulation. By using a fluorescence-activated cell sorting (FACS) method, they repeatedly separated and cultured high lipid-accumulating cells; 50 and 175 % lipid increases in two populations were ultimately observed. Proteomic analysis of time points provided insights into the carbon and nitrogen pathways that are upregulated in lipid over producers, thus uncovering potential metabolic engineering targets for increasing *C. reinhardtii* lipid production. Besides other ALE studies aiming for lipid enhancement (Shin *et al*., 2017; Kato *et al*., 2017), this technique has been applied to improve chloroplast protein expression (Fields *et al*., 2019) and to find spontaneous high light tolerant mutants (Shierenbeck *et al*., 2015). As of yet, no studies for improvement of carotenoid biosynthesis using ALE exist for *C. reinhardtii*.

1.7.3. Recombinant protein expression

Due to difficulties with heterologous gene expression in the nuclear genome, recombinant protein expression, in which the product is the protein itself, is generally performed in the *C. reinhardtii* chloroplast, where high protein titres are attainable (Potvin and Zhang, 2010). There are some examples, however, of successful expression of desired protein products from the nuclear genome.

The industrial enzyme xylanase1 was expressed in *C. reinhardtii* using the *ble*2A self-cleaving peptide expression system (See above); the xylanase1 protein was successfully targeted for secretion using a signal peptide, which aided recovery of the product (Rasala *et al*., 2012). Soon after, an ice-binding protein was transfected into the nuclear genome, overexpressed under the control of *Hsp70A-RbcS2* promoter, and also directed for secretion into the growth media (Lauersen *et al*., 2013b). In both experiments, the growth media was tested for enzymatic activity and both yielded high activity; this demonstrates *C. reinhardtii*'s potential value as a protein product secretion system.

Recently, the HIV antigen P24 was codon optimised for *C. reinhardtii* and transformed into the UVM4 strain (Barahimipour *et al*., 2016). Recombinant protein levels of up to 0.25 % of total cellular protein were reported, which are comparable to those achieved for the P24 antigen expressed in the established biotechnological host tobacco, suggesting *C. reinhardtii* can be competitive with seed plants as a recombinant protein production chassis.

1.7.4. Rate-limiting enzyme overexpression

A classic metabolic engineering strategy to improve flux through a metabolic pathway is to identify rate-limiting enzymes within a system, then to overexpress these enzymes to alleviate the bottle-

neck. This approach has been attempted with the isoprenoid and carotenoid biosynthetic pathways in *C. reinhardtii*.

The PSY gene from *D. salina* was cloned and overexpressed in *C. reinhardtii* under control of the *Hsp70A-RbcS2* overexpression system, which led to a maximum 2.6-fold increase in carotenoids compared to wild-type (WT) *C. reinhardtii* (Couso *et al*., 2011). A similar experiment, in which PSY from *Chl. zofingiensis* was heterologously expressed in *C. reinhardtii*, resulted in a 2.2-fold carotenoid increase (Cordero *et al*., 2011a). Although carotenoid levels are higher in the transformant strains described, they do not increase to a level that makes carotenoid production in *C. reinhardtii* economically viable. Furthermore, Kajikawa *et al*. (2015) attempted to improve production of the isoprenoid metabolite squalene in *C. reinhardtii* by overexpressing native squalene synthase under control of the PsaD promoter in the UVM4 strain; no increases in squalene were detectable, possibly due to stringent regulation at the post-transcriptional level.

The overexpression of bottle-neck enzymes is hence not a particularly effective means of improving product yield in *C. reinhardtii* due to as yet unknown regulatory mechanisms.

1.7.5. Competing pathway down-regulation

Vila *et al*. (2008) downregulated PDS protein levels in *C. reinhardtii* using antisense cDNA, although this did not lead to changes in the carotenoid levels in the cells. As part of a study that aimed to increase squalene synthesis in *C. reinhardtii* (Kajikawa *et al*., 2015), the squalene epoxidase gene that converts squalene into an epoxide, was knocked down by expression of an amiRNA in *C. reinhardtii*. This resulted in a reduction in transcript level of 59–76%, and accumulation of 0.9–1.1 micrograms (μg) of squalene mg^{-1} DCW, compared to no squalene detected in the WT. This suggests that knocking down competing metabolic enzymes is a more effective means of increasing metabolite production, however regulatory mechanisms within the cell can overcome this.

Recently, the ZEP gene in *C. reinhardtii* was knocked out using the CRISPR/ cas9 method, which eliminated enzymatic conversion of zeaxanthin to violaxanthin completely, resulting in constitutive zeaxanthin accumulation (Baek *et al*., 2016).

1.7.6. Introduction of new biosynthetic pathways

Attempts to expand the carotenoid profile of *C. reinhardtii* for the production of the commercial carotenoid astaxanthin have been undertaken by León *et al*. (2007). The BKT cloned from *H. pluvialis* was expressed in *C. reinhardtii* using the *RbcS2* promoter and chloroplast transit peptide. Neither

astaxanthin nor canthaxanthin, the keto-carotenoids produced by BKT in *H. pluvialis*, were produced; the novel carotenoid 4-keto-lutein was instead synthesised. 4-keto-lutein was produced in very small quantities and is not a desirable carotenoid in industry. This was however the first example of exogenous carotenoid genes being expressed within a microalga (Leon *et al*., 2007). Very recently, *C. reinhardtii* BKT was synthetically redesigned to remove a non-functional C-terminal domain and natively overexpressed, which led to 50% conversion of other ketocarotenoids into astaxanthin, and distinct reddish-brown coloured cultures (Perozeni *et al*., 2020).

A recent success in novel metabolite production in *C. reinhardtii* was its use as a production host for the sesquiterpenoid patchulol (Lauersen *et al*., 2016b). Patchulol synthase from the plant species *Pogostemon cablin* Benth was cloned and overexpressed in *C. reinhardtii* using the *Hsp70A-RbcS2* promoter system and the highest protein producing strains selected. Up to 1.03 mg per litre (L) of patchulol product could be produced. (E)-α-bisabolene, casbene, taxadiene, and 13R(+) manoyl oxide have since been non-natively synthesised in *C. reinhardtii* (Wichmann *et al*., 2018; Lauersen *et al*., 2018; Mehrshahi *et al*., 2020). This showcases *C. reinhardtii* as a sustainable terpenoid-producing microbial host.

1.7.7. Transcription factor engineering

Classical metabolic engineering strategies that have been applied to *C. reinhardtii* to improve production of metabolites have achieved limited success. Another strategy could be to manipulate several enzymes within a pathway from a single control point. Transcription factors (TFs) are regulatory elements that can positively or negatively affect target gene expression, or even entire pathways. Due to limited success in modifying individual genes, identifying TF targets and manipulating their expression could be a useful way to 'hack' metabolism and to circumvent complex regulatory processes (Bajhaiya *et al*., 2017).

This has been shown to be an effective strategy in *C. reinhardtii*. The gene for the *C. reinhardtii* PSR1 regulatory element was found to be missing in a mutant strain lacking the phosphorus (P)-deprivation response of starch and lipid synthesis (Bajhaiya *et al*., 2016). PSR1 was then overexpressed, inducing starch synthesis in P-replete strains. The transcriptome of the overexpression mutant was comparable to that in wild-type subjected to P-deprivation, indicating the effectiveness of PRS1 TF overexpression (Bajhaiya *et al*., 2016).

Examples of TF engineering in *C. reinhardtii* are few but show potential (Anderson *et al*., 2017; Salas-Montantes *et al*., 2017). The discovery and overexpression of an isoprenoid or carotenoid

biosynthetic regulator could vastly increase *C. reinhardtii*'s potential as a green producer of terpenoid compounds.

1.8. Thesis motivations

1.8.1. Strain and product selection

To develop a research question, a somewhat broader goal was first considered: to synthesise high-value products in microalgae using metabolic engineering. The high-value product and the microalgal strain had to first be selected, then more specific strategies and goals were developed.

As described above, microalgae are natural producers of carotenoids, which are high-value molecules with uses in several industries. The yellow carotenoid lutein has a rapidly growing market, and is mainly produced synthetically from fossil-fuel compounds or biologically from seasonal, energy-intensive marigold farming. Enhancing the natural production of lutein in a microalgal strain has the potential to improve the efficiency of lutein production for reasons discussed above. Furthermore, carotenoids are intimately associated with several important metabolic pathways and systems, meaning that increasing carotenoid production could have interesting and potentially useful knock-on effects. For example, light harvesting and non-photochemical quenching (NPQ) are mediated by carotenoid pigments. This means that a carotenoid over-producing strain could potentially be high-light or reactive oxygen species (ROS)-resistant, which in turn could be useful for various other biotechnological or research purposes. Moreover, improving carotenogenesis by engineering upstream isoprenoid pathways could broaden the applications of this work, by enhancing the potential of the strain to synthesise valuable terpenoids. For these reasons, lutein was selected as the target high-value product.

C. reinhardtii naturally synthesises lutein as a primary carotenoid to moderate levels. Although other microalgal strains possess superior lutein production capacities, *C. reinhardtii* was selected as the host strain due to the wealth of genetic engineering tools and -omics data available for *C. reinhardtii* compared to other microalgal species. The haploid genome of *C. reinhardtii* renders all mutations dominant, and the ability to mate *C. reinhardtii* with other strains in the laboratory enables classical genetics to be performed for strain genotyping, and for optimal strains to be generated by combining strain phenotypes. Genetically modified *C. reinhardtii* has also been successfully cultivated in industrial settings, for example Triton Algae Industries have commercialised the production of colostrum (breast milk proteins) from the *C. reinhardtii* chloroplast. Given these motivations, *C. reinhardtii* was selected as the host strain for carotenoid production.

1.8.2. Aims and objectives

After selecting lutein as the product of interest and *C. reinhardtii* as the biological cell factory, the following pieces of work were undertaken, as summarised in **Figure 1.6**.

The aim of the first experimental chapter of this thesis (**Chapter 3**) was to use available genetic engineering tools to overexpress a carotenoid pathway regulator in *C. reinhardtii*, with the goal of increasing carotenoid production. The classic metabolic engineering strategy of overexpressing bottle-neck enzymes in the carotenoid pathway has previously been attempted in *C. reinhardtii*; this was successful, but the lutein fold-increases were modest, at around 2-fold (Couso *et al*., 2011; Cordero *et al*., 2011a). Overexpressing a carotenoid pathway regulator could potentially target multiple enzymes as opposed to just one, and increase carotenoid levels further. Therefore, the *C. reinhardtii* homologue of ORANGE, a carotenoid pathway regulator in higher plants, was selected as the genetic engineering target for this chapter. This rationally-designed overexpression experiment resulted in the characterisation of a novel carotenoid regulator in *C. reinhardtii* and a 2.0-fold cellular increase in lutein.

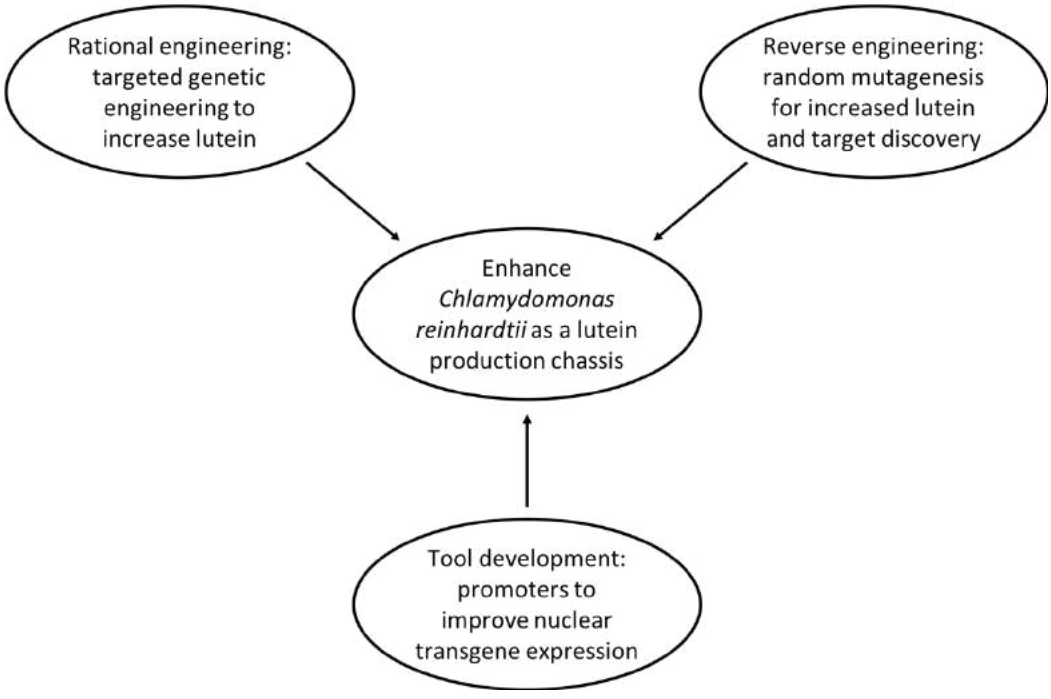


Figure 1.6: Summary of chapter aims and objectives with central goal. Three strategies were undertaken to achieve the overall goal of enhancing *C. reinhardtii* as a lutein production chassis. A rational engineering approach was first taken, in which a regulator of carotenoid biosynthesis was targeted for overexpression to

enhance lutein production. A reverse engineering approach was also applied by employing random chemical mutagenesis and selecting for strains with enhanced lutein production. The best strain was then subjected to proteomic analysis to give an in-depth picture of the mutant phenotype, to facilitate strain optimisation and to highlight potential genetic engineering targets for future experiments. The third methodology employed was to expand the genetic toolbox available to manipulate *C. reinhardtii*, with the goal of improving its capacity for metabolic engineering in the nuclear genome.

In contrast to **Chapter 3**, a reverse engineering approach was employed in **Chapter 4** to improve lutein production in *C. reinhardtii*, where a random mutagenesis and selection workflow was developed to isolate carotenoid-hyperproducing mutants. This random approach was selected due to its potential to override complex regulatory systems. Moreover, developing the phenotype first, then working backwards to characterise the desired strain could quickly generate an industry-ready strain while potentially revealing novel regulatory mechanisms and/or future metabolic engineering targets. Shotgun proteomics was employed to characterise the isolated strain, which provided a snapshot of the metabolic and redox state of the mutant line. This provided valuable insight into ways to optimize the growth of this strain, as well as for future characterisation and metabolic engineering experiments without having to perform NGS.

Despite the *C. reinhardtii* genetic engineering toolbox being relatively advanced in comparison to other microalgal species, issues still stand surrounding nuclear transgene expression, which impedes the usage of *C. reinhardtii* as an industrial biotechnology host. Nuclear transgene expression in *C. reinhardtii* is typically low and unstable due to silencing, and unreliable due to random genomic integration (**Section 1.6.1.**). However, this problem is not isolated to *C. reinhardtii*; other biotechnological chassis, such as Chinese hamster ovary (CHO) cells, face problems with unstable expression and random genomic integration that have been successfully tackled by developing high-expression synthetic promoters (Brown *et al*., 2014). Furthermore, very few DNA elements constituting high-expression nuclear promoters in *C. reinhardtii* have been characterised; known active DNA elements would be highly useful for the rational design of promoters for synthetic biology systems. In **Chapter 5**, a bioinformatic exploration of the promoters of highly expressed genes in *C. reinhardtii* was employed to extract potentially active DNA elements, then the motifs were tested *in vivo* using a fluorescent protein reporter and flow cytometry to calculate the median expression levels for each promoter element, thus revealing novel promoters/ promoter elements for improved metabolic engineering of the *C. reinhardtii* nuclear genome.

In summary, both forward and reverse engineering approaches were employed to improve lutein production in *C. reinhardtii*, while metabolic engineering tools were developed with the goal of improving *C. reinhardtii* as a chassis for the production of carotenoids and other high-value compounds.

Chapter 2: Materials and Methods

2.1. Standard Buffers, Reagents and Media

Deionised filtered water (dH_2O) was used for preparing buffers and media, and nuclease-free water (nfH_2O) for DNA preparations (Qiagen). All solvents used were high performance liquid chromatography (HPLC) grade. For sterilisation, growth media was autoclaved at 15 pound-force per square inch for minimum 20 minutes (min), or passed through 0.2 micrometre (µm) filters. All media was cooled to < 55°C before adding antibiotics. Chemicals and reagents were purchased from Sigma/ Merck unless stated otherwise. For media and buffer recipes not provided in the text, see **Appendix A**.

2.2. *E. coli*

2.2.1. Strains and Growth Conditions

All DNA cloning and preparation was completed in the host strain *E. coli* DH5α (Provided by Dr. S. Jaffé). *E. coli* strains were grown in Lysogeny Broth (LB; Thermo Scientific) at 37°C, and liquid suspensions agitated at 180 rotations per minute (rpm). Stocks were maintained at –80°C suspended in 25% LB-glycerol volume/ volume (v/v). Unless stated otherwise, antibiotic concentrations were as follows: Ampicillin at 50 µg per millilitre (mL; 50 mg mL^{-1} stock). All work was performed under sterile conditions with Bunsen and ethanol-sprayed surfaces. Growth was monitored by optical density (OD) at an absorbance wavelength (λ) of 600 nanometres (nm) using a spectrophotometer.

2.2.2. Creation of Chemically Competent *E. coli* Cells

10 mL DH5α cells were grown overnight in 50 mL Falcon tubes and 1 mL added to 100 mL LB-ampicillin; cells were cultivated for 1.5–2 h, or until OD_{600} reached 0.5 absorbance units (AU), and incubated on ice for 10–20 min, swirling occasionally. Cells were pelleted in 2x 50 mL Falcon tubes at 4,500 rpm, 10 min, 4°C; the supernatant was discarded and each pellet resuspended by swirling in 20 mL sterile 100 (millimoles) mM magnesium chloride ($MgCl_2$) solution. The cell suspension was centrifuged again at 4,500 rpm for 10 min, 4°C, and the supernatant discarded. Pellets were gently resuspended in 3 mL sterilised 100 mM calcium chloride ($CaCl_2$) solution and recombined into one Falcon tube to a total volume of 6 mL, which was placed on ice for 1.5 h. 1.8 mL sterile 50% glycerol (v/v) was subsequently added to cell solution and mixed to homogeneity. 100–200 microlitres (µL)

cells were aliquoted to cold sterile 1.5 mL Eppendorf tubes, snap-frozen in liquid nitrogen, and stored at −80°C until required.

2.2.3. Chemical Transformation of *E. coli*

0.1−500 nanograms (ng) DNA was added to 200 µL aliquots of competent DH5α cells and incubated on ice for 30 min, after which cells and DNA were heat shocked at 42°C, 1 min. Following incubation on ice for 2−5 min, 1 mL LB broth was added, and suspensions incubated at 37°C for 1−1.5 h to induce selective antibiotic production. Cells were then pelleted on a table-top microfuge at 4,000 rpm, 4 min, and resuspended in ~150 µL remaining LB to be spread on selective LB antibiotic agar plates, and incubated for < 24 h, 37°C.

2.3. *Chlamydomonas reinhardtii*

2.3.1. Strains and Growth Conditions

All *C. reinhardtii* culturing and media preparation was performed under an ethanol-sprayed laminar flow hood to maintain an aseptic environment. Two commonly used *C. reinhardtii* strains were selected for experimentation: the cell wall-deficient strain CC-4533 (Zhang *et al*., 2014; purchased from the Chlamydomonas Resource Centre https://www.chlamycollection.org/) and the cell wall-intact strain CC-125 (kindly provided by Dr. A. Sproles, University of California San Diego [UCSD]).

C. reinhardtii strain CC-4533 (hereon referred to as CC-4533) was chosen due to its lack of cell wall, flexible growth requirements, high transformation efficiency, comparatively normal swimming and lipid accumulation, its ability to mate and to recover from cryogenic storage (Zhang *et al*., 2014). Recent uses of this strain include creation of a mapped insertional mutant library using high-throughput procedures (Li *et al*., 2015), studies into pyrenoid and rubisco biology involving fluorescent protein expression to detectable levels (MacKinder *et al*., 2016; MacKinder *et al*., 2017), small RNA profiling (Cavaiuolo *et al*., 2017) and development of heterologous gene expression tools (Onishi and Pringle, 2016).

The cell wall-intact *C. reinhardtii* strain CC-125 was selected as it is more robust than CC-4533 due to its cell wall, and its use was recommended for electroporation (**Section 2.3.3.**) through personal communication with Dr A. Berndt, UCSD.

Standard *C. reinhardtii* growth conditions comprised culturing in tris-acetate-phosphate (TAP) media (for recipe, see **Appendix A**) at a temperature of 25°C, with continuous light from above at light intensity (LI) of 150 µmol m^{-2} s^{-1} (unless stated otherwise). Liquid cultures were agitated at 120

rpm to prevent shading of the interior of the flask; fixed cultures were grown on TAP 1.5% bactoagar (oxoid LP0011, Agar-bacterio). Stock 25 mL cultures were maintained in 50 mL Erlenmeyer flasks. For experiments on multiwell microtitre plates, the following volumes of liquid media were used per well: 96-well plates, 200 µL; 24-well plates, 0.5–1 mL; 6-well plates, 3 mL. Growth was monitored by measuring cell number via haemocytometer (Neubauer cell-counting chamber), OD at 750 nm (OD_{750}) using a spectrophotometer, or chlorophyll fluorescence by taking readings at excitation (Ex) 440 nm and emission (Em) at 680 nm using a Tecan Spark fluorescent plate reader. Chlorophyll fluorescence gains were manually set to 50. Fluorescence gains for mCherry and mVenus detection were set to 150. Unless otherwise stated, paromomycin was added to solid media at 20 µg mL^{-1} concentration during plasmid selection.

Specific growth rate (SGR) was calculated using the equation SGR = $(\ln x_2 - \ln x_1)/ (t_2 - t_1)$, in which x_1 and x_2 represent the number of cells at times t_1 and t_2, respectively. Doubling time (t_d) was calculated using the following equation: $t_d = ((t_2 - t_1) * \log 2)/ (\log x_2 - \log x_1)$, where t_1, t_2, x_1 and x_2 are equivalent to those described for the specific growth rate equation. Both t_d and SGR were calculated using cell numbers obtained during exponential growth.

2.3.2. Transformation of *C. reinhardtii* strain CC-4533 by electroporation

Fresh 100 mL cultures of CC-4355 cells were inoculated with 1 mL stock cell culture and grown for ~2 days to exponential phase, or 1 - 5 x 10^6 cells per mL. In 50 mL Falcon tubes, cells were harvested by centrifugation at 1,500 x gravity (x*g*), 18°C, 10 min, and resuspended in 1 mL TAP 60 mM sucrose (Sigma, filter sterilised). 250 µL concentrated cells were aliquoted into cold 0.4 centimetre (cm) gapped electroporation cuvettes (BioRad), and ~1 µg linearised plasmid DNA suspended in 10–20 µL nfH_2O was added to the cell suspensions; plasmid DNA is more readily integrated into the *C. reinhardtii* nuclear genome when linearised (Kindle, 1990). Following incubation on ice for ~20 min, cuvettes were subjected to electroporation with an exponential current at 800 volts (V), 25 microfarads (µF), no shunt resistance (BioRad GenePulser Xcell electroporator).

Transformed cells were quickly transferred to 10 mL TAP/ 60 mM sucrose solution in 15 mL Falcon tubes and incubated in the dark for 16–24 h at 25°C with 110 rpm agitation. Cells were harvested by centrifugation at 1500 xg, 18°C, for 10 min in 50 mL Falcon tubes and resuspended pellets in 500 µL fresh TAP. 200 µL recovered cells were gently spread on to selective TAP-agar/ 10 µg mL^{-1} paromomycin plates and dried under sterile conditions for 15–20 min, then transferred to a 25°C incubator under continuous irradiance at desired light intensity. Colonies appeared within 4–10

days. Transformants were sustained on agar or 96-well plates with appropriate antibiotic. Transformation efficiency was calculated by counting colonies using openCFU software. Method adapted from a protocol graciously provided by Dr M Scaife (Mara Renewables Corporation, Canada).

2.3.3. Transformation of *C. reinhardtii* strain CC-125 by electroporation

CC-125 cultures were grown to 0.5–1 x 10^6 cells mL^{-1} (early log phase) to a volume that provided ~1 x 10^7 cells per transformation, and harvested by centrifugation at 2,000 xg, 15°C, 5 min. Pellets were resuspended in 2–10 mL GeneArt Max Efficiency Transformation Reagent (Thermo Scientific) and transferred to a single 50 mL sterile Falcon tube, after which cells were pelleted again at 2,000 xg, 15°C, 5 min and the supernatant discarded. The pellet was resuspended in 10 mL of GeneArt Max Efficiency Transformation Reagent a second time and pelleted using the same centrifugation settings, discarding the supernatant afterwards. Cells were resuspended with GeneArt Max Efficiency Transformation Reagent to a final concentration of 1–2 x 10^8 cells mL^{-1}.

2–4 µg of linearised purified vector was then mixed by pipetting with 250 µL cells in sterile 1.5 mL Eppendorf tubes, and incubated on ice for 5–10 min. Cells were then transferred to prechilled 0.4 cm gapped electroporation cuvettes and electroporated using a BioRad GenePulser Xcell electroporator using the following parameters: 500 V, 50 µF, 800 Ohm shunt resistance, exponential decay. Immediately following transformation, cuvettes were transferred to a shallow 20–25°C water bath and allowed to recover in dim light for 15–20 min. Cells were then transferred to 10 mL filter sterilised TAP/ 60 mM sucrose in 50 mL sterile Falcon tubes and shaken overnight (14–18 h) in the dark, 25°C.

The next day, cells were harvested by centrifugation at 2,000 xg, 18°C, 5 min, and resuspended in 0.5–1 mL TAP. Cells were streaked out on to selective TAP-agar plates using disposable sterile loops and dried for 10 min under a sterile laminar flow hood. Plates were incubated under standard conditions for 5–10 days, depending on the downstream experiment. Method adapted from a protocol kindly provided by Dr A Berndt (UCSD).

2.3.4 Chemical mutagenesis of *C. reinhardtii* strains CC-4533 and CC-125

C. reinhardtii cultures were grown to early exponential phase (1–3 x 10^6 cells mL^{-1}) and harvested by centrifugation at 2,000 xg, 18°C, 5 min. Cultures were concentrated x10 in 0.1 molar (M) phosphate buffer (pH 6.8) inside sterile Falcon tubes, to which ethyl methansulphonate (EMS) was

added to a concentration of 0.27−0.3 M. Cells were incubated for 2 h, then the EMS mutagenesis stopped by adding 10 mL filter sterilised 5% sodium thiosulphate (w/v) followed by vortexing and centrifugation (same settings). Supernatant was discarded into a beaker containing sodium thiosulphate crystals, and the pellet was washed again with 10 mL 5% sodium thiosulphate weight/ volume (w/v). The pellet was washed once with 0.1 M phosphate buffer (pH 6.8) and with TAP buffer, each time with the same centrifugation settings and disposing of the supernatant in the sodium thiosulphate crystals, and lastly resuspended in TAP buffer to the desired cellular concentration. ~1 x 10^6 cells were spread on to TAP-agar with increasing concentrations of norflurazon and grown under the desired selection conditions.

2.4. Nucleic Acid Manipulation

2.4.1. Primers

Primers were designed using SnapGene software, and produced by Thermo Scientific. nfH_2O was added to primer DNA to achieve a 100 micromolar (µM) concentration; a 10-fold dilution was stored as a working stock for use in polymerase chain reactions (PCRs). Primers were stored at −20°C.

Table 2.1: Primers used in Chapter 3

Primer	Sequence 5' → 3'	Properties
crOR_F	TCATCATATGTCGCCGCTCCCCG	NdeI; T_m, 63°C
crOR_R	TCATGATATCTCAGAAGGGGTCAATGCGGGG	EcoRV; T_m, 63°C
crOR_N-His_F	TCACATATGCATCACCATCACCATCACTCGCCGCTCCCCGCGT GCAA	NdeI; T_m, 64°C
His_Tag_F	ATGCATCACCATCACCATCAC	T_m, 58°C
pOpt_insert_seq_R	GCGGGTGGCTCCAGAATTC	T_m, 60°C; Sequencing primer

Restriction enzyme cut sites underlined; T_m = Annealing temperature. T_m calculated in SnapGene. Histidine tag highlighted in grey.

Table 2.2: Primers used in Chapter 5

Primer	Sequence 5' → 3'	Properties

mCherry_NdeI_F	TCACATATGATGGTGAGCAAGGGCGAG	NdeI; T_m, 60°C
mCherry_EcoRI_R	TACGAATTCTTATTACTTGTACAGCTCGTCCATGCC	EcoRI; T_m, 60°C
iRbcS2_Amp_F	ACGTATCGATGTGAGTCGACGAGCAAGCC	ClaI; T_m, 60°C
pCore_Amp_F	ACGTTCTAGAAAGCCGAGCGAGCCC	XbaI; T_m, 60°C
pCore_Amp_R	ACGTATCGATGTGGGCACACGCTAAAAGAAAGA	ClaI; T_m, 59°C
pCRE_Amp_F	GGTTGCTGGGTGAGCTC	SacI; T_m, 59°C
pCRE_Amp_R	GCCGTGCTCTGCTCTAGA	XbaI; T_m, 58°C
mVenus_EcoRI_R	TGGCTCCAGAATTCGATATCCCTG	EcoRI; T_m, 59°C
pOpt_insert_seq_R	GCGGGTGGCTCCAGAATTC	T_m, 60°C; Sequencing primer
T7_F	TAATACGACTCACTATAGGG	T_m, 50°C; Sequencing primer
Mid_mt_F	ATGGCCTGTTTCTTAGCCAGG	T_m, 58°C; Screening primer
Mid_mt_R	CTACATGTGTTTCTTGACGCTGG	T_m, 57°C; Screening primer
mVen_Screen_F	ATCGAGGGCAGGGTGAG	T_m, 58°C; Screening primer
mVen_Screen_R	CCTGCCCTCGATCTTGTACAG	T_m, 59°C; Screening primer

Restriction enzyme cut sites underlined; T_m = Annealing temperature. T_m calculated in SnapGene.

2.4.2. DNA Fragments for vector constructs

Single Stranded DNA (ssDNA) for vector construction in **Chapter 5** was synthesised as primer DNA (ThermoFisher) and PCR amplified for insertion into appropriate vectors.

Table 2.3: ssDNA templates for construction of pCRE vectors in Chapter 5

Template	Sequence 5' → 3'	Properties

pCore	ACGTTCTAGAAAGCCGAGCGAGCCCGCTGCAGGTT AGTCTTTCTTTTAGCGTGTGCCCACATCGATACGT	Core_ssDNA amplified with pCore_Amp_F and pCore_Amp_R XbaI, ClaI 70 bp
pCRE-1	GGTTGCTGGGTGAGCTC**TCTCTCTCTT**TCTCTCTCTC TCTTTCTCTCTCTCTTTCTCTCTCTCTTTCTCTCTAGAGCAGAGCACGGC	91 bp SacI XbaI 66 bp digested
pCRE-2	GGTTGCTGGGTGAGCTC**GCCCCATGAGG**GCCCCAT GAGGGCCCCATGAGGGCCCCATGAGGGCCCCATGA GGTCTAGAGCAGAGCACGGC	90 bp SacI, XbaI 65 bp digested
pCRE-3	GGTTGCTGGGTGAGCTC**TTGGTCGCGATGG**TTGGT CGCGATGGTTGGTCGCGATGGTTGGTCGCGATGGT TGGTCGCGATGGTCTAGAGCAGAGCACGGC	100bp SacI, XbaI 75 bp digested
pCRE-4	GGTTGCTGGGTGAGCTC**GGGGTACTC**GGGGTACTC GGGGTACTCGGGGTACTCGGGGTACTCGGGGTACT CGGGGTACTCTCTAGAGCAGAGCACGGC	98 bp SacI, XbaI 73 bp digested
pCRE-5	GGTTGCTGGGTGAGCTC**TATGTAGG**TATGTAGGTA TGTAGGTATGTAGGTATGTAGGTATGTAGGTATGTA GGTCTAGAGCAGAGCACGGC	91bp SacI, XbaI 66 bp digested
pCRE-6	GGTTGCTGGGTGAGCTC**GCATGCATGCTG**GCATGC ATGCTGGCATGCATGCTGGCATGCATGCTGGCATGC ATGCTGTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested
pCRE-7	GGTTGCTGGGTGAGCTC**CATGGACCAGGA**CATGGA CCAGGACATGGACCAGGACATGGACCAGGACATGG ACCAGGATCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested
pCRE-8	GGTTGCTGGGTGAGCTC**CTCGGT**CTCGGTCTCGGTC TCGGTCTCGGTCTCGGTCTCGGTTCTAGAGCAGAGC ACGGC	77 bp SacI, XbaI 52 bp digested

pCRE-9	GGTTGCTGGGTGAGCTC**CGAGCGTTTTCT**CGAGCGTTTTCTCGAGCGTTTTCTCGAGCGTTTTCTTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested
pCRE-10	GGTTGCTGGGTGAGCTC**GTCCACCTGG**GTCCACCTGGGTCCACCTGGGTCCACCTGGGTCCACCTGGTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70bp digested
pCRE-11	GGTTGCTGGGTGAGCTC**CCCATGCGA**CCCATGCGACCCATGCGACCCATGCGACCCATGCGACCCATGCGATCTAGAGCAGAGCACGGC	98 bp SacI, XbaI 73 bp digested
pCRE-12	GGTTGCTGGGTGAGCTC**GGGCCCATTC**GGGCCCATTCGGGCCCATTCGGGCCCATTCGGGCCCATTCGGGCCCATTCTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested
pCRE-13	GGTTGCTGGGTGAGCTC**CGCATGGGGC**CGCATGGGGCCGCATGGGGCCGCATGGGGCCGCATGGGGCCGCATGGGGCTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested
pCRE-14	GGTTGCTGGGTGAGCTC**GGGCCAC**GGGCCACGGGCCACGGGCCACGGGCCACGGGCCACGGGCCACTCTAGAGCAGAGCACGGC	84 bp SacI, XbaI 59 bp digested
pCRE-15	GGTTGCTGGGTGAGCTC**TTTTTC**TTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTTTTTCTCTAGAGCAGAGCACGGC	77 bp SacI, XbaI 52 bp digested
pCRE-RM	GGTTGCTGGGTGAGCTC**CGAACCGGGC**CGAACCGGGCCGAACCGGGCCGAACCGGGCCGAACCGGGCTCTAGAGCAGAGCACGGC	95 bp SacI, XbaI 70 bp digested

Flanking DNA regions for restriction enzyme adherence/ amplification with primers pCRE_Amp_F and pCRE_Amp_R highlighted in grey. Restriction sites underlined. First repeat of pCRE motifs highlighted in bold font. Fragments lengths are noted, followed by lengths of the fragments following digestion.

2.4.2. Plasmids

Plasmids used throughout this project are listed in **Table 2.4** and **Table 2.5**. Plasmid maps throughout this work were created using SnapGene software.

Table 2.4: Plasmids used in Chapter 3

Plasmid	Properties	Source/ Reference
pOpt_mVenus_Paro	Hsp70A/rbcS2 promoter/ enhancer element; Ampicillin^R; Paromomycin^R;	Purchased from chlamycollection.org; Lauersen *et al*. (2015); **Figure 2.1.**
pOpt_crOR	pOpt_mVenus_Paro backbone containing crOR gene in place of mVenus	This work; **Chapter 3**; **Figure 3.2** for vector map

Table 2.5: Plasmids used in Chapter 5

Plasmid	Properties	Source/ Reference
pUC_mCherry	pUC19 vector containing mCherry gene	Courtesy of Dr S. Jaffe
pOpt_mCherry	pOpt vector carrying mCherry fluorescent protein under control of *Hsp70A-RbcS2* promoter/ enhancer	This work
pOpt_Core_mCherry	pOpt vector with mCherry reporter gene with Hsp70A/RbcS2 promoter replaced with core promoter and added restriction sites	This work
pOpt_Core_mVenus	See pOpt_Core_mCherry but with mVenus reporter	This work
pCRE-1_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-1 proximal promoter	This work
pCRE-2_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-2 proximal promoter	This work
pCRE-3_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-3 proximal promoter	This work

pCRE-4_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-4 proximal promoter	This work
pCRE-5_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-5 proximal promoter	This work
pCRE-6_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-6 proximal promoter	This work
pCRE-7_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-7 proximal promoter	This work
pCRE-8_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-8 proximal promoter	This work; linearised with BsaI
pCRE-9_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-9 proximal promoter	This work
pCRE-10_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-10 proximal promoter	This work
pCRE-11_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-11 proximal promoter	This work
pCRE-12_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-12 proximal promoter	This work
pCRE-13_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-13 proximal promoter	This work
pCRE-RM_mVenus	pOpt vector backbone with mVenus reporter driven by Core promoter and pCRE-RM proximal promoter	This work

The plasmid pOpt_mVenus_Paro (purchased from www.chlamycollection.org; Lauersen *et al*., 2015; **Figure 2.1**) was the main vector backbone for genetic engineering throughout this project. pOpt_mVenus_Paro contains an aminoglycoside 3'-phosphotransferase gene (*AphVIII*) from *Streptomyces rimosus*, which provides resistance to the antibiotic paromomycin when expressed in *C. reinhardtii* (Sizova *et al*., 2001). Upstream of *AphVIII* is the promoter region and first RbcS2 intron, and the promoter region of the constitutive expression promoter Hsp70A; these elements have been shown to increase transgene expression and reduce gene silencing following integration of foreign DNA in the nuclear genome (**Figure 2.1**; Sizova *et al*., 2001). The 3'-UTR of *RbcS2* is also included after *AphVIII* to further enhance gene expression (**Figure 2.1**; Sizova *et al*., 2001). The same regulatory elements (*Hsp70A* and *RbcS2*) encompass the expression cassette for the GOI, here mVenus (**Figure 2.1**). Additionally, an ampicillin resistance cassette and f1 origin of replication are incorporated within pOpt_mVenus_Paro for plasmid propagation and selection in *E. coli*.

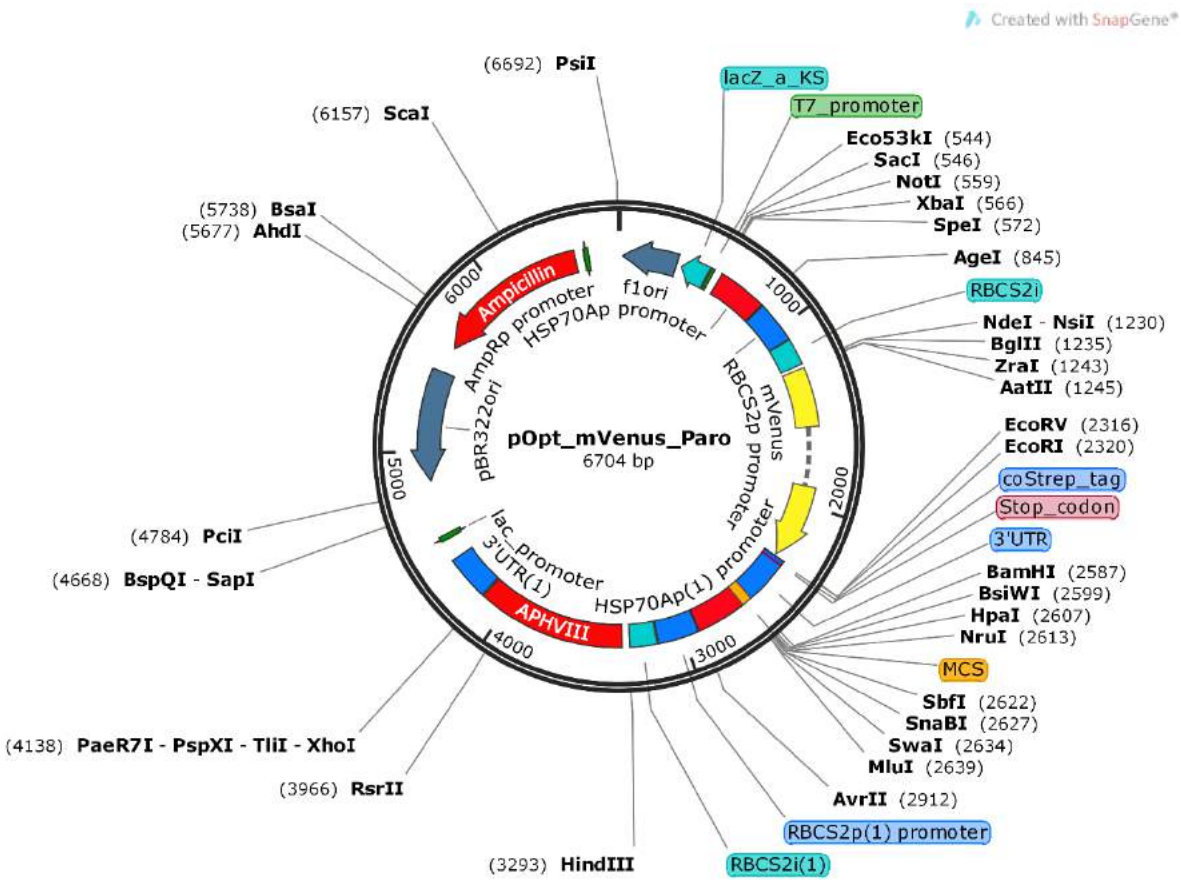


Figure 2.1: Plasmid map of pOpt_mVenus_Paro. Map highlights relevant features, such as *Hsp70A* and *RbcS2* promoters, *AphVIII* paromomycin resistance cassette, *RbcS2* 3'-UTR, ampicillin resistance gene, and unique restriction enzyme digestion sites. Plasmid created by Lauersen *et al*., (2015), as part of the publicly available pOptimised vector suite. Map created using SnapGene software.

2.4.3. Polymerase Chain Reaction (PCR)

DNA amplification was achieved via PCR, for which the enzymes Phusion High-Fidelity DNA polymerase (New England Biolabs; NEB) or Phire Hot Start II DNA Polymerase (Thermo Scientific) were used.

Phusion PCR mixtures are shown in **Table 2.6**. 50 µL reaction volumes were used for amplifying DNA fragments intended for vector insertion, sequencing or other downstream manipulations, whereas 25 µL reaction volumes were used for non-endpoint or diagnostic PCR reactions. Thermocycling conditions are shown in **Table 2.7**.

Table 2.6: Phusion High-Fidelity DNA Polymerase reaction mixture

Component	25 µL Reaction	50 µL Reaction	Final Concentration
5x Phusion HF or GC Buffer	5 µL	10 µL	1x
10 mM dNTPs	0.5 µL	1 µL	200 µM
10 µM Forward Primer	1.25 µL	2.5 µL	0.5 µM
10 µM Reverse Primer	1.25 µL	2.5 µL	0.5 µM
Phusion DNA Polymerase	0.25 µL	0.5 µL	1.0 units/ 50 µL PCR
Template DNA	variable	variable	1 pg – 250 ng
nfH_2O	to 25 µL	to 50 µL	

Table 2.7: Phusion high-fidelity DNA polymerase typical thermocycling conditions

Step	Temperature (°C)	Time
Initial Denaturation	98	30 s
35 Cycles	98	10 s
	45–72 (variable)	30 s
	72	15–30 s kbp^{-1}

Final Extension	72	5–10 min
Hold	4	∞

Due to its high resistance to PCR inhibitors in plants, ability to amplify GC-rich regions, and its accompanying dilution solution which enhances DNA release from polysaccharide-rich samples, Phire Hot Start II DNA Polymerase (Thermo Scientific) was chosen for PCR amplification of DNA direct from *C. reinhardtii* cell colonies/ pellets. Genomic template DNA (gDNA) was prepared by suspending a fresh single colony of *C. reinhardtii* in 20 µL dilution solution (Thermo Scientific). **Table 2.8** and **Table 2.9** show the Phire reaction mixture and thermocycling conditions, respectively.

Table 2.8: Phire Hot Start II DNA Polymerase reaction mixture

Component	**20 µL Reaction**	**50 µL Reaction**	**Final Concentration**
2x Phire plant PCR buffer	10 µL	25 µL	1x
10 µM Forward Primer	1 µL	2.5 µL	0.5 µM
10 µM Reverse Primer	1 µL	2.5 µL	0.5 µM
Phire Hot Start II DNA Polymerase	0.4 µL	1 µL	1.0 units/ 50 µL PCR
Template DNA	0.5 µL	1.25 µL	
nfH_2O	to 20 µL	to 50 µL	

Table 2.9: Phire Hot Start II DNA Polymerase thermocycling conditions

Step	**Temperature (°C)**	**Time**
Initial Denaturation	98	5 min
35 Cycles	98	5 s
	45–72 (variable)	5 s
	72	20 s kbp^{-1}
Final Extension	72	5–10 min

Hold	4	∞

2.4.4. Electrophoresis of DNA on Agarose Gel

To analyse DNA fragments with the size range 200–10,000 bp, DNA agarose gels were made by melting 1.0% (w/v) agarose powder, 1x tris-acetate-ethylenediaminetetraacetic acid (EDTA; TAE) buffer and 0.01% ethidium bromide, and poured into a mould with a well comb until set. Gels were then placed into an electrode bath filled with 1% TAE buffer. 5 µL 1 kbp HyperLadder™ (Bioline; **Figure 2.2A**) DNA marker was loaded to the gel for DNA concentration and fragment length estimation, and 6x purple loading dye (NEB) was added to the DNA samples prior to application. Gels underwent electrophoresis for 60 min, 120 miliamps, then DNA bands photographed with transilluminator and attached camera.

Smaller (< 200 bp) DNA fragments were separated using 2–3% (w/v) agarose/ TAE buffer gels, with 5 µl 25 bp HyperLadder™ (Bioline; **Figure 2.2B**) as a DNA marker; electrophoresis conditions were constant 75 V for 210 min.

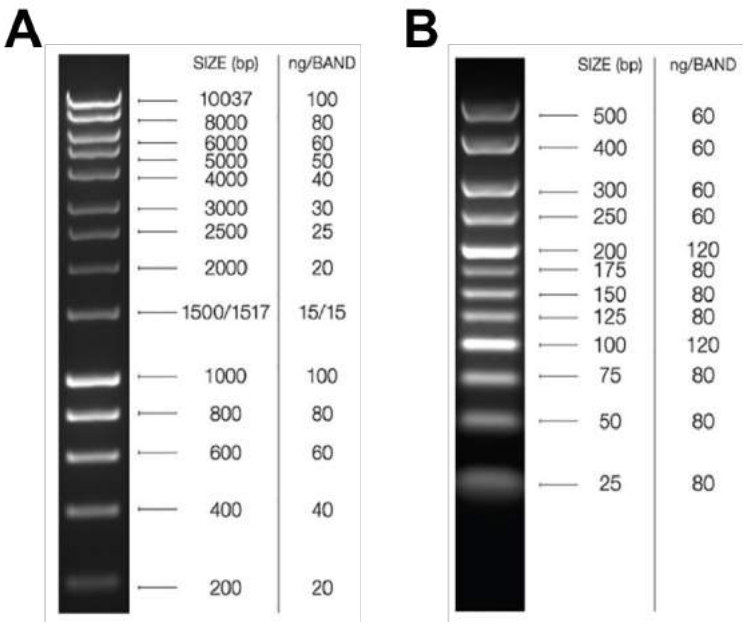


Figure 2.2: Bioline DNA ladders used for DNA agarose gel electrophoresis. A – 1 kbp HyperLadder. **B** – 25 bp HyperLadder. Images taken from Bioline website (https://www.bioline.com/us/ on 13/06/2020).

2.4.5. DNA extraction from agarose gels

DNA fragments of desired size were excised from the agarose gel with a sterile razor, and the DNA extracted using QIAquick Gel Extraction Kit (Qiagen). DNA was eluted using nfH_2O heated to 55 °C, and stored at −20°C.

2.4.6. DNA digestion by restriction endonucleases

All restriction endonuclease enzymes and buffers were purchased from NEB and stored at −20°C, keeping on ice during use. Suitable double digests and buffers were found using the following tool: https://www.neb.com/tools-and-resources/interactive-tools/double-digest-finder.

For diagnostic digestions, 1 µL miniprep DNA, 0.2 µL each restriction enzyme and 1 µL 10x CutSmart buffer were made up to a 10 µL solution with nfH_2O. Digestions of DNA intended for vector construction were performed at a total volume of 50 µL, containing vector or insert DNA, 1 µL each restriction enzyme per µg DNA, 5 µL 10x CutSmart buffer (unless otherwise stated) and the appropriate amount of nfH_2O. All mixtures were vortexed briefly, then incubated for > 2 h at 37°C.

Cut insert and single-digested vector were prepared for downstream manipulation by heat inactivation of restriction enzymes at 65°C for 20–30 min where appropriate. Double-digested vector was purified by gel electrophoresis and gel extracted.

2.4.7. Ligation of insert and plasmid DNA

Concentrations of DNA required for desired vector-insert ratios were calculated using http://www.insilico.uni-duesseldorf.de/Lig_Input.html tool; ratios of 1:3, 1:1 and 3:1 vector:insert were used unless otherwise stated. Restriction digested vector DNA was first treated with 1 µL alkaline phosphatase (NEB) to prevent self-ligation of the plasmid. Ligation reaction mixtures were then made up with nfH_2O to total volume 20 µL, containing 1 µL T4 DNA ligase (NEB), 2 µL 10x T4 DNA ligase buffer (NEB), and amounts of vector and insert calculated using ligation calculator (see above). Mixtures were vortexed briefly and incubated at room temperature for > 1 h (> 2 h if ligating blunt ends) before transforming into competent *E. coli* DH5α cells (**Section 2.2.3**.).

2.4.8. Small-scale preparation of plasmid DNA (Miniprep)

DH5α colonies harbouring desired plasmid were cultured to 10 mL overnight with appropriate antibiotic and pelleted at 4,000 rpm, 4°C, for 10 min. Pellets were resuspended in ~1 mL remaining supernatant, transferred to 1.5 mL Eppendorf tubes, and centrifuged at 13,000 rpm for 1 min.

Plasmid DNA was then extracted using QIAprep Spin Miniprep kit (Qiagen) and eluted with 50 µL of either EB buffer (Qiagen) or nfH_2O. DNA preparations were stored at −20°C.

2.4.9. Large-scale preparation of plasmid DNA (Maxiprep)

100 mL DH5α cells harbouring desired plasmid were grown overnight, then DNA extracted with Quick-StartProtocol Qiagen Maxiprep kit (Qiagen), following manufacturer's instructions and taking care to not disturb the DNA pellet at isopropanol and ethanol stages. DNA was eluted in 400 µL EB buffer (Qiagen) or nfH_2O, and stored at −20°C.

2.4.10. DNA sequencing

DNA samples were sequenced at the Core Genomics Facility, University of Sheffield. Plasmid DNA was made up to 100 ng μL^{-1} and PCR amplified fragments to 50 ng μL^{-1}, both to a total volume of 10 µL with nfH_2O. Results were retrieved by email, and sequences analysed using ebi global nucleotide alignment tool http://www.ebi.ac.uk/Tools/psa/ and SnapGene™ software.

2.4.11 DNA quantification

Rough estimates of DNA concentration were taken by comparing DNA gel electrophoresis bands with the DNA ladder (**Figure 2.2**). More accurate DNA measurements were performed using a NanoDrop™ 2000 spectrophotometer (Thermo); 1−2 µL DNA sample was measured per run.

2.5. Protein analysis

2.5.1. Protein preparations for SDS-PAGE and western blot (Chapter 3)

OD_{750} growth readings of *C. reinhardtii* cell cultures were taken, then 20 mL cell culture harvested at 4,000 xg, 4°C, 10 min. Lysis buffer A (0.8 M Tris-HCl pH 8.3, 0.2 M sorbitol, 1% β-mercaptoethanol [v/v], 1% sodium dodecyl sulphate [SDS; w/v]) was then added to resuspend the pellet; the OD_{750} of the sample was used to determine how many mL of solution to add, in order to standardise the samples. For instance, a pellet from a culture with an OD_{750} of 3.8 would be resuspended in 3.8 mL of lysis solution. Samples could then be aliquoted into LoBind 1.5 mL Eppendorf tubes and stored at −80°C for later use, or used immediately for SDS-polyacrylamide gel electrophoresis (SDS-PAGE) gel analysis.

2.5.2. Protein SDS-PAGE

Protein samples were heated to 98°C, 2 min, then cooled on ice, after which they were centrifuged 13,000 rpm, 2 min, and the supernatant transferred to fresh 1.5 mL LoBind Eppendorf tubes.

Samples were then loaded on to 12% bis-tris Novex NuPAGE precast mini-gels (Thermo) using 3-(N-morpholino)propanesulfonic acid (MOPS) running buffer (Thermo) at constant 200 V, 50 min. EZ-run prestained Rec protein ladder (Fisher; **Figure 2.3**) was used as a marker. Gels were washed with dH_2O and the bands visualised using InstantBlue protein stain (Expedeon) or used directly for Western blot.

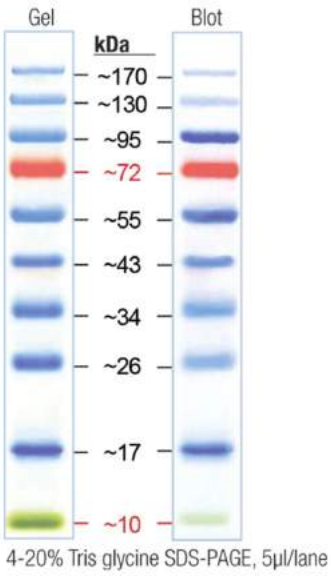


Figure 2.3: Protein ladder used as a marker for SDS-PAGE experiments. EZ-run prestained Rec protein ladder (Fisher). Image taken from Fisher website (https://www.fishersci.com/ on 13/06/2020).

2.5.3. Western blot

SDS-PAGE gels were stacked within iBlot transfer stacks according to manufacturer's instructions (Thermo), and iBlot dry blotting equipment (Thermo) employed to transfer the proteins to the Polyvinylidene fluoride (PVDF) membrane. The blotted PVDF membrane was incubated with ~20 mL Blocking Buffer (Phosphate-buffered saline solution, 0.05% Tween-20 [v/v], 5 % non-fat milk powder [w/v]) for 1 h with gentle rocking, and washed three times for 5 min with ~20 mL Wash Buffer (Phosphate-buffered saline solution, 0.05% Tween-20 [v/v]). 20 ml Blocking Buffer containing a 1:4000 dilution of horseradish peroxidase (HRP)-conjugated anti-6X Histidine tag antibody (Abcam, ab1187) was then added to the membrane for 1–2 h at room temperature, or overnight at 4°C. The membrane was then washed five times for 5 min with 20 mL Wash Buffer. ~10 mL 1-step TMB-blotting solution (Thermo) was then added until the bands developed sufficiently. Staining reaction was halted using dH_2O, and images taken using a digital camera. Densitometry plots of bands produced by Western blot were performed using ImageJ.

2.5.4. Protein extraction for label-free quantitative proteomics (Chapter 4)

20 mL cell culture was harvested by centrifugation (2000 xg, 18 °C, 5 min) and frozen at −20°C after removal of the supernatant. Samples were thawed and resuspended in 1 mL Lysis Buffer B (2% SDS [w/v], 40 mM Tris base, 60 mM Dithiothreitol [DTT]) with 10 µL Halt™ protease inhibitor cocktail (Thermo), frozen again at −80°C for ~24 h, then thawed quickly in a 37°C water bath. 500 µL of each sample was added to pre-chilled 2 mL Eppendorf tubes in triplicate, and ~500 µL glass beads were added to the samples. The samples were vortexed for 30 s then cooled on ice for 30 s for 10 cycles. Lysed samples were centrifuged at 18,000 xg, 4°C, 5 min, and left on ice for 20 min until the foam subsided. The green supernatant fraction was carefully transferred to 1.5 mL protein LoBind Eppendorf tubes and stored at −20°C. 100 µL of each sample was purified from lipid, pigment and other contaminants using a protein 2-D Clean-Up Kit (GE Healthcare) following manufacturer's instructions.

2.5.5. Protein quantification

Cleaned-up protein extract was quantified using a NanoDrop™ 2000 spectrophotometer (Thermo); 1−2 µL protein sample was measured per run.

2.5.6. Protein reduction, alkylation and digestion

The pellet from the 2-D protein clean-up was resuspended in 50 µL Urea Buffer (8 M urea; 100 mM Tris-HCl [pH 8.5]; 5 mM DTT) and placed in a sonication bath for 5 min, or until protein suspension became clear. Protein concentration was quantified, and ~50 µg protein was transferred to a fresh 1.5 mL protein LoBind Eppendorf tube. Protein samples were reduced by diluting up to 10 µL with Urea Buffer and incubating at 37°C for 30 min. Proteins were S-alkylated by adding 1 µL 100 mM iodoacetamide and incubating in the dark at room temperature for 30 min. 2 µg trypsin endoproteinase LysC enzyme mix (Promega) was added to the protein solution and incubated at 37°C for 3 h for LysC digestion, after which the solution was diluted with 75 µL 50 mM Tris-HCl (pH 8.5)/ 10 mM $CaCl_2$ and incubated overnight for trypsin digestion. The digestion was stopped by acidification by adding 0.05 volumes of 10% trifluoroacetic acid (TFA) to the peptide solution. Pierce® C18 spin columns (Thermo Scientific) were used according to the manufacturer's instructions; a yield of ~30 µg peptides is regularly achieved using this method. Samples were dried using a vacuum evaporator and stored at −80°C until ready for mass spectrometry (MS) analysis. Method was adapted from Hitchcock *et al*. (2016).

2.5.7. LC-MS/ MS for proteomics

Peptide sample pellets were thawed and resuspended in 15 µL Loading Buffer (97% acetonitrile, 3% H_2O, 0.1 % TFA v/ v) and sonicated in a water bath for 5 min until fully in suspension. Following 5 min centrifugation, 2 µL sample (~4 µg) was diluted 1 in 8 with loading buffer and transferred to a vial for liquid chromatography (LC)-MS/ MS analysis. 500 ng protein sample was analysed by nanoflow LC (Dionex UltiMate 3000 RSLCnano system) coupled online to a Q Exactive HF mass spectrometer (Thermo Scientific).

2.5.8. Proteomics data analysis

Raw MS data files were processed using MaxQuant version 1.5.2.8 software, using the MaxLFQ option. Data were searched against the *C. reinhardtii* proteome (UniprotKB proteome ID UP000006906, last modified December 2019; 18,829 proteins). MaxLFQ parameters were set accordingly: fixed modifications: Carbamidomethyl; variable modifications: Acetyl (Protein N-term), Oxidation; decoy mode: revert; peptide spectrum matches, protein, and site false discovery rates (FDRs): 0.01; Special amino acids (aas): arginine and lysine; MS/ MS tolerance (Ion trap MS): 0.5 Daltons (Da); MS/ MS tolerance (fourier transform MS): 20 ppm; MS/ MS tolerance (time of flight): 0.5 Da; Minimum peptide length: 7; minimum score for modified peptides: 40; peptides used for protein quantification: razor; minimum peptides: 1; minimum razor peptides: 1; minimum unique peptides: 0; minimum ratio count: 2. Principal component analysis (PCA) was performed using Perseus software v1.6.14.0 by uploading the ProteinGroups and Evidence .txt files obtained from the MaxLFQ analysis.

Statistical analysis of protein quantification was performed using the ProteoSign online program (Efstathiou *et al*., 2017). ProteinGroups and Evidence .txt files generated through the MaxLFQ analysis were uploaded directly to ProteoSign, and default settings were applied. Statistical plots were automatically generated by ProteoSign (**Figure 4.12**). Venn diagrams were generated from ProteoSign data using Venny 2.1 (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Differentially regulated proteins were mapped on to the KEGG *C. reinhardtii* metabolic model using the search and colour function https://www.genome.jp/kegg/tool/map_pathway2.html?cre). Enriched gene ontology (GO) terms associated with differentially regulated proteins were identified using the Panther GO database at http://pantherdb.org/ (Mi *et al*., 2019).

2.6. Pigment analysis

2.6.1. Pigment extractions for carotenoid and chlorophyll analysis

For batch cultures grown in Erlenmeyer flasks, 5 mL fresh *C. reinhardtii* culture was harvested in triplicate by centrifugation at 2000 xg, 5 min, 4°C. Supernatant was discarded and the pellets frozen at −20°C for up to 2 weeks. All following work was completed in the dark/ dim light, using ice wherever possible. Pellets were resuspended in 1 mL cold 80 or 100% acetone (v/v; specified in text), transferred to cold 1.5 mL Eppendorf tubes containing ~200 mL glass beads (Sigma), and incubated on ice for 10−15 min. Samples were then subjected to cycles of vortexing for 2 min, then incubation on ice for 2 min, for a total of five times, following which they were centrifuged at 10,000 xg, 5 min, 4°C. Pellets were checked to ensure they were completely white/ grey; if not, the batch was subjected to another round of 5x vortexing/ incubation on ice, and centrifuged again. The supernatant was transferred to a fresh, ice-cold Eppendorf and frozen at −80°C for storage.

For cultures grown in 24-well plates, 500 µL cells were pelleted in 1.5 mL Eppendorf tubes on a table top microcentrifuge at 13,000 rpm, 5 min. After discarding the supernatant, pellets were resuspended in 500 µL cold 100% acetone and incubated on ice for 30 min. Vortex cycles were then applied, with 2 min vortexing and 2 min on ice, a total of five times. Samples were centrifuged at 4°C, 10,000 xg, 5 min. Pigment-containing supernatants were transferred to fresh 1.5 mL Eppendorf tubes and frozen at −80°C for storage.

2.6.2. Estimating pigment concentrations using spectrophotometry

Pigment extracts were diluted in cold acetone (80 or 100% [v/v]) to an appropriate concentration and transferred to UV-transparent plastic cuvettes or solvent-resistant 96-well plates for spectrophotometric analysis. Wavelength scans of extracts were performed using a BioMate™ 160 UV-Vis Spectrophotometer (Thermo). Absorbance readings taken for each experiment are shown in brackets (A_λ) in the following equations that were used to estimate pigment concentrations for each sample.

For pigments extracted in 80% acetone (v/v) in **Chapter 3** for strain CC-4533:

Chlorophyll-*a* = $12.25(A_{664}) - 2.79(A_{647})$

Chlorophyll-*b* = $21.50(A_{647}) - 5.10(A_{664})$

Total carotenoids = $(1000[A_{470}] - 1.82[\text{chlorophyll-}a] - 85.02[\text{chlorophyll-}b])/198$

For pigments extracted in 100% acetone (pure solvent) in **Chapter 4** for strain CC-125:

Chlorophyll-*a* = $11.24(A_{662}) - 2.04(A_{645})$

Chlorophyll-*b* = 20.13(A_{645}) – 4.19 (A_{662})

Total carotenoids = (1000[A_{470}] – 1.90[chlorophyll-*a*] – 63.14[chlorophyll-*b*])/214

Units for estimated pigments are µg mL^{-1}. Equations were based on those reported by Lichtenthaler, (1987). 100% acetone was used in **Chapter 4** as this method more effectively extracted the pigments from the more robust CC-125 strain.

To calculate chlorophyll-*a*/ -*b* ratios, chlorophyll weight estimates were used to calculate picomoles (pmol) of chlorophyll per 1 mL culture. Total carotenoid molar estimates were calculated in pmol using HPLC results to determine the molar ratios of carotenoids per strain.

2.6.3. HPLC analysis of pigments

HPLC analysis was carried out on a Dionex UltiMate 3000 HPLC machine using a Hyperselect C_{18} reverse phase column (125 angstrom (Å) pore size, 5 µm particle size, 250 x 4.6 millimetre [mm]). The method performed was based on that stated in León *et al*. (2005), where solvent A was ethyl acetate, and solvent B was acetonitrile:water 9:1 (v/v). The separation program was as follows: 0–16 min, gradient from 0–60% solvent A; 16–28 min, 60% solvent A. Injection volume was 10 µL, and flow rate set to 1.0 mL min^{-1}. Carotenoids were detected at 450 nm wavelength. Carotenoid standards were obtained from Sigma-Aldrich in powder form, suspended in acetone (80 or 100%), and stored at –80°C.

2.7. Microscopy

2.7.1. Light microscopy

1 mL of cell culture was pelleted by centrifugation at 13,000 rpm, 5 min. Media was poured away, and the pellet resuspended in remaining supernatant. 20 µL concentrated sample was spotted on to a glass slide, and covered with a glass coverslip.

2.7.2. Confocal microscopy

In **Chapter 3**, images for live fluorescence cell microscopy were captured using Leica DM6B-Z-CS confocal fluorescent microscope; confocal images in **Chapter 5** were taken using a Leica SP8 TCS confocal fluorescent microscope. Images were analysed in real time using LasX software. The autofluorescence of chlorophyll was exploited to image the chloroplast by laser excitation at 488 nm and emission detection at 650–700 nm. mVenus was detected by excitation at 488 nm and detection at 500–550 nm. Numerical aperture 1.4. Pinhole 103.1 µm.

2.8 Flow cytometry

2.8.1. Flow cytometry

Fluorescence of individual cells was measured by flow cytometry using a BD FACS Melody cell sorter. 50,000 events were measured per run. Chlorophyll fluorescence was measured Ex488, Em710/ 45. mVenus was measured at the wavelengths Ex488, Em513/ 26.

2.8.2. Fluorescence activated cell sorting (FACS)

FACS was performed using a BD FACS Aria II cell sorter at Scripps Institute of Oceanography, La Jolla, US. Forward scatter and side scatter were measured by laser excitation at 488 nm. Chlorophyll fluorescence was measured Ex488, Em710/ 45. mVenus was measured at the wavelengths Ex488 Em513/ 26. 10–100 events falling within mVenus gated area (**Figure 5.17**) were sorted into 6-well plates.

2.8. Statistical analyses

Data was acquired using at least three biological replicates unless otherwise stated, and results displayed as a mean value of these replicates ± standard deviation (SD). Student's *t*-test, ANOVA, Kruskal-Wallis and Mann-Whitney tests were calculated using Graphpad Prism software, version 8. Statistical significance was attained when $P < 0.05$. Asterisks were used to display $*P < 0.05$, $**P < 0.01$, $***P < 0.01$, $****P < 0.001$.

2.9. Bioinformatic analyses

Individual FASTA sequences for DNA, RNA and proteins were obtained from the from the *C. reinhardtii* v5.5 genome within the Phytozome plant genomics resource (https://phytozome.jgi.doe.gov/pz/portal.html).

2.9.1. Protein bioinformatic tools

Protein sequence similarities and domains were analysed using BLASTP (National Center for Biotechnology Information; NCBI) using default parameters (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Chloroplast transit peptides were predicted using the PredAlgo server (http://lobosphaera.ibpc.fr/cgi-bin/predalgodb2.perl?page=main; Tardif *et al*., 2012), and transmembrane helices predicted using TMPRED (https://embnet.vital-it.ch/software/TMPRED_form.html). Isoelectric point (pI) was calculated using the ExPasy compute pI tool http://web.expasy.org/cgi-bin/compute_pi/pi_tool. TCoffee was used for multiple sequence

alignment (Notredame *et al*., 2000; Di Tommaso *et al*., 2011), and alignment images were produced in BioEdit.

2.9.2. DNA bioinformatic tools

Gene schematic diagrams were created using Wormweb Exon-intron (http://wormweb.org/exonintron). Random DNA sequences with specified GC contents were created using Random DNA Sequence Generator https://faculty.ucr.edu/~mmaduro/random.htm. Information about *C. reinhardtii* genes, such as GO terms, PFAM descriptions and identifiers, and promoter sequences were retrieved through Phytozome Biomart (Smedley *et al*., 2009) or UniProtKB (https://www.uniprot.org/).

2.9.3. Computational *de novo* motif discovery and analysis

All promoter sequences (−1000 bp from 5'-UTR) of the *C. reinhardtii* genome v5.5 were downloaded via Phytozome Biomart (Smedley *et al*., 2009).

Weeder version 2.0 (Pavesi *et al*., 2004; Zambelli *et al*., 2014) was downloaded from http://159.149.160.88/modtools/ and installed within a UNIX environment. All 17,741 available *C. reinhardtii* promoter sequences were formatted accordingly to create the frequency file for the Weeder background model. Default settings for Weeder2.0 were applied, where motif lengths considered were 6, 8 and 10 bp, with up to 3 substitutions per motif. The number of motifs generated was capped at 25.

HOMER v4.9 (Heinz *et al*., 2010) was downloaded from http://homer.ucsd.edu/homer/download.html and installed in UNIX. All 17,741 promoters (−1000 bp from TSS) in the *C. reinhardtii* genome v5.5 were used to generate an in-program background model, and default parameters were used for motif discovery.

To set up in-browser motif discovery using DREME, each 1000 bp promoter sequence of the top 267 high-expression promoter list was uploaded as 10 x 100 bp fragments in FASTA format. Data was submitted to the DREME Version 5.1.1 (Bailey, 2011) online server at http://meme-suite.org/tools/dreme, and 'discriminative' analysis was performed using the promoter FASTA sequences for the 300 lowest expressed genes in Mettler *et al*. (2014) instead of a background model.

MEME version 5.1.1 analysis was performed in-browser at http://meme-suite.org/tools/meme using default running parameters (Bailey and Elkan, 1995). Shuffled input sequences were used to generate the background model.

Motif clustering was performed using RSAT motif clustering software (Castro-Mondragon *et al*., 2017). The online server was accessed at http://rsat.sb-roscoff.fr/matrix-clustering_form.cgi. Position weight matrices (PWMs) were converted into compatible formats and inputted into the program. Default parameters were used. For motif enrichment, AME version 5.1.1 (http://meme-suite.org/tools/ame) and CentriMo version 5.1.1 (http://meme-suite.org/tools/centrimo) were used.

Searches for PLACE plant motifs in discovered motifs was completed by entering the consensus sequence of each motif individually into https://www.dna.affrc.go.jp/PLACE/?action=newplace, where each sequence was scanned for known plant TFBSs. Consensus sequences were scanned as 2x copies to cover potential TFBSs that could arise where the pCRE repeats join. For PLACE ID references: https://www.dna.affrc.go.jp/PLACE/place_seq.shtml.

Chapter 3: Enhanced lutein and β-carotene production in *C. reinhardtii* by overexpression of a putative post-translational regulator of phytoene synthase

3.1. Summary

In this chapter, a forward genetic engineering approach was applied to successfully enhance lutein biosynthesis in *C. reinhardtii*. The DnaJ-like ORANGE protein increases the catalytic activity of phytoene synthase, a rate-limiting enzyme in the plant carotenoid biosynthetic pathway, and triggers differentiation of chloroplasts to chromoplasts in higher plants; its overexpression has previously led to increased carotenoid accumulation in several plant species. Here, the *C. reinhardtii* protein CPL6 was discovered to be a homologue of ORANGE (crOR), and was subsequently cloned into an overexpression vector for transformation into *C. reinhardtii* by electroporation. Significant increases in lutein production were observed in crOR transformants; the two strains analysed, crOR-Mut-1 and crOR-Mut-2, produced 1.7-fold and 2.0-fold more lutein per cell than the wild-type strain, respectively. This work demonstrates the application of a novel protein tool for increasing lutein biosynthesis, which complements the traditional forward engineering approach of enzyme overexpression.

3.2. Introduction

Carotenoids are high-value isoprenoid pigment molecules that are naturally synthesised by photosynthetic organisms, where their functions include light harvesting and cellular protection from damaging excess light (Frank and Cogdell, 1996; Cazzonelli, 2011). The anti-oxidative properties of carotenoids make them excellent free-radical scavengers, capable of limiting toxic singlet oxygen species that are produced as by-products from photosynthesis and other cellular activity. The yellow carotenoid lutein is of particular value; it is a necessary metabolite for human health, and oral supplementation of lutein offers protection to the eyes and skin from damaging blue- and UV-light, as well as some degenerative age-related diseases (Bernstein *et al*., 2016; Tian *et al*., 2015; Bovier and Hammond, 2015; Grether-Beck *et al*., 2017). Lutein is also an important additive to poultry feed where it intensifies the yellow pigmentation of poultry egg yolks, fat and skin (Leeson and Caston, 2004).

Microalgae are natural carotenoid producers, often generating high amounts of pigments under certain conditions, such as high light or salt stress, making them ideal candidates for carotenoid

production. Examples of the successful use of microalgae as biofactories to produce carotenoids on a large scale include β-carotene production in *D. salina* (Lamers *et al*., 2010), and astaxanthin production in *H. pluvialis* and *Chl. zorifingiens* (Lorenz and Cysewski, 2000; Liu *et al*., 2014). Currently, microalgae are the dominant biological source of industrially produced carotenoids (Gong and Bassi, 2016).

Natural lutein is predominantly extracted from marigold oleoresin. Switching over to microalgal-based lutein production could present several advantages: microalgae can be more productive per unit of land due to their unicellularity and rapid growth, they are more photosynthetically efficient, can grow throughout the whole year, require less space for cultivation and do not compete with food crops for arable land (Chisti, 2008; Scaife *et al*., 2015). Microalgae also have the potential to be cultured on waste materials. Currently, microalgae do not produce lutein to high enough levels to be competitive with marigold, when extraction costs are considered (Fernández-Sevilla *et al*., 2010). Upstream cultivation conditions and microalgal strain improvement, as well as downstream processes such as carotenoid extraction and purification, can be enhanced to ensure the competitiveness of lutein production in microalgae. High microalgal producers of lutein include *Muriellopsis sp*. (4–6 mg g^{-1} DCW; Blanco *et al*., 2007) and *Scenedesmus almeriensis* (5.5 mg g^{-1} DCW; Sánchez *et al*., 2008) and several *Chlorella* species, including *Chl. minutissima* (8.24 mg g^{-1} DCW; Dineshkumar *et al*., 2016), *Chl. sorokiniana* (7 mg g^{-1} DCW; Cordero *et al*., 2010b) and *Chl. Vulgaris* UTEX 1803 (9.82 mg g^{-1} DCW; Gong and Bassi, 2017). However, comparatively slow growth and specific growth requirements can make these species difficult to culture on a large scale. Furthermore, employing a genetic engineering strategy to enhance lutein yields in the aforementioned species would be difficult, as genetic modification tools are currently lacking. An alternative approach could be to genetically enhance lutein production in the model green microalga *C. reinhardtii*, which has a relatively short doubling time (~7 h), naturally produces lutein to moderate levels, and has genomic and bioinformatic tools readily available to implement a metabolic engineering strategy towards improving lutein production (Cordero *et al*., 2011b; Jinkerson and Jonikas, 2015).

Attempts to increase production of carotenoids and other isoprenoid compounds in *C. reinhardtii* through overexpression of rate-limiting enzymes yielded modest or occasionally no increases in carotenoid accumulation. Experiments in which *D. salina*-derived and *Chl. Zorifingiens*-derived PSY (See **Figure 1.4** for the carotenoid biosynthetic pathway) were overexpressed resulted in 2.6-fold and 2.2-fold increases in lutein compared to WT *C. reinhardtii*, respectively (Couso *et al*., 2011;

Cordero *et al*., 2011a); no significant difference in squalene biosynthesis was observed following overexpression of a native squalene synthase (Kajikawa *et al*., 2015).

Alternative strategies to boost carotenoid levels further could include targeting the regulation of carotenoid biosynthetic enzymes, or introducing a metabolic sink into which carotenoids can accrue. The ORANGE protein has recently been discovered in higher plants, where certain ORANGE mutants can accumulate higher levels of carotenoids than their WT counterparts (Lu *et al*., 2006; Yuan *et al*., 2015; Tzuri *et al*., 2015). ORANGE was first identified in cauliflower, where mutants containing a retrotransposon insertion in the ORANGE gene are bright orange in colour due to hyper-accumulation of β-carotene (Li *et al*., 2001, 2003; Lu *et al*., 2006). ORANGE does not appear to increase expression of carotenogenic enzymes, but rather induces the differentiation of non-pigmented plastids into carotenoid-accumulating chromoplast organelles (Li *et al*., 2001; Lu *et al*., 2006). ORANGE has the additional function of post-transcriptionally stabilising PSY, a rate-limiting carotenogenic enzyme (**Figure 1.2**), by increasing its enzymatic availability, thus enhancing metabolic flux through the carotenoid pathway due to the enzyme's gate-keeper role (Li *et al.,* 2012; Zhou *et al*., 2015).

Overexpression of ORANGE in plant species has resulted in significant increases in carotenoid production; for example, a 6-fold increase in lutein was observed in sweet potato after native overexpression of its ORANGE gene (Kim *et al*., 2013). An ORANGE homologue (*crOR*) was identified in *C. reinhardtii* during a study examining the effects of light exposure on mRNA transcript abundances of carotenoid biosynthesis genes in *C. reinhardtii*; *crOR* displayed a 14-fold increase in mRNA abundance following high-light exposure, in coordination with the expression of other carotenogenic enzymes (Sun *et al*., 2010). As of the beginning of this project, the function of *crOR* in *C. reinhardtii* and its effect on carotenoid accumulation had not yet been explored. The aim of this chapter is to clone the putative *ORANGE* gene from *C. reinhardtii*, and to examine the effects of its endogenous overexpression on carotenoid sequestration and chloroplast morphology via HPLC analysis and confocal microscopy.

3.3. Results

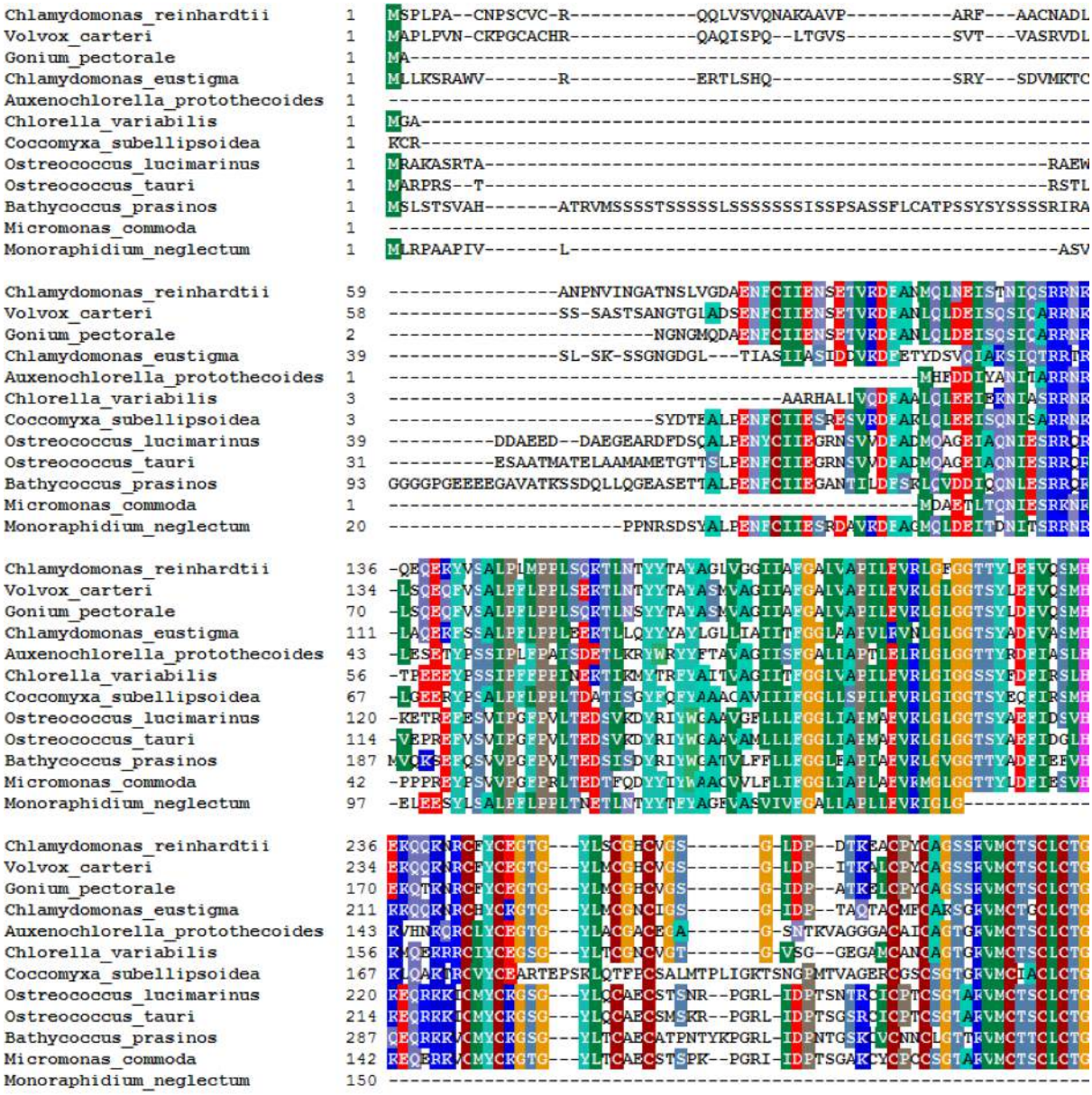
3.3.1. The putative ORANGE protein, CPL6, in *C. reinhardtii*

The putative *C. reinhardtii* ORANGE protein (CPL6) is 302 aa in length with an approximate molecular weight of 32.5 kilodaltons (kDa) and a pI of 6.86 (Merchant *et al*., 2007). CPL6 is predicted to contain two transmembrane domains between amino acids 163–182 and 213–230 and an N-terminal

chloroplast transit peptide of approximately 29 aa, suggesting localisation of CPL6 in the thylakoid membrane. CPL6 contains a putative C-terminal DnaJ-like cysteine-rich zinc-finger domain, which is characteristic of the ORANGE proteins found in plants (Lu *et al*., 2006; Zhou *et al*., 2015; Park *et al*., 2016). The highly conserved DnaJ-like domain comprises 4 repeats of a CxxCxGxG motif (Lu *et al*., 2006). DnaJ-like proteins are co-chaperones that function by activating Hsp70A ATPase domains, and which are involved in protein translation, translocation, folding and unfolding, stabilisation, and degradation (Cheetham and Caplan, 1998; Hartl and Hayer-Hartl, 2002).

A BLASTP search (National Center for Biotechnology Information) using the CPL6 protein sequence as the query (Merchant *et al*., 2007) resulted in several hits, including a hypothetical protein in the closely-related green multicellular alga *V. carteri* (74% similarity), as well as the characterised plant ORANGE proteins of *A. thaliana* (59%), sweet potato (54%), tomato (53%) and cauliflower (55%). Several green algal species, or Chlorophyta, also harboured sequences with high sequence similarity to ORANGE, such as *Chl. variabilis* (64% similarity), *Chl. sorokiniana* (60%), *D. salina* (62% similarity), *H. lacustris* (41%), *Ostreococcus tauri* (35%), *Bathycoccus prasinos* (44%) and *Monoraphidium neglectum* (64%). **Figure 3.1** shows multiple sequence alignment analyses between CPL6 and other homologous ORANGE proteins obtained using BLASTP. The C-terminal region of each sequence selected shared strong sequence homology corresponding to the DnaJ-like domain, which in *C. reinhardtii* spans aa 231–290. *C. reinhardtii*, however, appears to have lost a 16 aa region within the DnaJ region that is present in plant species. The N-terminal region of CPL6 appears to share less sequence homology with the plant species, although this is also the case between the plant species themselves. Despite this, from 70 aa onward in *C. reinhardtii* CPL6, strong homology is evident, suggesting that CPL6 is in fact the *C. reinhardtii* ORANGE protein, crOR.

A


Chlamydomonas_reinhardtii
Volvox_carteri
Gonium_pectorale
Chlamydomonas_eustigma
Auxenochlorella_protothecoides
Chlorella_variabilis
Coccomyxa_subellipsoidea
Ostreococcus_lucimarinus
Ostreococcus_tauri
Bathycoccus_prasinos
Micromonas_commoda
Monoraphidium_neglectum

B

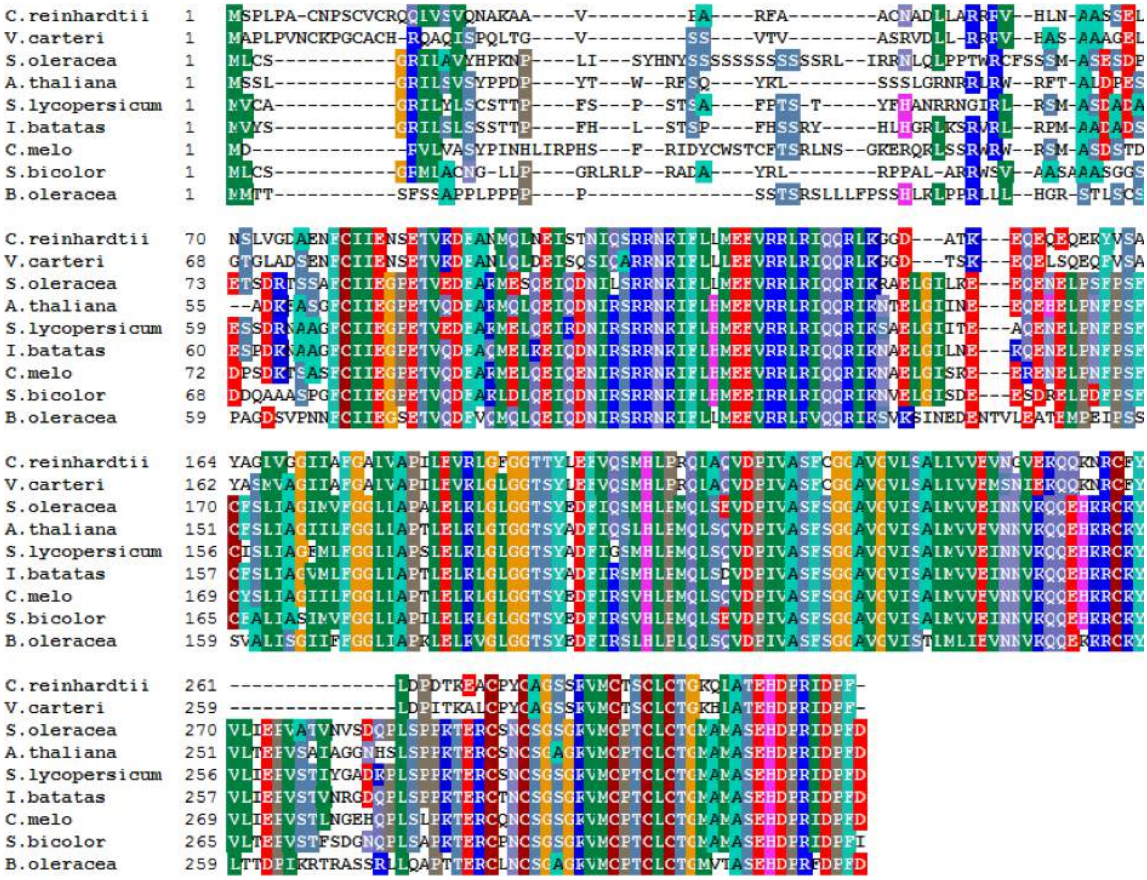


Figure 3.1: Multiple sequence alignment of the *C. reinhardtii* putative ORANGE protein with algal Chlorophyte homologous proteins and established plant ORANGE proteins. Multiple sequence alignment of *C. reinhardtii* putative ORANGE protein and indicated Chlorophyte (**A**) and plant (**B**) and species ORANGE amino acid sequences. Amino acid matches are coloured similarly. GenBank Accession numbers for each of the protein hits for **A** are as follows: *C. reinhardtii,* XP_001695304.1; *V. carteri*, XP_002946319.1; *Gonium pectoral*, KXZ44851.1; *C. eustigma*, GAX82776.1; *Auxenochlorella protothecoides*, XP_011397208.1; *Chl. variabilis*, XP_005843417.1; *Coccomyxa subellipsoidea* C-169, XP_005643465.1; *Ostreococcus lucimarinus* CCE9901, XP_001422032.1 predicted; *Ostreococcus tauri*, CEG00612.1; *Bathycoccus prasinos*, XP_007512757.1; *Micromonas commoda*, XP_002502675.1; *Monoraphidium neglectum*, XP_013899468.1. GenBank Accession numbers for each of the protein hits for **B** are as follows: *C. reinhardtii,* XP_001695304.1; *V. carteri*, XP_002946319.1; *Spinacia oleracea*, XP_021865700.1; *Arabidopsis thaliana*, NP_200975.1; *Solanum lycopersicum*, NP_001315338.1; *Ipomoea batatas*, APG21184.1; *Cucumis melo*, XP_008467325.1; *Sorghum bicolor*, XP_002452827.1; *Brassica oleracea var. oleracea*, XP_013620420.1.

3.3.2. Cloning and expression of ORANGE in CC-4533

The sequence for the putative ORANGE protein in *C. reinhardtii* (CPL6, Cre06.g279500, protein NCBI accession number XP_001695304.1) was acquired following a BLASTP search (See above) with the query comprising the *Brassica oleracea* OR peptide sequence (XP_013607043.1), followed by a search for the recovered *cpl6* gene in the *C. reinhardtii* genome v5.5 (**Figure 3.2A**). The candidate *C. reinhardtii* ORANGE gene, *clp6*, was cloned directly from *C. reinhardtii* gDNA using primers crOR_F and crOR_R (**Table 2.1**); the *clp6* gene was cloned in its entirety, including intronic regions, due to evidence that introns can act as positive enhancers for gene expression (Lumbreras *et al*., 1998; Eichler-Stahlberg *et al*., 2009). 5'- and 3'-UTRs were excluded. The *cpl6* gene including introns comprises 2557 bp with 6 exonic regions. Restriction sites *Nde*I and *EcoR*V were included in the primer design for insertion of the cloned fragment into the pOpt_mVenus_Paro vector (**Figure 2.1**), as well as an N-terminal poly-6-histidine tag (His_6-tag) to allow differentiation between native and re-introduced *cpl6* in downstream experiments (**Figure 3.2B**). The resulting DNA fragment, of a total size of 2593 bp including the His_6-tag and restriction sites, was amplified via PCR and inserted into the pOpt_mVenus_Paro vector by digestion and ligation reactions to create the vector pOpt_CrOR (**Figure 3.2**). The His-tagged *cpl6* gene fragment (His_6-*cpl6*) was inserted downstream of the *Hsp70A-RbcS2* hybrid promoter region plus RbcS2 intron1 in order to drive its constitutive expression (**Figure 3.2B**; Sizova *et al*., 2001; Lauersen *et al*., 2015). **Figure 3.2C** depicts the full vector map. For associated gels for all steps of pOpt_crOR construction, see **Appendix Figure B1**.

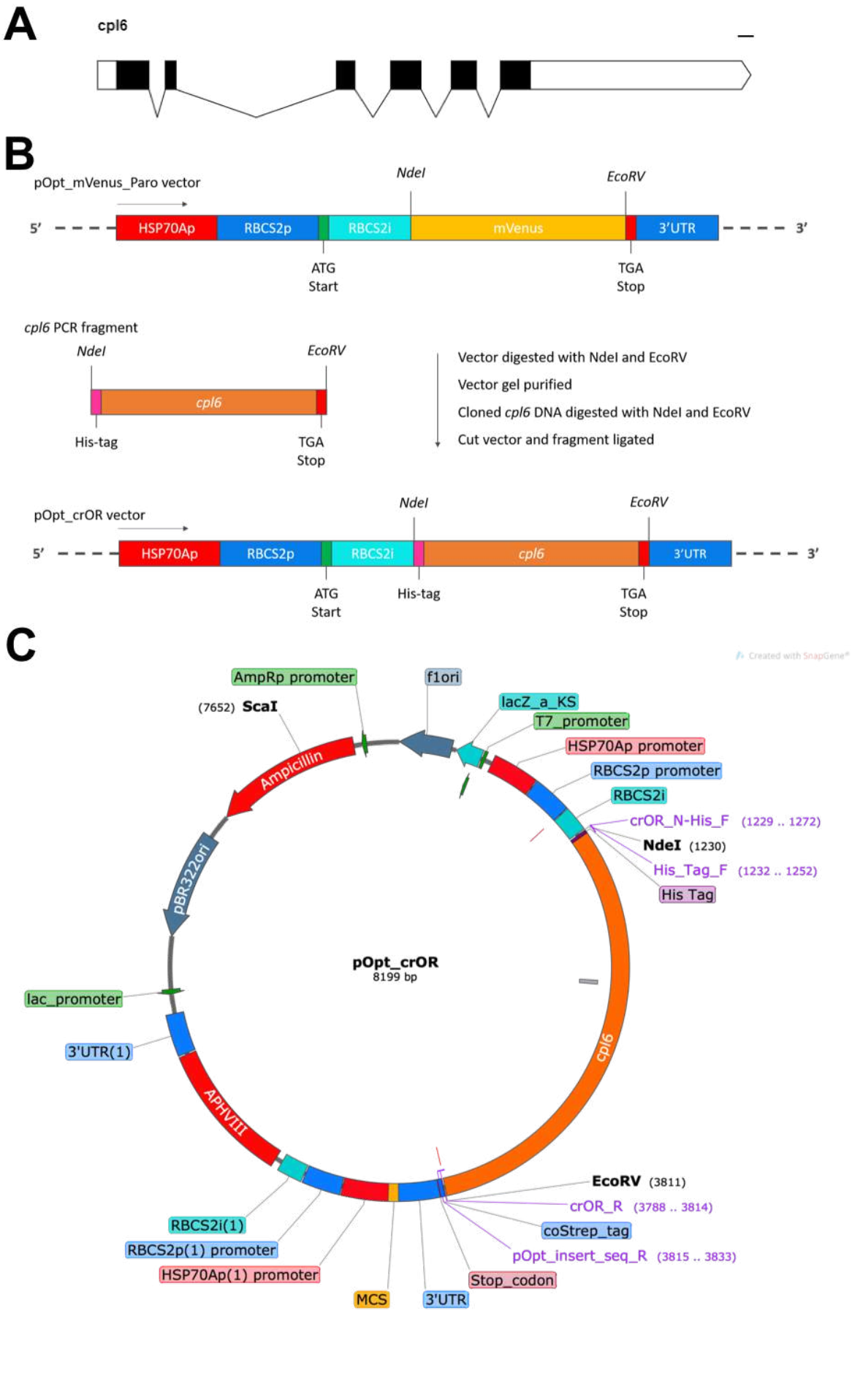

A
cpl6
B
pOpt_mVenus_Paro vector
NdeI
EcoRV
5′
HSP70Ap
RBCS2p
RBCS2i
mVenus
3′UTR
3′
ATG
Start
TGA
Stop
cpl6 PCR fragment
NdeI
EcoRV
cpl6
His-tag
TGA
Stop
Vector digested with NdeI and EcoRV
Vector gel purified
Cloned cpl6 DNA digested with NdeI and EcoRV
Cut vector and fragment ligated
pOpt_crOR vector
NdeI
EcoRV
5′
HSP70Ap
RBCS2p
RBCS2i
cpl6
3′UTR
3′
ATG
Start
His-tag
TGA
Stop
C
Created with SnapGene®
AmpRp promoter
f1ori
lacZ_a_KS
(7652) ScaI
T7_promoter
HSP70Ap promoter
RBCS2p promoter
RBCS2i
crOR_N-His_F (1229 .. 1272)
NdeI (1230)
His_Tag_F (1232 .. 1252)
His Tag
Ampicillin
pBR322ori
pOpt_crOR
8199 bp
lac_promoter
3'UTR(1)
APHVIII
cpl6
EcoRV (3811)
crOR_R (3788 .. 3814)
coStrep_tag
pOpt_insert_seq_R (3815 .. 3833)
Stop_codon
3'UTR
MCS
HSP70Ap(1) promoter
RBCS2p(1) promoter
RBCS2i(1)

Figure 3.2: Cloning strategy for pOpt_crOR construction and vector map of pOpt_crOR. A – Schematic image of the full genomic *cpl6* gene. 5' to 3' direction. White boxes at left and right terminals represent 5'-UTR and 3'-UTR, respectively. Black boxes represent exons, and lines conjoining black boxes represents intronic regions. Diagram to scale, scale bar in top right = 100 bp. **B** – Cloning strategy to replace mVenus gene with *cpl6* in pOpt_mVenus_Paro expression vector to create pOpt_crOR. mVenus was digested out of the pOpt_mVenus_Paro vector using NdeI and EcoRV restriction enzymes, after which the empty vector was gel purified; *cpl6* was similarly digested with NdeI and EcoRV, and the empty vector plus digested *cpl6* fragment ligated together and transformed into *E. coli* DH5α for plasmid propagation and testing. Genetic elements shown include *Hsp70A-RbcS2* hybrid promoter (HSP70Ap and RBCS2p), RbcS2 intron I (RBCS2i), RbcS2-3'UTR (3'UTR), translation start site (ATG start), translation stop site (Stop TGA), poly-6-histidine tag (His-tag, pink), *Nde*I and *EcoR*V restriction sites. Arrows show direction of transcription. See **Appendix Figure B1** for gels. **C** – Vector map of pOpt_crOR. Primers shown in purple. *cpl6* gene is shown in orange, downstream of the *Hsp70A-RbcS2* promoter-enhancer region.

pOpt_crOR was linearised using the restriction enzyme ScaI (NEB) and transformed into *C. reinhardtii* CC-4533 by electroporation. Successful transformants were selected for using paromomycin, then colony screened by PCR using primers specific for His_6-tag DNA and *cpl6* to identify positive pOpt_crOR transformed mutants (Forward primer, 'His_Tag_F'; Reverse primer, 'crOR_R'; **Table 2.1; Figure 3.2C**). An example agarose gel of a colony screening PCR is shown in **Figure 3.3**, where positive transformants exhibited a band at ~2500 bp, as indicated by the red arrow. Lanes 8, 10, 11 and 16 represent positive transformants; no 2500 bp band is present in lane 3, which represents the wild-type (WT) negative control. An average of 349 colonies were formed following selection with paromomycin; of 78 colonies screened from a randomly selected transformation agar plate, 11 colonies contained the inserted His_6-*cpl6* gene, giving a co-transformation efficiency of both the *AphVIII* resistance gene and the His_6-*cpl6* gene of ~14%.

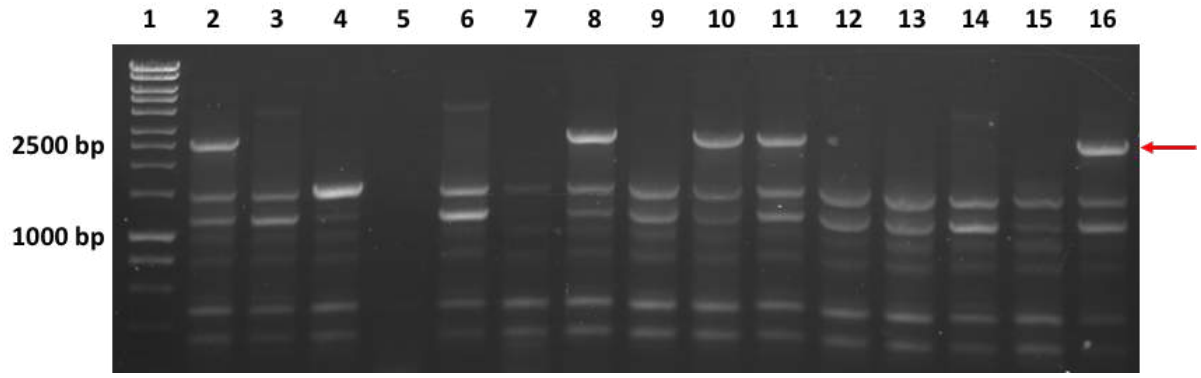


Figure 3.3: Colony PCR screening for positive *C. reinhardtii cpl6* nuclear genomic transformants. Paromomycin resistant *C. reinhardtii* colonies that were transformed with the vector pOpt_crOR containing His_6-*cpl6* gene were suspended in Dilution Solution (Phire Plant Direct PCR Kit, Thermo Fisher Scientific) and genomic DNA amplified using primers His_Tag_F and crOR_R (**Table 2.1**). Lane 1, 1 kb DNA ladder; lane 2, confirmed positive transformant (positive control); lane 3, wild-type (negative control); lanes 4–16, paromomycin resistant colonies. Red arrow indicates fragments at ~2500 bp in length.

3.3.3. Growth and physiology of parental and pOpt_crOR-transformed strains

Two randomly selected positive *crOR* transformants, crOR-Mut-1 and crOR-Mut-2, were grown under standard conditions alongside the parental strain CC-4533. All three strains appeared to grow similarly, reaching stationary phase after 91 h, and remaining stable for the remainder of the experiment (**Figure 3.4**). **Table 3.1** displays the growth kinetics of the three strains; crOR-Mut-2 had a significantly higher specific growth rate and doubling time than CC-4533, whereas there was no significant difference between crOR-Mut-1 and the other strains.

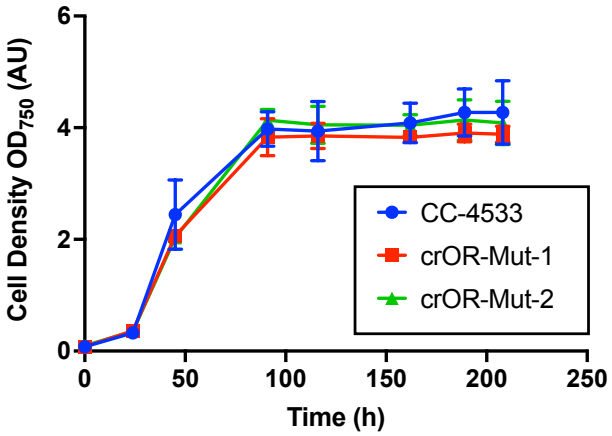


Figure 3.4: Growth kinetics of wild-type CC-4533 and *cpl6*-positive transformant strains grown under standard conditions. Data shown are taken from three biological replicates. Error bars represent standard deviation, all of which are < 10% of the mean. Growth was measured by optical density at 750 nm.

Table 3.1: Growth kinetics of parental CC-4533 and *cpl6*-positive transformant strains grown under standard conditions

Strain	Specific growth rate (SGR, h^{-1})	Doubling Time (h)
CC-4533 (control)	0.062 ± 0.008	11.298 ± 1.585
crOR-Mut-1	0.081 ± 0.018	8.939 ± 2.255
crOR-Mut-2	0.091 ± 0.002**	7.627 ± 0.159*

Growth rate and doubling time calculated using cell counts taken at 24 h and 45 h time points. Data expressed as means ± standard deviation (SD), number of replicates (n) = 3. Significant differences were determined using a student's *t*-test. *$P < 0.05$, **$P < 0.01$.

Confocal fluorescence images were taken of CC-4533, crOR-Mut-1 and crOR-Mut-2 strains grown to stationary phase in order to examine the effect of CPL6 expression on chloroplast morphology; evidence suggests that the overexpression of ORANGE in plant species induces chloroplast differentiation into carotenoid-rich chromoplast structures (Lu et al., 2006; Lopez *et al*., 2008; Yuan *et al*., 2015; Chayut *et al*., 2017). Chloroplasts were imaged by exploiting natural chlorophyll autofluorescence; **Figure 3.5** shows the confocal images taken. Cells from each strain are roughly spherical in shape, at around 10 μm in diameter. Cup-shaped fluorescence is visible within the cells, corresponding to the characteristic shape of the *C. reinhardtii* chloroplast. No discernible differences can be observed between the chloroplasts of each strain; the lack of obvious differences

in chloroplast morphology between the CC-4533 and *crOR* transformant strains crOR-Mut-1 and crOR-Mut-2 suggests that the crOR protein may not be involved in chloroplast differentiation, at least in terms of structure and chlorophyll fluorescence.

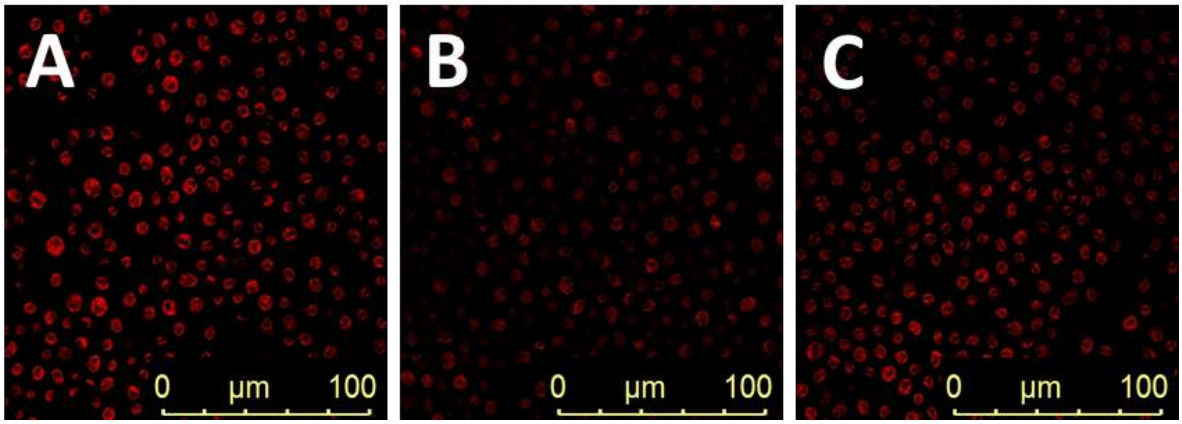


Figure 3.5: Confocal images showing chloroplast fluorescence of *cpl6*-transformant strains. Strains shown are **A**, CC-4533; **B**, crOR-Mut-1; **C**, crOR-Mut-2. Chlorophyll autofluorescence was imaged by laser excitation at 488 nm and emission detection at 650–700 nm. Magnification X40. Cells harvested on Day 9.

3.3.4. Protein expression analysis of *cpl6*-transformed strains

In order to confirm the overexpression of recombinant crOR in the transformed cells, proteins were extracted from strains crOR-Mut-1, crOR-Mut-2 and CC-4533 and a Western immunoblot was performed. The re-introduced version of the crOR protein could be distinguished from native crOR via the His_6-tag insert at the N-terminus of the protein; this was exploited by blotting with an anti-His_6 HRP-conjugated antibody. **Figure 3.6A** displays a PVDF membrane blotted with stained protein bands with anti-His_6 antibody affinity. Non-specific binding to a protein 43–55 kDa in length is apparent across all strains; given its presence in each of the samples, this band was used as a natural internal control to measure protein band density. A clear protein band sized between 34–43 kDa can be seen for mutant crOR-Mut-2 (**Figure 3.6A** Lanes 7, 8, & 9), and the same band but more diminished for crOR-Mut-1 (Lanes 4, 5 & 6); this band is not present in the wild-type CC-4533 strain (Lanes 2 & 3). The predicted size of the crOR protein with attached His_6 tag is 33.3 kDa; it is likely that the bands indicated by the red rectangle in **Figure 3.6A** correspond to this protein. Densitometry analysis of the highlighted bands revealed a significant difference (P = 0.0063)

between the crOR-Mut-1 and crOR-Mut-2 band densities; no density peak at the same molecular weight as the putative crOR band was detected in the CC-4533 control sample.

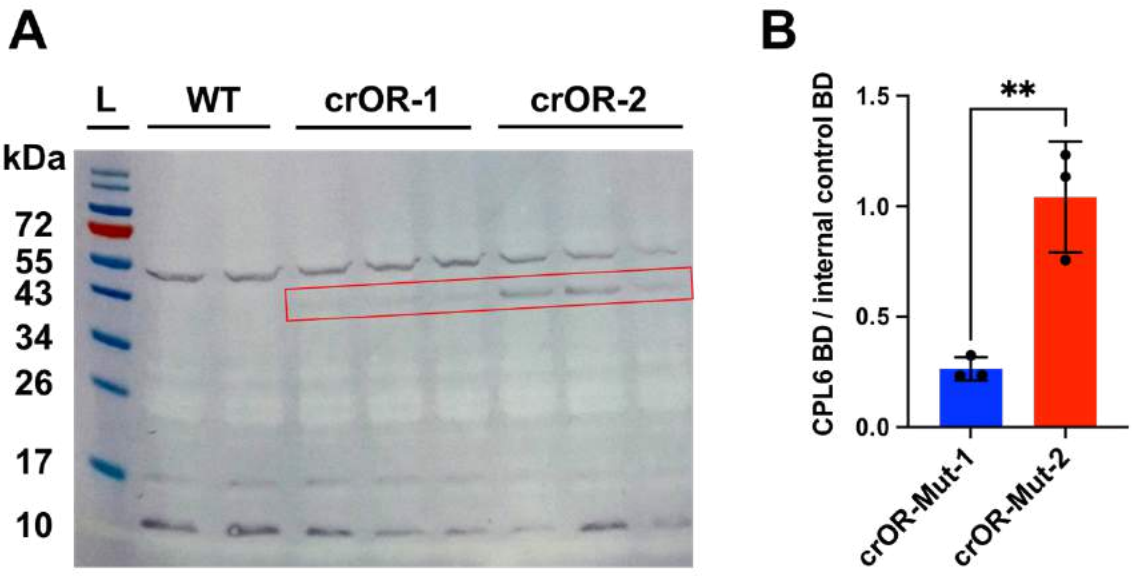


Figure 3.6: Western immunoblot showing protein bands with corresponding epitopes to anti-6-histidine antibody and densitometry calculations. A – Western blot showing protein bands stained with HRP-linked anti-6-his antibodies and TMB blotting solution. L = protein ladder. The strains from which protein samples were extracted are indicated above the lanes and abbreviated as follows: WT, untransformed CC-4533; crOR-1, crOR-Mut-1; crOR-2, crOR-Mut-2. Red rectangle highlights the putative His-CPL6 protein. **B** – The mean normalised densities of the putative His-CPL6 bands are plotted here. BD = band density. The protein band at ~50 kDa was present in all samples, and was hence used as an internal control to normalise the His-CPL6 band. Densities were measured using ImageJ. The mean normalised density values for crOR-Mut-1 and crOR-Mut-2 were compared using a student's t-test (**$P < 0.01$).

3.3.5. Analysis of pigment profiles of parental and *cpl6*-transformed strains

Parental CC-4533 and *cpl6*-transformed strains were grown under standard conditions and their pigments extracted with 80% acetone at late log phase (68 h). Total pigment concentrations were calculated using spectrophotometric measurements, and the values obtained are depicted in **Table 3.2**. The chlorophyll *a* and total carotenoid contents of the crOR-Mut-2 transformant strain were significantly higher than CC-4533; however, no other significant differences in total pigment levels were observed for crOR-Mut-1. Similarly, the chlorophyll-*a* to -*b* ratios and chlorophyll-to-carotenoid ratios did not significantly differ between strains.

Table 3.2: Pigment contents of parental CC-4533, crOR-Mut-1 and crOR-Mut-2 strains

Strain	Chl *a*/cell (pg)	Chl *b*/cell (pg)	Car (pg)	Chl *a*/*b*	Chl/Car
CC-4533 (WT)	1.39 ± 0.09	0.63 ± 0.08	0.45 ± 0.01	2.22 ± 0.16	4.50 ± 0.38
crOR-Mut-1	1.66 ± 0.25	0.78 ± 0.13	0.53 ± 0.08	2.13 ± 0.05	4.57 ± 0.16
crOR-Mut-2	1.84 ± 0.24*	0.84 ± 0.14	0.59 ± 0.05**	2.19 ± 0.09	4.50 ± 0.28

Data expressed as mean ± SD. Asterisks indicate significantly different values (one-way ANOVA; *$P < 0.05$, **$P < 0.01$; n = 3) from the control strain CC-4533. Following abbreviations used: Chl, Chlorophyll; Car, total carotenoids; pg, picograms. For calculations, see **Section 2.6.2**.

Extracted pigments were subjected to HPLC separation analysis. Lutein, β-carotene and chlorophyll-*a* were identified by comparison to known standards (**Figure 3.7A–C**; **Figure 4.9**). The other peaks were tentatively assigned using previous results from the literature obtained using the same HPLC protocol (Couso *et al*., 2011; **Figure 3.7D**). **Figure 3.7E** shows representative chromatograms following HPLC separation of pigment extractions from strains CC-4533, crOR-Mut-1 and crOR-Mut-1; pigments 3, 5 and 7 represent lutein, chlorophyll-*a* and β-carotene, respectively. By comparison to **Figure 3.7D**, pigments 1, 2 and 4 of **Figure 3.7E** appear to represent neoxanthin, violaxanthin and chlorophyll-*b*, respectively, and will from hereon be denoted with 'p' for putative. For each of the strains examined, pNeoxanthin was the first pigment to elute (~5.8 min), followed by pViolaxanthin (~6.5 min), lutein (~9.2 min), chlorophylls *b* and *a* (~14.3 and ~15.7 min), then β-carotene (~20 min). The β-carotene peak at ~20 min was not detectable during every run of the HPLC. By comparison with the standard spectra, a peak between 17–18 min, which appeared in some but not all chromatograms, is likely a degradation product of β-carotene (**Appendix Figure B3**).

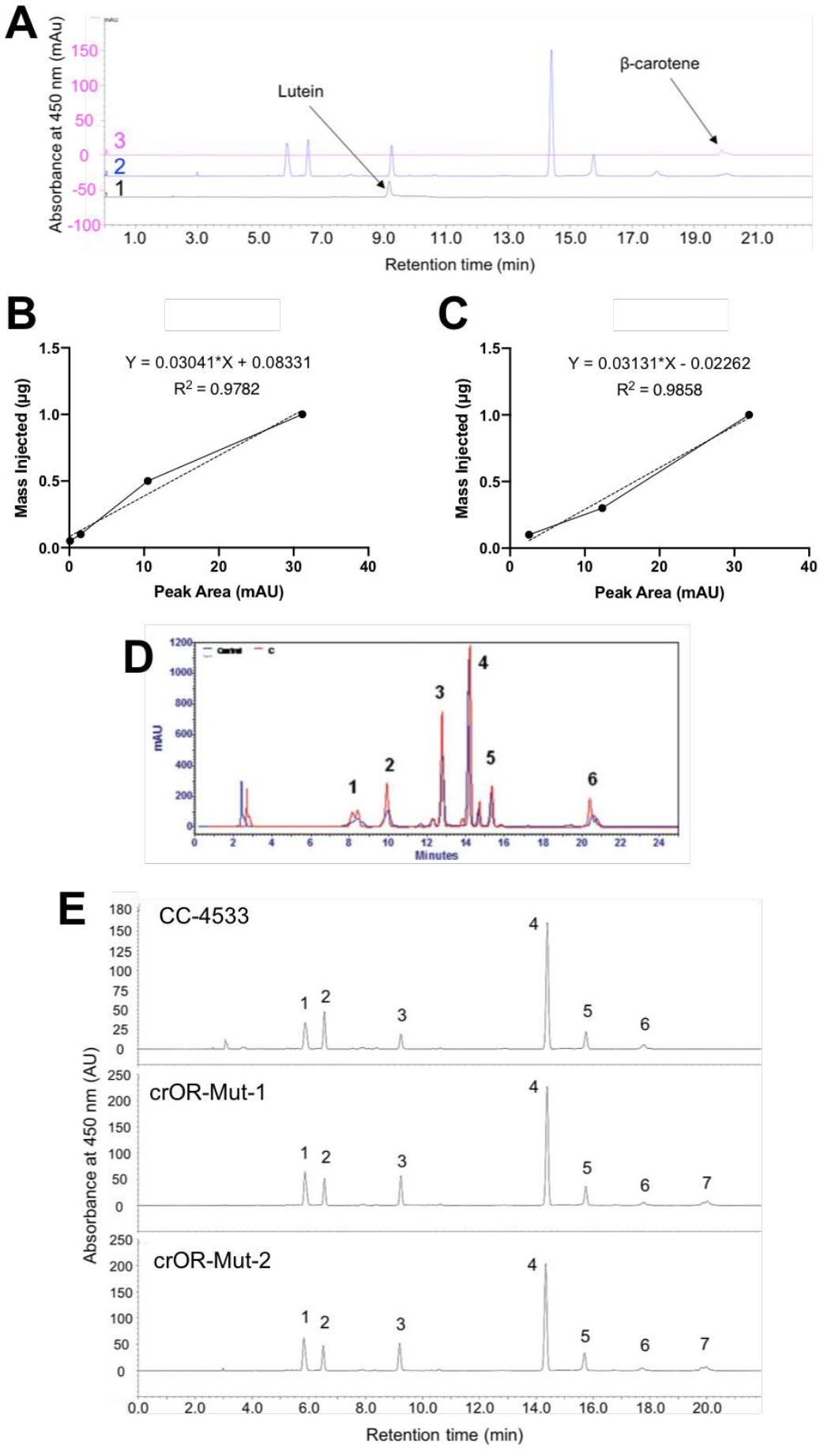

A
Absorbance at 450 nm (mAu)
Lutein
β-carotene
Retention time (min)
B
Y = 0.03041*X + 0.08331
R2 = 0.9782
Mass Injected (µg)
Peak Area (mAU)
C
Y = 0.03131*X - 0.02262
R2 = 0.9858
Mass Injected (µg)
Peak Area (mAU)
D
mAU
Minutes
E
CC-4533
crOR-Mut-1
crOR-Mut-2
Absorbance at 450 nm (AU)
Retention time (min)

Figure 3.7: HPLC separation of *C. reinhardtii* pigments. A – Overlay chromatogram showing HPLC separation of lutein (**1**; 0.1 μg, black) and β-carotene (**2**; 0.1 μg, magenta) pigment standards. A representative chromatogram of acetone-extracted pigments from strain CC-4533 are shown in blue (**2**). Y-axis corresponds to β-carotene spectrum (shown in magenta). **B** – Calibration curve for lutein analytical standard. Calculated per mL injected sample. **C** – Calibration curve for β-carotene analytical standard. Calculated per mL injected sample. **D** – Previously published HPLC chromatogram showing the carotenoid profiles of a *C. reinhardtii* wild-type control (blue) and a mutant *C. reinhardtii* strain transformed with *Dunaliella salina* PSY (red), following a similar HPLC program to that applied in this study. Numbers correspond to: **1**, neoxanthin; **2**, violaxanthin; **3**, lutein; **4**, chlorophyll-*b*; **5**, chlorophyll-*a*; **6**, β-carotene. Image taken with permission from Couso *et al*. (2012). **E** – Representative chromatograms of pigments extracted from the CC-4533 control strain and two overexpression strains (crOR-Mut-1 and crOR-Mut-2). Numbers correspond to the following pigments: **1**, putative neoxanthin; **2**, putative violaxanthin; **3**, lutein; **4**, chlorophyll-*b*; **5**, chlorophyll-*a*; **6**, β-carotene degradation product; **7**, β-carotene. See **Section 2.6.3***.* for separation program. All spectra for each strain are shown in **Appendix Figure B2**.

Figure 3.8 shows the abundance of each carotenoid in the CC-4533 parental strain, as well as the *crOR*-transformed strains crOR-Mut-1 and crOR-Mut-2, per 10^7 cells. A significant increase in all carotenoids was observed in both transformant strains crOR-Mut-1 and crOR-Mut-2 when compared to the CC-4533 parental strain, except for β-carotene. Of particular interest, the greatest increase of detected carotenoids was seen in the relative lutein content of transformants crOR-Mut-1 and crOR-Mut-2, in which lutein contents were 1.5- and 1.8-fold higher per cell than the control, respectively (**Figure 3.8**). There were also significant increases in pNeoxanthin (1.5- and 1.6-fold) and pViolaxanthin (1.1- and 1.8-fold) in crOR-Mut-1 and crOR-Mut-2, respectively (**Figure 3.8**).

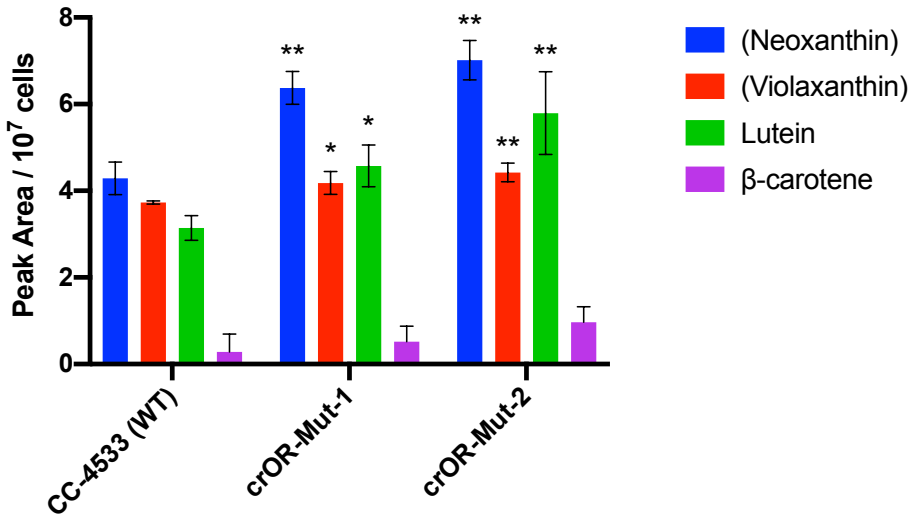


Figure 3.8: Peak area per 10^7 cells for each carotenoid detected via HPLC. Peak area calculated as follows: (Peak Area / cell number) x 10^7. Statistically significant differences from the CC-4533 control mean were calculated using student's *t*-test. *$P < 0.05$, **$P < 0.01$. n = 3. Error bars = SD. Carotenoids in brackets have been tentatively identified using previously published work, as opposed to analytical standards.

Lutein and β-carotene analytical standards were injected in increasing concentrations to produce calibration curves for quantitative analysis (**Figure 3.7B, C**). The contents of lutein and β-carotene per L culture, per cell, and per g of dried biomass are shown in **Figure 3.9**. Volumetric and cellular lutein content increased significantly in both transformant strains. crOR-Mut-1 produced 1.7-fold more mg lutein per L culture than CC-4533, and crOR-Mut-2 demonstrated a 2.0-fold increase at 4 mg L^{-1} (**Figure 3.9A**). The same fold changes in femtogram (fg) lutein per cell were observed for crOR-Mut-1 (1.7-fold) and crOR-Mut2 (2.0-fold); CC-4533 produced 161.0 ± 16.0 fg lutein $cell^{-1}$, whereas 252.7 ± 29.6 and 325.7 ± 61.5 fg lutein $cell^{-1}$ were observed for crOR-Mut-1 and crOR-Mut-2, respectively. 6.1 ± 1.3 mg lutein g^{-1} DCW was recorded for crOR-Mut-2, which is almost double that recorded for CC-4533 (3.3 ± 0.3 mg lutein g^{-1} DCW), however this was not significant. β-carotene appears to increase slightly in each measured parameter, but is only significantly higher when calculated in fg per cell in crOR-Mut-2, which produced 1.3-fold more β-carotene than CC-4533 (**Figure 3.9B**). Large variations in peak area were observed for β-carotene (**Appendix Tables B3 & B4**), which likely contributed to the lack of significance for these measurements.

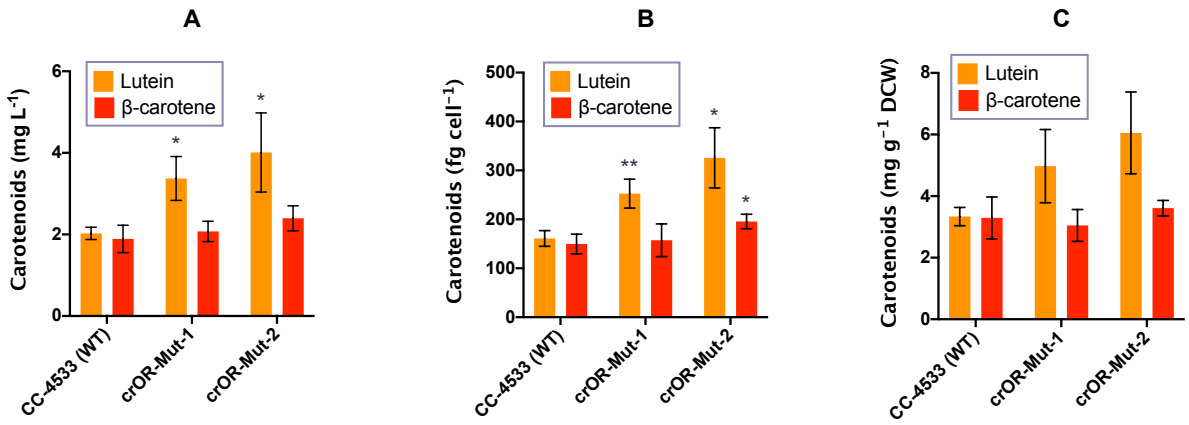


Figure 3.9: Lutein and β-carotene contents of parental CC-4533 and *cpl6*-transformed strains. Lutein and β-carotene contents expressed as mg per L of culture (**A**), fg per cell (**B**) and mg per g of dry biomass (**C**). Values calculated using standard curves generated from known amounts of pigment standard (**Figure 3.7B, C**). Biomass measurements used to calculate **C** can be found in **Appendix Figure B4**. Statistically significant differences from the CC-4533 (control) mean were calculated using student's *t*-tests. **P* < 0.05, ***P* < 0.01. n = 3. Error bars = SD. Values calculated in triplicate, with the exception of CC-4533 in (**C**), where n = 2.

3.4 Discussion

ORANGE proteins have been identified, isolated and characterised in higher plant species such as cauliflower (Lu *et al*., 2006), *Arabidopsis thaliana* (Bai *et al*., 2014), melon (Tzuri *et al*., 2015), sweet potato (Kim *et al*., 2013), and *Sorghum bicolor* (Yuan *et al*., 2015). As shown in **Figure 3.1**, each of the ORANGE proteins share sequence homology, particularly in the C-terminal DnaJ-like domain. Cloning and overexpression experiments of ORANGE proteins in higher plants gave rise to significant increases in carotenoids. Particularly successful examples include the native overexpression of the sweet potato *OR* gene, which resulted in a 6-fold increase in lutein compared to the WT (Kim *et al*., 2013), the heterologous overexpression of cauliflower *ORANGE* gene in white potato, producing coloured potato flesh with 6-fold augmentation of overall carotenoids (Lu *et al*., 2006), and the heterologous overexpression of *A. thaliana ORANGE* gene in rice which resulted in a 2.2-fold increase in carotenoids compared to rice strains that that had been engineered to overexpress carotenoid biosynthetic enzymes (Bai *et al*., 2014). Increasing the expression of ORANGE has important applications in food biotechnology, such as improving the nutritional value of crop plants.

Given the sequence similarity of crOR with the plant ORANGE proteins, as well as the ability of crOR to increase lutein production 2.0-fold per cell in *C. reinhardtii* when overexpressed (**Figure 3.9**), it is

highly probable that crOR is the *C. reinhardtii* equivalent of ORANGE. During the course of this project, Morikawa *et al*. (2018) conducted a similar experiment, in which they successfully cloned and overexpressed the crOR protein in *C. reinhardtii* in order to enhance carotenoid accumulation; they observed 1.9-fold increases in fg lutein per cell, which is consistent with the results of this work; this strengthens the hypothesis that crOR is the *C. reinhardtii* ORANGE protein.

Co-transformation efficiencies of the antibiotic resistance gene *AphVIII* and *cpl6 (crOR)* were relatively low, at ~14%. An improved version of the 'pOptimised' series of vectors are now available from the Chlamydomonas Resource Centre (Wichmann *et al*., 2018), and vectors that can express a heterologous gene bicistronically using the FMDV peptide (**Section 1.6.5.8.**) have become commercially available since the beginning of this project. Using these improved vector systems could reduce the number of colonies that must be screened to obtain an antibiotic-resistant strain containing the GOI. Using traditional cloning methods to isolate the coding sequence of the *crOR* gene from mRNA to produce cDNA, as opposed to cloning the gene in its entirety from gDNA, could also improve transformation efficiency by reducing the size of the DNA fragment transformed (Zhang *et al*., 2014). The second intron of *crOR* is disproportionately long at 993 bp, contributing almost 40% to the total gene from start to stop codon **Figure 3.2**; minimising the length of the DNA cassette transformed can improve both transformation efficiency and gene expression (personal communication with Dr A. Berndt). Essentially, keeping the introns intact during cloning with the aim of improving gene expression may have ultimately hindered it, as seen by the relatively low protein levels in the Western immunoblot (**Figure 3.6**). In future experiments, these large endogenous introns could be replaced with RbcS2 introns to facilitate higher protein expression (Baier *et al*., 2018; Jaeger *et al*., 2019). Further improvements to the expression system could also be considered, such as optimising the promoter used to reduce transgene silencing. The development of novel synthetic promoters will be explored in **Chapter 5**.

Confocal microscopy was used to examine chloroplast morphology in *C. reinhardtii*, as it was hypothesised at the beginning of this chapter that the overexpression of crOR may promote the accumulation of carotenoids within the thylakoid membrane to a greater extent in *C. reinhardtii*, which could in turn alter the size and shape of the chloroplast. As stated earlier, ORANGE is a known trigger for chloroplast differentiation into chromoplasts in plants (Li *et al*., 2001; Lu *et al*., 2006). The confocal images did not reveal noticeable changes in chloroplast architecture between the strains examined as detectable by chlorophyll fluorescence (**Figure 3.5**). This, however, does not mean to say that there are no differences at all between WT and crOR-overexpression strains. It could prove

worthwhile to repeat the confocal imaging process at a higher magnification at multiple stages of growth. Using a stronger microscopy technique such as electron microscopy could be another solution for imaging the chloroplast in more detail. It could also simply be the case that *C. reinhardtii* does not have the cellular machinery to drastically alter its chloroplast morphology; higher plants tend to have several chloroplasts per cell, as well as the capacity to harbour different species of plastid such as etioplasts, amyloplasts and chromoplasts (Sun *et al*., 2018), whereas *C. reinhardtii* contains just one large cup-shaped chloroplast. For this reason, it seems unlikely that the entire chloroplast would differentiate into a chromoplast structure, as this could be detrimental to the cell. Nevertheless, crOR may increase the capacity for carotenoid storage in chloroplastidic organelles such as plastoglobules and this should be investigated further.

Expression of the His_6-tagged crOR protein was not detected in the parental strain CC-4533, and was tentatively detected in both transformant strains (**Figure 3.7**). Bands at ~43 kDa were present in the crORANGE transformants, albeit faintly for strain crOR-Mut-1. The predicted crOR protein size is 33.3 kDa, which is smaller than the detected band. It is, however, still likely that the detected protein is crOR; SDS micelles can bind to hydrophobic regions of transmembrane proteins, thus causing the protein to run aberrantly on SDS-PAGE gels and appear to be of a higher molecular weight than expected (Rath *et al*., 2009). An additional purification step prior to the Western blot, such as histidine-tagged protein purification by nickel affinity column, could be applied in future work to confirm crOR overexpression; this would separate natively expressed crOR from tagged crOR, enabling MS analysis to confirm the identity of the protein. Assuming the highlighted band in **Figure 3.6A** is crOR, the relative amounts of crOR protein expressed by strains crOR-Mut-1 and crOR-Mut-2 appears to reflect the carotenoid yields from each strain, in that crOR-Mut-2 exhibited significantly higher crOR expression than crOR-Mut-1 (**Figure 3.6B**), alongside higher carotenoid levels. This suggests a causal relationship between crOR expression and carotenoid abundance. A repetition of this study with several transformant lines could establish this relationship further. Moreover, the number of recombinant *crOR* integration sites in each strain should be determined, enabling further examination into the relationship between gene copy number, protein expression and carotenoid production.

The N-terminus of crOR was selected for His_6-tagging as it appeared to be the lesser of two evils; the C-terminal region of crOR is highly conserved across species and was therefore regarded as a potential active site with which a peptide tag could interfere (**Figure 3.1**), whereas the N-terminus contains a putative chloroplast signal transit peptide. This may have impeded crOR detection, as the

N-terminal signal is likely cleaved upon integration into the thylakoid membrane, thus only allowing detection of uncleaved, unlocalised protein. It could also be possible that the histidine tag blocked the signal peptide and caused mislocalisation of the protein. Morikawa *et al*. (2018) added a histidine tag to the C-terminus of crOR and obtained very similar carotenoid yields to those in this work; this suggests that neither tag (or less likely, both tags) had a detrimental effect on the function of crOR.

As well as a general increase in all carotenoids, the ratios of pViolaxanthin and lutein appear to be altered in transformant strains compared to the control (**Figure 3.8**). More pViolaxanthin than lutein was produced in the control strain, whereas the opposite was the case in both *crOR*-expressing strains, with more lutein being produced than pViolaxanthin. This may not in fact reflect an actual switch in metabolism from pViolaxanthin to lutein, as the value obtained for pViolaxanthin is only relative and not absolute; comparison of each of the carotenoids measured to their corresponding known standard would permit accurate quantification and comparison between carotenoids produced. Investigation into the protein and transcript levels of the carotenoid pathway in *crOR* overexpression strains could potentially reveal any alterations in metabolic flux.

Observations during ORANGE studies in higher plants suggest that the ORANGE protein acts as a post-translational regulator of PSY, a known rate-limiting enzyme in the plant carotenoid biosynthetic pathway, by improving its stability and increasing its availability in the chloroplast (Li *et al.,* 2012; Zhou *et al*., 2015). The overexpression of PSY cloned from other microalgal species in the *C. reinhardtii* nuclear genome led to an increase in lutein compared to wild-type: PSY from *D. salina* gave an increase of 2.6-fold, and PSY from *Chl. zofingiensis*, 2.2-fold (Couso *et al*., 2011; Cordero *et al*., 2011a). Given the clear gate-keeper role of PSY in carotenogenesis, and the potential role of crOR in the stabilisation of PSY, it could be a profitable next step to overexpress transgenic PSY and crOR simultaneously with the hope of creating a synergistic effect, driving metabolism towards increased carotenoid synthesis.

Further studies and improvements to *crOR*-expressing strains such as crOR-Mut-2 should also be pursued. An obvious next step for exploring the functions of crOR would be to perform physiological and biochemical studies using a *crOR* knock-out strain, such as strain LMJ.RY0402.222814_1 (Li *et al*., 2019). Determining the parameters under which crOR is active/necessary, and conducting a complementation assay, would enhance our understanding of the functions of crOR. Moreover, employing quantitative proteomics and computational analyses to examine the effects of CPL6 overexpression on metabolism could reveal unknown functions of crOR, and potentially identify

new targets for metabolic engineering *C. reinhardtii* for carotenoid production. Applying mutagenesis to strain crOR-Mut-2 and selecting for high-carotenoid producing strains could be another technique for strain enhancement.

In this chapter, CPL6, or crOR, has been demonstrated to be the *C. reinhardtii* ORANGE protein equivalent. Mutant crOR-Mut-2, with its particularly high production of lutein, could be a useful strain to take forward for further experimentation, and perhaps for industrial cultivation. Currently, to the author's knowledge, *C. reinhardtii* is not an industrial producer of carotenoids; with further optimisation and improvements perhaps crOR-Mut-2 could become the first *C. reinhardtii* strain to be employed by industry to synthesise lutein. Nevertheless, these insights into carotenoid metabolism within the model organism *C. reinhardtii* could be applied directly to other industrially relevant microalgae, thus improving the prospects for microalgal production of lutein.

Chapter 4: Development of a mutant selection workflow for improved carotenoid production with mutant characterisation using comparative shotgun proteomics

4.1. Summary

In this Chapter, a semi high-throughput reverse genetic engineering pipeline was established, in which 658 *C. reinhardtii* mutants were screened for increased carotenoid biosynthesis using a combination of a carotenoid biosynthesis inhibitor and strong light. This generated 5 mutant strains that produce significantly higher total carotenoids per cell than the parental strain. The most promising mutant line (EMS-Mut-5) produced 5.4-fold more lutein per cell than the wild-type strain. EMS-Mut-5 was phenotypically characterised using a label-free quantitative shotgun proteomics workflow, which revealed prospective metabolic engineering targets for augmenting carotenoid synthesis and strategies for optimising EMS-Mut-5 for maximal lutein production.

4.2. Introduction

In **Chapter 3**, *C. reinhardtii* was genetically modified to produce higher levels of lutein and β-carotene than the WT strain through overexpression of the putative carotenogenic enzyme regulator, crOR. Despite achieving a 2.0-fold increase in fg lutein cell^{-1} and ~6 mg lutein g^{-1} DCW, further increases in carotenoid levels would be required to enable *C. reinhardtii* to be competitive with other lutein-producing species for commercial lutein production (Fernández-Sevilla *et al*., 2010; **Section 1.3.5.**).

Random mutagenesis is a fast and effective method for generating strains with improved traits, and has successfully been applied to several microalgal species to enhance production of high-value carotenoids. Volumetric lutein production was increased 2.0-fold following chemical mutagenesis in *Chl. sorokiniana* (Cordero *et al*., 2011b), *D. tertiolecta* mutants produced 10–15% more zeaxanthin per cell than the WT (Kim *et al*., 2017), and chemical mutagenesis of *P. tricornutum* produced a mutant exhibiting 69.3% more fucoxanthin (Yi *et al*., 2018).

The random nature of mutagenesis also provides an opportunity to discover novel characteristics within metabolic pathways and their regulation, and possibly new targets for metabolic engineering. Another important advantage to using random mutagenesis for strain development is that strains generated are not subject to the same cultivation restrictions and regulations as targeted GMOs, at

least in the European Union (EU European Parliament and the Council of The European Union, 2001); this could open up access to health food markets as consumer scepticism about GMO products would be avoided, and the potential for large-scale cultivation (Beacham *et al*., 2017).

To the author's knowledge, random mutagenesis has not yet been attempted in *C. reinhardtii* with the goal of increasing carotenoid biosynthesis. Although *C. reinhardtii* is not credited as a high producer of lutein compared to other species (Cordero *et al*., 2011b; **Section 1.3.5.**), this model alga has the benefit of having fast growth, genetic tractability and several decades' worth of research and -omics data. Furthermore, the *C. reinhardtii* genome is haploid, meaning that all mutations are dominant, and *C. reinhardtii* can reproduce both asexually and sexually, enabling genetic crosses between strains exhibiting desirable characteristics *i.e.* selective breeding. Targets identified in *C. reinhardtii* could also potentially be applied to other algal strains, given its model status.

One of the pitfalls of metabolic engineering in *C. reinhardtii* is that it has notoriously stringent metabolic pathway regulation at multiple levels, including feedback inhibition (**Section 1.7.4.**; Kajikawa *et al*., 2015; Ravina *et al*., 2002; Vasileuskaya *et al*., 2005; Sun *et al*., 2010; Ramundo *et al*., 2013). Applying random mutagenesis to generate improved carotenoid-producing strains could circumvent such difficulties, and phenotypic characterisation of high-carotenoid strains could provide key insights into the regulation of carotenoid biosynthesis in *C. reinhardtii*, as well as potentially highlight new targets for genetic engineering that are less susceptible to regulation or feedback inhibition.

The development of an effective screening strategy is essential to isolate mutants with specific phenotypic traits. Enzymatic inhibitors that disrupt the pathway of a desired metabolite can be applied as a selective pressure following mutagenesis, where surviving mutants are likely to carry genetic changes that enable them to overcome the inhibition through increased synthesis of the desired product. Norflurazon is an inhibitor of PDS, a rate-limiting enzyme in the *C. reinhardtii* carotenoid pathway (highlighted in **Figure 1.4**). Post-mutagenesis exposure to sub-lethal concentrations of norflurazon to algal species has successfully generated *H. pluvialis* and *Chl. zofingiensis* mutants that exhibit increased astaxanthin production by 210% and 44%, respectively (Chen *et al*., 2003; Liu *et al*., 2010), and *Chl. sorokiniana* mutants that synthesise 2.0-fold increased lutein (Cordero *et al*., 2011b).

Comparative proteomics can be a useful tool to rapidly enable large-scale phenotyping of mutant strains, providing leads for ways to optimise product yield and highlighting targets for metabolic

engineering (Wang *et al*., 2012; Choi *et al*., 2013; Baek *et al*., 2016c; Sithtisarn *et al*., 2017). Proteomics has the advantage over other -omics techniques of directly revealing protein abundance; post-transcriptional regulation is rife within *C. reinhardtii*, and transcript abundance does not necessarily reflect protein abundance, particularly in the case of photosynthesis-related proteins (Lumbreras *et al*., 1998; Kong *et al*., 2015; Floris *et al*., 2013).

The work in this chapter aims to develop a semi high-throughput method for generating and isolating microalgal strains with improved carotenogenic properties, followed by phenotypic characterisation of high carotenoid-producing mutants by label-free quantitative (LFQ) shotgun proteomics. The workflow for this section is summarised in **Figure 4.1.**

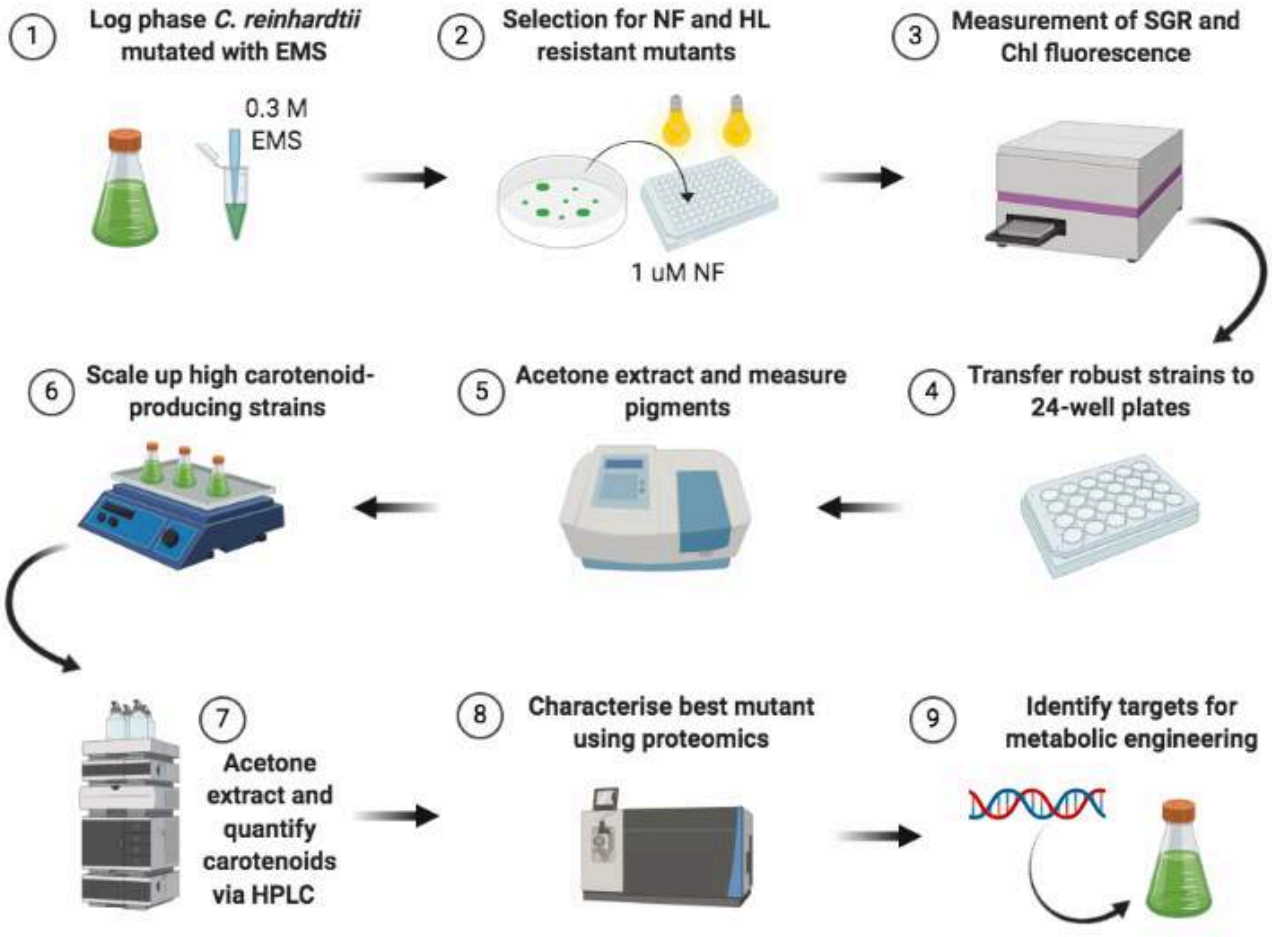


Figure 4.1: Schematic flow diagram depicting semi high-throughput process for generation, selection and characterisation of improved carotenoid-producing strains. EMS = ethyl methanesulfonate; NF = norflurazon; HL = high light; SGR = specific growth rate; Chl = chlorophyll. Image created using BioRender.

4.3. Results

4.3.1. Mutagenesis and screening

4.3.1.1. Determining the minimal inhibitory concentration of norflurazon for mutagenesis selection

Norflurazon resistant *C. reinhardtii* mutants have previously been generated (Liu *et al*., 2013; Suarez *et al*., 2014); however, the aim of these experiments was to find and study mutations in the PDS enzyme that confer norflurazon resistance. Increases in carotenoids were observed in strains with PDS mutations that are resistant to lethal levels of norflurazon, but the increases were small, at ~1.3-fold (Liu *et al*., 2013; Suarez *et al*., 2014). To avoid restricting mutations to PDS only, sub-lethal concentrations of norflurazon that confer some but not total inhibition of PDS were applied in this study. This way, novel mechanisms that increase carotenoid production more substantially could be revealed.

To find the sublethal concentration of norflurazon for *C. reinhardtii* mutant selection, strain CC-4533 was cultured in microtitre plates in increasing concentrations of norflurazon, and cell density measured daily using a microplate reader. **Figure 4.2** shows the growth curves of cultures grown in 6 concentrations of norflurazon. 0, 0.1 and 0.3 μM norflurazon appear to display similar growth traits, but with 0.3 μM norflurazon exhibiting a slightly lower cell density between Days 1 and 5. Norflurazon concentrations of 1 μM and above clearly have a negative effect on growth, as confirmed when the growth rates for each condition were compared statistically (**Table 4.1**). Concentrations surrounding 1 μM were therefore chosen as the sublethal concentrations for mutant selection.

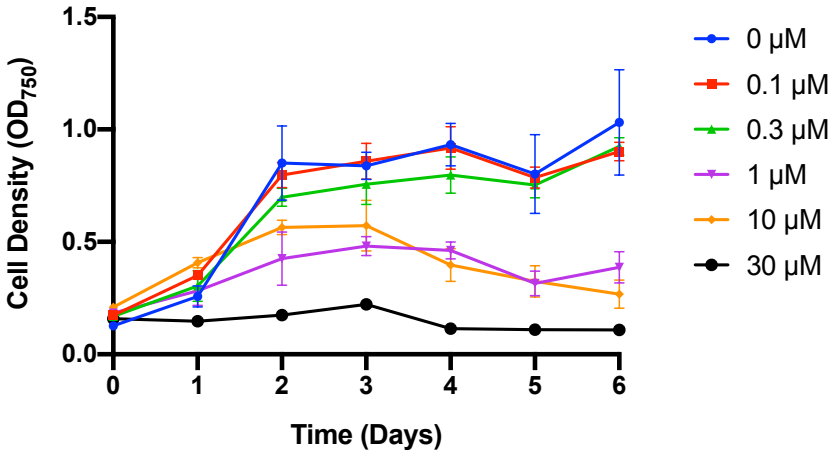


Figure 4.2: Growth of *C. reinhardtii* strain CC-4533 cultured in increasing concentrations of norflurazon. CC-4533 cultured in 96-well plates under standard conditions in triplicate, measured using a microplate reader. Error bars = SD. Error bars too small to be shown for the 30 μM norflurazon concentration.

Table 4.1: Growth measurements of CC-4533 grown in increasing concentrations of norflurazon

Norflurazon Concentration (μM)	Specific growth rate (SGR, h^{-1})	Doubling Time (h)
0	0.050 ± 0.007	14.13 ± 1.912
0.1	0.034 ± 0.001*	20.26 ± 0.7698
0.3	0.035 ± 0.009	20.54 ± 5.258
1	0.017 ± 0.005****	44.62 ± 14.34**
3	0.005 ± 0.005****	89.49 ± 22.05****
10	0.014 ± 0.001****	51.28 ± 3.756***
30	0.007 ± 0.0005****	100.3 ± 6.495****

Mean of calculated specific growth rates for 3 independent replicates (except for 3 μM where $n = 2$) shown ± SD. Asterisks show values that differ significantly from the mean of the control concentration of NF (0 μM), calculated via one-way ANOVA and Tukey's multiple comparisons test. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

4.3.1.2. Mutagenesis and round one of mutant selection for high carotenoid producing strains

Initial norflurazon experiments were carried out with *C. reinhardtii* strain CC-4533 due to its fast growth rate (see above), however the following experiments were completed using the cell wall-intact strain CC-125, so that mutant strains produced would be sufficiently robust for potential scale-up in larger bioreactors.

The combined effects of high light (> 800 μmol photons m^2 s^{-1}) and norflurazon (> 0.3 μM) have previously been shown to enhance their individual negative effects on growth (Fischer *et al*., 2010). Carotenoids play a vital role in protecting cells from strong irradiation, which damages cells through generation of singlet oxygen species (Niyogi *et al*., 1997; Baroli *et al*., 2003). Norflurazon, being an inhibitor of carotenoid biosynthesis, acts to block this protective mechanism, thus lowering cellular

defence against oxidation under saturating light. This phenomenon was exploited for mutant selection with the goal of combining high light and norflurazon to more rigorously select for carotenoid over-producing strains following mutagenesis.

Chemical mutagenesis has the advantage of potentially conferring mutations that improve the function of a protein or its expression through bp changes; insertional mutagenesis, while being more convenient for identifying the mutation site, is restricted to gene disruption only (**Section 1.6.3.**). EMS was selected as the chemical mutagen for this experiment. Guanine alkylation is the predominant mechanism of mutagenesis by EMS (Sega, 1984). The atypical base O^6-ethylguanine is generated through interaction of guanine with the ethyl group of EMS, which leads to the replacement of cytosine with thymine as the matching base for O^6-ethylguanine during DNA replication; the result is a point mutation, where GC pairs are replaced with AT. EMS mutagenesis has been applied to generate microalgal mutants with high levels of carotenoids (Yi *et al*., 2018; Kim *et al*., 2017), as well as to produce other types of metabolic mutants in *C. reinhardtii* (Loppes, 1968; McCarthy *et al*., 2004; Lee *et al*., 2014; Xie *et al*., 2014). UV mutagenesis is another popular method for mutant generation in *C. reinhardtii*, but due to equipment availability, chemical mutagenesis with EMS was selected.

Following previous methods for EMS mutagenesis of *C. reinhardtii* (Loppes, 1968; McCarthy *et al.,* 2004, Xie *et al*., 2014), CC-125 was exposed to 0.27 M and 0.3 M concentrations of EMS; this was followed by selection on TAP-agar plates supplemented with 1, 2 or 3 μM norflurazon. Colonies were picked from selective norflurazon-TAP agar plates into liquid culture on 96-well plates and grown for 5 days in either 0.5 μM (Plates A–D) or 1 μM (Plates E–H) norflurazon-TAP in high light (HL; 900–1200 μmol photons m^2 s^{-1}) conditions, then used to inoculate fresh 96-well plates containing 1 μM norflurazon-TAP which were grown under HL for another 5 days.

Chlorophyll fluorescence was measured daily using a plate reader to track growth; SGRs for each individual mutant were calculated from chlorophyll fluorescence measurements taken for each well. Chlorophyll fluorescence, as opposed to OD_{750}, was selected for growth calculations and screening selection based on previous work that correlated high chlorophyll abundance with high carotenoid levels in *D. salina* and *P. tricornutum*, and exploited this for selection of mutants with improved carotenoid content (Mendoza *et al*., 2008; Yi *et al*., 2018). This relationship was confirmed in the work from **Chapter 3**, where carotenoid content shows positive correlation with total chlorophyll (r^2 = 0.9145; **Appendix Figure C2**).

Figure 4.3 shows the screening criteria for the first mutant elimination step. For each plate, the mean SGR was calculated, and the mean + 1 SD used as the first constraint in the screen for viable mutants (**Figure 4.4**). Each plate was considered individually, due to variations in light intensity within the HL box, which ranged from 900–1200 μmol photons m^2 s^{-1}. The chlorophyll fluorescence readings for strains exhibiting SGRs > 1 SD from the mean were then considered (**Figure 4.5**); this was to eliminate cultures with seemingly high SGRs between Days 1 and 2, but poor overall growth as measured by comparatively low chlorophyll measurements after Day 2. Strains with chlorophyll fluorescence less than the average chlorophyll fluorescence of its respective plate measured on Day 4 were therefore discounted. Lastly, strains that exhibited particularly high chlorophyll fluorescence, with fluorescence > mean + 1 SD from the plate mean as measured on Day 4, were included in the catchment criteria for this screen. In total, 648 *C. reinhardtii* EMS mutants were screened in the initial stage, of which 144 mutants fit the criteria for the second round of screening (**Appendix Table C1**).

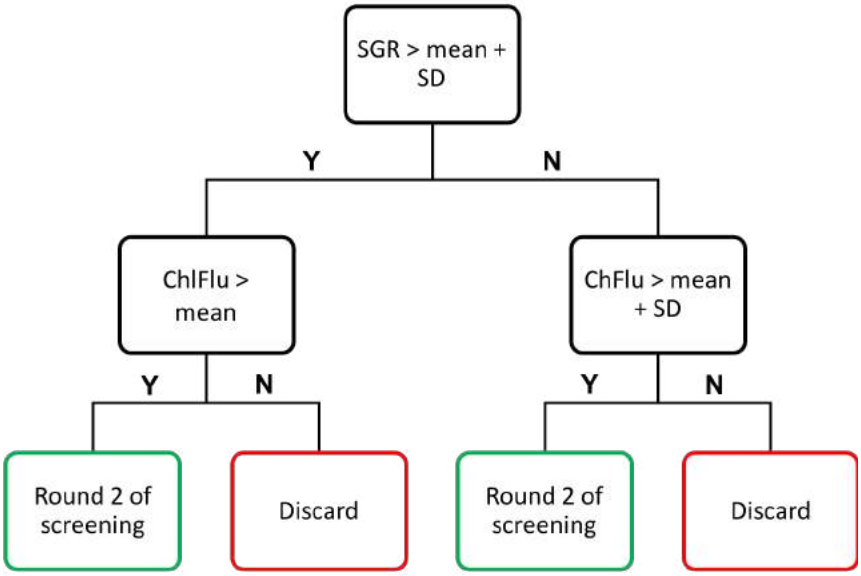


Figure 4.3: Decision tree for initial round of mutant screening. SGR = specific growth rate; mean = mean value for plate on which mutant was grown; SD = standard deviation; ChlFlu = chlorophyll fluorescence; Y = yes, N = no. Strains that fit the criteria for the green boxes were sub-cultured into 24-well plates for the second round of screening.

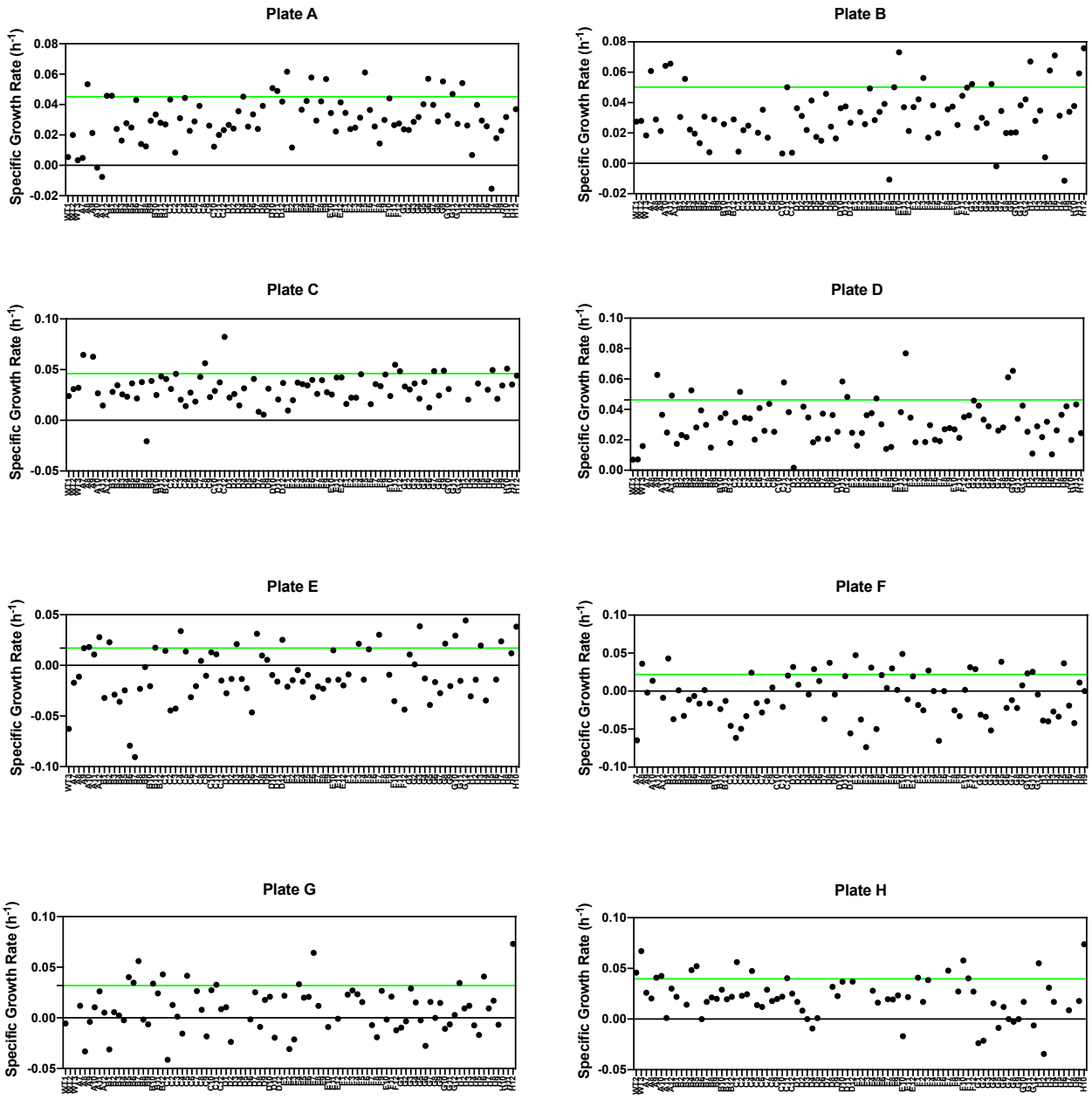


Figure 4.4: Specific growth rates of mutants grown for first round of selection. Specific growth rates were calculated from chlorophyll fluorescence measurements taken by plate reader for each well between Days 1 and 2. Each black dot represents an individual mutant strain in an individual well of a 96-well plate. Labels on x-axis correspond to wells on 96-well plate. Green line represents the mean growth rate of each plate + 1 SD. Mutants with growth rates above the green line were considered for the second round of selection (**Figure 4.3**).

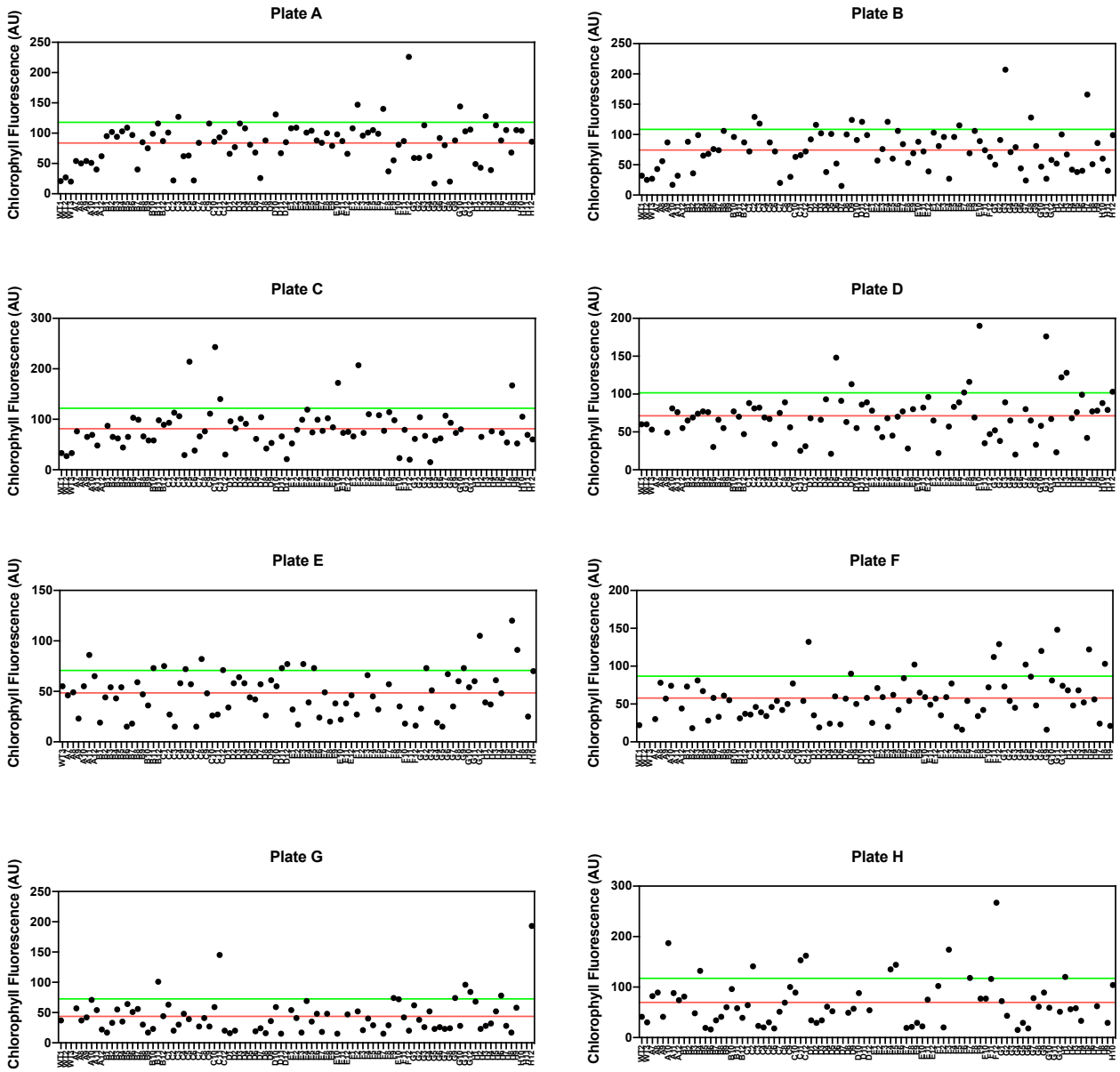


Figure 4.5: Chlorophyll fluorescence of mutant strains grown for first round of selection. Chlorophyll fluorescence measurements taken by plate reader after 4 days' growth with the following parameters: Ex440 nm, Em680 nm, gains 50. Each black dot represents an individual mutant strain. x-axis labels indicate the well position for each reading. Red line shows the mean chlorophyll fluorescence for each plate; green line shows mean chlorophyll fluorescence for each plate + 1 SD.

4.3.1.3. Round two of mutant selection for high carotenoid producing strains

Mutant strains that passed the initial elimination test (**Figure 4.3**) were sub-cultured in 24-well plates in non-selective TAP media to examine their total carotenoid content under standard conditions. After 84 h growth, 0.5 µL of each culture was collected for pigment extraction, followed by spectrophotometric total pigment analysis conducted using a microplate reader. **Figure 4.6**

shows the total carotenoids extracted for each individual mutant (full pigment analyses in **Appendix Figure C3**). After adjusting the values to be proportional to cell density (OD_{750}), 9 mutants had a total carotenoid content higher than the CC-125 control mean (**Figure 4.6**). The CC-125 control strain exhibited comparatively high pigment readings, which is likely due to a difference in initial growth conditions for this strain. Fresh CC-125 stock culture was used to inoculate the 24-well control plate for this experiment, whereas the mutant strain inoculates had been grown in norflurazon-treated media in 96-well plates, likely giving the mutants an initial disadvantage in growth and pigment production. Although this potentially added bias to the experiment, this screen was ultimately developed to identify high carotenoid producers, so this CC-125 mean was still used for the mutant selection criteria and considered to add stringency to the screen; the 9 norflurazon-burdened mutants that could outcompete the healthier WT strain are likely able to produce even more carotenoids in more amenable conditions, therefore they were selected for batch culture growth and HPLC analysis.

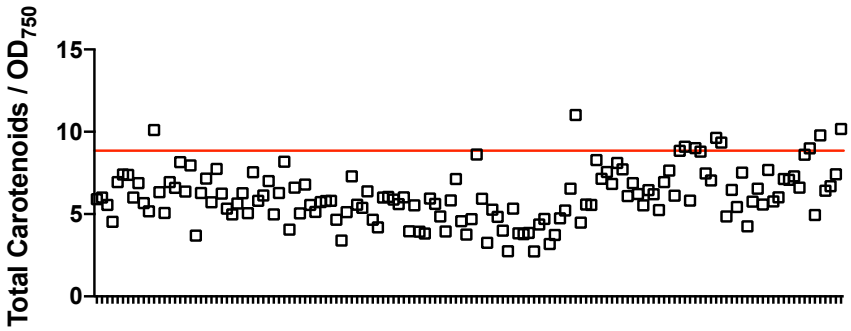


Figure 4.6: Total carotenoid content of 144 mutant *C. reinhardtii* strains adjusted to OD_{750}. Total carotenoids as calculated following pigment extraction in acetone and subsequent spectrophotometer analysis (See **Section 2.6.2.** for calculations). Total carotenoid content adjusted to cell density at OD_{750}. Each square represents an individual mutant strain. Red line shows mean value for control strain CC-125.

4.3.2. Growth and pigment analysis of selected CC-125 mutants

4.3.2.1. Growth analysis of mutant strains

The 9 mutants identified in the previous screen plus WT CC-125 were scaled up to 25 mL batch cultures in conical flasks and harvested after 96 h. **Figure 4.7** shows the growth rates and doubling times for each of the 9 mutant strains plus CC-125. EMS-Mut-1 appeared to grow the fastest, followed by EMS-Mut-9 and EMS-Mut-3; however, after comparing the growth rates for each of the

mutant strains with CC-125, no significant difference was observed (**Figure 4.7**). Significant differences were present between the final cell densities at the end of the 96 h growth period; strains EMS-Mut-1, -2, -5 and -6 exhibited significantly lower cell differences compared to the CC-125 control, while EMS-Mut-9 was significantly more dense, exhibiting the highest cell density at the end of the period.

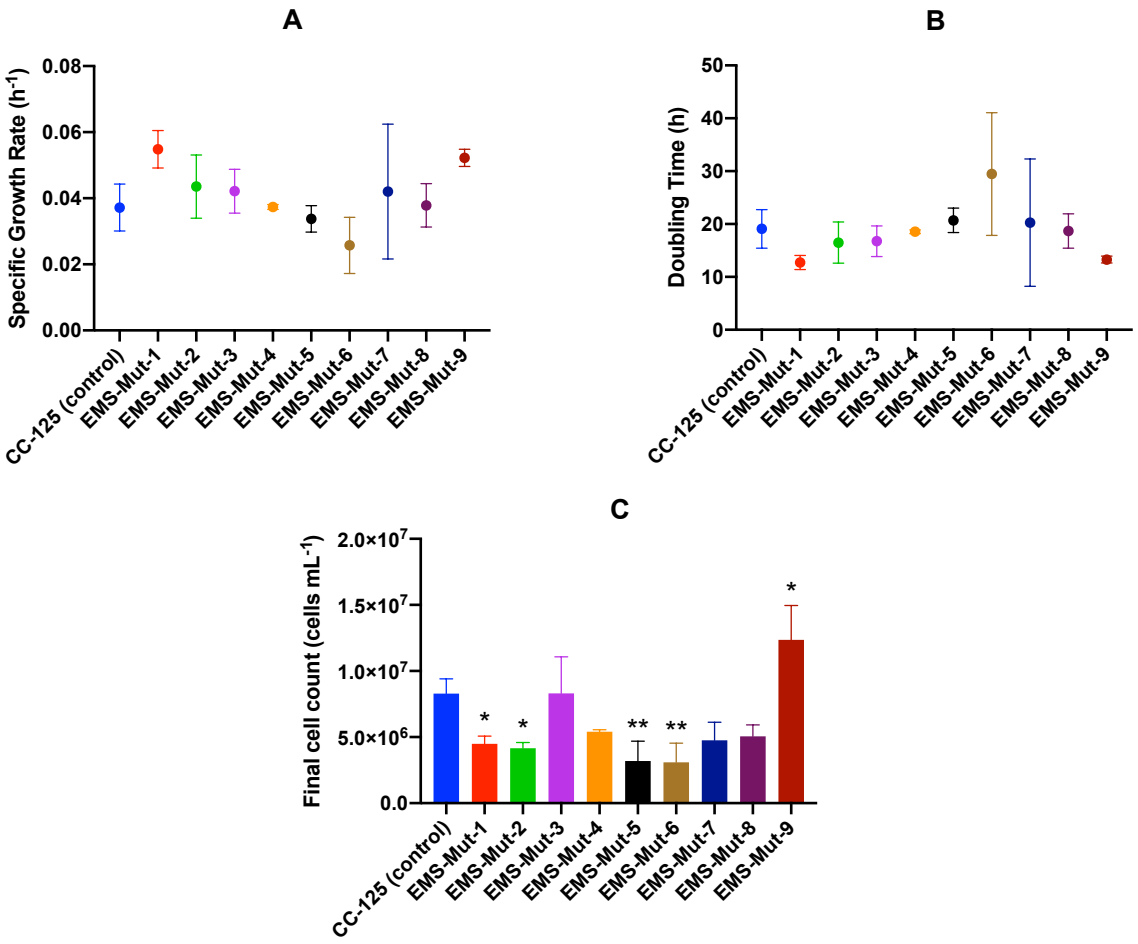


Figure 4.7: Growth rates for 9 EMS mutants and CC-125 control strain. A – Specific growth rate. **B** – Doubling time. Calculated from cell count data taken between 48 and 96 h. No significant difference between EMS mutants and the control strain was found for the growth rates shown in **A** or **B** ($P > 0.05$ for one-way ANOVA). **C** – Final cell density after 96 h growth. Significant difference from the mean cell density of the CC-125 control was calculated using one-way ANOVA and a Dunnett's multiple comparisons test. *$P < 0.05$, **$P < 0.01$, n = 3.

4.3.2.2. Pigment analysis of mutants by spectrophotometry

Pigments were extracted from each strain with 100% acetone after 4 days; wavelength scans for each strain are shown in **Appendix Figure C4**. 5 out of the 9 mutant strains, EMS-Mut-2, -4, -5, -6 and -7, harboured significantly more total carotenoids per cell than CC-125, the highest being EMS-Mut-5 which produced 3.6-fold higher total carotenoids (**Figure 4.8**). EMS-Mut-5 also had 3-fold more chlorophyll *a* and *b* than CC-125. The chlorophyll-*a* / chlorophyll-*b* ratios were similar across strains, except for strains EMS-Mut-8 and -9. The carotenoid-to-chlorophyll ratios were significantly higher than CC-125 in all mutant strains besides EMS-Mut-5 and -6. These combined results suggest that the screening method was successful in identifying mutants with significantly higher total carotenoids than the basal strain.

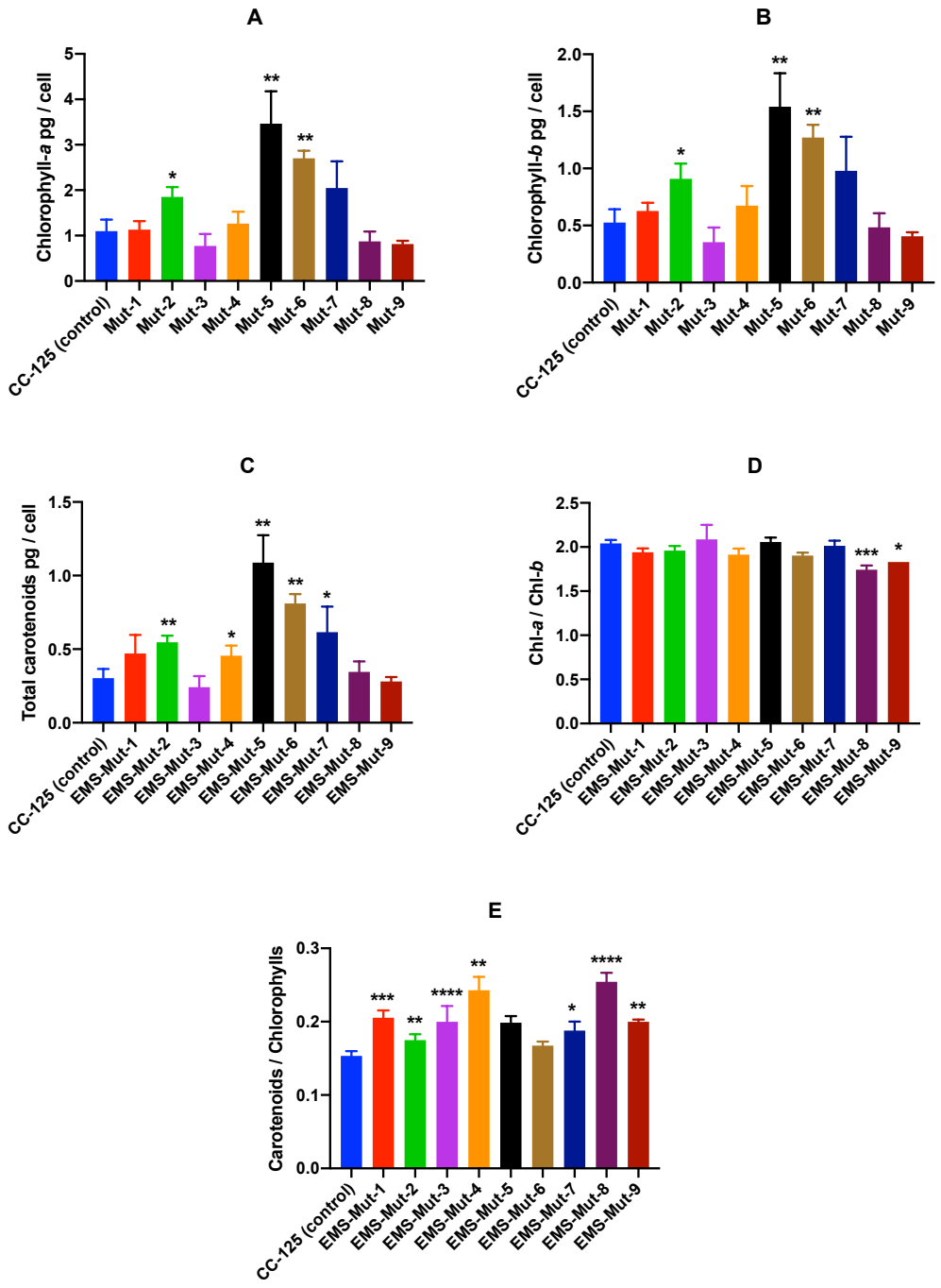
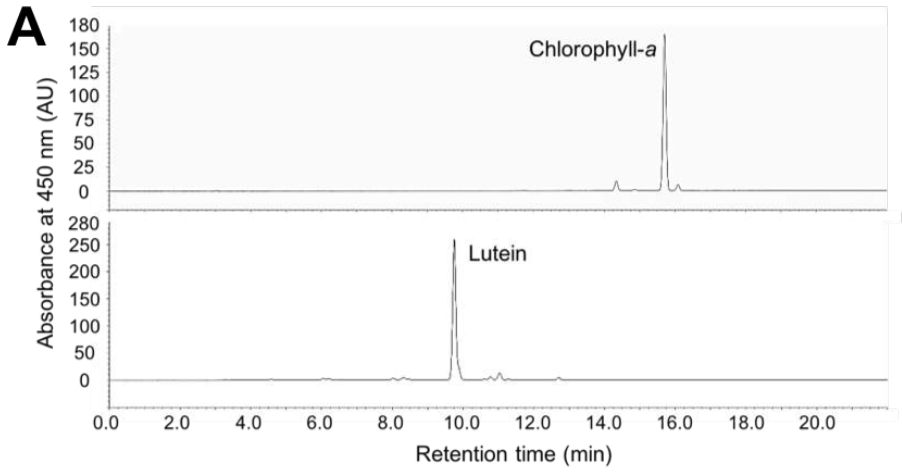


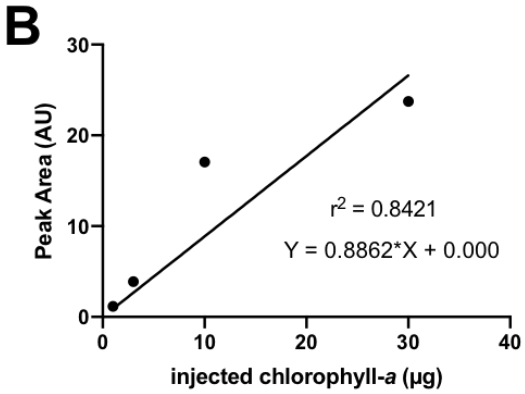
Figure 4.8: Pigment amounts and ratios of strains CC-125 (control) and EMS mutants 1–9 measured by spectrophotometer. Total chlorophyll and carotenoid amounts were calculated using the spectrophotometric assay described in **Section 2.6.2**., then normalised to their respective cell number to obtain values in pg / cell (**A**, **B** and **C**). The relative ratios of chlorophylls *a* to *b* were calculated (**D**). Carotenoid/ chlorophyll ratios were calculated in **E**. Statistically significant samples calculated using student's t-tests. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001, n = 3.

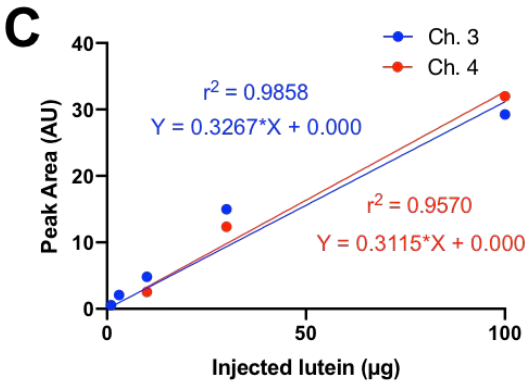
4.3.2.3. Pigment analysis of mutants by HPLC

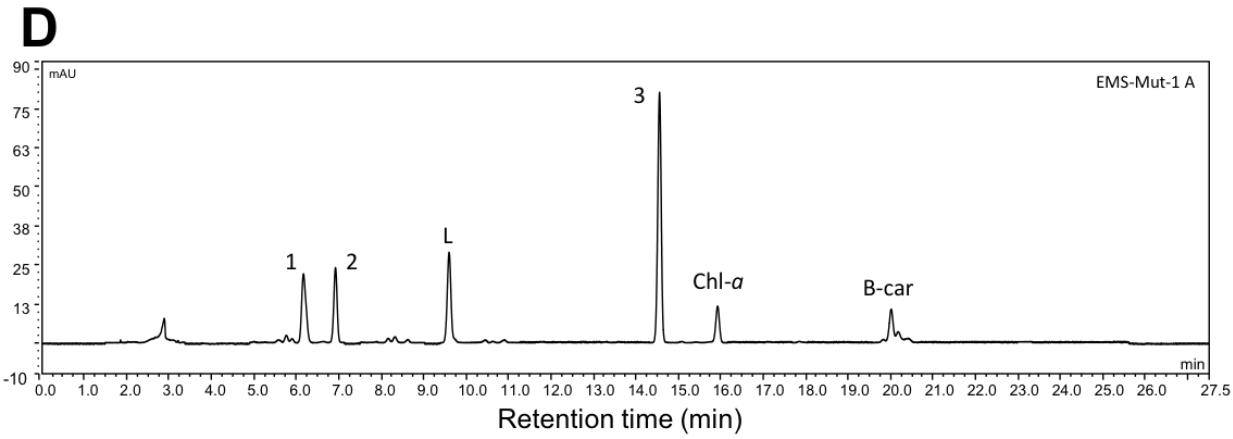
The pigment extractions examined in **Figure 4.8** were analysed by HPLC for separation, identification and quantification. Pigment standards for lutein, β-carotene and chlorophyll-*a* (**Figure 3.7; Figure 4.9**) were used to identify their respective peaks within spectra obtained for each strain; other peaks within the spectra were tentatively identified using previously published work in which the same HPLC method was applied (See **Figure 3.7**). **Figure 4.9D** shows an example chromatogram obtained following HPLC separation of the pigment extracts. The elution profiles for wild-type strains CC-4533 and CC-125, as well as their respective mutants, were similar (**Figure 3.7**): pigment 1 (putative neoxanthin; pNeoxanthin) eluted first, followed by pigment 2 (putative violaxanthin; pViolaxanthin), lutein, pigment 3 (putative chlorophyll-*b*; pChlorophyll-*b*), chlorophyll-*a* and lastly, β-carotene. EMS-Mut-5 produced the most of each carotenoid per 10^6 cells than all other strains, followed by EMS-Mut-6 (**Figure 4.9E**); this result complies with the spectrophotometric analysis (**Figure 4.8**). Compared to CC-125, EMS-Mut-5 and -6 produced 5.4 and 3.1-fold more lutein, and 3.1 and 2.7-fold more β-carotene, respectively. EMS-Mut-6 produced ~3-fold more of each carotenoid than the control, and EMS-Mut-2, -4 and -7 ~2-fold more, with pViolaxanthin being the carotenoid exhibiting the highest increase in these mutants (**Figure 4.9E**). EMS-Mut-5 had a disproportionately large increase in lutein, but more modest -fold increases in other pigments (**Appendix Table C6**), suggesting that an alternative, lutein-specific mechanism is responsible for the observed increase in carotenoids in EMS-Mut-5.

A calibration curve (**Figure 4.9C**) was generated from known quantities of lutein standard to quantify lutein produced by each EMS mutant. Lutein was measured volumetrically (mg lutein mL^{-1}), cellularly (fg lutein $cell^{-1}$) and mg lutein per g dry biomass (mg lutein g^{-1} DCW; **Figure 4.10**). EMS-Mut-5 produced significantly more lutein than the control strain in all 3 analyses (**Figure 4.10**). 7 out of the 9 strains produced significantly more lutein per g of dried biomass than the control strain, although the comparative increases were less dramatic per g DCW than per cell. Overall, these results validate the method developed in this chapter for generating *C. reinhardtii* strains with high-carotenoid levels, with the potential to generate high-lutein phenotypes.


A
Chlorophyll-a
Lutein
Absorbance at 450 nm (AU)
Retention time (min)


B
Peak Area (AU)
r2 = 0.8421
Y = 0.8862*X + 0.000
injected chlorophyll-a (µg)


C
Ch. 3
Ch. 4
r2 = 0.9858
Y = 0.3267*X + 0.000
r2 = 0.9570
Y = 0.3115*X + 0.000
Peak Area (AU)
Injected lutein (µg)


D
mAU
EMS-Mut-1 A
1
2
L
3
Chl-a
B-car
min
Retention time (min)

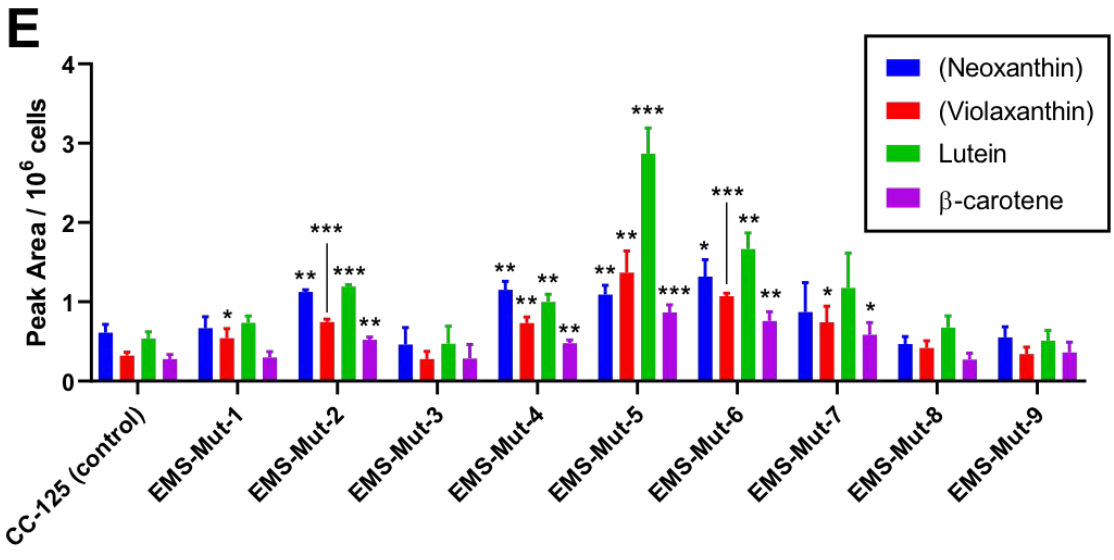

E
(Neoxanthin)
(Violaxanthin)
Lutein
β-carotene
Peak Area / 10^6 cells
CC-125 (control)
EMS-Mut-1
EMS-Mut-2
EMS-Mut-3
EMS-Mut-4
EMS-Mut-5
EMS-Mut-6
EMS-Mut-7
EMS-Mut-8
EMS-Mut-9

Figure 4.9: Pigment standards and mutant profiles measured by HPLC. A – HPLC chromatograms showing chlorophyll-*a* and lutein pigment standards measured at absorbance 450 nm. **B –** Chlorophyll-*a* standard curve. **C** – Lutein standard curves used for the HPLC analysis in **Chapter 3** (Ch. 3, blue) and in **Chapter 4** (Ch. 4, red). Following a comparison of the two slopes, they were not significantly different from one another (*P* = 0.06). **D** – Example absorbance spectrum at 450 nm for HPLC of *C. reinhardtii* strains, taken from EMS-Mut-1 biological replicate A. L = lutein, Chl-*a* = chlorophyll-*a*, B-car = β-carotene. All chromatograms shown in **Appendix Figure C6**. **E** – Peak area per 10^7 cells for carotenoids detected by HPLC for each mutant strain. Peak area calculated as follows: (Peak Area / cell number) x 10^6. Error bars = SD. Carotenoids in brackets have been tentatively identified using previously published work, as opposed to analytical standards. Significant differences from the CC-125 control mean for each carotenoid were calculated using a one-way ANOVA. **P* < 0.05, ***P* < 0.01, ****P* < 0.001, n = 3. Error bars = SD. Peak areas and retention times can be found in **Appendix C**.

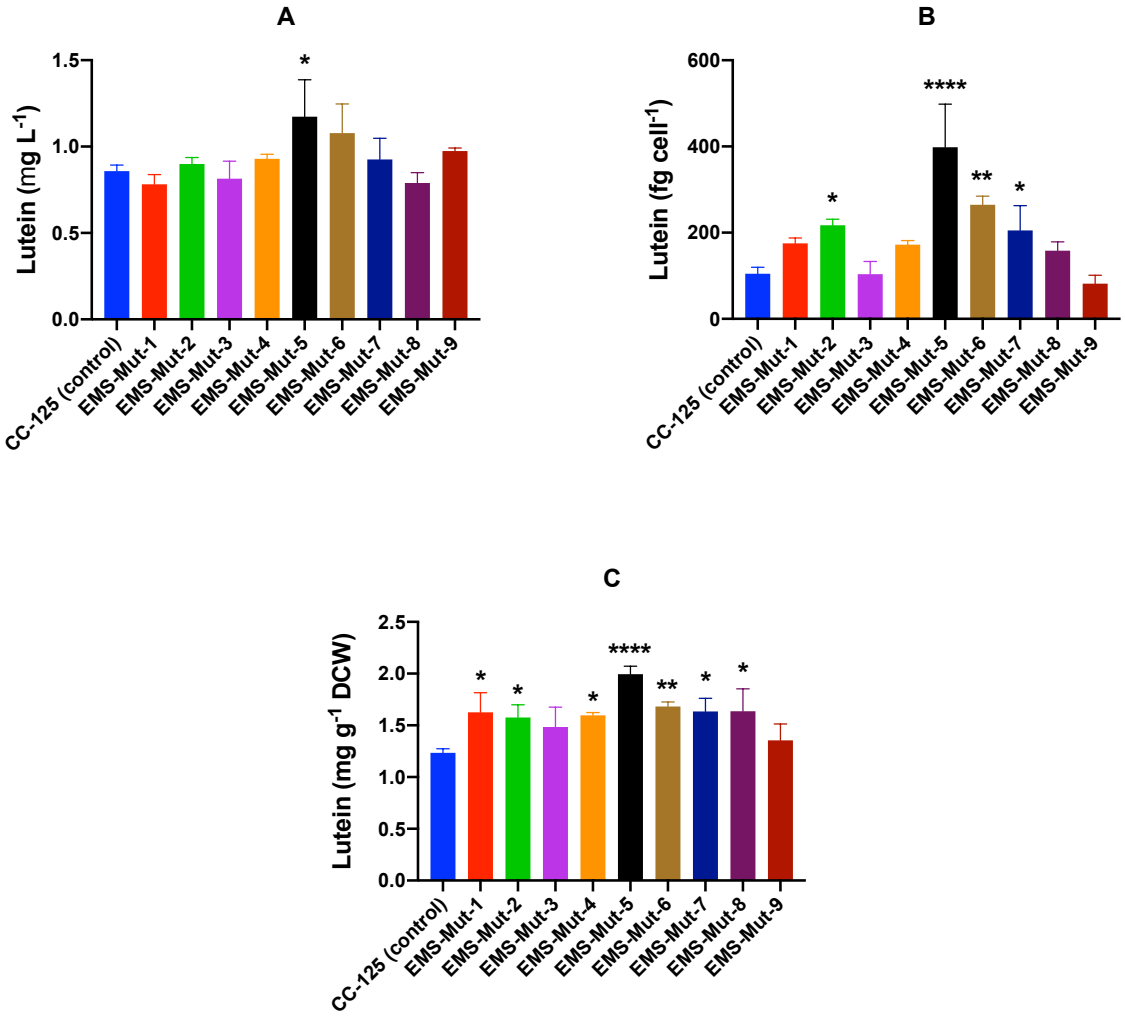


Figure 4.10: Lutein contents of CC-125 and EMS mutant strains. Lutein contents for CC-125 and EMS mutants expressed as mg L^{-1} culture (**A**), fg per cell (**B**) and as mg g^{-1} DCW (**C**). Significant differences from the CC-125 control mean were calculated via one-way ANOVA. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, n = 3. Error bars = SD.

4.3.3. Characterisation of EMS-Mut-5 by quantitative shotgun proteomics

EMS-Mut-5 produced 5-fold more lutein than the CC-125 control strain, and the highest levels of total carotenoids of the mutants generated in this work (**Figure 4.9, 4.10**). EMS-Mut-5 also appeared to exhibit different carotenoid ratios to each of the other strains generated (**Figure 4.10**), suggesting it could carry a novel mutation of interest. EMS-Mut-5 was selected for further characterisation by proteomics, with the aim of uncovering the metabolic mechanisms behind its superior lutein production.

4.3.3.1. Time point selection for proteomics

An 8-day growth and pigment study of CC-125 and EMS-Mut-5 was conducted to determine the optimal time point at which to harvest each strain. EMS-Mut-6 was included in the study, as it produced the second-highest amount of lutein and total carotenoids (**Figure 4.9E, 4.10**), and could be used for comparison. **Figure 4.11A** shows the growth curves for CC-125, EMS-Mut-5 and EMS-Mut-6, which reached stationary phase at Days 4, 5 and 6, respectively. The growth rates of EMS-Mut-5 and EMS-Mut-6 were lower than CC-125, however this was not statistically significant (**Table 4.2**). Total pigments were measured daily from Day 3 onwards using the spectrophotometric method applied previously (**Figure 4.11)**. EMS-Mut-5 consistently produced more carotenoids per cell than the control strain throughout the growth study (**Figure 4.11**). Despite the lower cell count, both the carotenoid and chlorophyll levels of EMS-Mut-5 surpassed CC-125 after Day 5 of growth. EMS-Mut-6 displayed a faster growth rate but lower carotenoid and chlorophylls than EMS-Mut-5. Pigment levels for EMS-Mut-6 were higher than in WT. Chlorophyll-*a*/ *b* and carotenoid-to-chlorophyll ratios were similar in all strains tested.

Table 4.2: Growth rates of candidate strains for proteomics

Strain	Specific growth rate (SGR, h^{-1})	Doubling Time (h)
CC-125 (Control)	0.0387 ± 0.0092	18.7 ± 4.86

EMS-Mut-5	0.0272 ± 0.0042	25.9 ± 3.77
EMS-Mut-6	0.0330 ± 0.0018	21.1 ± 1.09

To 3 significant figures. Values ± SD. Differences between SGR and doubling times were not significant, as calculated using student's t-test ($P < 0.05$).

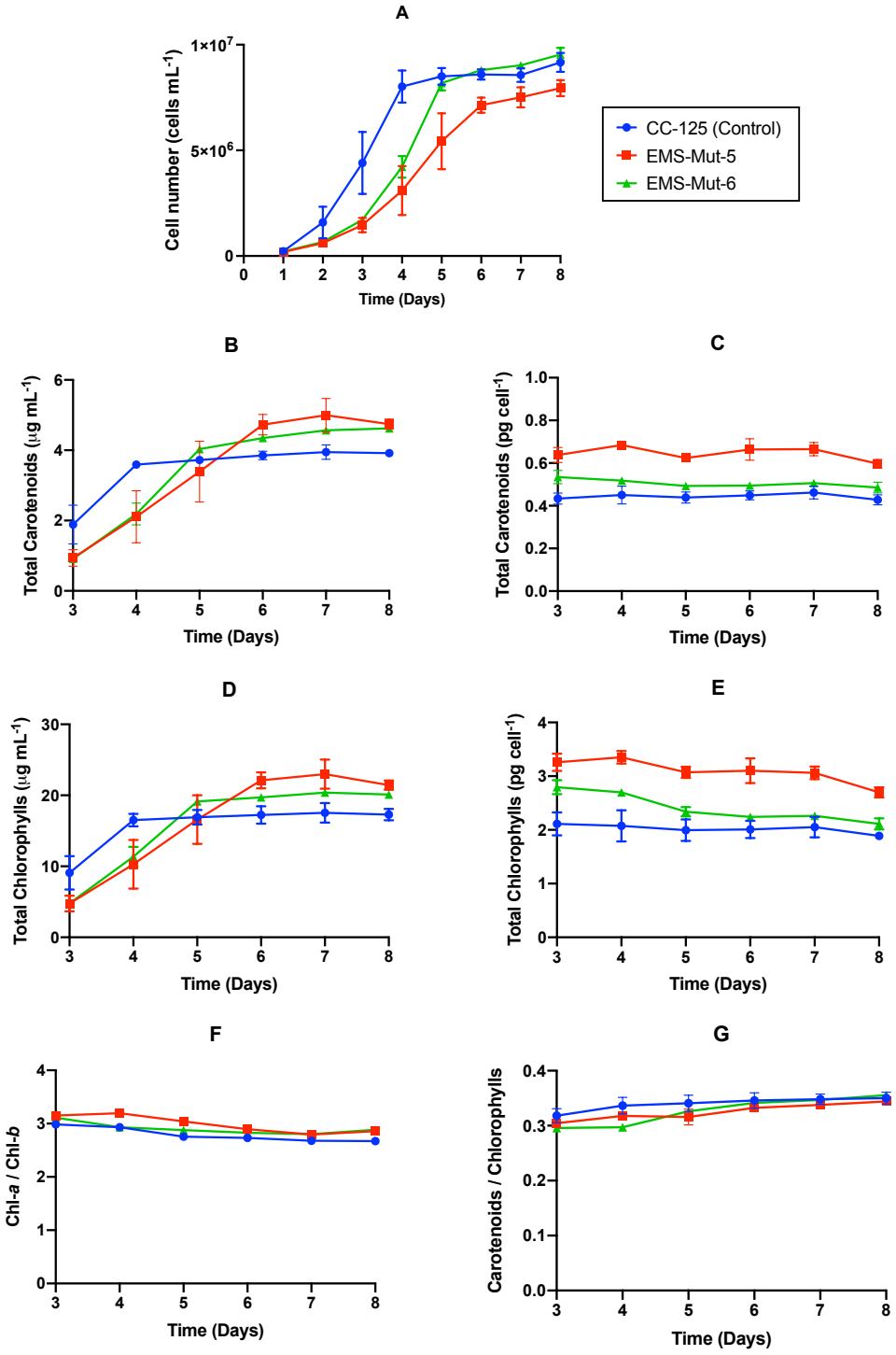
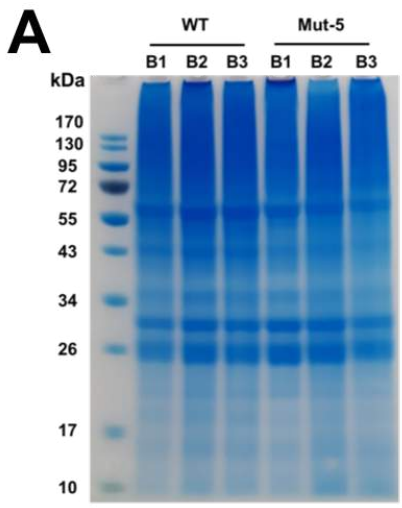


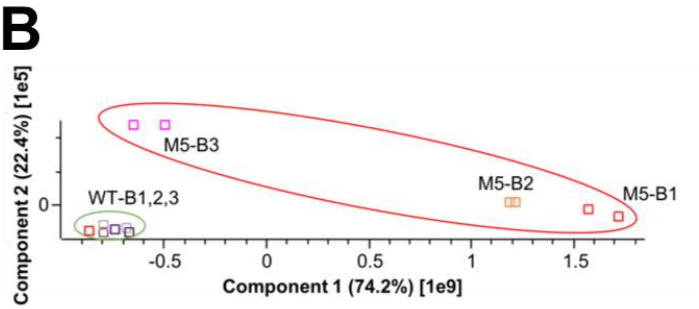
Figure 4.11: Growth and pigment study of candidate strains for proteomic analysis. A – Growth curves for CC-125, EMS-Mut-5 and EMS-Mut-6. Cell number values obtained via calibration curve from OD_{750} measurements (**Appendix Figure C6**). **B** – µg total carotenoids per mL culture per day. **C** – pg total carotenoids per cell per day. **D** – µg total chlorophyll (chlorophyll-*a* + chlorophyll-*b*) per mL culture. **E** – pg total chlorophyll per cell per day. **F** – Chlorophyll-*a* to chlorophyll-*b* ratio (pmol) per day. **G** – carotenoids-to-chlorophylls ratio (pmol) per day. Total pigments calculated following 100% acetone extraction and measurement by spectrophotometer. Error bars = SD.

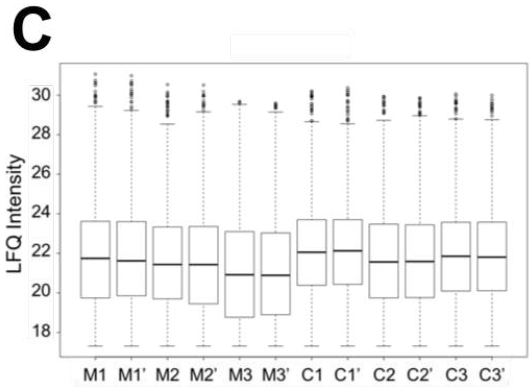
Given this data, the control strain CC-125 was harvested at Day 4, and EMS-Mut-5 harvested at Day 6 for the proteomics experiment (See **Appendix Figure C7** for growth curves). These time points were selected as they represent late-log/ early stationary phase for both strains, with no significant difference in cells mL^{-1}. The carotenoids per cell for both strains remains relatively stable across all time points, only dropping off after Day 7, suggesting that enzymes involved in pigment production and storage would continue to be expressed at the time points selected.

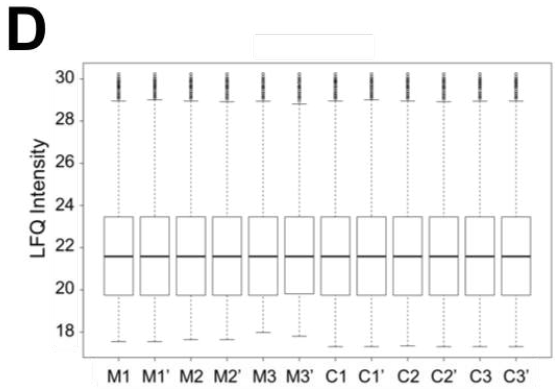
4.3.3.2. Protein quantification and statistical analysis

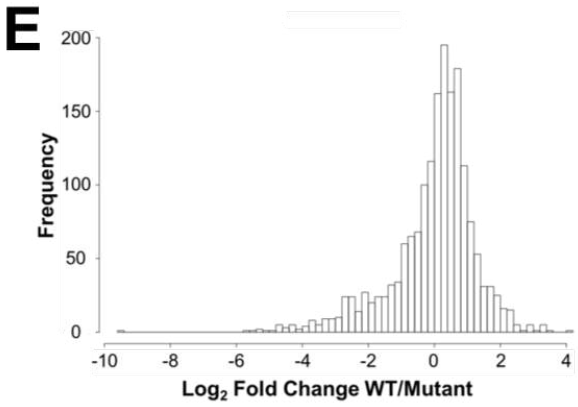
Following WT and EMS-Mut-5 culturing in triplicate, proteins were extracted (**Figure 4.12A**) and prepared for LC-MS/ MS analysis. A LFQ shotgun proteomics approach was applied to compare EMS-Mut-5 to WT. LFQ proteomics is a fast and cost-effective technique enabling rapid comparative global analysis of two phenotypes or conditions (Wang *et al*., 2012). Recent advances in HPLC resolution, MS mass accuracy, and bioinformatic tools allow complex samples to be analysed directly by nanoflow LC-MS/ MS without labelling or multiple fractionation steps (Cox *et al*., 2014). Although label-free approaches can have the disadvantage of being less sensitive to low abundance proteins (Zhu *et* al., 2010), it is possible to capture more peptides with LFQ than with iTRAQ, potentially leading to a more in-depth study while avoiding the use of expensive regents (Wang *et al*., 2012). The raw MS data files were processed using the label-free quantification option in MaxQuant software (MaxLFQ), which matched MS peaks to peptides in the *C. reinhardtii* proteome using the Andromeda search engine, and the LFQ intensities for identified peptides were calculated (Cox *et al*., 2014). As part of the MaxLFQ analysis, common contaminants and non-unique/ razor peptides were filtered out. **Appendix Table C8** contains a summary of the results from MaxLFQ.

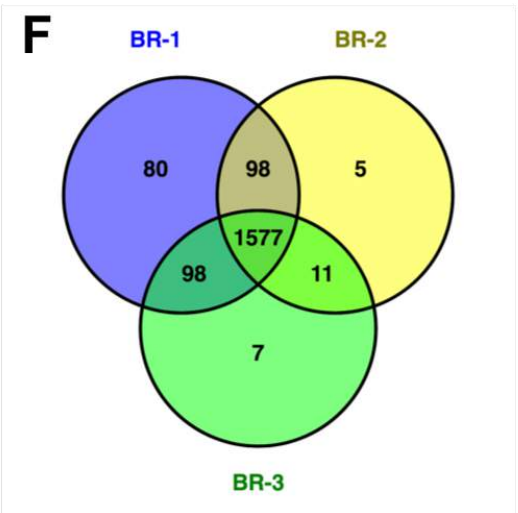

A
WT
Mut-5
B1 B2 B3 B1 B2 B3
kDa
170
130
95
72
55
43
34
26
17
10


B
M5-B3
M5-B2
M5-B1
WT-B1,2,3
Component 2 (22.4%) [1e5]
Component 1 (74.2%) [1e9]


C
LFQ Intensity
M1 M1' M2 M2' M3 M3' C1 C1' C2 C2' C3 C3'


D
LFQ Intensity
M1 M1' M2 M2' M3 M3' C1 C1' C2 C2' C3 C3'


E
Frequency
Log2 Fold Change WT/Mutant


F
BR-1
BR-2
BR-3
80
98
5
1577
98
11
7

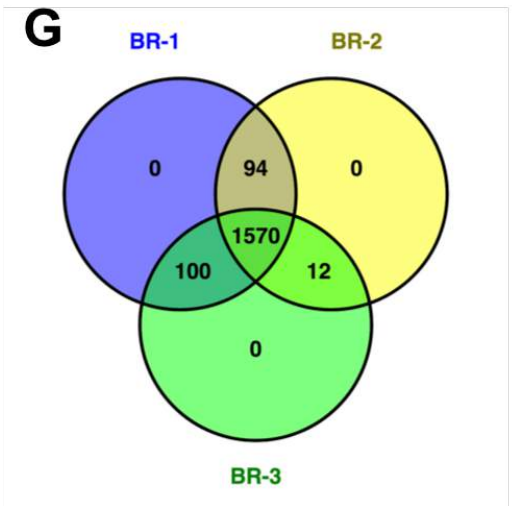

G
BR-1
BR-2
BR-3
0
94
0
1570
100
12
0

Figure 4.12: Proteomics data analysis and quality control. A – SDS-PAGE gel showing ~50 µg 2D cleaned-up protein extracted from CC-125 (WT) and EMS-Mut-5 (Mut-5). Biological replicates are numbered B1–3. **B** – Principal component analysis comparing the proteomes of WT and EMS-Mut-5. Samples are clustered according to strain: CC-125 (WT), green; EMS-Mut-5 (M5), red. Biological replicates are numbered B1–3; technical replicates of the same biological replicate are coloured similarly. The MaxQuant output files Proteingroups.txt and evidence.txt were uploaded to Perseus to perform the PCA. **C** – Boxplots showing median LFQ intensity data for proteins in each sample before quantile normalisation using Proteosign. Samples are labelled as follows: 'M' for mutant, 'C' for control; numbers represent biological sample; apostrophes represent technical replicates. Central line represents the median. **D** – Boxplots showing median LFQ intensity data for proteins in each sample after quantile normalisation using Proteosign. Samples are labelled as follows: 'M' for mutant, 'C' for control; numbers represent biological sample; apostrophes represent technical replicates. Central line represents the median. **E** – Histogram showing the frequencies of Log2 fold changes in protein intensities in CC-125 (WT) compared to EMS-Mut-5 (mutant). Generated using Proteosign. **F**, **G** – Venn diagrams showing identified (**F**) and quantified (**G**) proteins for each biological replicate (BR) in both the WT and EMS-Mut-5 strains. Data for Venn diagrams was generated by ProteoSign.

A PCA of the MaxLFQ proteomics data was performed using Perseus software (**Figure 4.12B**). All WT replicates, both technical and biological, clustered together closely. EMS-Mut-5 biological replicates 1 and 2 (M5-B1 and M5-B2) clustered together away from the WT samples, indicating that their proteomic profiles were similar to each other but distinct from those of the WT. EMS-Mut-5 biological replicate 3 (M5-B3) clustered away from both the WT and the two other biological replicates of EMS-Mut-5; the M5-B3 biological replicate was still clearly different to the WT replicates, and was hence included in downstream analyses.

Statistical analysis of the quantified proteins was carried out using ProteoSign, an open source platform that performs differential expression/ abundance analysis of LFQ proteomics data using the Linear Models for Microarray data (LIMMA) statistical methodology (Efstathiou *et al*., 2017). Of the 1876 proteins identified by the MaxLFQ analysis, 98 proteins were filtered out (not identified in at least 2 biological replicates), leaving 1776 quantifiable proteins. The LFQ intensities calculated for each protein were Log_2 transformed within the Proteosign software, and quantile normalised; LFQ intensities before and after quantile normalisation can be seen as boxplots in **Figures 4.12C** and **4.12D**. Following normalisation, the WT/EMS-Mut-5 Log_2 fold change was calculated for each protein. The histogram shown in **Figure 4.12E** shows the distribution of the WT/EMS-Mut-5 Log_2 fold change values; the Log_2 fold changes tended to cluster around 0 with a slightly positive centre

of the distribution, indicating a higher number of upregulated proteins in the WT strain compared to EMS-Mut-5 (or downregulated proteins in EMS-Mut-5 compared to the WT). Fold changes were present with Log_2 values far above and below 0, suggesting that several proteins were differentially regulated in the mutant strain relative to the WT.

Biological replicates 1–3 of EMS-Mut-5 and WT were compared in the LIMMA analysis conducted by the Proteosign software. **Figure 4.12F** shows the number for proteins identified in each of the biological replicates of WT and EMS-Mut-5; a total of 1577 of the 1776 proteins were common to all biological replicates. **Figure 4.13B** shows the number of quantified proteins for each biological replicate; this indicates that proteins found only in one biological replicate were filtered out of the analysis. Of the 1776 proteins, 1570 were quantified in all biological replicates.

To examine whether the WT/EMS-Mut-5 Log_2 fold changes observed in **Figure 4.12E** were statistically significant, *P*-values were calculated for each protein within Proteosign and plotted against Log_2 fold change (**Figure 4.13**). Of the 1776 quantifiable proteins, 960 (~50%) were found to be significantly differentially expressed in EMS-Mut-5 compared to the WT ($P < 0.05$), with 393 upregulated and 567 downregulated in EMS-Mut-5 compared to WT (full lists shown in **Appendix Tables C9** and **C10**). **Figure 4.13** shows the volcano plot generated from the ProteoSign analysis, where each point represents a quantifiable protein; the proteins of relative increased abundance in EMS-Mut-5 are fewer, but their fold-change (Log_2) appears to be higher overall compared to the Log_2 of proteins with decreased abundance. When more stringent trimming of the protein lists was applied to include only those with $Log_2 > 1$ (or fold-change > 2) and *P*-values < 0.01, 226 proteins displayed increased expression and 172 decreased (**Appendix Tables C9** and **C10**).

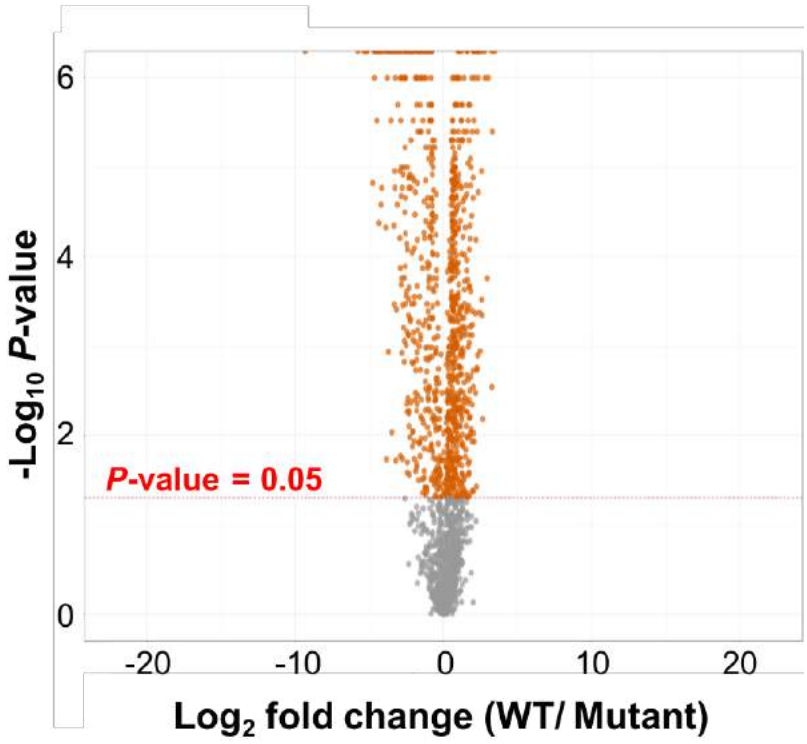


Figure 4.13: Volcano plot showing *P*-values vs Log_2 fold change of quantified proteins. Log_2 is presented as WT/mutant. Differentially regulated proteins ($P < 0.05$) are depicted in orange; grey dots represent proteins that are not differentially regulated in EMS-Mut-5. Protein lists for the rest of this chapter will be presented in terms of mutant/ WT.

4.3.3.3. Metabolic pathway analysis and protein functional enrichment

To explore global metabolic changes in EMS-Mut-5 compared to the WT strain, metabolic mapping was conducted using KEGG pathway analysis. 605 proteins could be assigned to KEGG IDs, and 73 proteins were mapped to *C. reinhardtii* metabolism, as shown in **Figure 4.14**. Several key pathways appear to be downregulated, including carbon fixation, oxidative phosphorylation, glycolysis/ gluconeogenesis, glycoxylate and dicarboxylate metabolism, and ribosomal subunits (See **Appendix Figure C9** for figures of differentially regulated KEGG pathways).

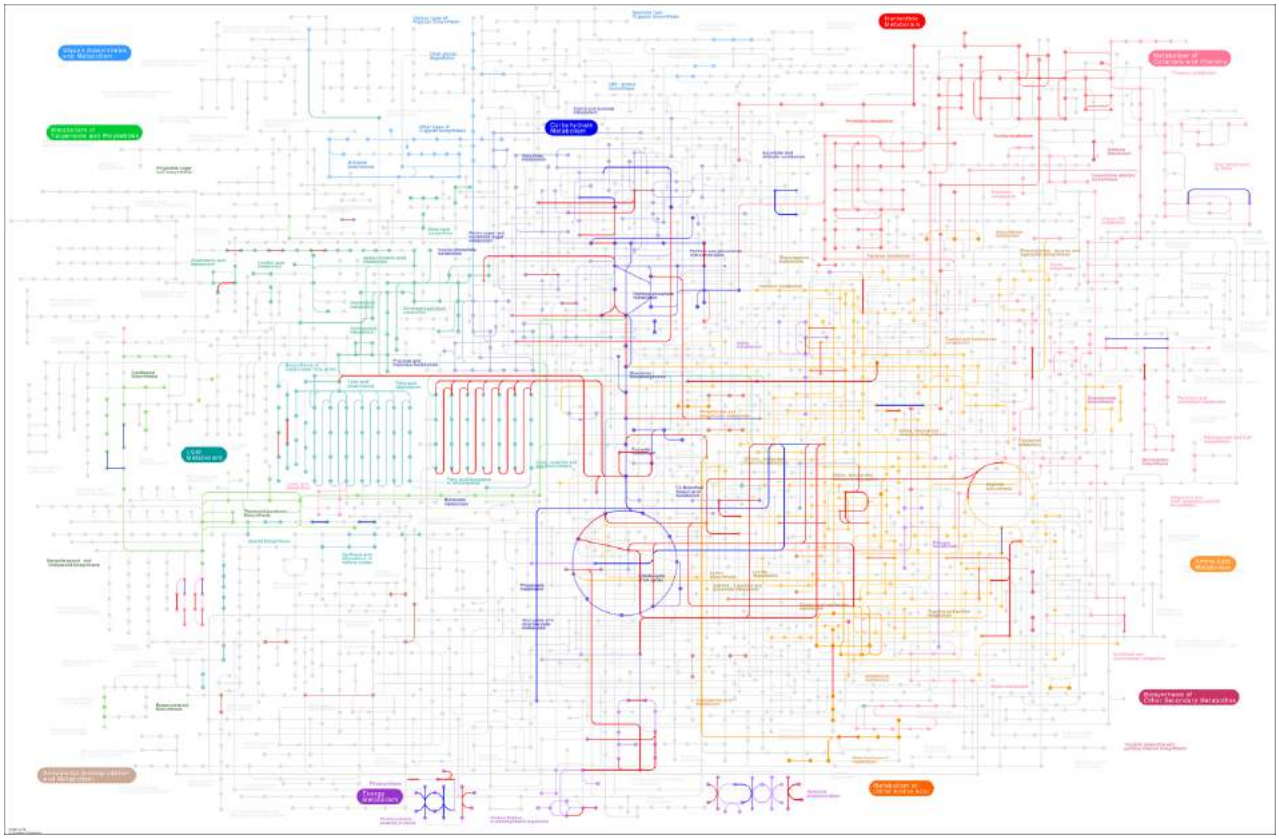
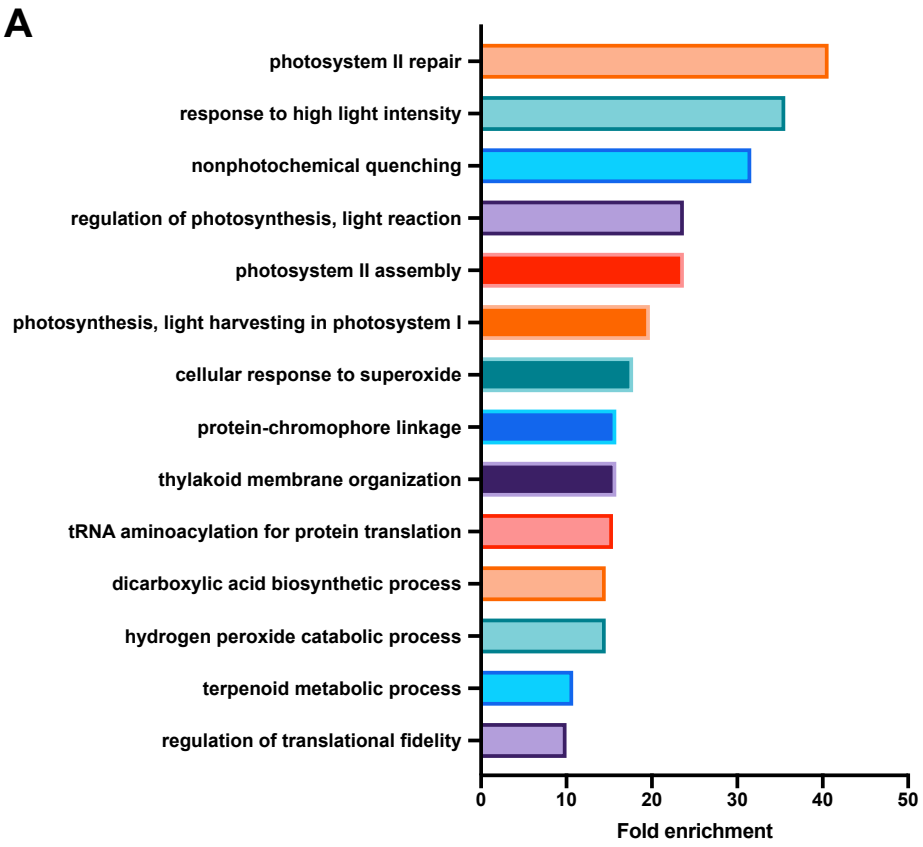


Figure 4.14: Metabolic pathway diagram from KEGG showing differentially regulated proteins mapped to *C. reinhardtii* metabolism. Provides a general overview of *C. reinhardtii* metabolism and pathways affected. Pathways upregulated in EMS-Mut-5 with respect to WT highlighted in blue, pathways downregulated in EMS-Mut-5 with respect to WT highlighted in red.

Functional analysis and enrichment were performed using the Panther classification system, which searched for significantly enriched gene ontology (GO) terms within the lists of differentially expressed proteins when compared to the *C. reinhardtii* reference genome. **Figure 4.15** shows the most enriched specific subclasses of GO terms in proteins presenting positive and negative Log_2 fold change in relative abundance in EMS-Mut-5 compared to WT. Given the increased carotenoid content of EMS-Mut-5, the GO term 'terpenoid metabolic process' was, as expected, enriched (**Figure 4.15**). Other enriched terms associated with biological processes in proteins of higher abundance in EMS-Mut-5 relative to WT were primarily involved in light and oxidative stress responses, as well as photosynthesis and photosystem assembly and repair (**Figure 4.15**). Biological process GO terms enriched within the list of lower abundance proteins in EMS-Mut-5 relative to WT include amino acid biosynthesis and ribosomal assembly, suggesting protein synthesis is downregulated in EMS-Mut-5 (**Figure 4.15**). Central energy metabolic processes such as glycolysis,

TCA cycle and gluconeogenesis also appear to be downregulated; the combination of decreased translation and carbon metabolism could begin to explain the extended lag phase and slower growth rate of EMS-Mut-5 (**Table 4.2**; **Figure 4.11**).



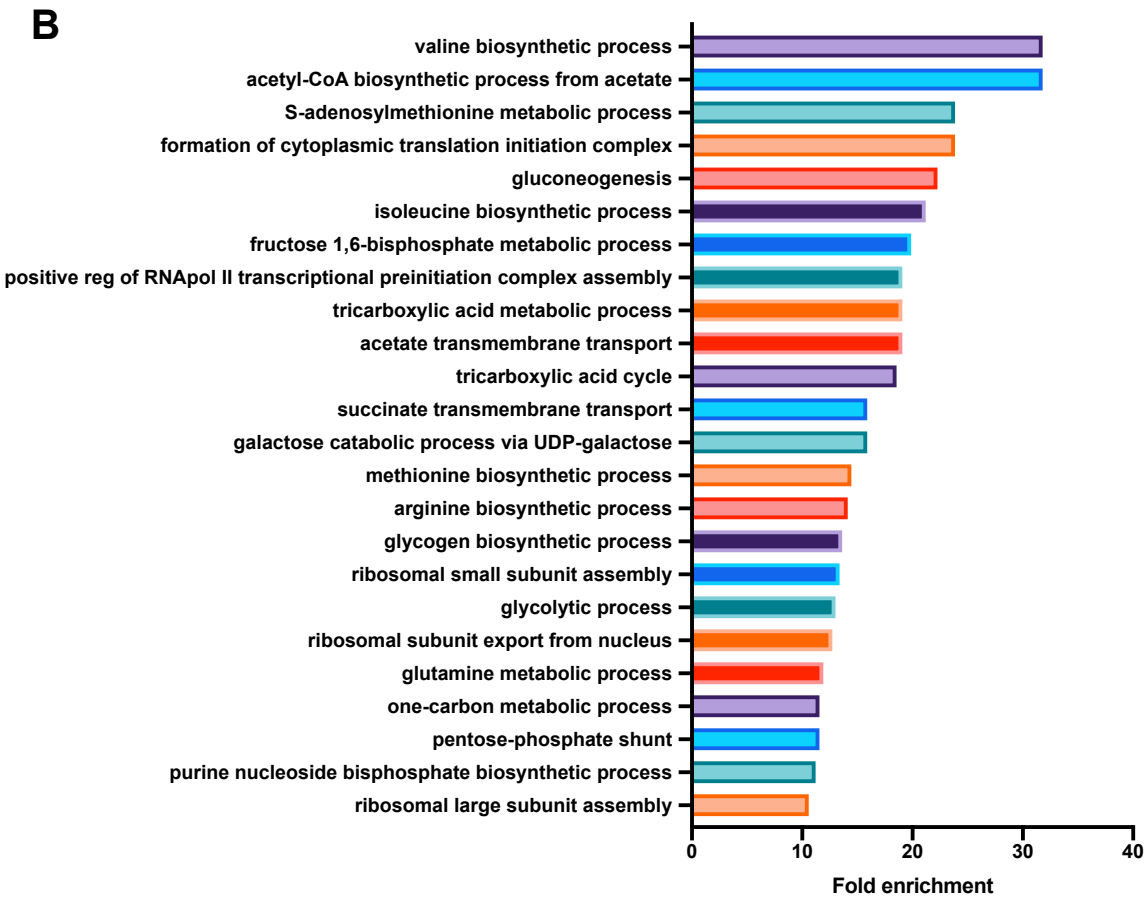


Figure 4.15: Enriched biological process GO terms in differentially expressed proteins. A – GO terms enriched in proteins with significantly higher ($P < 0.05$) relative abundance in EMS-Mut-5 compared to WT. **B** – GO terms enriched in proteins with significantly lower ($P < 0.05$) relative abundance in EMS-Mut-5 compared to WT. GO terms with > 10-fold enrichment shown. Only most specific subclass of each cluster shown. RNApol II = RNA polymerase II.

4.3.3.4. Analysis of individual proteins of higher relative abundance in EMS-Mut-5 with respect to WT

The greatest difference in protein expression observed in EMS-Mut-5 relative to WT was for light harvesting complex stress-related protein 1 (LHCSR1; **Table 4.3**), which had a Log_2 fold change of 9.3 in EMS-Mut-5 relative to WT. Similarly, light harvesting complex stress-related protein 3 (LHCSR3) was exhibited an EMS-Mut-5/WT Log_2 fold change of 4.32. The LHCSR proteins are essential components of energy-dependant (qE) non-photochemical quenching (NPQ) in photosystem II (PSII) under HL stress (Peers *et al*., 2009). LHCSR3 contains two xanthophyll binding sites: one for lutein and one for violaxanthin (Bonente *et al*., 2011); although LHCSR1 contains carotenoid binding sites, its precise pigment stoichiometry and carotenoid identities remain unclear

(Bonente *et al*., 2011). PSBS, another protein crucial for qE NPQ, had the second highest relative abundance in EMS-Mut-5 (Log_2 = 5.77; **Table 4.3**). The precise function of PSBS is currently unknown in *C. reinhardtii*, but its expression is necessary for LHCSR3-mediated qE NPQ and survival in HL (Peers *et al.,* 2009; Correa-Galvis *et al*., 2016; Redekop *et al*., 2020). An uncharacterised protein (Cre13.g586050) with homology to the *A. thaliana* protein SOQ1 (BLAST E-value 3e-138, 47% identity) had an EMS-Mut-5/WT Log2 fold change of 4.1. SOQ1, or suppressor of quenching, prevents slowly reversible PSII antenna quenching in *A. thaliana* (Brookes *et al*., 2013), suggesting the involvement of Cre13.g586050 in NPQ.

Table 4.3: 50 proteins with highest positive log_2 values in EMS-Mut-5 compared to WT

Grouping	Protein name	Description	Log_2	*P*-value
NPQ	P93664 (Cre08.g365900)	Light-harvesting complex stress-related protein 1 (LHCSR1)	9.33	0
	A8HPM5 (Cre01.g016750)	Photosystem II protein PSBS	5.77	0
	P0DO19 (Cre08.g367400)	Light-harvesting complex stress-related protein 3 (LHCSR3.1)	4.32	0
	A0A2K3D0R0 (Cre13.g586050)	Uncharacterised; homology to *A. thaliana* SOQ1	4.1	0
Carotenoid biosynthesis	A0A2K3DUD0 (Cre04.g221550)	Violaxanthin de-epoxidase	4.27	0
	A0A2K3D0U9 (Cre13.g587500)	Phytoene desaturase	3.77	0
PAP fibrillin/ Fasciclin-like	A8JDR9 (Cre12.g492600)	Fasciclin-like domain (FAS3)	4.68	0
	A0A2K3D1W6 (Cre12.g492650)	Fasciclin-like domain (FAS2); SAK1 regulated	4.67	1.00E-06
	A0A2K3DYR0 (Cre03.g197650)	PAP-fibrillin domain-containing protein	3.09	2.00E-06
PSII (maintenance/ assembly/ repair)	A8I686 (Cre07.g315150)	Rubredoxin (RBD1)	4.76	1.50E-05
	Q5W9T4 (Cre09.g393173)	Early light-inducible protein (ELIP2); induced under high light stress	3.88	4.70E-05
	A8HYP8 (Cre06.g251150)	Low CO2 and stress-induced OHP1	3.49	0
	A0A2K3E2M0 (Cre02.g105650)	Uncharacterised; low PSII accumulation homologue	3.47	9.21E-03
	A0A2K3E132 (Cre02.g088400)	DegP-type protease DEG1A	3.32	0

	A8HR79 (Cre13.g562850)	Thylakoid formation protein THF1; PSII RC PSB29 homologue	3.2	0
	A8IAE5 (Cre02.g078507)	Photosystem II Pbs27	2.99	0
PS! Assembly	A8J5N6 (Cre12.g524300)	Thylakoid membrane protein involved in PSI assembly CGL/1	3.78	1.00E-06
Chlorophyll metabolism	A0A2K3D5N5 (Cre12.g558550)	Uncharacterised; homologue of *A. thaliana* chlorophyll dephytylase	3.27	1.70E-05
	A8JDK2 (Cre02.g142146)	Divinyl chlorophyllide-*a* 8-vinyl-reductase	3.01	0
Chloroplastidic/ photosynthesis related	A8J6G0 (Cre17.g721700)	Uncharacterised; thylakoid lumenal protein homologue of AT1G12250	3.74	0
	A8J230 (Cre06.g281800)	Chloroplastidic protein with domain of unknown function (DUF1995)	3.05	0
	A8JF72 (Cre16.g666050)	CPLD49 required for cytochrome b_6f accumulation in high light	2.93	1.10E-05
ROS stress	A8IWW7 (Cre03.g197750)	Glutathione peroxidase (GPX3)	5.37	0
	A0A2K3DMK2 (Cre06.g258600)	Dienelactone hydrolase family protein (SAK1 induced)	4.57	0
	A0A2K3CTK5 (Cre16.g661750)	Calcium/calmodulin dependent protein kinase II association domain (SAK1 induced)	4.48	3.00E-06
	A0A2K3DAQ7 (Cre10.g444550)	Uncharacterised; predicted signal peptide peptidase (SAK1 induced); upregulated in HL and ROS	4.04	0
	A0A2K3D3L4 (Cre12.g513750)	Glutaredoxin, CPYC type (GRX1)	3.79	0
	A8J3M8 (Cre16.g676150)	Chloroplast Mn superoxide dismutase MSD3 (SAK1 induced)	3.28	0
Redox	A8IYH1 (Cre12.g550400)	Glutaredoxin, CPYC type (GRX2)	4.5	0
	A0A2K3DQI7 (Cre06.g294450)	Aldo/keto reductase	4.4	0
	A0A2K3CXI9 (Cre14.g615000)	Methionine Sulfoxide Reductase	3.71	1.16E-03
	A8IQW5 (Cre17.g697050)	Thioredoxin superfamily protein homologue	3.35	4.50E-05
	A0A2K3DBZ5 (Cre10.g461900)	Aldo/ keto reductase; homologous to AtbZIP11/ ATB2	3.28	0
	A0A2K3DMQ1 (Cre06.g261500)	Glutaredoxin	2.98	0
Lipid metabolism	A0A2K3DL78 (Cre07.g349700)	Similar to 3-beta hydroxysteroid dehydrogenase/isomerase	5.18	0
	A0A2K3CVL1 (Cre16.g673001)	Δ-3 palmitate desaturase, crFAD4	4.67	0

	O48663 (Cre13.g590500)	Omega-6-fatty acid desaturase, chloroplast isoform (crDES6; light regulated)	4.17	1.70E-05
	A8IWN6 (Cre03.g195200)	Haloalkane dehalogenase-like hydrolase (Putative lysophospholipase)	3.42	0
RNA binding/ translation	A0A2K3E0P9 (cre02.g082877)	Seryl-tRNA synthetase	3.74	0
	A0A2K3DG80 (Cre08.g358540)	Ribonuclease P	3.56	0
	A8J2S3 (Cre03.g148950)	Putative organellar polyribonucleotide phosphorylase/ nucleotidyltransferase CGL43 (HL induced)	3.53	0
	A0A2K3D9Z9 (Cre10.g433000)	glycyl-tRNA synthetase	3.25	1.00E-06
	A8JCK3 (Cre07.g335200)	Elongation factor-type GTP-binding protein (chloroplastidic; involved in ROS response)	3.17	3.90E-04
Carbon metabolism	A8HQC2 (Cre01.g028600)	Alcohol dehydrogenase / Aldehyde reductase	5.22	0
	A0A2K3D5Y3 (Cre12.g554100)	Putative inorganic carbon transporter	4.38	4.20E-05
	A0A2K3DIB0 (Cre08.g384750)	Alpha-amylase; HL induced	3.12	2.60E-05
Protein folding/ chaperone	A0A2K3DCA6 (Cre10.g466850)	Uncharacterised; peptidylprolyl isomerase (gun4 regulated)	4.44	0
	A0A2K3CPW9 (Cre17.g715000)	Heat shock protein 33 (light stress response)	3.3	1.10E-05
Sterol synthesis	A0A2K3CR99 (Cre17.g734644)	Squalene monooxygenase / Squalene epoxidase (SQE)	2.95	6.20E-05
General stress response	A0A2K3CN34 (Cre24.g755497)	Uncharacterised; putative tryptophan-rich sensory protein	3.74	0

Log_2 represents the protein intensity ratio in EMS-Mut-5/WT; high positive Log_2 values equate to higher relative protein abundance in EMS-Mut-5 relative to WT. Proteins with $P > 0.01$ and orphan proteins with no characterisation or orthologues in related species were discounted from the list; full list of proteins with increased expression can be found in **Appendix Table C9**. Protein name contains the UniProtKB identifier, followed by the Phytozome gene identifier in brackets. Grouping was determined through Panther GO terms, or otherwise predicted through domain homology using BLAST and Phytozome database. SAK1-regulated proteins noted. *P*-values < 1.00E-06 show as 0.00 as an artefact of the ProteoSign software.

Increased carotenoid pathway proteins in EMS-Mut-5 include a violaxanthin de-epoxidase (VDE) with an EMS-Mut-5/WT Log_2 fold change of 4.27, and a PDS with an EMS-Mut-5/WT Log_2 fold change of 3.77 (**Table 4.3**). This VDE is unique to *C. reinhardtii* and was only recently characterised (Li *et al*.,

2016). Lycopene β-cyclase activity and involvement in the biosynthesis of lutein and β-carotene are predicted for VDE. **Table 4.4** shows other carotenoid biosynthesis-related proteins that display increased expression in EMS-Mut-5; 7 carotenoid biosynthetic enzymes in total were found to be increased in EMS-Mut-5, many of which are uncharacterised but predicted computationally using BioCyc, along with 2 enzymes in the isoprenoid biosynthetic pathway and an ABC1 kinase (Cre13.g581850) that is predicted to be a positive regulator of carotenoid biosynthesis in *C. reinhardtii* (**Table 4.4**). The ABC1 kinase was recently identified as an essential gene for *C. reinhardtii* photosynthesis (Li *et al*., 2019) and has 15 associated GO terms, mainly related to stress responses, photosynthesis and pigment metabolism (**Table 4.4**); its *A. thaliana* homologue is necessary for antioxidant biosynthesis, primarily of lutein, β-carotene and tocopherol (Martinis *et al*., 2014). No proteins associated with the carotenoid pathway were found to be of lower abundance in EMS-Mut-5 than in WT. **Figure 1.4** shows the isoprenoid and carotenoid pathways in *C. reinhardtii*.

Table 4.4: Proteins involved in carotenoid biosynthesis with increased expression in EMS-Mut-5 relative to WT

Protein ID	Description	GO terms	Log_2	*P*-value
A0A2K3DUD0 (Cre04.g221550)	Violaxanthin de-epoxidase	-	4.27	0
A0A2K3D0U9 (Cre13.g587500)	Phytoene desaturase	-	3.77	0
A8J3K3 (Cre07.g314150)	Prolycopene isomerase	-	1.95	2.83E-04
A0A2K3D5G7 (Cre12.g560900)	Phytoene desaturase	-	1.81	1.00E-06
A0A2K3CSZ6 (Cre16.g651923)	Phytoene desaturase	carotenoid biosynthetic process [GO:0016117]; etioplast organization [GO:0009662]	1.62	2.82E-04
Q6J213 (Cre12.g509650)	Chloroplast phytoene desaturase (PDS)	carotenoid biosynthetic process [GO:0016117]	1.37	0
A8I647 (Cre07.g314150)	Zeta-carotene desaturase	9,9'-di-cis-zeta-carotene desaturation to 7,9,7',9'-tetra-cis-lycopene [GO:0052889]; carotene biosynthetic process [GO:0016120]; carotenoid biosynthetic process [GO:0016117]; lycopene biosynthetic process [GO:1901177]	1.01	8.00E-06
O81954 (Cre07.g356350)	1-deoxy-D-xylulose-5-phosphate	terpenoid biosynthetic process [GO:0016114]	0.79	5.00E-05

	synthase (DXS)			
A8IX41 (Cre03.g207700)	Farnesyl diphosphate synthase (FPPS)	farnesyl diphosphate biosynthetic process [GO:0045337]	0.51	2.47E-02
A0A2K3D0E7 (Cre13.g581850)	ABC1 kinase	cellular response to nitrogen starvation [GO:0006995]; chlorophyll catabolic process [GO:0015996]; photosynthetic electron transport chain [GO:0009767]; plastoglobule organization [GO:0080177]; positive regulation of carotenoid biosynthetic process [GO:1904143]; positive regulation of chlorophyll biosynthetic process [GO:1902326]; regulation of anthocyanin biosynthetic process [GO:0031540]; regulation of photosynthesis [GO:0010109]; regulation of tocopherol cyclase activity [GO:1902171]; response to blue light [GO:0009637]; response to paraquat [GO:1901562]; response to photooxidative stress [GO:0080183]; response to red light [GO:0010114]; response to water deprivation [GO:0009414]; thylakoid membrane organization [GO:0010027]	1.96	1.84E-02

Several chloroplast-localised proteins, particularly associated with PSII assembly and repair, were discovered within the top 50 proteins of increased expression in EMS-Mut-5 (**Table 4.4**). Examples include rubredoxin (RBD1), which protects PSII complexes undergoing *de novo* repair during photooxidative stress (García-Cerdán *et al*., 2019) and an uncharacterised *A. thaliana* early light inducible protein (ELIP2) homologue (Cre09.g393173) which is a hypothesised pigment binding protein that is induced under HL (Adamska *et al*., 1999; Hayami *et al*., 2015). Other predicted pigment binding and thylakoid biogenesis proteins had increased relative abundance in EMS-Mut-5, but they are yet to be properly characterised (**Table 4.3**).

Proteins involved in reactive oxygen species (ROS) stress and redox displayed particularly high fold increases in EMS-Mut-5 compared to WT (**Table 4.3**). Of note, several glutaredoxins were upregulated and a superoxide dismutase (Cre16.g676150). An uncharacterised aldo/ keto reductase (Cre10.g461900) that shares homology with the *A. thaliana* protein ATB2 (BLAST E-value 3e-26, 30% identity), a light-regulated bZIP TF (Rook *et al*., 1998), was also identified.

Interestingly, a plastid lipid-associated protein (PAP) was discovered in the increased abundance proteins (Cre03.g197650); PAP-domain proteins in plants have roles in carotenoid storage during chromoplast development, and potentially play a similar role in carotenoid sequestration in *C.*

reinhardtii (Leitner-Dagan *et al*., 2006). Fascilin-like proteins are similar membrane-bound peptides associated with the carotenoid-rich eye-spot (Eitzinger *et al*., 2015), two of which were upregulated in EMS-Mut-5 (**Table 4.3**). Two other PAP fibrillin-domain containing proteins, Cre03.g197650 (Log_2 = 3.01) and Cre01.g039550 (Log_2 = 2.19), were identified within the extended list of proteins of higher relative abundance in EMS-Mut-5 (**Appendix Table C9**).

Two fatty acid desaturases, Cre16.g673001 and Cre13.g590500, were upregulated in EMS-Mut-5, the latter of which is predicted to be light regulated (Romero-Campero *et al*., 2016). Cre07.g349700, which has an EMS-Mut-5/WT Log_2 fold change of 5.18, is an uncharacterised protein similar to 3-beta hydroxysteroid dehydrogenase/ isomerase that is thought to be involved in fatty acid biosynthesis (Eitzinger *et al*., 2015). Conversely, the upregulated Cre03.g195200 is a haloalkane dehalogenase-like hydrolase that is likely a TAG-hydrolysing lipase due to its increased expression following nitrogen (Tsai *et al*., 2018), suggesting lipid degradation may have been engaged in EMS-Mut-5. This suggestion, however, is tentative as the function of Cre03.g195200 has not been experimentally confirmed.

Other EMS-Mut-5 upregulated proteins of note include an Elongation factor-type GTP-binding protein (Cre07.g335200), which bears sequence similarity with the *A. thaliana* protein AT5G13650, otherwise known as happy on norflurazon 23 (hon23; Saini *et al*., 2011). A mutation in a conserved region of hon23 conferred *A. thaliana* with norflurazon resistance and prevented photobleaching in HL by disrupting chloroplast gene expression.

4.3.4.5. Analysis of individual proteins of decreased relative abundance in EMS-Mut-5 with respect to WT

The most heavily downregulated proteins in EMS-Mut-5 compared to WT were those involved in acetate transport, with 3 acetate uptake transporters (Cre17.g702950, Cre17.g702900 and Cre17.g700750) displaying reduced relative abundance (**Table 4.5**). These acetate transporters are members of the GPR1/FUN34/YaaH (GFY) superfamily of transporters, and their expression increases under light-limiting conditions (Durante *et al*., 2019). Mixotrophic growth using acetate as a carbon source is reliant upon isocitrate lyase (ICL; Plancke *et al*., 2014), which is downregulated in EMS-Mut-5. Carbon metabolism from acetate is localised to peroxisomes (Lauersen *et al*., 2016a), and a peroxisomal biogenesis factor (Cre12.g540500) showed decreased expression in EMS-Mut-5. Together, this suggests that acetate metabolism was downregulated, which is in accordance with the KEGG pathway analysis (**Appendix Figure C9**). Furthermore, 5 glycolysis and gluconeogenesis

proteins (including ICL1) were observed to be lower in EMS-Mut-5 (**Table 4.5**), suggesting downregulated central carbon respiration.

Table 4.5: 50 proteins with lowest negative log_2 values in EMS-Mut-5 compared to WT

Grouping	Protein name	Description	Log_2	*P*-value
Acetate metabolism	A0A2K3CP19 (Cre17.g702950)	Putative acetate uptake transporter GFY5	-3.46	0
	A0A2K3CP17 (Cre17.g702900)	Acetate uptake transporter GFY4	-3.26	2.88E-03
	A8IQG4 (Cre17.g700750)	Acetate uptake transporter GFY3	-2.64	6.55E-03
	A8J244 (Cre06.g282800)	Isocitrate lyase (ICL)	-3.24	0
Carbon metabolism/ glycolysis/ gluconeogenesis	A8J0N7 (Cre02.g141400)	Phosphoenolpyruvate carboxykinase (PCK1a)	-2.25	4.00E-06
	A8JHR9 (Cre12.g485150)	Glyceraldehyde 3-phosphate dehydrogenase (GAP1a)	-2.11	1.00E-06
	A0A2K3DKE9 (Cre07.g338451)	Uncharacterised; fructose-bisphosphatase/ Hexose diphosphatase	-2	2.87E-03
	A0A2K3DPM5 (Cre06.g280950)	Pyruvate kinase	-1.87	8.50E-03
C4 photosynthesis	A7UCH9 (Cre09.g405750)	Carbonic anhydrase (CAH8)	-1.99	1.89E-03
Translation factors	A8JGK5 (Cre12.g498100)	Eukaryotic translation initiation factor 3 subunit E (eIF3e)	-3.29	4.00E-06
	A8IP53 (Cre06.g298100)	Translation initiation protein (SUI1A)	-2.77	0
	Q8VZZ5 (Cre13.g587050)	Eukaryotic release factor 1 (ERF1)	-2.25	0
Translation/ ribosome associated	A0A2K3D4E7 (Cre12.g525200)	NOP56 ribosome biogenesis factor	-3.01	1.00E-06
	A0A2K3DLU5 (Cre06.g249250)	Ribosomal protein L7Ae	-2.59	4.51E-04
	A0A2K3DAI9 (Cre10.g441400)	NOP58 ribosome biogenesis factor	-2.07	0
	A8JF66 (Cre16.g666301)	40S ribosomal protein S30 (RPS30)	-1.94	5.80E-04
	A0A2K3CZF8 (Cre13.g567850)	H/ACA ribonucleoprotein complex subunit 1	-1.81	8.00E-06
RNA binding/ splicing	A0A2K3DZM1 (Cre03.g199647)	Uncharacterised; eukaryotic translation initiation factor 4A; potential splicing factor	-2.79	1.00E-06

	A0A2K3DNK6 (Cre06.g275100)	Uncharacterised; nucleolin (NCL, NSR1); SAK1 regulated	-2.34	1.00E-06
	A0A2K3CX88 (Cre14.g611150)	small nuclear ribonucleoprotein B and B' (SNRPB, SMB)	-2.28	0
	A0A2K3CV63 (Cre16.g679600)	U2 small nuclear ribonucleoprotein A' (SNRPA1)	-1.93	4.49E-03
	A0A2K3CTM9 (Cre16.g662702)	RNA-binding protein Musashi (MSI)	-1.8	6.05E-04
	A8HQ72 (Cre01.g026450)	Serine/arginine-rich pre-mRNA splicing factor SRP35	-1.77	4.50E-05
Transcription factors	A8J3F0 (Cre16.g672300)	HMG group, predicted YABBY TF	-2.94	1.75E-04
	A8HXE1 (Cre06.g261450)	HMGB1 TF	-2.42	4.23E-04
Chromatin associated	A8HRZ0 (Cre13.g567450)	Histone H1 (HON1); Gun4 regulated	-2.3	0
Nucleotide biosynthesis	A0A2K3CZ88 (Cre13.g565450)	Adenylosuccinate lyase (ASL)	-1.93	4.00E-06
Lipid metabolism	A8JHJ5 (Cre15.g641200)	Predicted mitochondrial fatty acid carrier	-2.56	1.10E-05
	A0A2K3DTB8 (Cre04.g216950)	Predicted beta-ketoacyl-acyl-carrier-protein synthase III	-2.04	3.00E-06
	A0A2K3DE56 (Cre09.g390615)	Triacylglycerol lipase – DAG degradation	-1.99	1.30E-05
ECM/ Cell wall/ secreted	A0A2K3CY13 (Cre14.g620702)	Pherophorin	-2.39	0
	A0A2K3DT03 (Cre05.g238650)	Cell wall protein pherophorin-C5	-2.33	1.90E-05
	A8I1E4 (Cre10.g463350)	Hydroxyproline-rich glycoprotein HRP3	-1.84	2.18E-03
	A0A2K3D8J9 (Cre11.g479250)	Ran GTPase-activating protein 1	-1.83	1.60E-05
	A8IZV1 (Cre09.g393700)	Matrix metalloproteinase MMP3	-1.79	6.26E-03
Flagella	A0A2K3DLB7 (Cre07.g351650)	Flagellar Associated Protein (FAP20)	-2.47	1.14E-03
	A0A2K3DCH9 (Cre09.g392867)	Uncharacterised; flagella membrane glycoprotein	-2.19	6.50E-05
Cell signalling/ kinase/ phosphatase	A0A2K3D2I5 (Cre12.g511850)	Glycogen Synthase Kinase	-2.56	0
	Q9XGU3 (Cre06.g292550)	Serine/threonine protein phosphatase PP1 (Flagellar)	-1.99	2.20E-05
	A0A2K3DE67 (Cre09.g391023)	Serine/threonine protein phosphatase PP2A, metallophosphoesterase	-2.03	0

	A8J2Q0 (Cre03.g150300)	calmodulin (CALM)	-1.81	9.37E-03
Protein folding/ chaperone activity	A8HXD3 (Cre06.g261650)	Uncharacterised; prefoldin molecular chaperone PFD1, subunit 3	-2.21	0
	A8HUK0 (Cre13.g586300)	Peptidyl-prolyl cis-trans isomerase FKB12 (Rapamycin sensitivity)	-1.93	4.03E-04
Protein degradation	A8IRB7 (Cre17.g706800)	Uncharacterised; isochorismatase family protein	-2.3	1.27E-03
Phosphate metabolism	A8J133 (Cre09.g387875)	Soluble inorganic pyrophosphatase IPY3	-2.33	0
Iron uptake	A8HYQ6 (Cre06.g251000)	Uncharacterised; putative ferroportin	-1.92	7.58E-04
Thiamine biosynthesis	A8J841 (Cre05.g240850)	Hydroxymethylpyrimidine phosphate synthase THICb	-1.91	3.50E-05
ROS stress	A8JBB4 (Cre16.g682725)	Glutathione S-transferase GSTS2; upregulated under ROS stress	-2.24	1.75E-03
Vacuolar	Q5VLJ9 (Cre12.g549300)	Aquaporin, glycerol transport activity MIP1	-2.15	4.30E-04
Peroxisomal	A0A2K3D6X8 (Cre12.g540500)	Peroxisomal Membrane Protein PMP27 (light regulated)	-1.78	6.30E-05

Log_2 represents the protein intensity ratio in EMS-Mut-5/ WT; low negative Log_2 values equate to lower relative protein abundance in EMS-Mut-5 relative to WT. Proteins with $P > 0.01$ and orphan proteins with no characterisation or orthologues in related species were discounted from the list; full list of proteins with increased expression can be found in **Appendix Table C10**. Protein name contains the UniProtKB identifier, followed by the Phytozome gene identifier in brackets. UC = uncharacterised. Grouping was determined through Panther GO terms, or otherwise predicted through domain homology using BLAST and Phytozome database. *P*-values < 1.00E-06 show as 0.00 as an artefact of the ProteoSign software.

About a third of the 50 proteins with the lowest negative Log_2 values (EMS-Mut-5/ WT) have nucleotide binding domains, particularly for RNA (**Table 4.5**). The translation factor eIF3e (Cre09.g405750) had an EMS-Mut-5/WT Log2 fold change of -1.99; its *A. thaliana* homologue At3g57290 is light responsive and plays a (currently unclear) role in photomorphogenesis through controlling ribosome occupancy of mRNAs (Wu *et al*., 2012). A eukaryotic release factor (ERF; Cre13.g587050) was discovered in the decreased expression proteins; interestingly, the ERF1 homologue in plants is supressed by ORANGE, the carotenoid-inducing regulatory protein whose *C. reinhardtii* equivalent was overexpressed in **Chapter 3** (Zhou *et al*., 2011).

Two putative transcription factors were identified as having lower relative abundance in EMS-Mut-5 compared to WT. Cre16.g672300, a putative YABBY TF with two high-mobility group domains, is postulated to be involved in cell cycle regulation (Romero-Campero *et al*., 2016) and is reduced in low CO_2 conditions (Arias *et al*., 2020). The transcriptional regulator Cre06.g261450 was also found to have lower expression, however its function is unknown. The abundance of chromatin associated protein HON1 (Cre13.g567450) was lower in EMS-Mut-5 than WT. HON1 transcripts are 4-fold higher in *gun4* mutants; Gun4, which is involved in chloroplast-to-nucleus retrograde signalling using singlet oxygen, is downregulated 28% in EMS-Mut-5. The downregulation of HON1 further suggests photosynthesis and ROS signalling pathways have been affected by the mutation in EMS-Mut-5.

Other comparatively downregulated proteins in EMS-Mut-5 include secreted and extracellular matrix (ECM) proteins, such as two pherophorins (Cre14.g620702 and Cre05.g238650), a hydroxy-proline rich glycoprotein (HRP3) and a matrix metalloproteinase (MMP3). Flagella associated proteins Cre07.g351650 and Cre09.g392867, as well as a kinase which regulates flagella length (Liang *et al*., 2018) were also down regulated.

4.4. Discussion

In this chapter, a semi high-throughput platform for generating enhanced carotenoid-producing *C. reinhardtii* mutants was developed, where 658 norflurazon-resistant mutants were screened to yield 9 candidate strains, 7 of which (EMS-Mut-1, -2, -4, -5, -6, -7 and -8) produced significantly more lutein per g DCW than the wild-type (**Figure 4.10**). Mutant EMS-Mut-5 exhibited > 5-fold increase in lutein, which to the author's knowledge, is the highest fold increase in *C. reinhardtii* carotenoids recorded in the literature by genetic engineering alone. LFQ shotgun proteomics enabled a comparison of the carotenoid-accumulating mutant strain with WT, and provided insight into how the lutein is overproduced and where it could accumulate. The proteomics additionally revealed ways in which EMS-Mut-5 growth conditions could be optimised, and exposed potential genetic engineering targets for future experiments, which will be discussed below.

The increase in carotenoid production in EMS-Mut-5 can likely be attributed to the upregulation of several carotenoid biosynthetic enzymes (**Table 4.4**). The increase in lutein specifically may be linked to the increased expression of VDE in EMS-Mut-5 compared to WT (**Table 4.4**), which is predicted to be involved in lutein biosynthesis. The increased abundance of the LHCSR proteins, amongst other predicted pigment binding proteins, could provide a metabolic sink for lutein within the

thylakoid membrane. Sequestration of carotenoids in general could tentatively be linked to the increase of membrane proteins such as PAP-fibrilin domain containing proteins, which have been linked to carotenoid storage in plants (Pozueta-Romero *et al*., 1997; Singh and McNellis, 2011). The function of the PAP-fibrillin domain-like protein Cre03.g197650 would therefore be worth investigating further.

Alongside an increase in lutein production, a plethora of metabolic pathways were affected in EMS-Mut-5. Of the 50 proteins with the greatest positive fold-change in EMS-Mut-5 compared to WT (**Table 4.3**), 31 were involved with photosynthesis or ROS stress. Many high abundance proteins in **Table 4.3** are upregulated under HL, ROS stress or both (Barth *et al*., 2014), and many (some overlapping) are regulated by singlet oxygen kinase 1 (SAK1; Wakao *et al*., 2015). High light stress responses in *C. reinhardtii* include induction of NPQ, changes in electron transport and thylakoid membrane ultrastructure, altered stoichiometry of PSI:PSII, accumulation of xanthophylls and antioxidants such as tocopherol, and photosynthetic apparatus degradation and repair (Erikson *et al*., 2015); GO term enrichment (**Figure 4.15A**) and individual protein analysis (**Table 4.3**) revealed involvement of several of these mechanisms in EMS-Mut-5. Taken together, these data suggest that EMS-Mut-5 harbours a mutation that somehow affects the regulatory pathway governing HL and/ or ROS stress responses, leaving the cell in a perpetually stressed state. This overzealous stress response was likely responsible for the survival of EMS-Mut-5 during the norflurazon screening step.

The *C. reinhardtii* qE NPQ proteins LHCSR1, LHCSR3 and PSBS, which are notably increased in EMS-Mut-5, have recently been subjects of avid investigation in *C. reinhardtii* photosynthesis research (Dinc *et al*., 2016*;* Kosuge *et al*., 2018; Tokutsu *et al*., 2019; Aihara *et al*., 2019; Gabilly *et al.,* 2019). Under ambient conditions, the LHCSR and PSBS proteins are virtually undetectable (Strenkert *et al*., 2019), and the discovery that they are in fact vital components of the qE NPQ HL stress response was relatively recent (Peers *et al*., 2009). Previous studies show that these proteins are coregulated (except in altered CO_2 conditions [Maruyama *et al*., 2014]), and their expression increases under high light, blue light and UV irradiation (Peers *et al*., 2009; Petroutsos *et al*., 2016; Allorant *et al*., 2016). The UV light receptor UVR8 was not detected in the proteomics analysis of EMS-Mut-5, whereas very slight downregulation of blue light photoreceptor PHOT (Log_2 = -0.4) was detected. Given that LHCSR3 is reliant upon PHOT for expression (Aihara *et al*., 2019), it is likely that any mutations in this light-induced pathway have emerged further downstream in the signalling cascade.

Recent examination of a *C. reinhardtii* mutant that overexpresses LHCSR1 and PSBS (similarly to EMS-Mut-5) was shown to retain a missense mutation in a component of a SPA1-COP1 E3 ubiquitin ligase complex, which supresses qE protein expression (Gabilly *et al.,* 2019). Similarly, another ubiquitin ligase complex CUL4-DDB1^{DET1} was recently found to supress induction of LHCSR and PSBS proteins (Aihara *et al*., 2019). This suggests that part of an E3 ubiquitin ligase complex may have been disrupted in EMS-Mut-5, given its similar phenotype. High carotenoid phenotypes, however, were not reported in the publications presenting these results (Gabilly *et al*., 2019; Aihara *et al*., 2019). Furthermore, no differential expression was detected for any of the proteins known to be involved in either of the E3 ubiquitin ligase complexes or their TF targets in EMS-Mut-5, although a mutation that sterically disrupts complex formation but not protein expression may have arisen. Many proteins with the GO term 'protein ubiquitination' but with unknown functions are present in the list of downregulated genes (**Appendix Table C10**), so there is the potential that a novel ubiquitin ligase complex was contributing to the high carotenoid and qE phenotype in EMS-Mut-5. Altered expression of other constituents of the light stress signalling pathway could alternatively have been modified; one such candidate is the ABCK1 kinase Cre13.g581850 (**Table 4.4**), whose long list of GO terms suggests its involvement in several biological processes related to stress responses and photosynthesis.

Many other proteins implicated (or predicted to be implicated) in gene expression were differentially expressed in EMS-Mut-5, most of which were translation factors that present decreased relative abundance compared to WT. It is possible that one or more of these regulatory factors contributed to the phenotype of EMS-Mut-5. Their dominance in the downregulated list more broadly suggests that post-transcriptional regulation plays an important role in gene expression in *C. reinhardtii*.

An alternative explanation for the induction of HL and ROS stress responses that must be considered is that the mutation itself may be causing ROS stress; examining the redox state of the cell could help to answer this question biochemically. This could be achieved by measuring the reduced glutathione pool by use of the non-fluorescent probe monochlorobimane (Fritzsche and Mandenius, 2010), or by measuring lipid peroxidation using lipophilic fluorescent dyes (Melegari *et al*., 2013).

The combination of reduced acetate uptake and glycoxylate metabolism (**Figure 4.15**) alongside increased expression of photosynthetic proteins suggests a metabolic shift from mixotrophic growth using acetate as a carbon source to photoautotrophic growth that requires CO_2 in EMS-Mut-5; this could be contributing to the slow growth rate observed in **Figure 4.11**. The CO_2-limiting conditions

used to grow EMS-Mut-5, i.e. on a shaker in an Erlenmeyer flask without additional CO_2, may have restricted the ability of the mutant to flourish; likewise, the low light conditions (150 μmol photons m^2 s^{-1}) may have further hampered its growth. Growing EMS-Mut-5 photoautotrophically in high light conditions with CO_2 bubbling will likely increase its growth rate and biomass accumulation, potentially leading to improved volumetric lutein and biomass yields. A phenotypically similar mutant generated by Aihara *et al*. (2019) exhibited slower growth than WT under low light conditions but faster growth rates when grown under saturating light, and another high light resistant mutant was found to grow optimally under photoautotrophic conditions at 600 μmol photons m^2 s^{-1} (Forster *et al*., 1999); these findings from the literature provide support for the hypothesis that simple changes to the growth conditions could improve the performance of EMS-Mut-5. There is still the problem, however, of downregulated protein expression and central carbon metabolism through the glycolysis pathway, but this may be an effect of the sub-optimal growth conditions that will potentially be rectified through environmental manipulation. If this is the case, then EMS-Mut-5 could be an extremely valuable biotechnological strain, for both its enhanced lutein production and ability to survive fluctuating and extreme light conditions (Day *et al*., 2012).

Although the precise location of the mutations in EMS-Mut-5 were not identified, some potential leads for targeted metabolic engineering for enhanced lutein production can be inferred from the proteomics study. For example, manipulating the expression of components of the HL or ROS signalling cascade could prove an effective method for increasing carotenoid synthesis; this way, the altered expression of just one protein could elicit global changes within the cell, enabling the simultaneous upregulation of the carotenoid biosynthetic pathway and of proteins involved in carotenoid storage with minimal genetic alterations. The ABCK1 kinase identified in the 50 most upregulated proteins (**Table 4.3**) is one such candidate for overexpression, as its relatively long list of associated GO terms suggests it has multiple targets in stress response and photosynthetic pathways related to carotenoid biosynthesis. It would also be interesting to examine whether carotenoid biosynthesis was increased in E3 ubiquitin ligase complex mutants with increased LHCSR protein expression; if this is the case, RNAi silencing of a component of the Cul4-$DDB1^{DET1}$ or COP1-SPA1 complexes could achieve improvement of carotenoid yield (Aihara *et al*., 2019; Gabilly *et al*., 2019).

Another interesting lead for metabolic engineering is the PAP-fibrillin domain protein (**Table 4.3**). As discussed above, PAP-fibrillin domain proteins are localised to plastoglobules in plants and perform roles in lipid and carotenoid storage (Bréhélin *et al*., 2007; Nacir and Bréhélin, 2013).

Interestingly, three PAP-fibrillin proteins were upregulated in EMS-Mut-5 (**Appendix Table C9**). Functionally characterising these proteins in *C. reinhardtii* could be worthwhile, as they could potentially act as a metabolic sink for carotenoids and prevent metabolite-induced feedback inhibition.

The label-free shotgun proteomics strategy applied in this work enabled the characterisation of EMS-Mut-5, helping to derive the mechanisms behind the high lutein accumulation, as well as other metabolic pathways affected by the mutagenesis. The proteomics experiment also offered an explanation for the slower growth rate of EMS-Mut-5, and ways in which to improve this in future experiments. Although interesting hypotheses were generated, they should ideally be complemented with other biochemical assays for validation.

Repeating the proteomics experiment at multiple timepoints could indicate whether the HL and ROS responses discussed above were triggered or constitutive in EMS-Mut-5. This, coupled with additional HPLC analysis timepoints, would also reveal a) whether the increased lutein production is sustained throughout the growth cycle of the mutant, and b) the optimal time at which to harvest. Repetition of the proteomics study would also clarify whether biological replicate 3 of EMS-Mut-5 impacted the statistical and pathway analyses, as the PCA plot (**Figure 4.12B**) indicated that this replicate differed from the two other EMS-Mut-5 biological replicates analysed. Despite this, the high number of differentially regulated proteins in EMS-Mut-5 with respect to WT that present low *P*-values suggests that the inclusion of this outlying replicate has not significantly influenced the analysis.

The upregulation of photosynthetic and NPQ proteins in EMS-Mut-5 strongly suggests that photosynthetic mechanisms are affected. Conducting fluorometry measurements would facilitate examination of qE NPQ, electron transfer rates and the quantum fluorescence yield of PSII in EMS-Mut-5. This could provide evidence for the functional increase in qE NPQ, as well as determine whether there are significant changes in the EMS-Mut-5 photosynthetic apparatus. It would also be interesting to compare these measurements to the LHCSR mutants discussed earlier (Aihara *et al*., 2019; Gabilly *et al*., 2019). The sunlight-to-biomass energy conversion efficiency could also partially be determined using the fluorometry data, which would ensure that the mutant is efficient enough for biotechnological exploitation. Furthermore, examination of the thylakoid membrane ultrastructure in EMS-Mut-5 could reveal further effects of the mutation related to photosynthesis, as other forms of NPQ such as state transitions (qT) may be affected by the mutation, as qT and qE are intimately linked (Allorent *et al*., 2013).

Mutant selection could potentially be streamlined from semi-high throughput to high-throughput using FACS. Chlorophyll-based cell-sorting strategies to isolate carotenoid-enriced *D. salina* and *P. tricornutum* strains have been developed (Mendoza *et al*., 2008; Yi *et al*., 2018). Given that mutant selection was partially based on chlorophyll fluorescence, perhaps similar results could be achieved by sorting mutant cell lines based solely on chlorophyll fluorescent readings after growth on norflurazon, rather than SGR. A downside to this strategy would be that the increased chlorophyll could reduce the photon conversion efficiency of the strain, as less light can penetrate the culture (Beckmann *et al*., 2009). The workflow applied in this work, although not fully high-throughput, could be reapplied to permit the inclusion of low chlorophyll/ high carotenoid strains with minor alterations, and could be an interesting investigation for the future.

Given the pervasive effects of the mutagenesis in EMS-Mut-5, it is difficult to pinpoint the precise genetic occurrence (or occurrences) that confer its altered phenotype. To draw any conclusion about the precise mechanisms affected, it would first need to be determined whether the mutation is a singular event, or if there are alterations to multiple loci in the EMS-Mut-5 genome. Backcrossing the mutant with the WT strain could provide answers. PCR amplification and sequencing of potential regulators of HL and ROS response could reveal the mutation site; however, potentially numerous genes have been mutated, and the mutation could lie within an uncharacterised gene, rendering this approach ineffective. Whole genome sequencing has successfully been applied to identify deletions and SNPs within novel genetic elements in high-light tolerant strains (Schierenbeck *et al*., 2015; Garbilly *et al*., 2019), hence this should be the next step towards genotyping EMS-Mut-5.

Chapter 5: Synthetic promoters to expand the range of recombinant protein expression levels from the *C. reinhardtii* nuclear genome

5.1. Summary

In this chapter, novel genetic devices for metabolic engineering were developed, due to the need for a diverse and effective toolset for nuclear recombinant protein expression in *C. reinhardtii*. *Cis*-regulatory DNA elements (CREs) that instigate high transcriptional activity in *C. reinhardtii* promoters were searched for computationally using *de novo* motif discovery software to analyse a publicly available transcriptomics dataset. Thirteen of the identified putative CREs (pCREs), plus a random DNA sequence control of similar length, were synthesised as multiple repeats attached to a common minimal core promoter, then cloned into test vectors upstream of a yellow fluorescent protein reporter gene. Following transformation of the vectors into *C. reinhardtii* by electroporation, *in vivo* measurements of yellow fluorescent protein expression by flow cytometry revealed that 10 out of the 14 DNA motifs analysed displayed significantly higher fluorescent protein expression compared to the core promoter control. Strains transformed with promoter 12 (pCRE-12) exhibited the most robust expression levels of those tested, and in some instances pCRE-12 displayed higher fluorescence intensities than the commonly used *Hsp70A-RbcS2* hybrid promoter. In this work, the *C. reinhardtii* genetic engineering toolkit was expanded with the addition of a new set of synthetic promoters with a range of expression levels. This analysis provides insight into *C. reinhardtii* promoter structure and gene regulation, as well as new DNA modules for developing second generation synthetic promoters in future experiments.

5.2. Introduction

Advancing the molecular toolkit for recombinant protein expression from the *C. reinhardtii* nuclear genome is vital for its development as a biotechnological host. Several recent successes in *C. reinhardtii* transgene expression have targeted the chloroplast genome, where 0.2–5.0% total soluble protein (TSP) can be expected for recombinant protein expression (Manuell *et al*., 2007; Rasala and Mayfield, 2015). Despite these advancements, transgene expression from the nuclear genome is necessary for more complex metabolic engineering workflows that involve multigene pathways, or require the recombinant protein to be post-translationally modified, localised to a specific organelle or secreted, as the chloroplast lacks the necessary machinery to accomplish such tasks. At present, nuclear recombinant protein expression levels in *C. reinhardtii* lag behind those

of the chloroplast, peaking at ~0.25% TSP (Rasala *et al*., 2012). This is largely due to transgene silencing (Cerutti *et al*., 1997), and a lack of strong and reliable gene promoters for the expression of nuclear transgenes is in part responsible for this setback. The *Hsp70A-RbcS2* promoter (AR-1) is regarded the strongest constitutive expression promoter available for *C. reinhardtii* (Schroda *et al*., 2000; Sizova *et al*., 2001), although it is still susceptible to transgene silencing, and expression levels are unpredictable and low compared to the chloroplast. This was observed in **Chapter 3**, where crOR expression was inconsistent (**Figure 3.6**). Additionally, relying on one good promoter to express multiple transgenes in the same organism can be problematic, as sequence-specific silencing mechanisms occurring as a result of introducing multiple copies of the same promoter can come into effect through homology-based gene silencing (Meyer and Saedler, 1996).

Synthetic promoters have overcome similar problems in several host cell systems, including bacteria (Johnson *et al*., 2018), yeast (Gertz *et al*., 2009), CHO cells (Brown *et al.* 2014) and plants (Koschmann *et al*., 2012). Promoters that are not found in nature can have several advantages, including a reduced propensity for homology-based silencing, and the potential to push gene expression to higher levels than natively-derived promoters (Venter, 2007). A recent attempt at creating synthetic promoters in *C. reinhardtii* was by and large a success (Scranton *et al*., 2016). In their study, the promoter regions of genes were scanned by POWRS motif discovery software to identify common motifs and patterns, as well as to estimate their positional biases relative to the transcription start site (TSS). The motif data was then used to generate 500 bp synthetic promoters *in silico*, which were subsequently synthesised and tested *in vivo* (Scranton *et al*., 2016). Promoters driving high levels of expression were discovered, as well as one potential *cis*-regulatory motif. This method for developing synthetic promoters was effective, but the individual promoter components that were responsible for eliciting improved expression were not identified in this study.

A common method for designing synthetic promoters involves combining known DNA *cis*-regulatory elements (CREs) that are known to recruit TFs (Ventor, 2007). This level of understanding is not currently available for *C. reinhardtii*; very few CREs in *C. reinhardtii* have been identified and characterised, and the ones that have are generally involved in inducible protein expression under nutrient-limiting conditions, as opposed to constitutive expression (Scaife *et al*., 2015). Advancing our understanding of individual CREs in *C. reinhardtii* opens up the opportunity to produce bespoke synthetic promoters with interchangeable parts, allowing tailored expression levels by combining high, low, and potentially inducible DNA motifs to optimise nuclear transgene expression. This would have the added bonus of increasing our understanding of general promoter characteristics

in algal systems. Although Scranton *et al*. (2016) did isolate a putative regulatory motif within their strongest synthetic promoter, this was discovered through laborious promoter deletion analysis.

Alternatively, identifying and testing a small set of motifs individually could quickly provide insight into which DNA sequences can be incorporated into synthetic promoters to induce transcriptional activation, and those that could be used as novel promoters in isolation. This would enable better control over synthetic promoter design in *C. reinhardtii* through understanding individual promoter components, facilitating the production of modular promoter 'building blocks' for predictable and more precise protein expression.

The aim of this chapter was to first identify putative *cis*-regulatory elements (pCREs) within the promoter regions of highly expressed genes using previously published transcriptomics data and open source motif discovery software, and then to screen these motifs *in vivo* for promoter activity and assess their suitability as standalone synthetic promoters and as modules for use in future synthetic promoter design for microalgal systems.

5.3. Results

5.3.1. Identification and bioinformatic analysis of putative *cis*-regulatory elements (pCREs) in *C. reinhardtii* promoters

The aim of this section is to identify potential TF binding sites (TFBSs), or pCREs within strong constitutive promoters in *C. reinhardtii*. The depth of knowledge regarding transcriptional regulation in *C. reinhardtii* lags behind other organisms, despite there being a wealth of transcriptomics data freely available, alongside open-source bioinformatics programs capable of detecting novel TF binding motifs within DNA sequences. These factors were exploited here by applying *de novo* motif discovery software to interrogate the promoter sequences of highly expressed genes found in a previously published RNA microarray dataset; the analytic workflow is summarised in **Figure 5.1**.

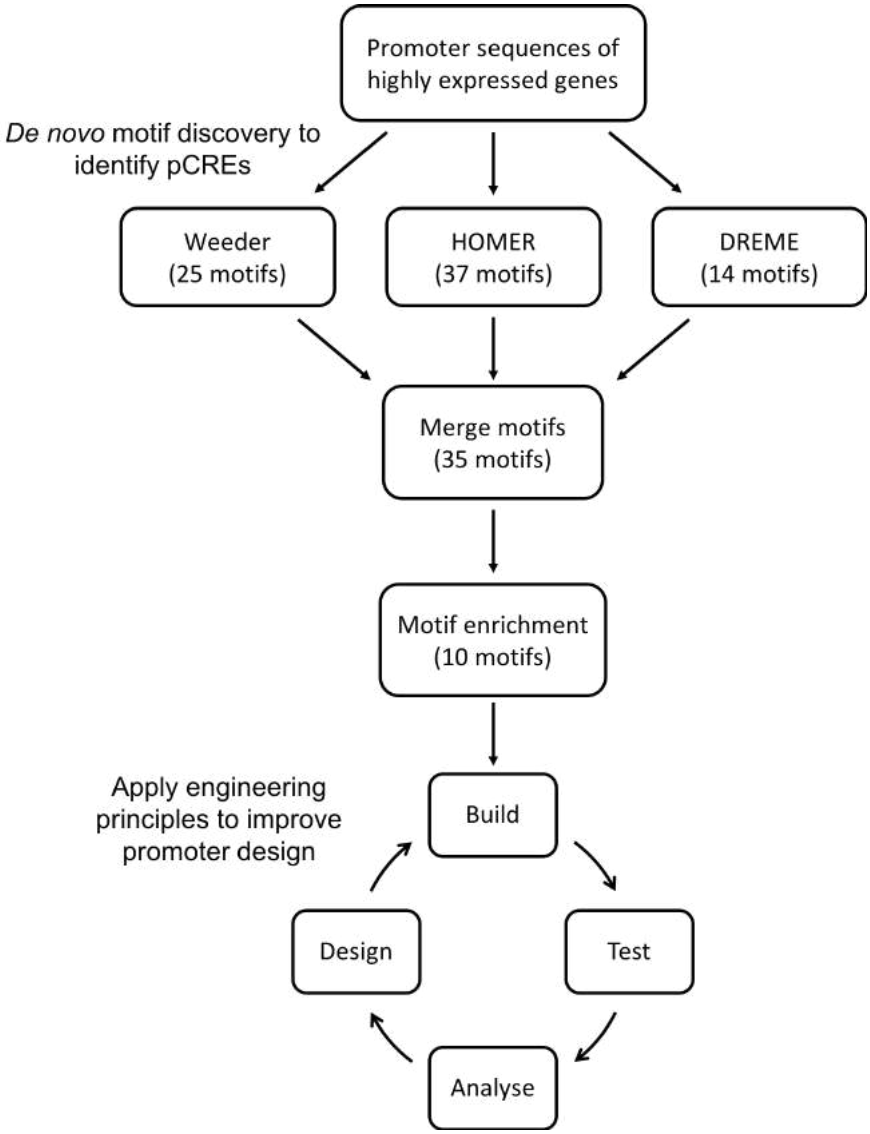


Figure 5.1: pCRE discovery and testing workflow. Figure depicts the promoter analysis pipeline applied in this chapter to identify and test putative *cis*-regulatory elements (pCREs). Motifs are discovered then refined computationally before *in vivo* testing.

5.3.1.1. Selecting a dataset for *de novo* discovery of pCREs

The first step taken towards discovering novel pCREs in *C. reinhardtii* was to identify genes that are constitutively expressed at high levels, as they are likely to contain DNA motifs within their promoter regions that strongly induce gene expression. This was achieved through analysis of a previously published microarray dataset that quantified mRNA transcripts for 11,455 nucleus-encoded genes to examine changes in gene expression between two limiting light conditions (Mettler *et al*., 2014). This study was chosen as it allows light-inducible motifs to be disqualified from the study, and the second light condition (145 μmol photons m^2 s^{-1}) closely resembles the standard conditions applied

in this work (150 μmol photons m^2 s^{-1}). The 300 genes with the most abundant RNA transcripts were selected for analysis. Scranton *et al*. (2016) considered the 50 most highly expressed genes for their analysis, which could have missed some important motifs; Hamaji *et al*. (2016) examined 300 highly expressed genes in their motif analysis, and successfully discovered the CRE responsible for zygotic gene expression, hence 300 was selected as the number of promoters to include. Genes with > 5% difference in abundance between 0 min and 480 min timepoints were regarded as differentially expressed and thus excluded from the analysis (**Figure 5.2**; Mettler *et al*., 2014).

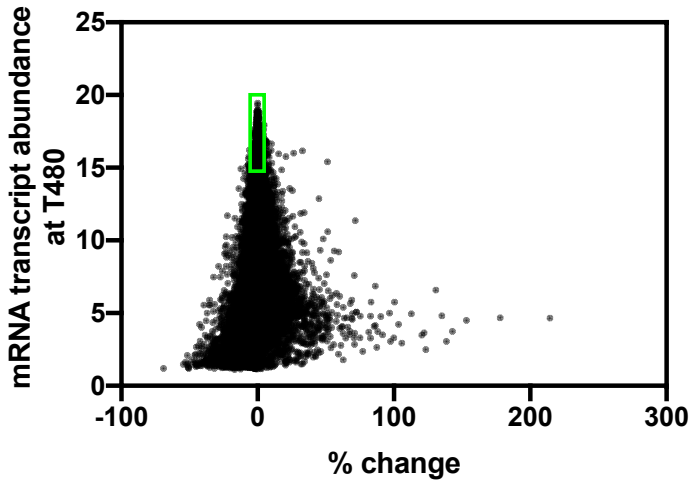


Figure 5.2: Selection of genes for promoter analysis. Scatter plot showing transcript abundance of 11,455 genes after 480 min of growth vs percentage change in transcript abundance between 0 min and 480 min timepoints. Data taken from Mettler *et al*. (2014). Green box contains data points representing genes selected for promoter analysis; the 300 genes with the highest expression that show < 5% change between time points fall within this region.

The promoter region for the selected genes was defined as −1000 bp from the TSS; of the 300 genes selected, 267 unique nucleotide sequences with characterised 5'-UTRs (from here referred to as top 267 promoters) were retrieved from Phytozome Biomart for computational analysis (For list of genes see **Appendix Table D1**). The majority of genes captured by this analysis were for ribosomal subunit proteins, and genes involved in photosynthesis and light harvesting. The promoter sequences for the 215 least abundant genes were also retrieved for comparison.

5.3.3.2. *de novo* motif discovery

The top 267 promoters were analysed using *de novo* motif discovery software to find pCREs. In simplistic terms, motif discovery programs search for short enriched sequences within a given set of oligonucleotides that could be similar enough to be recognised by the same transcription factor (D'haeseleer, 2006). Three *de novo* motif discovery programs - Weeder, HOMER and DREME - were selected for CRE discovery; using multiple programs with complementary algorithms increases the probability of finding positive hits (Tompa *et al*., 2005; Zou *et al*., 2011; Munusamy *et al*., 2017).

Weeder has long been established as one of the highest performing motif discovery programs available and was therefore selected for this study (Pavesi *et al*., 2004; Zambelli *et al*., 2014; Tompa *et al*., 2005). Weeder employs a consensus-based algorithm that attempts to identify statistically significant similarities within input nucleotide sequences when compared to random sequences. Twenty-five motifs were found to be enriched within the top 267 promoters (**Appendix Table D2**). Redundancy between motifs is apparent, for example Weeder motifs 1, 4, 5 and 13 each contain the embedded sequence 'CTCTTT'.

The HOMER (hypergeometric Optimization of Motif EnRichment) program was selected for a second round of motif discovery, as it has successfully identified motifs in several cell types including plants and *C. reinhardtii* (Heinz *et al*., 2010; Hetzel *et al*., 2016; Lu *et al*., 2017; Romero-Campero *et al*., 2016). HOMER searches for motifs of a particular length that are overrepresented in a given set of promoters when compared to a background promoter set; the background used for this experiment was built in-program using the promoter regions for all 17,743 genes for *C. reinhardtii* available from Phytozome Biomart. **Appendix Table D3** shows the 37 motifs obtained using HOMER; the top results are similar to those found using the Weeder program, with a TC-rich motif containing 'CTCTTTC' having the highest *P*-value, followed by a sequence containing the 'GCCCCATG' motif. A 'CATG' palindromic sequence frequently appears within motifs discovered by both Weeder and HOMER.

DREME (Implemented novel Discriminative Regular Expression Motif Elicitation algorithm) uses a discrete motif discovery algorithm specifically designed to find short TFBSs in large datasets. It can also be run discriminatively, where it compares two promoter sets to find motifs unique to one specified set (Bailey, 2011; Bailey *et al*., 2015). For discriminative motif selection, the promoters of the 300 lowest expressed genes from Mettler *et al*. (2014) were uploaded in lieu of a background model containing all promoters; 14 motifs were discovered (**Appendix Table D4**).

5.3.3.3. Motif clustering

In total, 76 motifs were found by the three motif finder programs, varying in length from 5–14 bp. As highlighted previously, many of the motifs are redundant, in that the same short sequences occur repeatedly within longer motifs; these short sequences are likely to represent true TFBSs, but need to be identified to prevent redundant motif testing *in vivo* (Pavesi *et al*., 2004; Zambelli *et al*., 2014). In order to reduce redundancy and condense the motifs to their core sequences, the position weight matrices (PWMs) generated for each motif were phylogenetically compared and aggregated into motif sub-clusters using Regulatory Sequence Analysis Tool (RSAT) matrix-clustering software (Castro-Mondragon *et al*., 2017). The 76 motifs were reduced to 35 sub-clusters, and a new PWM was generated for each merged motif representing the 'root motif' for each sub-cluster, which was calculated by the RSAT program through averaging the probability values for each aligned PWM within each sub-cluster. **Figure 5.3** shows an example of a motif sub-cluster tree generated by the RSAT motif-clustering program, in which six related DNA motifs found by DREME, Weeder and Homer are merged together to produce a root motif (cluster_1). Consensus sequences for 20 root motifs are listed in **Table 5.1**.

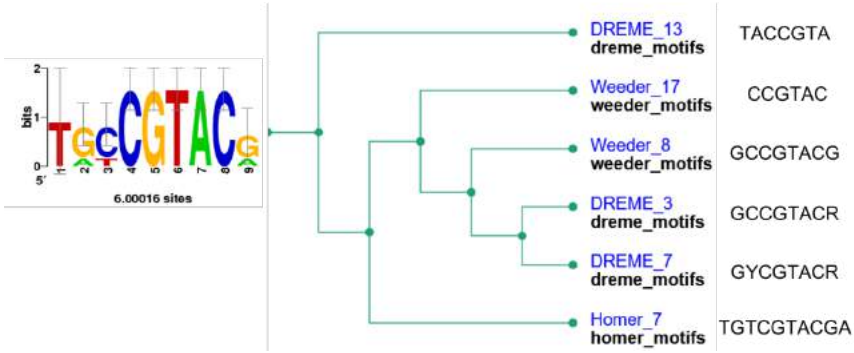


Figure 5.3: Example motif clustering tree. The motif clustering tree for Cluster_1. Six similar motifs were merged to generate a motif that represents each consensus sequence. The PWM for the root motif is represented by a sequence logo. Tree and sequence logo generated by RSAT motif clustering software (Castro-Mondragon *et al*., 2017).

Several motifs found across the three programs exhibited redundancy, and were clustered together. Clusters 1, 2, and 4 were found by all three motif discovery programs, which strongly suggests that these motifs have a role to play in gene structure and/ or regulation. Homer picked up 21 unique

motifs not found by either Weeder or DREME, whereas only two unique motifs were discovered by Weeder. All motifs found by DREME were similar to those found by Weeder and Homer.

Table 5.1: Top 20 clustered motifs

Motif Name	Motif Forward	Motif Reverse	Motifs merged list
cluster_1	TGCCGTACGA	TCGTACGGCA	DREME_13, Homer_7, Weeder_17, Weeder_8, DREME_3, DREME_7
cluster_2*	GCCCCATKCAGG	CCTGMATGGGGC	Homer_6, Homer_3, DREME_8, Weeder_2, Weeder_3, Weeder_14, Weeder_20, Homer_2, Weeder_10, Weeder_6
cluster_3	CGAGAGVC	GBCTCTCG	Weeder_18, Weeder_21, Weeder_11, Weeder_12, Weeder_9
cluster_4*	GHGAAAGARRGAGA	TCTCYYTCTTTCDC	DREME_10, DREME_2, Homer_29, Homer_1, Weeder_1, Weeder_13, DREME_4, Weeder_4, Weeder_5
cluster_5	CCTSGCC	GGCSAGG	DREME_12, DREME_5
cluster_6	SRGTMCCCC	GGGGKACYS	Homer_36, Weeder_16, Homer_28, Weeder_15, Weeder_7
cluster_7	CTCCAGGKTA	TAMCCTGGAG	DREME_6, Homer_10
cluster_8	TGTAGSCAGG	CCTGSCTACA	Homer_35, Weeder_23, Weeder_25
cluster_9*	TRTGYAGG	CCTRCAYA	DREME_14, DREME_1, DREME_11, Weeder_24
cluster_10*	CTCGGT	ACCGAG	Weeder_22
cluster_11	CRGTWCSGTGTG	CACACSGWACYG	Homer_21, Homer_34
cluster_12*	CCMTCKCGMSCVA	TBGSKCGMGAKGG	Homer_18, Homer_16, Homer_4
cluster_13*	GTATGCHTGCTG	CAGCADGCATAC	Homer_21, Homer_34
cluster_14	CCMTCKCGMSCVA	TBGSKCGMGAKGG	Homer_18, Homer_16, Homer_4
cluster_15	ACGCGGGGTA	TACCCCGCGT	Homer_13
cluster_16	AACCASGGYTAG	CTARCCSTGGTT	Homer_31
cluster_17*	GTCCACCTGG	CCAGGTGGAC	Homer_30
cluster_18	SATSSACCAGGW	WCCTGGTSSATS	Homer_8
cluster_19	GCCCTYCCAAGG	CCTTGGRAGGGC	DREME_9, Homer_9

cluster_20*	CGAGCGTTTTCT	AGAAAACGCTCG	Homer_20

All 35 motif clusters are listed in **Appendix Table D5**. Motifs taken forward for further analysis starred with an asterisk. See **Appendix Table A6** for IUPAC nucleotide base nomenclature system.

5.3.3.4. Motif enrichment

Motif clusters 1–20 (**Table 5.1**) were tested for enrichment within the promoter sequences of the top 267 genes; this was to ensure that the computationally-generated merged motifs have retained their biological relevance, and to eliminate false positives. Taking forward only the most enriched sequences for *in vivo* testing narrows the design space, and increases the likelihood of discovering a genuine TFBS. Two online programs were used to test for motif enrichment: Analysis of Motif Enrichment (AME) and CentriMo (McLeay and Bailey, 2010; Bailey and Machanick, 2012).

AME identifies user-provided motifs that are relatively enriched in a given set of promoter sequences compared to a control set (McLeay and Bailey, 2010). For this experiment, the inputted promoter sequences were shuffled to create the control. Seven of the merged motifs were shown to be enriched within the highly expressed promoter set relative to the shuffled control, with *P*-values < 0.05 according to Fisher's exact test (**Table 5.2**).

Table 5.2: AME Results

Rank	Cluster ID	Consensus	*P*-value
1	cluster_4	GHGAAAGARRGAGA	2.84E-20
2	cluster_12	CCMTCKCGMSCVA	7.65E-12
3	cluster_5	CCTCGCC	1.74E-10
4	cluster_2	GCCCCATGCARG	1.28E-07
5	cluster_9	TRTGYAGG	4.73E-06
6	cluster_13	GTATGCHTGCTG	7.49E-05
7	cluster_18	SATGSACCAGGW	1.31E-04

Inputted motifs: clusters 1–20. Only shows clusters with an AME enrichment score of $P < 0.05$, as generated using Fisher's exact test. Used shuffled top267 sequences FASTA file as control.

CentriMo, similarly to AME, identifies relatively enriched motifs within a sequence set, but additionally determines whether a motif has a particular bias towards a location within a given set of sequences of the same length (Bailey and Machanick, 2012). Five merged motifs were found to be enriched with a positional bias relative to the TSS by CentriMo, two of which (Clusters 2 and 4) were also found to be enriched using the AME program. Cluster_2 has a strong positional bias around −118 bp from the TSS, whereas Cluster_4 is highly likely to be found −37 bp from the TSS (**Figure 5.4**). Clusters 10, 17 and 20 were found to have statistically significant positional biases further upstream from the TSS within promoter regions of highly expressed genes (**Table 5.3**).

Table 5.3: CentriMo Results

Cluster ID	Consensus	*E*-value	Bin Centre from TSS
cluster_2	GCCCCATGCARG	9.40E-12	−118.5
cluster_4	GHGAAAGARRGAGA	6.40E-19	−36.5
cluster_10	CTCGGT	3.80E+00	−233.5
cluster_17	GTCCACCTGG	8.90E+00	−762
cluster_20	CGAGCGTTTTCT	7.00E+00	−305

Inputted motifs: clusters 1−20. Only shows motifs with an *E*-value < 10. Bin centre from TSS represents the centre of the site where the motif can be found with the highest probability.

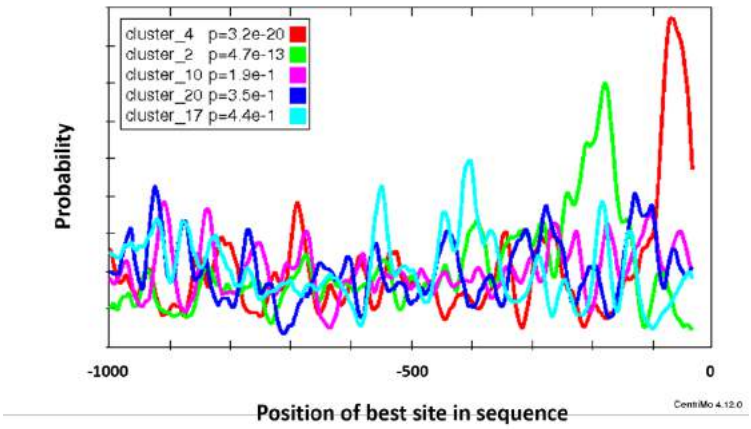


Figure 5.4: CentriMo positional biases of five enriched motif clusters. Shows the probability of each motif being found in each site within the 1000 bp promoter region, relative to the TSS. Figure generated using CentriMo (Bailey and Machanick, 2012).

5.3.3.5. Analysis of previous synthetic promoters

Scranton *et al*. (2016) successfully produced synthetic promoters. The promoters were computer-generated, and although this was based on found consensus sequences, the individual motifs discovered within this study were not isolated and characterised *in vivo*. A quick analysis of the 25 highest expression promoters created by Scranton *et al*. (2016) using Multiple Expectation maximization for Motif Elicitation (MEME) motif discovery (Bailey and Elkan, 1994) highlighted some of the common motifs present (**Table 5.4**). S2016_MEME motifs 2, 4 and 5 appear to be redundant, and to closely resemble cluster_2. There is also strong similarity between S2016_MEME_3 and cluster_4.

Table 5.4: Common motifs in Scranton *et al*. (2016) synthetic promoters

Motif name	Motif Forward	Motif Reverse	E-value
S2016_MEME_1	GGGCCCATTC	GAATGGGCCC	4.40E-10
S2016_MEME_2	CGCATGGGGC	GCCCCATGCG	7.80E-09
S2016_MEME_3	TCTCTTTCTCTT	AAGAGAAAGAGA	2.40E-06
S2016_MEME_4	TGCATGGGGC	GCCCCATGCA	8.80E-06
S2016_MEME_5	GCCCCATGCA	TGCATGGGGC	1.80E-08
S2016_MEME_6	GAGCGAGCGC	GCGCTCGCTC	8.20E-01
S2016_MEME_7	GCAAGCAAGT	ACTTGCTTGC	8.40E+00

5.3.3.6. Motif selection for *in vivo* analysis

Guided by the enrichment results, the ten motif clusters found to be significantly enriched within the highly expressed promoter sequences were selected for *in vivo* analysis (**Figure 5.5**). In addition to these, the 'CCCATGCGA' motif discovered by Scranton *et al*. (2016) was selected for individual motif analysis, as well as S2016_MEME_1 and _2 (**Table 5.4**). A random 10 bp DNA sequence with a similar GC content to the other motifs was generated to use as a control; this motif is not significantly enriched in the top promoter sequences (AME *P*-value = 1, no sequence matches). **Figure 5.5** displays all PWMs of motifs selected for *in vivo* testing in the form of sequence logos. All

motifs were renamed putative *cis*-regulatory element (pCRE) 1–13, and will be referred to as such for the rest of this chapter (**Figure 5.5**).

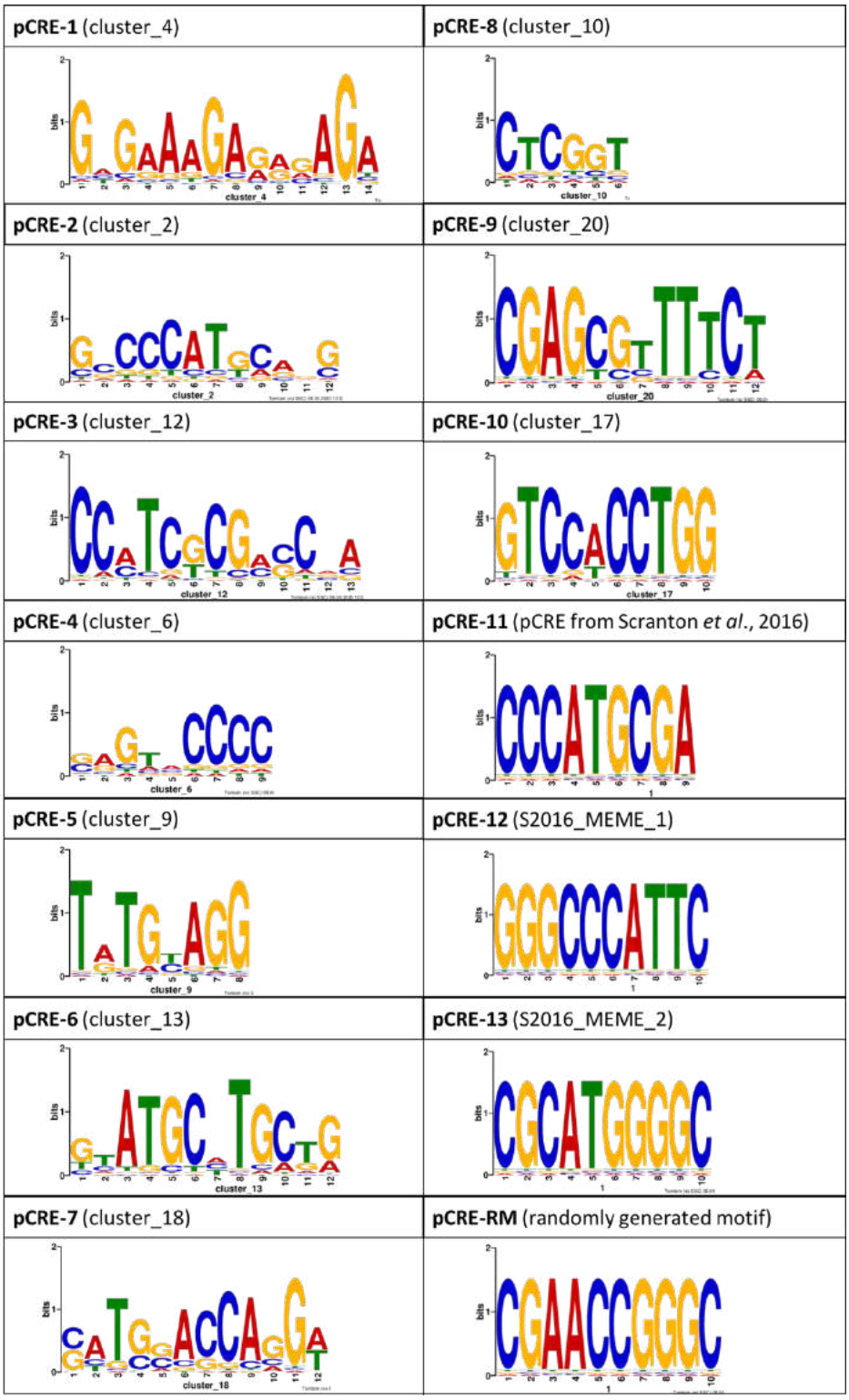


Figure 5.5: pCRE motifs selected for *in vivo* analysis. Position weight matrices are represented by sequence logos that were produced using MEME suite (Bailey *et al*., 2015).

5.3.3.7. Comparison of selected pCREs to previously reported motifs

The motifs listed in **Figure 5.5** were compared to known plant TFBSs, as CREs are often well conserved across related species and even different kingdoms of life (Patikoglou *et al*., 1999; Burgess and Freeling, 2014). PLACE is a database of motifs found in plant *cis*-acting regulatory DNA elements compiled from previously published data (Higo *et al*., 1999). **Table 5.5** shows the PLACE TFBSs found within each pCRE motif. The CTRMCAMV35S motif in CRE_1 is a TC-rich motif found downstream of the TSS in the cauliflower mosaic virus promoter sequence that can enhance gene expression in plants (Pauli *et al*., 2004). SORLIP2AT motif, which stands for Sequences Over-Represented in Light-Induced Promoters (SORLIPs) in *A. thaliana* (Hudsen and Quail, 2003), is present in CRE-2, -12 and -13. CURECORECR, present in CRE-4, is the core motif in a *C. reinhardtii* that elicits responses to copper and oxygen deficiency (Kropat *et al*., 2005).

Table 5.5: Search results for each pCRE in PLACE database

Motif name	PLACE TF site	Signal Sequence	PLACE ID
CRE-1	CTRMCAMV35S	TCTCTCTCT	S000460
	NODCON2GM	CTCTT	S000462
	OSE2ROOTNODULE	CTCTT	S000468
	DOFCOREZM	AAAG	S000265
	POLLEN1LELAT52	AGAAA	S000245
CRE-2	SORLIP2AT	GGGCC	S000483
CRE-3	BOXLCOREDCPAL	ACCWWCC	S000492
	MYBPZM	CCWACC	S000179
CRE-4	CACTFTPPCA1	YACT	S000449
	CURECORECR	GTAC	S000493
CRE-5	No result	-	-
CRE-6	RYREPEATVFLEB4	CATGCATG	S000102
	RYREPEATLEGUMINBOX	CATGCAY	S000100

	RYREPEATGMGY2	CATGCAT	S000105
	RYREPEATBNNAPA	CATGCA	S000264
CRE-7	No result	-	-
CRE-8	SURECOREATSULTR11	GAGAC	S000499
	DRE1COREZMRAB17	ACCGAGA	S000401
CRE-9	POLLEN1LELAT52	AGAAA	S000245
CRE-10	SITEIOSPCNA	CCAGGTGG	S000224
	EBOXBNNAPA	CANNTG	S000144
	MYCCONSENSUSAT	CANNTG	S000407
	RAV1BAT	CACCTG	S000315
CRE-11	No result	-	-
CRE-12	SORLIP2AT (x2)	GGGCC	S000483
	SITEIIATCYTC	TGGGCY	S000474
CRE-13	SORLIP2AT	GGGCC	S000483
CRE-RM	SORLIP2AT	GGGCC	S000483

See **Section 2.9.3.** for running details and links to PLACE IDs. IUPAC nomenclature table in **Appendix Table A6**. Consensus sequences were identified based on exact matches using a signal scan program.

5.3.2. Method development and optimisation for *cis*-regulatory element testing

5.3.2.1. Reporter selection

To test the pCRE motifs selected in **Figure 5.5** for promoter activity, a fluorescent protein reporter system was designed. Fluorescent proteins have been extensively applied as simple but effective tools for measuring gene expression (Fuhrmann *et al*., 1999; Onishi and Pringle, 2016; Scranton *et al*., 2016; Dong *et al*., 2017), and are relatively easy to detect without requiring expensive reagents, such as luciferase and secreted alkaline phosphatase assays. For this experiment two fluorescent proteins, mCherry and mVenus, were considered for measuring promoter strength. **Figure 5.6** shows the fluorescence spectra for mCherry and mVenus fluorescent proteins alongside chlorophyll-*a* and chlorophyll-*b*. Excitation and emission of both mVenus and mCherry fall between the ranges of chlorophyll-*a* and -*b*, so both are theoretically suitable for expression and detection in *C. reinhardtii*.

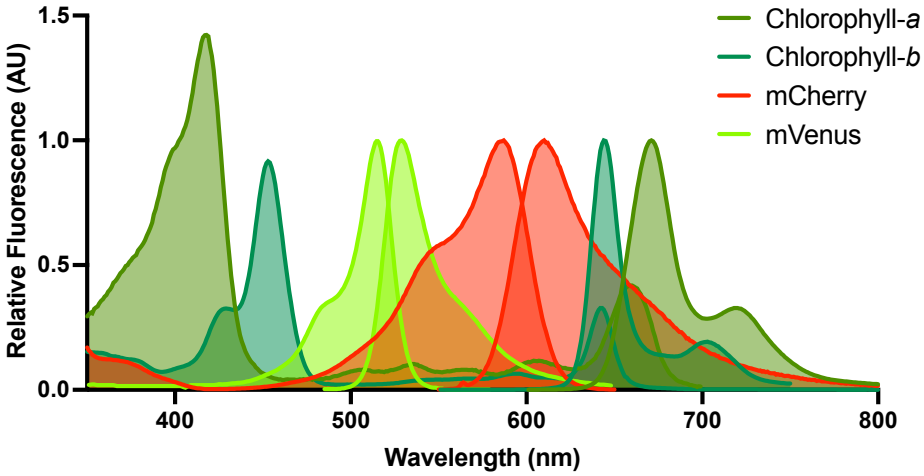


Figure 5.6: Fluorescence spectra for fluorescent proteins and chlorophyll-*a* and -*b*. Excitation and emission spectra. mVenus excitation and mission maxima: 515 nm and 529 nm. mCherry excitation and emission maxima: 587 nm and 610 nm .csv file containing spectral data downloaded from https://www.fpbase.org/spectra/ 06-04-2020

The mCherry protein is relatively small in size (236 aa) and has recently been used as a reporter to quantify transgene expression in *C. reinhardtii* (Rasala *et al*., 2014; Lauersen *et al*., 2015; Lauersen *et al*., 2016; Onishi and Pringle, 2016; Scranton *et al*., 2016). The distance between the excitation and emission peaks for mCherry (23 nm) is larger than that of mVenus (14 nm); with equipment set-up in mind, where 20 nm bandpass filters are present in the plate reader available, mCherry was initially selected as the protein reporter for fluorescent activity.

mRNA was not selected as the readout for this study, as previous reports have shown that despite high mRNA readouts, protein expression can still be poor (Lumbreras *et al*., 1998; Kong *et al*., 2015). It therefore seemed reasonable to bypass this step, as the desired outcome of the synthetic promoter systems is for robust and stable protein expression, not just transcription.

5.3.2.2. mCherry vector construction

The mCherry expression vector pOpt_mCherry (**Figure 5.7**) was created by replacing the mVenus gene in pOpt_mVenus_Paro (**Figure 2.1**) with mCherry. Although pOpt_mVenus_Paro already contained the fluorescent report gene mVenus, mCherry was inserted to expand the range of reporter vectors available for this project. For details about vector construction, see **Appendix Figure D1**. The correct insertion of the mCherry gene was confirmed by sequencing.

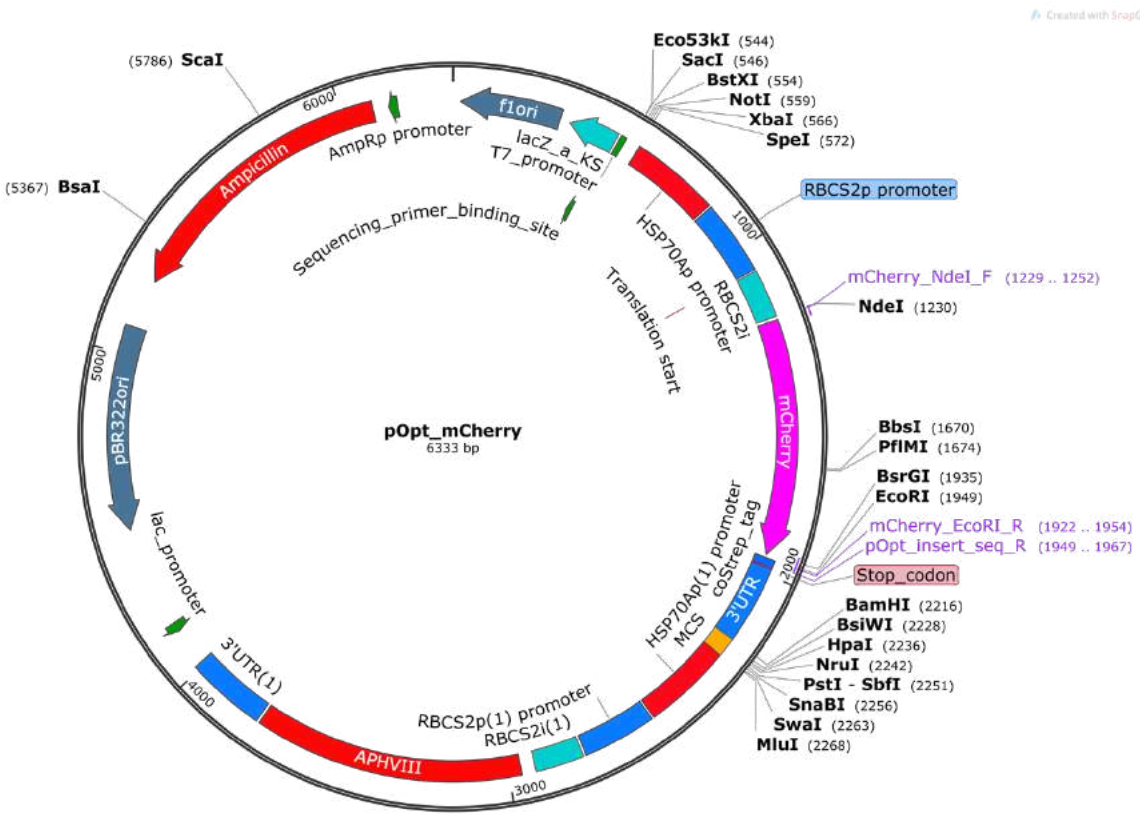


Figure 5.7: Vector map of pOpt_mCherry. Plasmid map showing insertion of the mCherry gene (pink) between *NdeI* and *EcoRI* restriction sites. Translation start site is depicted at the 5'-end of the *RbcS2* promoter region (turquoise). Primers shown in purple. Image created using SnapGene.

5.3.2.3. mCherry fluorescent reporter is not suitable for promoter testing

pOpt_mCherry was linearised with the ScaI restriction enzyme and 1 µg DNA was transformed into *C. reinhardtii* strain CC-4533 (wild-type; WT). Paromomycin resistant colonies were picked and resuspended individually in 96-well plates, alongside untransformed WT colonies as a control. Following 5 days' growth, this 'master plate' was used to inoculate three identical 96-well plates, which, after 5 more days of growth under standard conditions, were examined for chlorophyll and mCherry fluorescence. **Figure 5.8** shows the mCherry fluorescence readings relative to chlorophyll fluorescence. Surprisingly, the median mCherry fluorescence for the transformant strains was lower than that of WT, despite there being some outliers with higher fluorescence in the mCherry transformant pool. This could indicate low co-transformation efficiency of paromomycin and mCherry (as observed in **Section 3.3.2.**), low expression levels of the mCherry protein, an ineffective detection method, or a combination of the three. Large differences were observed between

readings of replicate samples despite the replicates having been grown under the same conditions, even after normalising values by growth changes based on chlorophyll fluorescence.

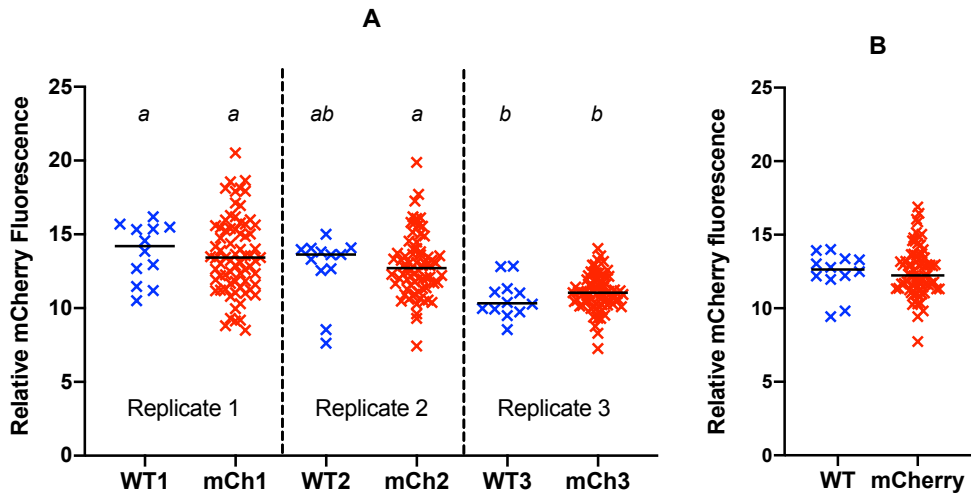


Figure 5.8: mCherry fluorescence readings for test transformation of pOpt_mCherry grown in 96-well plates. A - Relative mCherry fluorescence readings for each 96-well plate replicate. Populations that are significantly different are denoted *a* and *b* (Tukey's test $P < 0.05$). **B** – Average relative mCherry fluorescence readings for WT and mCherry transformants, calculated from the three replicates shown in **A**. Readings from individual wells are indicated by crosses. Bar shows median value for population. Relative fluorescence was calculated by subtracting blank readings (TAP media only) from mCherry fluorescence measurements; the resulting mCherry value was divided by chlorophyll fluorescence to correct for differences in growth between replicates. mCherry measured at Ex561 Em610. Chlorophyll measured at Ex440 Em680.

To explore this problem in more detail, the co-transformation efficiency of the AphVIII and mCherry genes was examined by PCR screening 92 paromomycin-resistant transformant colonies; 16 were shown to contain the mCherry gene (See **Appendix Figure C2** for gels), giving a co-transformation efficiency of ~17%, which is comparable to the 14% co-transformation efficiency observed for pOpt_crOR (**Section 3.3.2.**). Each of the 16 positive mCherry transformants were grown on 96-well plates in triplicate for 3 days; as shown in **Figure 5.9**, the strains did not exhibit higher relative mCherry fluorescence compared to the WT. The highest fold change observed here is in mCh-4, which exhibits 1.7-fold higher fluorescence than WT, but this is still very low and not significantly higher.

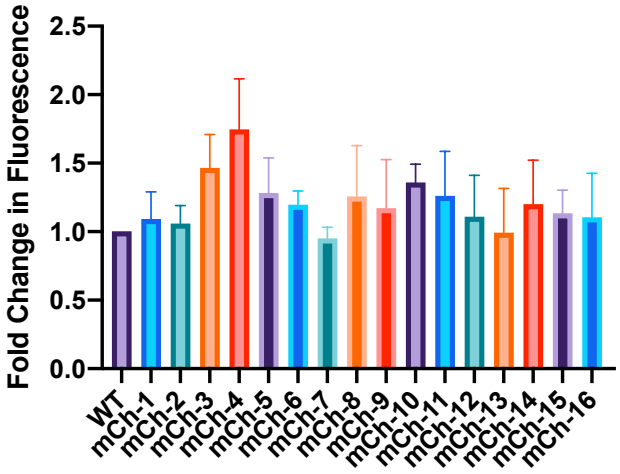


Figure 5.9: Sixteen positive mCherry transformants grown on 96-well plates. mCherry was measured at wavelengths Ex561/Em610 nm, gains 150. No significant difference was observed between means.

Despite mCherry expression being driven by the robust AR-1 promoter, its expression was not high enough to detect potentially subtle differences in expression driven by pCRE motifs. This could be because the mCherry gene used for this study was not codon optimised for *C. reinhardtii* (mCherry was codon optimised for expression in *E. coli*), and did not contain the RbcS2 intron (iRbcS2), which is a basic tool for increasing gene expression in *C. reinhardtii* (Lumbreras *et al*., 1998; Baier *et al*., 2019), embedded within the gene. This variant of mCherry was therefore rejected as the reporter protein for this study.

5.3.2.4. Replacement of reporter gene

The original pOpt_mVenus_Paro vector was tested for its suitability as a reporter for motif analysis, as the mVenus gene here is codon optimised and contains iRbcS2 (Lauersen *et al*., 2015). WT *C. reinhardtii* was transformed with 1 µg of ScaI-linearised pOpt_mVenus_Paro. Paromomycin resistant colonies were picked on to 96-well plates into non-selective TAP media, and measured directly after 2 days' growth to get a rough indication of transformation efficiency and ability to detect mVenus using a fluorescent plate reader. **Figure 5.10** shows the fold difference in mVenus fluorescence from WT for each individual transformant. Of 203 transformants picked, 10 strains (~5%) exhibited > 2.5-fold higher mVenus fluorescence. These strains were scaled up to 6-well plates and measured again using a plate reader; 7 of the strains retained their high fluorescence (**Appendix C**), and following PCR screening were shown to be positive transformants for mVenus (**Appendix C**).

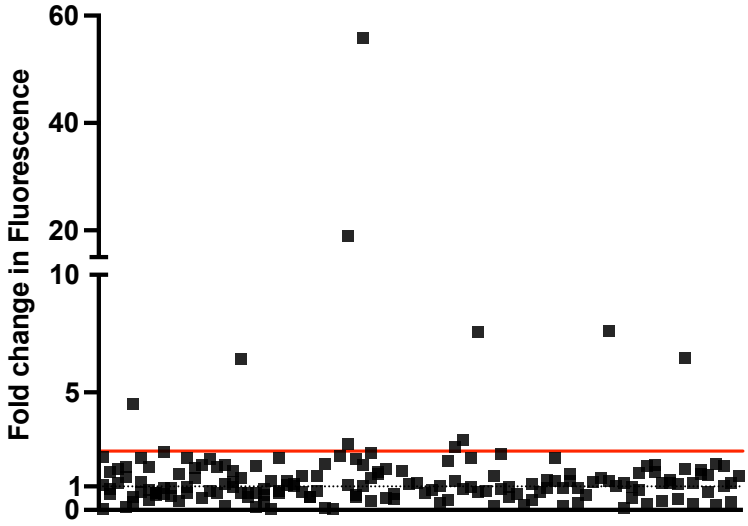


Figure 5.10: Fold change in mVenus fluorescence compared to WT for individual pOpt_mVenus_Paro transformants. mVenus detected at wavelengths Ex500 nm Em550 nm. Chlorophyll detected at Em440/Ex680 nm. mVenus fluorescence normalised to chlorophyll fluorescence to counteract growth differences. Results displayed as fold change from the average readout for WT. Red line shows 2.5-fold increase from WT mean.

The two highest expressing mVenus-positive strains were examined under a confocal microscope to validate mVenus expression at the single-cell level (**Figure 5.11**)**.** mVenus fluorescence was detected for individual cells that had been transformed with the pOpt_mVenus_Paro vector, but not in the WT strain.

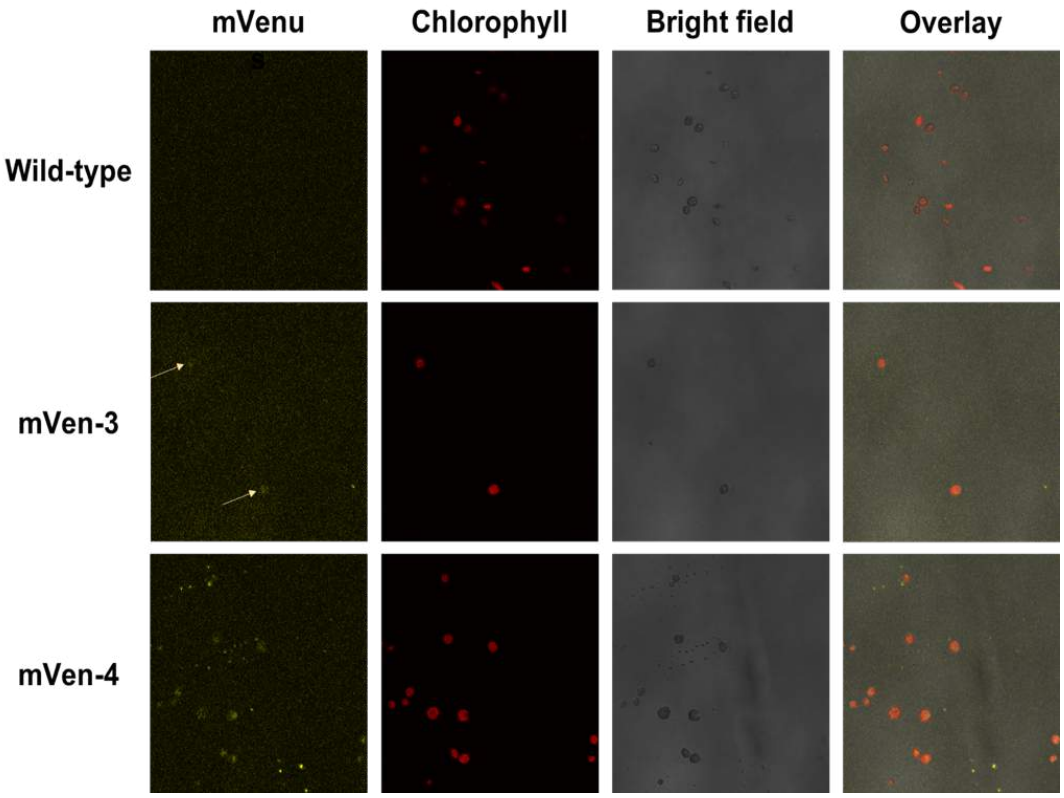


Figure 5.11: Confocal images showing mVenus expression in of two pOpt_mVenus_Paro transformant strains compared to WT. Laser wavelengths for excitation and detection were as follows: Ex488 nm; channel 1 mVenus detection, Ex500–550 nm; channel 2 chlorophyll detection, Em650–700 nm. X40 magnification.

mVenus expression could be detected at the population level (plate reader) and at the single cell level (confocal), unlike mCherry, and mVenus transformants retained their expression after sub-culturing. mVenus was therefore selected as the reporter gene for measuring synthetic promoter pCRE activity in this work.

This study, however, suggested that using a plate reader alone to measure fluorescence for synthetic promoter activity could be problematic. The number of transformants expressing mVenus to significantly detectable levels (1 SD above the WT mean) was low, at approximately 5%. Given that 10–50 clones of each promoter type should be measured to overcome expression variation occurring as a result of random genomic insertion (Scranton *et al*., 2016), this would require picking around 1000 colonies per promoter, amounting to around 12x 96-well plates. In order to test the 14 motifs in **Figure 5.5** plus two controls, this would require picking 16,000 colonies on to 192x 96-well plates; without the help of a robot, this would be an extremely arduous, expensive and time-consuming task that is beyond the scope of this project. It was concluded that solely using a plate reader to measure fluorescence was unsuitable for this experiment.

Flow cytometry is a high-throughput method for rapidly measuring fluorescence signals of single cells, and is capable of processing 1000s of cells per minute. This technique has the potential to overcome the issues highlighted above: 100s of individual clones can be pooled together, and fluorescence signals representing the collective transformant population can be rapidly assessed and averaged, thus reducing the noise generated through random transgene integration. Flow cytometry has recently been applied to measure fluorescent protein expression in *C. reinhardtii* (Rasala *et al*., 2013; Scranton *et al*., 2016; Barjona do Nascimento Coutinho *et al*., 2017), and was therefore selected as the fluorescence measurement method.

5.3.3. Design and construction of synthetic promoter vectors

5.3.3.1. Overall design of synthetic promoter vectors

This section describes the design and construction of the synthetic promoter reporter vectors. The basal vector pOpt_mCherry (**Figure 5.7**) was modified to remove the AR-1 promoter region upstream of mCherry, and to replace this with a core promoter region and a site for insertion of each of the pCRE motifs listed in **Figure 5.5** upstream of the core promoter. The RbcS2 intron 1 was retained upstream of the mCherry gene to enhance gene expression. This section of work was completed before and during the suitability testing for mCherry as a reporter (**Section 5.3.2.3.**), and it was at this time unknown that mCherry would be unsuitable for this experiment. The subsequent alteration of the vectors to include the mVenus reporter gene is included in this section.

5.3.3.2. Core promoter selection

The core promoter contains the minimal DNA elements required to initiate transcription; this forms the basic component of the synthetic promoters designed in this section. For the purpose of this study, the core promoter is defined as the DNA region –50 bp upstream of the TSS. The first 50 bp (from 3' end) of the SAP-11 synthetic promoter generated by Scranton *et al*. (2016) was used as the core promoter in the following vectors; their work shows that this region drives mCherry expression to detectable levels, but that are lower than when the rest of the promoter is present. This should therefore contain the minimum sequences necessary to recruit RNA polymerase II and the rest of the preinitiation complex, and was selected as a measure for baseline reporter protein expression. This core promoter is not found in nature, and would theoretically be less susceptible to homology-induced transgene silencing. The core promoter DNA sequence can be found in **Table 2.3**.

5.3.3.3. Building core vectors

pOpt_mCherry (**Figure 5.7**) was used as a scaffold for building the synthetic promoter vectors. The vector pOpt_Core_mCherry (**Figure 5.12**) was created using the workflow depicted in **Figure 5.13**, where the AR-1 promoter region was removed and replaced with the core promoter, with new cut sites included for the insertion of proximal promoter elements upstream of the core region. Gels for this section can be found in **Appendix D**.

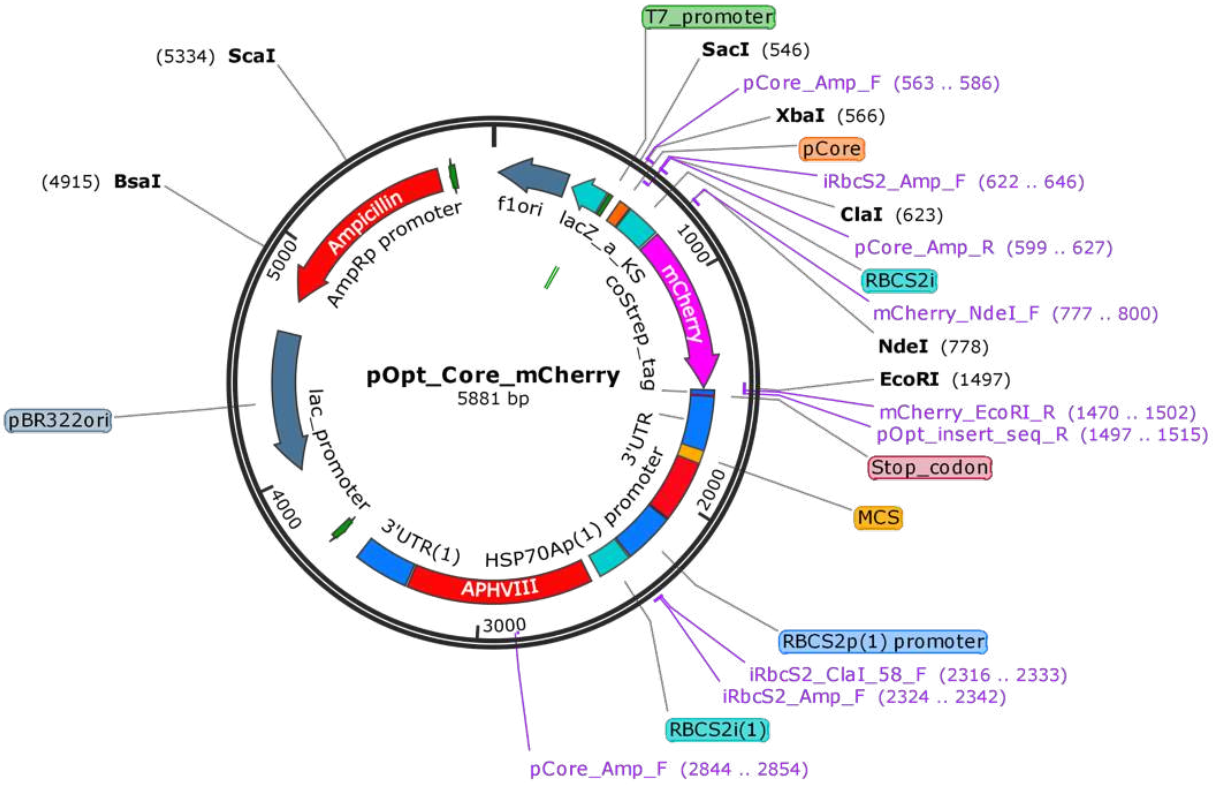


Figure 5.12: Vector map of pOpt_Core_mCherry. mCherry gene highlighted in pink. Primers and their binding sites are denoted in purple. Restriction sites are shown in bold.

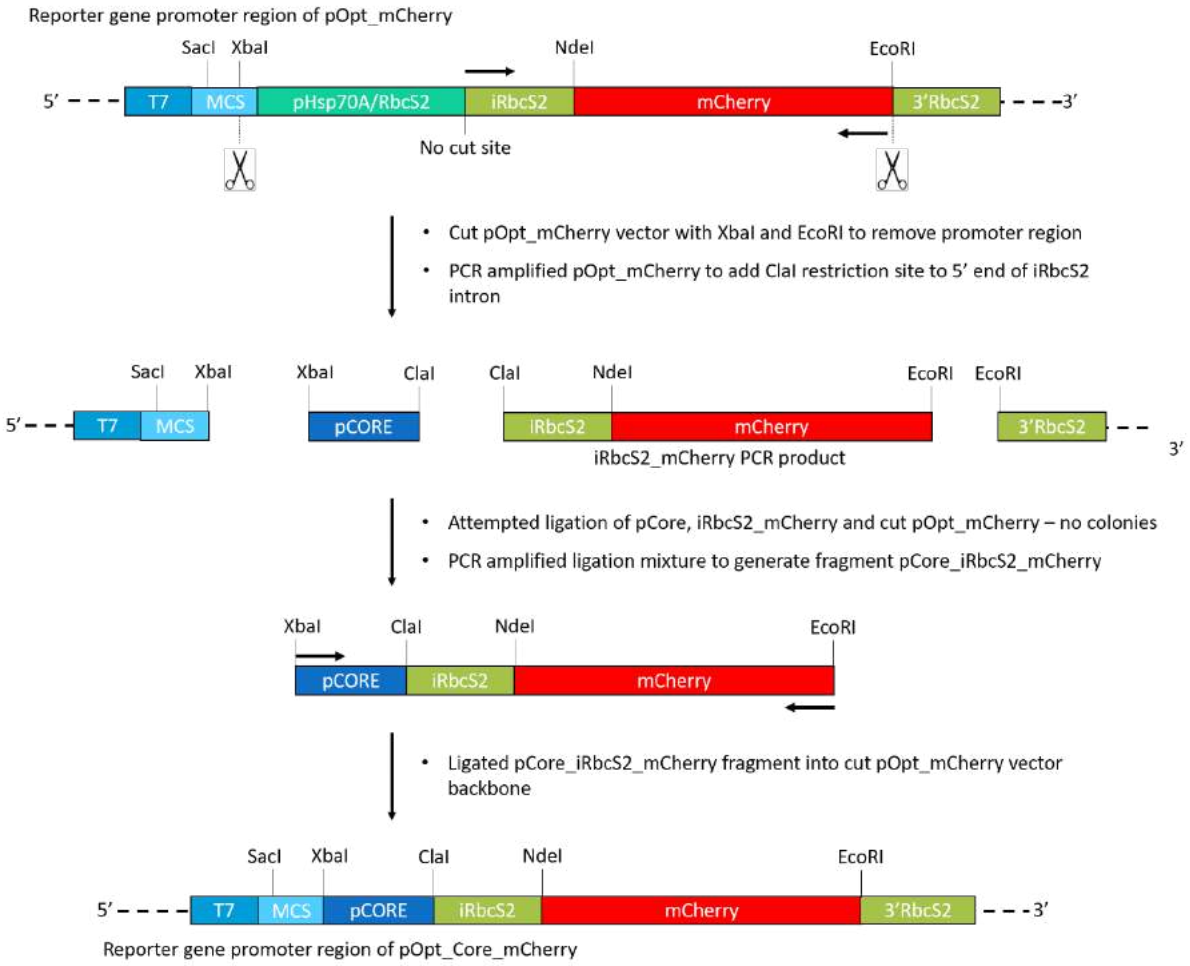


Figure 5.13: Cloning strategy to create pOpt_Core_mCherry vector. T7 = bacterial T7 promoter; MCS = multiple cloning site; pHsp70A/RbcS2 = AR-1 promoter; iRbcS2 = RbcS2 intron; mCherry = mCherry CDS; 3'RbcS2 = RbcS2 3'-untranslated region; pCORE = core promoter. Horizontal black arrows show primer annealing sites. The bacterial T7 promoter was noted at the 5' end, as a forward T7 primer was used for vector sequencing. Fragment sequences in **Appendix D**.

For improved expression, the iRbcS2 intron was kept within the construct upstream of the reporter gene *Nde*I insertion site (Lumbreras *et al*., 1998); however, no cut site was present between the iRbcS2 intron and the pRbcS2 promoter region. This was resolved by PCR amplifying pOpt_mCherry with primers iRbcS2_Amp_F and mCherry_EcoRI_R (**Table 2.2**) to introduce a *Cla*I restriction site upstream of the iRbcS2 intron, resulting in the iRbcS2_mCherry DNA fragment (**Figure 5.13**).

To generate the pCore fragment (**Figure 5.13**), a 50 bp ssDNA template containing the pCore sequence (**Table 2.3**) was PCR amplified using primers pCore_Amp_F and pCore_Amp_R (**Table 2.2**) which introduced *Xba*I and *Cla*I cut sites, and the 70 bp pCore PCR product was gel purified.

To generate the pOpt_Core_mCherry vector, pOpt_mCherry was digested with XbaI and EcoRI to remove the entire AR-1 promoter and mCherry region (**Figure 5.13**), leaving a 4950 bp vector backbone. The iRbcS2_mCherry PCR fragment was digested with ClaI and EcoRI, and the pCore fragment digested with ClaI and XbaI, for insertion into the pOpt vector backbone (**Figure 5.12**).

Ligation reactions were attempted with the three cut fragments shown in **Figure 5.13** in a single step; following transformation into competent *E. coli* DH5α cells, no colonies were present. Assuming part of the ligation mixture contained correctly ligated pCore + iRbcS2_mCherry fragments, 0.3 µL of the ligation mixture was PCR amplified using primers pCore_Amp_F and mCherry_EcoRI_R (**Table 2.2**) to generate the new fragment pCore_iRbcS2_mCherry (**Figure 5.13**). pCore_iRbcS2_mCherry was gel purified and digested with XbaI and EcoRI; following PCR clean-up of the digested fragment, pOpt_mCherry cut with XbaI and EcoRI was ligated with similarly digested pCore_iRbcS2_mCherry to generate the pOpt_Core_mCherry vector (**Figure 5.12**). Minipreps of the constructed vector were confirmed by sequencing.

The core promoter vector pOpt_Core_mCherry was then ready for use as a control vector, and as the foundation for inserting each pCRE from **Figure 5.5** into the proximal promoter region.

4.3.3.4. Building pCRE reporter vectors

DNA fragments of ~70 bp, each containing repeats of motifs from **Figure 5.5**, were then amplified for insertion upstream of the core promoter in pOpt_Core_mCherry to produce individual reporter vectors for testing (DNA sequences of blocks listed in **Table 2.3**). The closest consensus sequence for each pCRE motif (**Figure 5.5**) was synthesised as ssDNA in as many copies as could fit within the 70 bp size limit, to a maximum of seven repeats (**Table 5.6**). In yeast, increasing the copy number of motifs has a positive effect on transcription, then tends to saturate after ~4 copies (Sharon *et al*., 2012); assuming this is also the case in *C. reinhardtii*, increasing the number of motifs past 4 would enable optimal expression, while not affecting expression significantly through the addition of more motif copies. A 70 bp limit was placed on the motif fragment size, as this is long enough to allow 5–7 repeats of each motif but also to control for promoter size. The 70 bp limit was also chosen for practical purposes, as purchasing ssDNA < 100 bp in size was more cost effective, and DNA sequences much smaller than 70 bp can be difficult to purify without a specialised kit.

Table 5.6: Motif consensus sequences used in synthetic promoters

Motif	Consensus F	Consensus R	Motif length (bp)	# repeats	Length of promoter (bp)
pCRE-1	TCTCTCTCTT	AAGAGAGAGA	10	6	66
pCRE-2	GCCCCATGAGG	CCTCATGGGGC	11	5	65
pCRE-3	TTGGTCGCGATGG	CCATCGCGACCAA	13	5	75
pCRE-4	GGGGTACTC	GAGTACCCC	9	7	73
pCRE-5	TATGTAGG	CCTACATA	8	7	66
pCRE-6	GCATGCATGCTG	CAGCATGCATGC	12	5	70
pCRE-7	CATGGACCAGGA	TCCTGGTCCATG	12	5	70
pCRE-8	CTCGGT	ACCGAG	6	7	52
pCRE-9	CGAGCGTTTTCT	AGAAAACGCTCG	12	5	70
pCRE-10	GTCCACCTGG	CCAGGTGGAC	10	6	70
pCRE-11	CCCATGCGA	TCGCATGGG	9	7	73
pCRE-12	GGGCCCATTC	GAATGGGCCC	10	6	70
pCRE-13	CGCATGGGGC	GCCCCATGCG	10	6	70
pCRE-RM	CGAACCGGGC	GCCCGGTTCG	10	6	70

Consensus F = forward consensus sequence (DNA sequence used to construct test vectors). Consensus R = Reverse complement of consensus sequence. # repeats = number of motif repeats in the test promoter sequence.

Figure 5.14A shows the design of the ssDNA motif templates. Multiple copies of each motif are flanked by restriction sites for their insertion into pOpt_Core_mCherry, upstream of the pCore region (**Figure 5.14B; Figure 5.12**). These restriction sites are flanked at the 5' and 3' ends with common extension sequences, enabling amplification of each ssDNA pCRE template by the same primer set: pCRE_Amp_F and pCRE_Amp_R (**Table 2.2**).

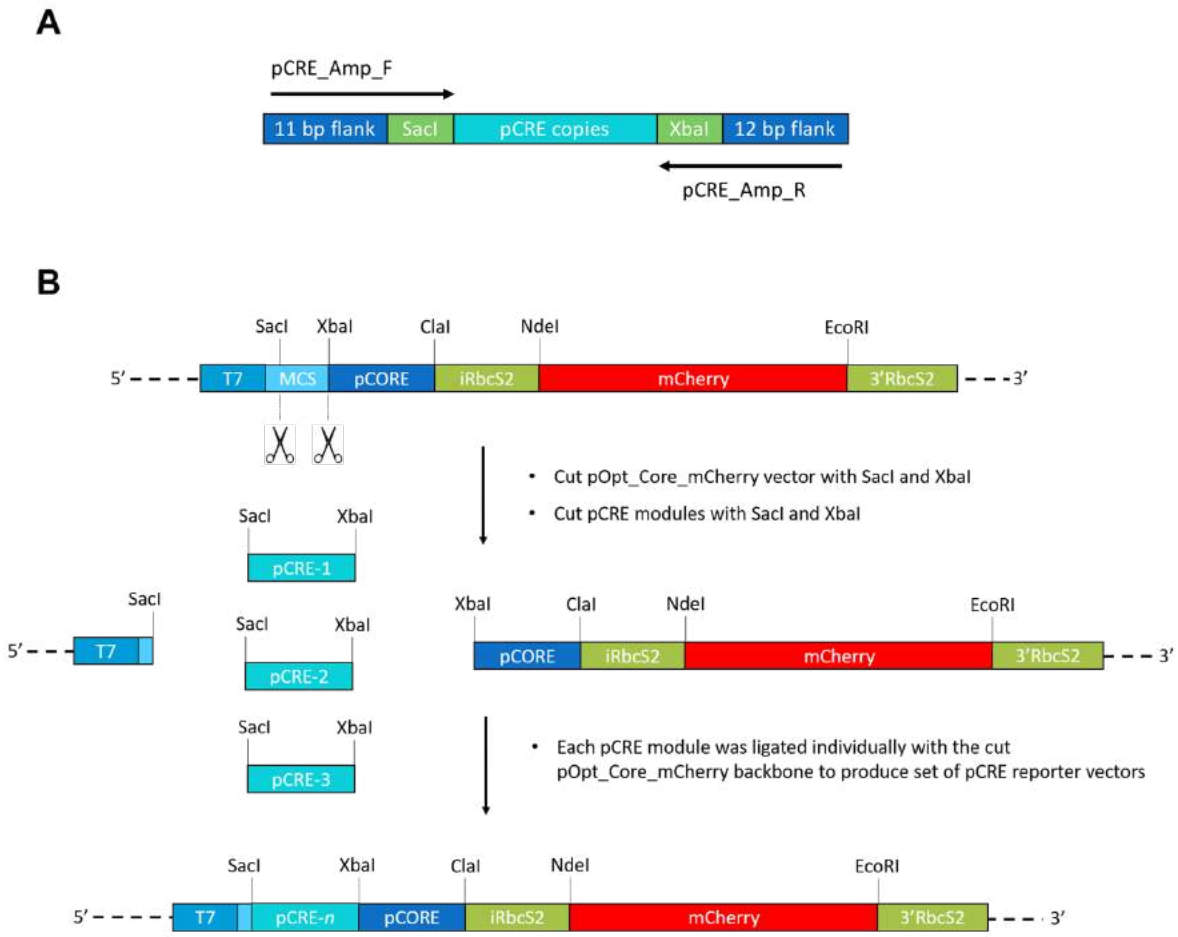


Figure 5.14: PCR design for amplification of pCRE modules for synthetic promoter reporter vectors. A - ssDNA template design for each individual pCRE module. The 11 bp and 12 bp flanks, as well as the restriction sites SacI and XbaI are common to each of the pCRE modules; this facilitated amplification of each module with the same pair of primers. **B** - a schematic diagram of the construction of the pCRE_mCherry reporter vector suite.

Following amplification, digestion (with enzymes SacI and XbaI), and gel purification of the individual pCRE modules, pOpt_Core_mCherry was digested with the same restriction enzymes and ligated to each individual pCRE, forming 14 pCRE_mCherry vectors (**Figure 5.14**). For a list of the vectors created, see **Appendix D**). Each vector contains the common core promoter region upstream of the mCherry reporter gene, and a distinct proximal promoter sequence upstream of the promoter core. For gels associated with vector construction, see **Appendix D**.

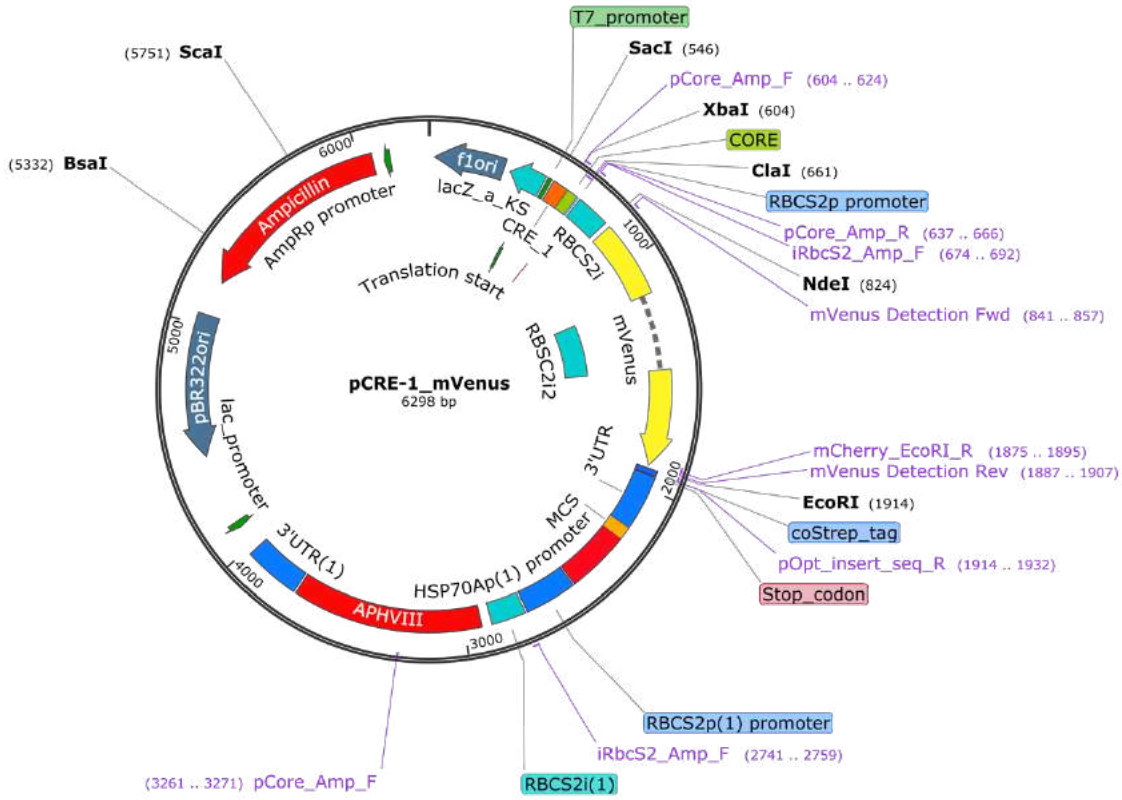


Figure 5.15: Example vector map of a pCRE vector with mVenus reporter gene. pCRE-1_mVenus is here used as an example to show pCRE_mVenus vector structure. In the other vectors listed in **Table 2.5**, CRE_1 (orange) is replaced with repeats of motifs CRE-2–CRE-13 and CRE-RM.

Following the decision not to use mCherry as the reporter, pOpt_Core_mVenus and pCRE-mVenus vectors were generated by amplifying the pOpt_mVenus_Paro vector with primers iRbcS2_Amp_F and mVenus_EcoRI_R (**Table 2.2**; **Figure 2.1**), resulting in a 1260 bp fragment with *Cla*I and *Eco*RI restriction sites. This fragment was digested with ClaI and EcoRI, along with pOpt_Core_mCherry and the pCRE_mCherry vectors; each cut vector was ligated with the cut iRbcS2_mVenus fragment to produce a new library of pCRE-reporter vectors, this time with mVenus as the reporter vector. An example pCRE_mVenus vector is shown in **Figure 5.15**. In total, 30 novel vectors were generated; the 15 mVenus plasmids tested in this study are listed in **Table 2.5**.

5.3.4. *In vivo* testing of pCRE modules in *C. reinhardtii*

5.3.4.1. Experimental design

The controls for the flow cytometry experiments were as follows:

1. **Wild-type (negative)** – untransformed *C. reinhardtii* was used as a negative control, and as a measure for baseline chlorophyll autofluorescence

2. **AR-1 (positive)** – the pOpt_mVenus_Paro vector, containing the original AR-1 promoter, was used as a positive control to verify fluorescence detection
3. **Core promoter (baseline)** – cells transformed with the core promoter vector (pOpt_Core_mVenus) were measured as a baseline for transcriptional activity, to which pCRE vectors were compared
4. **Randomly generated motif (negative)** – cells were transformed with pCRE-RM_mVenus as a negative control to determine whether increases in promoter activity are due to the DNA sequences themselves, or are due to other structural differences in the promoter region

5.3.4.2. Preliminary results

As a preliminary test, WT strain CC-125 was transformed with the vectors listed in **Table 2.5**, with the exception of pOpt_pCRE-11, -12, and -13. Transformant cells were selected for on two strengths of paromomycin: 20 and 50 $\mu g\ mL^{-1}$. The rationale behind using the increased paromomycin concentration was that cells surviving on high concentrations of paromomycin may be more likely to have integrated the vector in an optimal genomic position for high expression.

Colonies formed on the 20 $\mu g\ mL^{-1}$ plates after 5 days, and were prepared for FACS analysis; FACS uses flow cytometry to measure single cell events, and can then isolate cells that exhibit specified characteristics. FACS was used here in an attempt to collect high mVenus-expressing strains. WT and AR-1 strains were used to select the gating controls for cell sorting (**Figure 5.16, 5.17**). At the final elimination step, a gate was drawn to isolate high mVenus-expressing cells by comparing the WT and AR-1 populations. Cells that fell within this final gate (**Figure 5.17**) were sorted into 6-well plates; 10–100 cells transformed with each vector were sorted into separate wells.

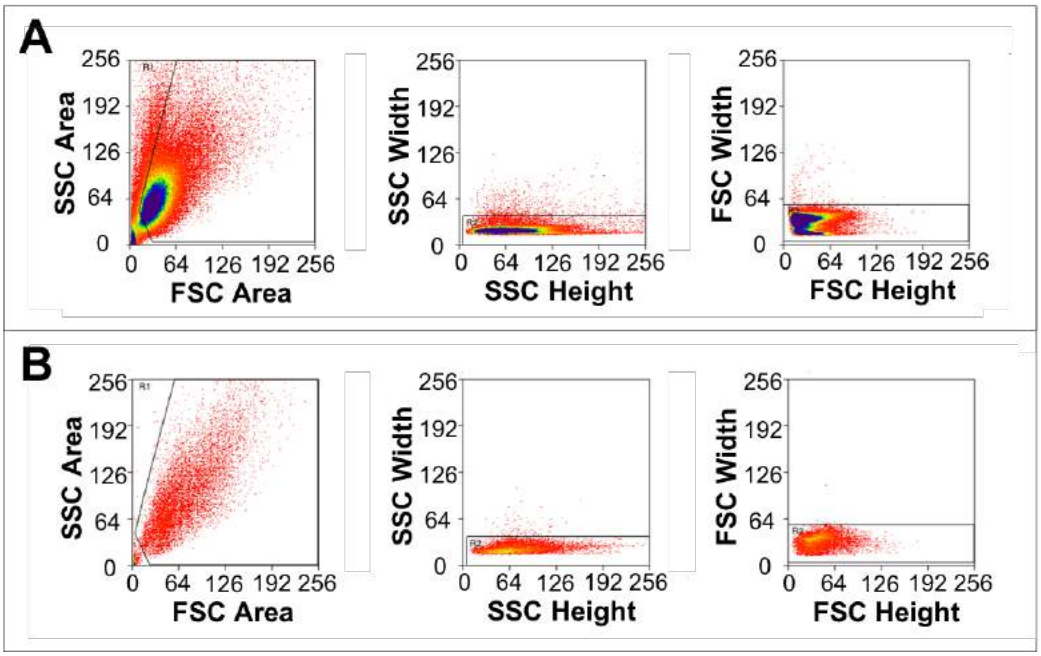


Figure 5.16: Gating selection for AR-1 and WT strains preliminary FACS study. A – AR-1 positive control. **B** – WT negative control. Depicts initial gates drawn to eliminate cellular debris. First scatter plot for **A** and **B**, forward scatter area (x-axis) vs side scatter area (y-axis); second scatter plot for **A** and **B**, side scatter height (x-axis) vs side scatter width (y-axis); third scatter plot for **A** and **B**, forward scatter height (x-axis) vs forward scatter width (y-axis). Both forward and side scatter were measured by measured by excitation at 488 nm. Work was undertaken with the guidance of Dr. A. Sproles, UCSD, US.

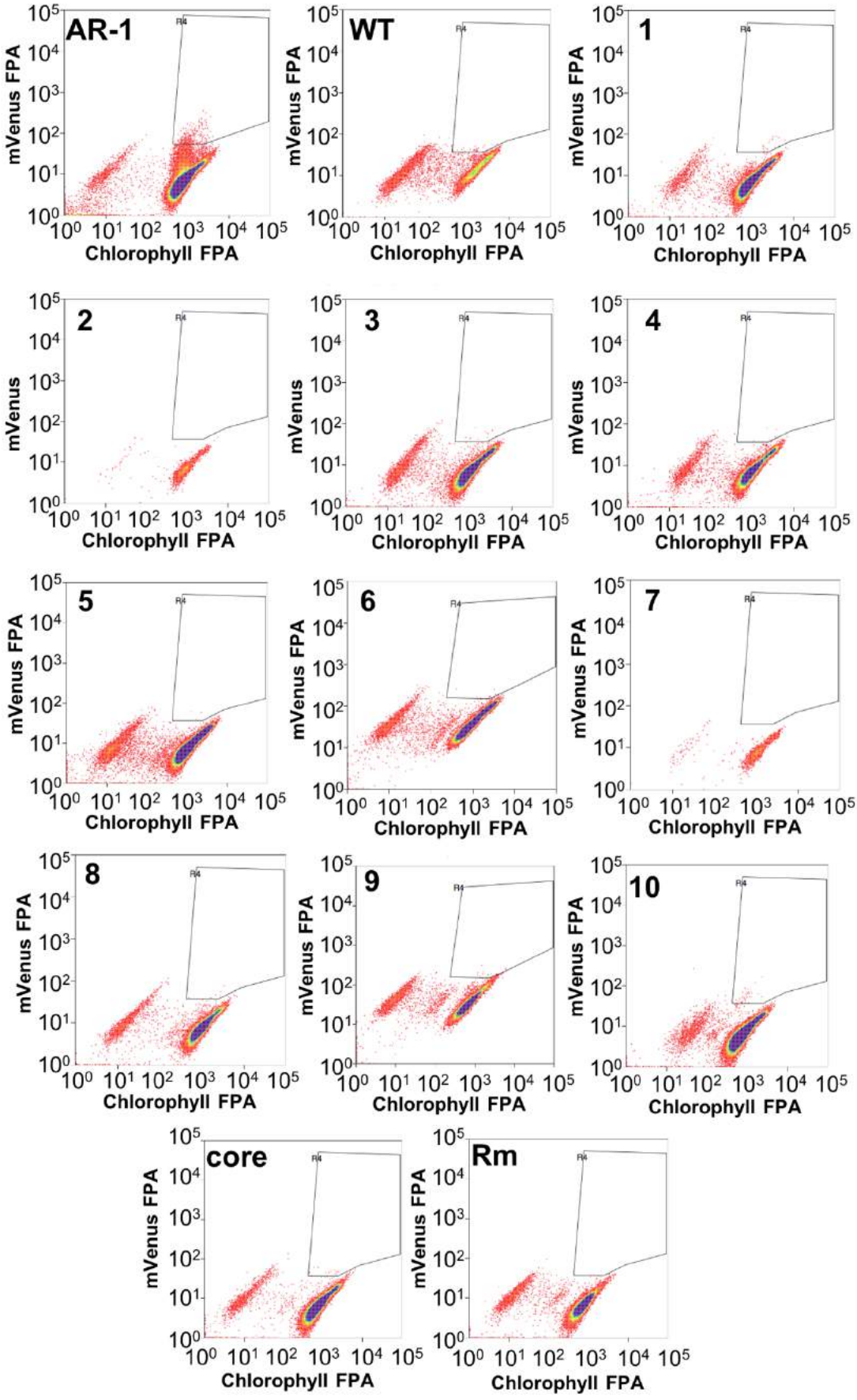

AR-1
WT
1
2
3
4
5
6
7
8
9
10
core
Rm
mVenus FPA
mVenus
Chlorophyll FPA
R4

Figure 5.17: Preliminary FACS scatter plots with gating for pCRE vectors with controls. Gate was designed to isolate individual events with strong mVenus expression. Scatter plots were generated by analysis of the following cell lines: AR-1, pOpt_mVenus_Paro-transformed positive control; WT, negative control (no vector); 1–10, cultures transformed with pCRE-mVenus vectors 1–10 respectively (**Table 2.5**); Core, cells transformed with pOpt_Core_mVenus; Rm, cells transformed with pCRE-RM_mVenus. x-axis, chlorophyll fluorescence (Ex488 nm, Em513/26 nm); y-axis, mVenus fluorescence (Ex488 nm, Em710/45 nm). FPA, fluorescence peak area.

Two populations exhibiting different chlorophyll fluorescence intensities are present in each run (**Figure 5.17**); the lower chlorophyll fluorescence population was excluded from the analysis as it likely represents dead or dying cells (personal communication with Dr A. Sproles, UCSD). There is a sizeable population of events above the higher chlorophyll group in the AR-1 positive control that is absent in the WT plot; this population was assumed to represent mVenus-fluorescent cells, and the sorting gate was therefore drawn to encompass this population only for sorting. The gate was drawn high relative to mVenus fluorescence, so that only the cells expressing mVenus to the highest levels were captured. This approach proved to be overly ambitious; as seen in **Figure 5.17**, very few events for the transformant strains, if any, were captured by this gate. Furthermore, there were issues with the health of the cells subjected to FACS analysis; many of the outgrowth cultures prepared for FACS, notably 2 and 7 but including the others, were visibly stressed and clumped together prior to analysis. The study was hence repeated in **Section 5.3.4.3**.

Meanwhile, the population of transformants selected for using the higher paromomycin concentration (50 $\mu g\ mL^{-1}$) were picked into individual wells of 96-well plates after the colonies appeared 7 days later. 24 individual colonies of each promoter type were picked and analysed via plate reader after 2 days of growth in liquid TAP media. **Figure 5.18** shows the plate reader results.

pCRE-9 produced the highest fold increase in mVenus fluorescence with respect to the WT strain, with a median 7.7-fold increase (**Figure 5.18A**). Other pCREs that displayed significantly higher expression were pCRE-2, -1, -6, -8 and -3. The AR-1 strain, which was expected to exhibit strong mVenus fluorescence being the positive control, had a median fluorescence difference of 1.6-fold with respect to WT, which was surprisingly low (**Figure 5.18A**); at least 1 in 20 pOpt_mVenus_Paro transformants were expected to display > 2.5-fold higher fluorescence than WT (**Figure 5.10**), and potentially even more given the higher paromomycin selection concentration in this experiment. Looking at the individual data points for each transformant strain, no AR-1 clones displayed high

fluorescence (**Figure 5.18B**). This validates earlier findings (See **Section 5.2.3.4.**) that 24 is not a high enough number of colonies to guarantee selection of enough co-transformant strains for a robust analysis. This in itself suggests scepticism should be applied to the results in **Figure 5.18**, and that further confirmation analysis is required. Another reason for repeating this analysis is that the chlorophyll fluorescence signals for some strains were particularly low (**Appendix Figure D8**), which affected the mVenus results after normalisation of mVenus fluorescence to chlorophyll fluorescence. This could be due to slowed growth as a result of mVenus over-production, but is more likely an artefact of the normalisation method applied.

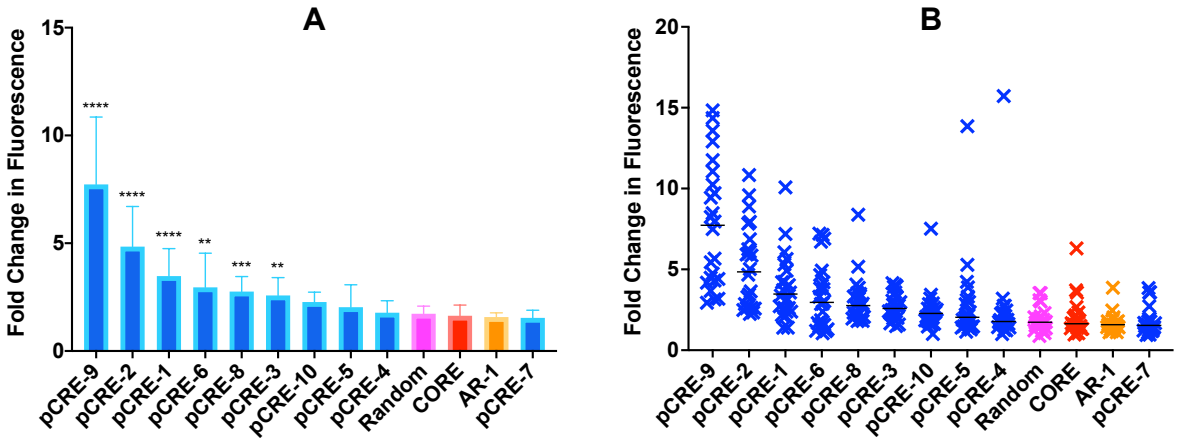


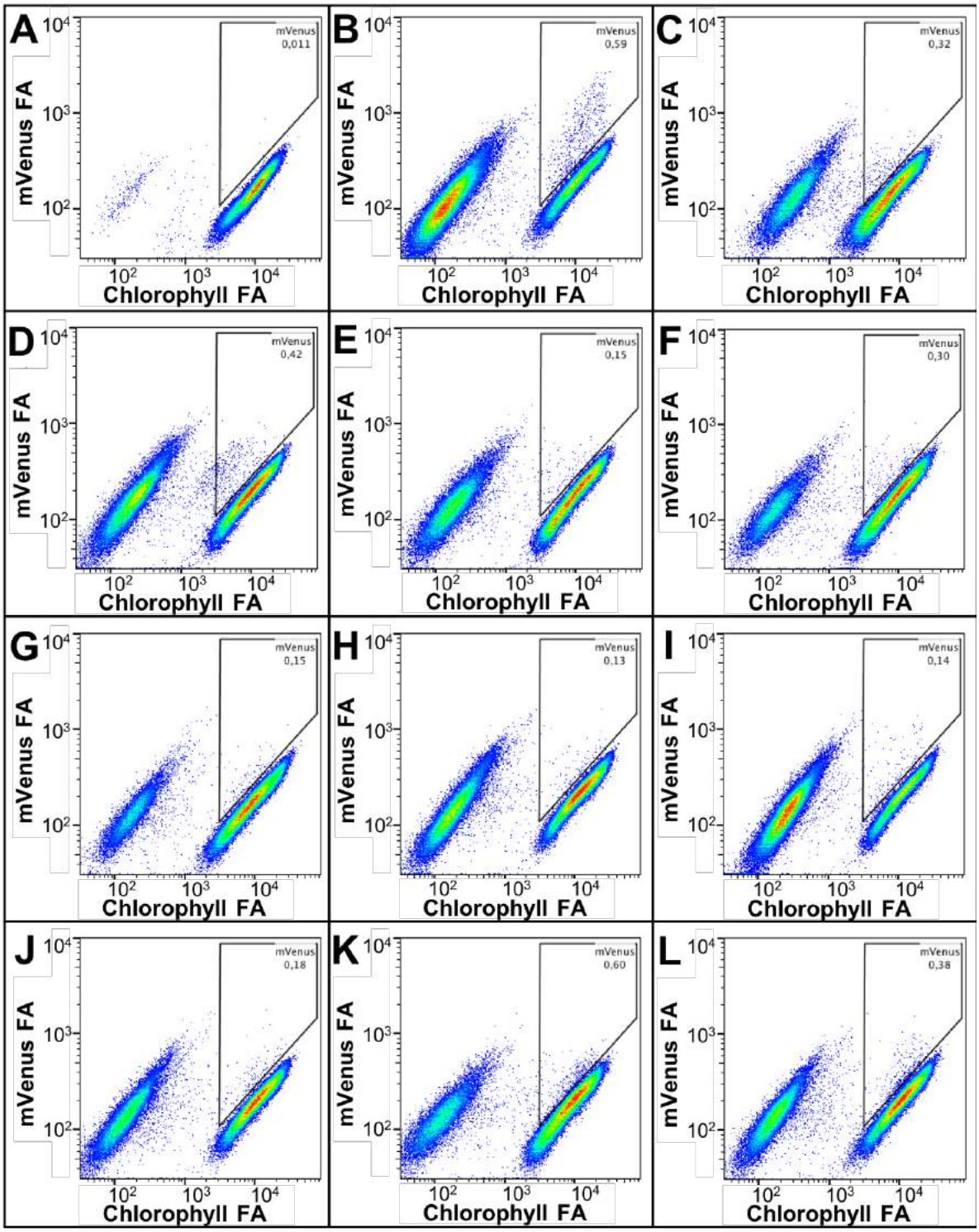
Figure 5.18: Fold difference in mVenus fluorescence with respect to WT CC-125 for pCREs 1–10, AR-1 promoter and core promoter as measured by plate reader. A – Median fold change in mVenus fluorescence with respect to WT. Error bars represent interquartile range. Significant difference from the core promoter was calculated using a Kruskal-Wallis non-parametric test (n = 24, **P* < 0.05, **P* < 0.01, ****P* < 0.001 *****P* < 0.0001). **B** – Scatter plot showing fold change for each of the 24 individuals wells measured per promoter variant, with respect to WT. Black bar represents median. Raw readings in **Appendix Figure D8**.

5.3.4.3. Flow cytometry results

The flow cytometry analysis was repeated to get a clearer picture of transcriptional activation by the discovered set of pCREs. pCREs -11 to -13 were also included in this analysis (**Figure 5.5**), and the initial gating strategy shown in **Figure 5.16** was reapplied. The mVenus gate was modified for this experiment, again by comparing WT and AR-1 strains (**Figure 5.19A, B**); the difference here was that the gate drawn is more lenient, selecting for all possible mVenus expressers, not just the highest (**Figure 5.19**). The cells for this experiment were healthier after modifications to the preparation

protocol; cells were incubated in almost complete darkness following scraping into 10 mL TAP for overnight preparation for flow cytometry, as opposed to in the light for the previous experiment. More events were captured within this gate for each promoter type, which can be seen in graphical format in **Figure 5.20**.

The low chlorophyll population is essentially absent from the WT culture, suggesting that this batch of cells was less stressed and represents a healthier population (**Figure 5.19A**). It could also be due to the fact that there is less genetic variation within the WT population; each of the transformed cells are different mutants in themselves due to random genomic integration of the vectors, whereas the WT population was grown from colonies of the same parental background. The mVenus population is pronounced in the AR-1 strain as expected, as well as for pCRE-12 and -13 (**Figure 5.19A, O, P**). The mVenus population is visibly present within all other strains (except WT), but to a lesser extent.



Continued on next page.

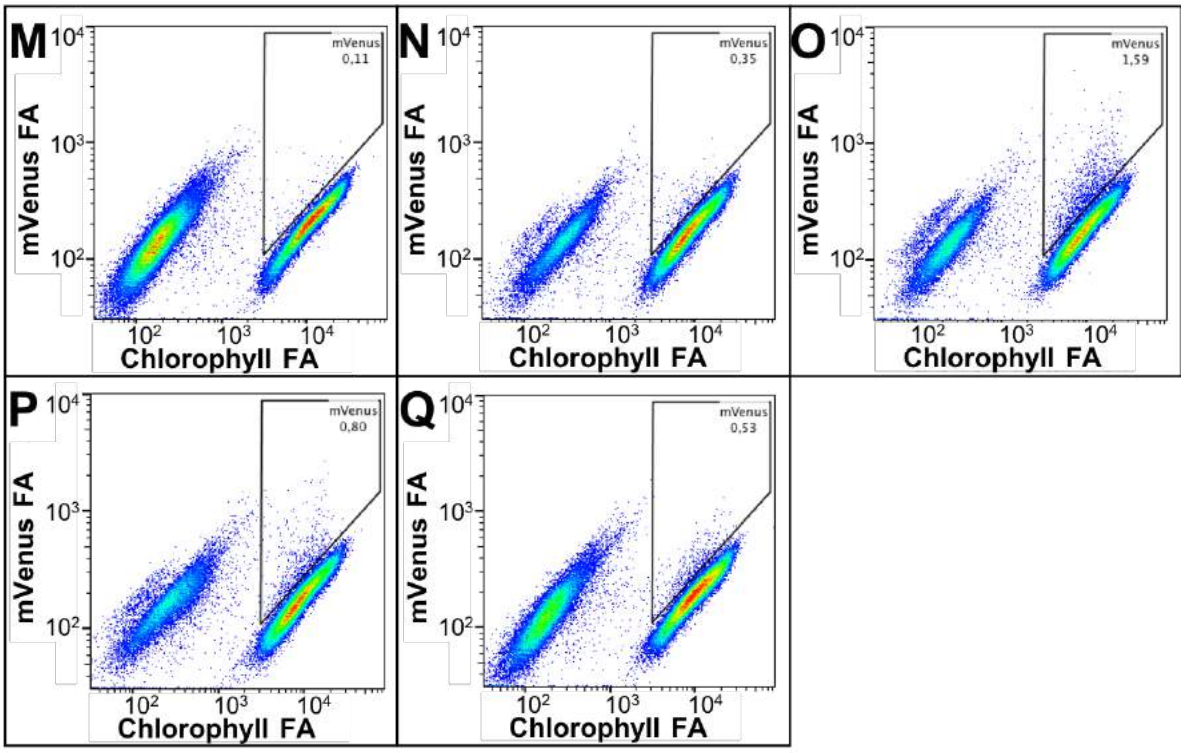


Figure 5.19: Flow cytometry analysis of mVenus expression driven by different promoter elements. x-axis = chlorophyll fluorescence area (FA). Y-axis = mVenus FA. Vector-less negative control (WT; **A**), AR-1 (**B**), Core promoter (**C**), pCRE-1 (**D**), pCRE-2 (**E**), pCRE-3 (**F**), pCRE-4 (**G**), pCRE-5 (**H**), pCRE-6 (**I**), pCRE-7 (**J**), pCRE-8 (**K**), pCRE-9 (**L**), pCRE-10 (**M**), pCRE-11 (**N**), pCRE-12 (**O**), pCRE-13 (**P**), Random CRE (**Q**). Number in corner of gated region = % of total events inside mVenus gate.

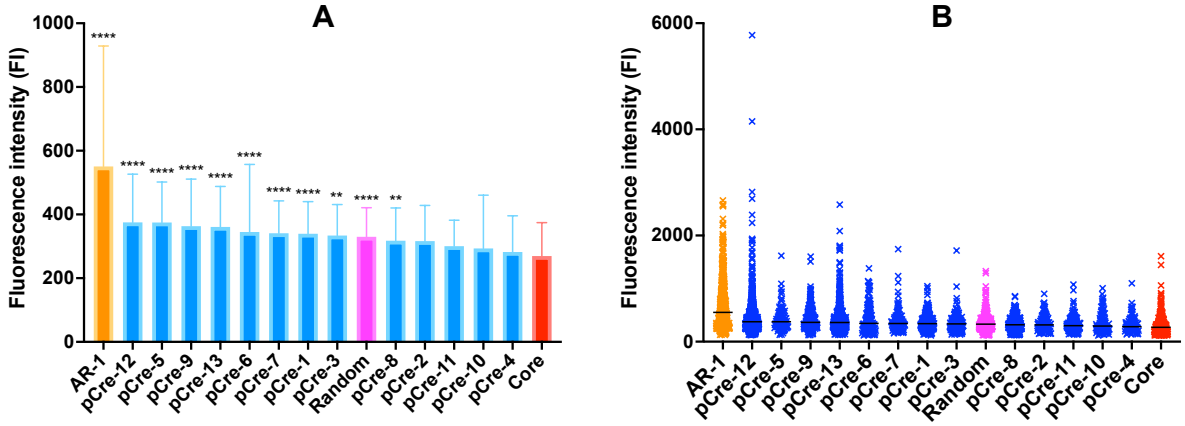


Figure 5.20: Fluorescence intensities of events detected within mVenus gate. A - Bars represent median value, error bars represent interquartile range. **B** – individual data points captured from within the mVenus gate are plotted, with median shown as a black bar. No datasets passed normality tests ($P > 0.05$ for Anderson-Darling, D'agostino and Pearson, Shapiro and Wilks, and Kolmogorov-Smirnov tests), hence the non-parametric Kruskal-Wallis test was applied to compare medians for each population to fluorescence

exhibited by the core promoter. Asterisks represent significant difference in median fluorescence from the core promoter (red). *$P < 0.05$, *$P < 0.01$, ***$P < 0.001$ ****$P < 0.0001$. n range: 93–1052. Test was adjusted to accommodate large difference in *n* values.

Ten out of 14 pCREs tested exhibited significantly higher mVenus fluorescence intensity (FI) than the core promoter alone. AR-1, unlike in the plate reader analysis (**Figure 5.18**), has by far the highest overall expression (median FI 550.4). However, the pCRE-12 promoter drove expression in some individuals to be higher than AR-1; the highest AR-1 value was 2660 FI, whereas the highest for pCRE-12 was more than double that at 5775 FI. pCRE-12 had the highest median mVenus FI of all the motifs tested (median = 375.1). This strongly suggests that pCRE-12 elicits transcriptional activation. pCRE-13 displays a similar range in FI to AR-1, and is its median is significantly higher than the core promoter alone, thus demonstrating that pCRE-13 is another strong candidate for promoter activity. The other motifs that display significantly higher mVenus fluorescence compared to the core are pCRE-5, -9, -13, -6, -7, -1, -3 and -8. pCRE-9 was also highlighted in the preliminary study as a strong TFBS candidate (**Figure 5.18**), and so is likely to also play a functional role in transcription. The randomly generated motif pCRE-RM also displayed promoter activity; this could either call into question the validity of the experiment, or the random motif could induce promoter activity itself. There is also the chance that the promoter structure enhances transcription, in that a repeated motif of a particular GC content and length upstream of the core could in itself enhance transcriptional activity. This could be explored further by running the test again with a different randomly generated motif of the same length.

5.4. Discussion

Individual putative transcription factor binding sites, or pCREs, were predicted using motif discovery software, and subsequently tested for activity *in vivo* by measuring expression of an mVenus reporter. Ten pCRE candidates exhibited statistically higher mVenus expression than the core promoter baseline, two of which (pCRE-12 and pCRE-13) were comparable to the strongest constitutive promoter, AR-1. This work has shown that through interrogation and usage of open source data and software, genomic data can be mined for regulatory motifs that can effectively drive protein expression.

pCRE-12 produced the highest median fluorescence intensity of the pCRE suite, and the highest overall single fluorescence intensity event of all the promoters tested, including AR-1. This motif

was discovered by scanning the strongest 25 synthetic promoters found by Scranton *et al*. (2016) for the most common sequences with MEME. This validates their work and provides more evidence for pCRE-12 being a genuine TFBS.

The SORLIP2AT motif 'GGGCC', which is present within pCRE-12 and -13, may be responsible for their high expression (**Table 5.5**). SORLIP2AT is a light inducible motif in *A. thaliana*, which suggests that this analysis likely captured some light inducible motifs. The complete exclusion of light regulated CREs would have been difficult to achieve, as ~80% of *C. reinhardtii* genes are differentially expressed throughout a 24 h photoperiod (Zones *et al*., 2015), meaning the majority of gene expression is directly or indirectly influenced by light. Given that for many biotechnological purposes, algal cells will be grown in light conditions, this is not too problematic. Comparing mVenus expression in pCRE-12 and -13 transformants grown under different trophic conditions, *i.e.* heterotrophically in the dark, and mixotrophically or autotrophically in light, could reveal the conditions under which these motifs exert influence. Furthermore, examining the ability of SORLIP2AT to drive protein expression alone in *C. reinhardtii* could help determine whether this is the active motif in pCRE-12 and -13.

pCRE-2 also contains the SORLIP2AT motif, and although this was the second highest performing motif in the initial plate reader analysis (**Figure 5.18**), pCRE-2 did not elicit significantly higher mVenus expression in the flow cytometry study (**Figure 5.20**). This could mean that the SORLIP2AT motif is not solely responsible for enhanced expression, or possibly that other surrounding bases form the consensus sequence for the TF that in part binds to the SORLIP2AT sequence. This could be examined further by changing individual bps surrounding the common consensus sequence to find the optimal motif.

pCRE-11 (CCCATGCGA), a motif discovered by Scranton *et al*. (2016) to be essential for high expression in their strongest promoter, was expected to induce strong mVenus expression, but instead mVenus levels were not significantly higher than the core (**Figure 5.20**). Their analysis demonstrated that the pCRE-11 motif is *required* for high expression in that particular promoter, but not that this motif is alone *responsible* for high levels of transcription; the pCRE-11 motif was also present in their non-functional synthetic promoters, suggesting other genetic components are necessary for increased transcriptional activity (Scranton *et al*., 2016). The modest mVenus expression observed in the flow cytometry analysis for pCRE-11 in this study (**Figure 5.20**) was likely due to the requirement of this motif to have other DNA elements present that were missing in the

reporter vectors. Combining pCRE-11 with other CREs could perhaps determine other elements to which this motif can be functionally paired.

The 'CCCAT' motif (present within pCRE-11 and also common to pCRE-2, -12 and -13), was presented as a core motif in *C. reinhardtii* (Scranton *et al*., 2016), but only half of the pCREs containing this motif (pCRE-12 and pCRE-13) drove high mVenus expression (**Figure 5.20**). It could be that the bp flanking this motif influence TF binding; pCRE-12 and -13 retain the common sequence GNCCATNC, which could be explored further in variations. Another commonly found short motif, 'CATG' found in pCRE-2, -6, -7, -11 and -13, is an N^6-adenine methylation (6mA) site in *C. reinhardtii* (Fu *et al*., 2015). 6mA can be a marker for active and high gene expression, and is thought to contribute to gene activity through preventing nucleosome binding near the TSS, which could partially explain the significantly higher mVenus expression from the CATG-containing pCREs (**Figure 5.20**; Fu *et al*., 2015).

The TC-rich pCRE-1 motif is localised around the TSS (–36 bp) in high expression constitutive promoters, according to the Centrimo results (**Figure 5.4)**. This validates the findings made by Scranton *et al*. (2016), where they discover that a TC-rich region often lies immediately upstream of the TSS in *C. reinhardtii* highly expressed genes. A similar TC-rich region occurs in the plant viral overexpression promoter CaMV35 (Pauli *et al*., 2004) and in many *A. thaliana* promoters (Bernard *et al*., 2010); TC-rich motifs in close proximity to the TSS in plants and *C. reinhardtii* could function similarly to TATA boxes and elicit protein expression. mVenus expression driven by pCRE-1 was significantly higher than that from the core promoter only, but only by 26% (**Figure 5.20**). The potential positive effects that pCRE-1 could have on transcription may have been weakened, as this motif is located further upstream to its optimal position (–36 bp from TSS) in the reporter vectors used in this study. The precise biological function of TC-rich motifs in *C. reinhardtii*, as well as in plants, is yet to be determined.

pCRE-9, one of the best performing promoter elements (**Figure 5.19, 5.20**), contains the 'TTTTC' motif, which is similar to the high light-inducible GT1 motif 'TTTTTC' that drives high expression under medium and high light (300 and 600 μmol photons m^2 s^{-1}, respectively) in *C. reinhardtii* (Baek *et al*., 2016a). This motif could potentially be recognised by the same TF, although this would require further experimental evidence.

A potential problem with the experimental design was overlooking the presence of the PLACE motif SORLIP2AT in pCRE-RM (**Table 5.5**); PLACE analysis for the randomised motif was completed

retrospectively. This motif is very similar to the SORLIP 'GGGCCAC' found and tested by Baek *et al*. (2016a). Although it is in the opposite orientation (AACCGGG), it is still possible that this motif is causing its unexpectedly high mVenus expression; Lis and Walther (2016) demonstrate that, at least in *A. thaliana*, the orientation of a TFBS motif does not have a significant effect on transcription. Although this somewhat alters the results as there is no randomised negative control, this does provide more evidence for the functionality of SORLIP2AT in *C. reinhardtii*. The influence of the CRE-RM motif could also be coincidental, or be a structural enhancer of gene expression, as discussed in the final paragraph of the results section. At least two randomised motifs should be incorporated as negative controls in future experiments. Additional reporter vectors were made during this project carrying repeats of the 'TTTTTC' and 'GGGCCAC' light inducible motifs from Baek *et al*. (2016a), deemed pCRE-14 and pCRE-15, respectively, however due to time constraints these vectors were not transformed and analysed in *C. reinhardtii*.

The definition of the promoter region used in this study was −1000 bp relative to the TSS (or the 1000 bp upstream flank of the 5'-UTR), and has been defined as such in several other large-scale promoter analyses (Zou *et al*., 2011; Koschmann *et al*., 2012). This definition, however, may have prevented detection of important motifs that lie immediately downstream of the TSS; the 5'-UTR, for example, can harbour TFBSs that largely influence transcriptional activation (López-Paz *et al*., 2017). For this reason, some promoter analyses take into account the 1000 bp upstream of the *translation* start site, as opposed to the *transcription* start site (Hu *et al*., 2014a; Ding *et al*., 2012). Conducting this analysis from the translation start site was considered, but decided against for this work, as some *C. reinhardtii* 5'-UTRs are longer than 1000 bp in length whereas other 5'-UTRs are completely absent; this would mean the promoter regions analysed would differ in size for each gene with relation to the TSS. Repeating the search to include the +100 bp from the TSS could potentially capture motifs that were otherwise missed in this work.

Three algorithmically different *de novo* motif discovery programs were used to mine potential TFBSs from *C. reinhardtii* promoters in **Section 5.3.3.2.** Given that only a small subset of the motifs found would be tested *in vivo*, and that Weeder and Homer had detected unique motifs (2 and 21, respectively), the three programs chosen were deemed sufficient. Conversely, in some plant promoter studies, at least 5 motif discovery programs were integrated for motif discovery (Zou *et al*., 2011; Koschmann *et al*., 2012). Although this was in part due to the scale of their studies, in which they were scanning the promoter sequences of thousands of co-regulated genes, perhaps

including more discovery programs in this pipeline could increase the number and quality of motifs found.

The most effective motifs (pCRE-12 and pCRE-13) were discovered by scanning the strongest computationally generated promoters created by Scranton *et al*. (2016). One reason for this could simply be that POWRS software that they used for motif discovery is more effective at identifying motifs (Davis *et al*., 2012). Another more likely reason is that through scanning their best synthetic promoters, whose motifs were computationally designed to mimic common consensus sequences but with slight variations, the PWMs for the motif consensus sequences identified in the MEME search in this study were refined and optimised for increased promoter activity. It would be interesting to test all 7 motifs found from the MEME scan of the Scranton *et al*. (2016) top synthetic promoters shown in **Table 5.4** individually, as they are likely to drive high expression.

Discovering novel CREs is a difficult challenge in computational biology due to the short length and degeneracy of TFBSs; D'Haeseleer (2006) describes *de novo* motif discovery as "searching for imperfect copies of an unknown pattern". Without prior knowledge, either experimental or from closely related species, computational *de novo* motif discovery can be limited in its power to provide biologically meaningful predictions (Simcha *et al*., 2012). Many active sequences will have been overlooked and likewise some false-positives retrieved, neither of which can be verified through computational modelling alone. Furthermore, a potential problem arises from integrating PWMs from multiple *de novo* discovery programs using clustering software, where multiple biologically distinct motifs may have been merged together, thus unwittingly nulling their biological effects. Essentially, computational motif discovery can be a useful ancillary tool, but only alongside other analytical techniques.

A potential improvement to this study could have been to control for the number of repeats of each CRE, as opposed to promoter length. In yeast promoters, a positive relationship between TFBS copy number and protein expression can be observed for up to ~4 repeats, after which it becomes saturated and no further improvement can be observed (Sharon *et al*., 2012). Given this information, 4–7 repeats of each promoter were used to build each pCRE reporter vector, controlling for the length of the promoter region instead. Baek *et al*. (2016a), however, found that the number of repeats of their light inducible motif was not proportional to protein expression levels; duplicating the 'GGGCCAC' motif within the promoter caused an 8-fold increase in luciferase expression compared to a single copy of the motif, whereas when 3x copies of the motif are present,

luciferase expression is only 6-fold higher than the single repeat. This suggests that that the number of motif repeats should be optimised before making claims about their effectiveness as CREs, as more repeats does not necessarily equate to higher expression.

Previous studies have used a plate reader to analyse promoters without resorting to flow cytometry (Lauersen *et al*., 2015, 2016b; Blaby-Haas *et al*., 2018). This was sufficient for these studies, in which one of the two high expression UVM strains, UVM4 or UVM11, were used. UVM strains have co-transformation efficiencies and protein expression levels that greatly exceed those of WT strains (Neupert *et al*., 2009). Although these strains would have facilitated this analysis, they were not used as they are cell wall-deficient, making them susceptible to damage during cell sorting and less likely to survive (personal correspondence with Dr. A. Sproles). It is also thought that the mutations responsible for the superior expression of the UVM strains somehow reduce nucleosome occupancy at transgene insertion sites (Barahimipour *et al*., 2015); such mutations could exaggerate the effectiveness of motifs, which could potentially be ineffective in other laboratory *C. reinhardtii* strains. Performing this analysis in the popular strain CC-125 avoided this possibility, but brought with it the problems of low co-transformation efficiency and gene expression (**Section 5.3.2.4.**). It would be worthwhile to repeat this study using a UVM strain for comparison.

Flow cytometry was an effective alternative to plate reader analysis, circumventing low expression and co-transformation efficiency by rapidly measuring 1000s of individual cells representing different genomic integration sites to build a broad assessment of each promoter's activity. One setback to the flow cytometry method is the potential overrepresentation of rapidly growing strains. Given that each individual transformant is a mutant in itself, some strains will inevitably grow more slowly as a result of the genomic integration site, and be underrepresented in the population analysed. Furthermore, strains that accumulate recombinant protein to higher levels are more likely to have a slower growth rate. This was accounted for by incubating the pooled mutants in liquid culture for less than the CC-125 doubling time in standard conditions (~14 h), but this will not have been perfect. Hopefully, the large number of events assessed (50,000) will have accounted for this. Another issue is that gene expression may have been reduced through the preparation method for flow cytometry. The pooled paromomycin-resistant colonies were grown in the dark overnight in liquid culture prior to analysis, as cultures grown in standard light conditions had previously flocculated and were unfit for flow cytometry analysis (**Section 5.3.4.2.**). Incubation in the dark may have prevented full gene activation, given that comparison of the pCREs tested revealed strong similarity to light inducible motifs in plants (**Table 5.5**).

Chapter 6: Final discussion and future work

The principal aim of this thesis was to develop metabolic engineering tools to improve *Chlamydomonas reinhardtii* as a biotechnological chassis, with a focus on the production of high-value carotenoids. This was achieved using three different but complementary genetic engineering approaches, as detailed in Chapters 3, 4, and 5. In this chapter, the main findings of this thesis will be summarised and their context within the field of algal biotechnology will be discussed, alongside proposals for future research.

6.1. Enhanced lutein and β-carotene production in *C. reinhardtii* by overexpression of a putative post-translational regulator of phytoene synthase

In **Chapter 3**, the *C. reinhardtii* CPL6 protein (crOR) was identified as a homologue of the plant carotenoid biosynthesis regulator, ORANGE. crOR was cloned and overexpressed from the *C. reinhardtii* nuclear genome for the first time, and HPLC analysis of the pigment profiles of crOR-overexpressing mutants revealed a 2.0-fold increase in lutein and a 1.3-fold increase in β-carotene per cell compared to the CC-4533 WT strain. The enhanced carotenoid production in crOR-overexpressing strains provides evidence for the regulatory role of crOR in carotenoid biosynthesis and its potential value as a genetic engineering tool for increasing carotenoid bioproduction.

This work has contributed to our understanding of the regulation of pigment production in microalgae. The C-terminal DnaJ-like domain and predicted protein disulphide isomerase activity of crOR are suggestive of its role as a protein chaperone and/or post-translational regulator. Plant homologues of crOR act as post-translational activators of the carotenoid pathway enzyme PSY. Violaxanthin de-epoxidase in *Arabidopsis* contains 6 disulphide bonds, and its zeaxanthin-producing activity is sensitive to redox conditions (Simionato *et al*., 2015); it is possible that crOR controls the activity of PSY or other carotenoid biosynthetic enzymes in *C. reinhardtii* by altering the redox state of disulphide bridges. Further investigations into crOR protein interactions would provide more insight into this hypothesis. Performing a co-immunoprecipitation assay coupled to MS using crOR as bait could reveal interacting proteins and complexes. Furthermore, obtaining the crystal structure of crOR and its partner proteins could shed light on its activity. Nevertheless, the crOR protein could be the first post-translational regulator of carotenoid biosynthesis identified in green algae.

Interestingly, a recent targeted transcriptomics experiment revealed that crOR regulates the expression of carotenoid-related genes in *C. reinhardtii* at the mRNA level, suggesting a further role for crOR in the transcriptional regulation of carotenoid biosynthesis (Yazdani *et al*., 2020); however, this article has not been subjected to peer review. This finding could be further explored by performing comparative multi-level discovery -omics experiments with crOR-overexpressing and knockout strains. A localisation assay conducted using a crOR-fluorescent protein fusion could both confirm its localisation to the chloroplast and expose potential nuclear localisation, which would provide further evidence for its role as a transcriptional regulator. These analyses could compliment structural analyses, revealing whether crOR modulates gene expression directly, indirectly, or at not all. The prospect of crOR playing regulatory roles at both the transcriptional and post-translational level is interesting and warrants further exploration.

Overexpressing crOR in other microalgal species related to *C. reinhardtii* could also be profitable. For example, expressing crOR in a microalga with multiple chloroplasts could have a more pronounced effect on carotenoid sequestration. Conversely, cross-species studies recombinantly expressing crOR homologues from plants or other algal species in *C. reinhardtii* could further boost carotenoid production. crOR could also be modified to increase its effectiveness. *A. thaliana* mutants expressing ORANGE with a R90H substitution produce 7-fold more carotenoids than their wild-type counterparts (Yuan *et al*., 2015); using site-directed mutagenesis to introduce an equivalent mutation in crOR could augment its activity even further. However, since the initial submission of this thesis, an R106H mutation was introduced into crOR and overexpressed in *C. reinhardtii*, yielding no significant changes in carotenoid production (Kumari *et al*., 2020). In the same study, the crOR equivalent from *Brassica oleracea* was also overexpressed, resulting in an increase in carotenoid production and chloroplast enlargement; this supports the hypothesis proposed in **Chapter 3** that crOR overexpression could elicit morphological changes in the chloroplast, although the confocal microscopy images presented in this work (**Figure 3.6**) were insufficient to draw such conclusions. The role of crOR in chloroplast development should therefore be examined further.

The overexpression of rate-limiting enzymes to increase the production of secondary metabolites has been somewhat successful, but not comparable to other biotechnology hosts, such as yeast or prokaryotes. This is due to stringent and complex regulatory systems exhibited by green algae, which have likely contributed to their evolutionary success and enabled them to thrive in a multitude of environments. In **Chapter 3**, the *C. reinhardtii* ORANGE homologue was selected as a

metabolic engineering target due to its role as a carotenoid regulator in higher plants; the hypothesis was that targeting a carotenoid regulator could perhaps circumvent or mitigate these complex processes, driving carotenoid production to higher levels than enzyme overexpression alone. Although the increase in lutein achieved by crOR overexpression (2.0-fold) was comparable to that of PSY overexpression (2.6-fold and 2.2-fold; Couso *et al*., 2011; Cordero *et al*., 2011a), this study has nevertheless succeeded in doubling the cellular lutein content of *C. reinhardtii*, while in the process revealing a novel carotenoid enzyme regulator. This research has both identified a new genetic engineering target to enhance the biosynthesis of high-value carotenoids in green algae, and raised interesting questions regarding the regulation of pigment biosynthesis.

6.2. Development of a mutant selection workflow for improved carotenoid production with mutant characterisation using comparative shotgun proteomics

In **Chapter 4**, a reverse genetic engineering approach was applied to generate and isolate lutein-hyperaccumulating *C. reinhardtii* mutants. Of the nine mutants isolated, five produced significantly more carotenoids per cell than the WT, one of which produced ~5-fold more lutein per cell than the parental strain. The proteomes of the mutant and WT strains were then compared using LFQ shotgun proteomics, which enabled the generation of various hypotheses regarding the optimisation of growth conditions, affected pathways, and target identification for future metabolic engineering experiments.

The mutagenesis experiment was designed to capture interesting, novel metabolic mutants; sublethal concentrations of norflurazon, a herbicide which inhibits the carotenoid biosynthetic enzyme PDS, were applied to chemically-mutated cells, with the hope of capturing previously-uncharacterised mutations that can increase carotenoid production. The markedly different pigment and proteomic profiles of EMS-Mut-5 suggest that this goal was achieved; the mutant produced 5-fold more lutein than the WT, and ~50% of the identified proteins were differentially regulated. **Chapter 4** was the first reported mutagenesis selection workflow for isolating carotenoid-rich *C. reinhardtii* strains. This process could be tailored to isolate lutein-producers in other biotechnologically useful algal species.

The generation and isolation of the EMS-Mut-5 strain emphasises the utility of the workflow developed in **Chapter 3**; this mutant produced ~5-fold more lutein than its WT counterpart, which is the highest fold-change per cell of lutein in *C. reinhardtii* reported to date. Although it exhibited a relatively slow growth rate, the proteomics data could be used as a diagnostic tool to optimise the

growth conditions of EMS-Mut-5 in future experiments. By increasing the light intensity and CO_2 availability, perhaps EMS-Mut-5 could become a competitive lutein production strain and be scaled up for industrial exploitation.

One of the goals of **Chapter 4** was to identify potential targets for future genetic engineering to rationally improve lutein production in microalgae. The proteomics data highlighted several potential proteins of interest that could be overexpressed and their effect on lutein production examined. Several proteins involved in qE NPQ (LHCSR1, LHCSR3 and PSBS) were found to be highly expressed in EMS-Mut-5 (**Table 4.3**). Both LHCSR proteins contain lutein binding domains, with LHCSR1 exhibiting a higher affinity for lutein (Perozeni *et al*., 2020b); LHCSR1 could potentially be overexpressed in *C. reinhardtii* to act as a metabolic sink for lutein, thus enhancing accumulation in the chloroplast. Another class of highly upregulated in EMS-Mut-5 were fascilin-domain containing proteins FAS3 and FAS2, and the uncharacterised PAP-fibrilin domain containing protein Cre03.g197650. These proteins are enriched within the carotenoid-abundant *C. reinhardtii* eye-spot (Eitzinger *et al*., 2015), and are implicated in carotenoid storage in plants and algae (Kawasaki *et al*., 2013; Leitner-Dagan *et al*., 2006). Overexpressing one or more of these proteins could further improve carotenoid sequestration in *C. reinhardtii*. Possibly the most promising target was a predicted ABC1 kinase (Cre13.g581850), which has the associated GO term 'positive regulation of carotenoid biosynthesis' and was upregulated in EMS-Mut-5 (**Table 4.4**). Overexpressing this kinase may be an effective means of upregulating carotenoid biosynthetic enzymes and storage proteins with minimal metabolic engineering effort. Furthermore, investigating its function and interacting proteins could reveal more interesting biotechnological targets, while at the same time unravelling the complex regulatory mechanisms underpinning algal stress responses. Several other potential overexpression targets were present within the proteomics dataset, including a violaxanthin de-epoxidase (Cre04.g221550), and several uncharacterised proteins involved in high light stress, ROS stress, and thylakoid biogenesis (**Table 4.3**). This discovery experiment has provided many promising leads for future investigations into the mechanisms driving carotenoid synthesis in *C. reinhardtii*.

The EMS-Mut-5 mutant is also potentially of value to photosynthesis researchers. Such dramatic increases in both LHCSR1 and LHCSR3 protein levels in a single strain have never been reported. A novel regulatory component of qE NPQ could be responsible for this phenotype, and sequencing analysis should be performed in this strain. Assuming multiple mutations are present within EMS-Mut-5, several rounds of genetic backcrossing of the mutant with the parental strain CC-125 should first be performed to isolate the mutation loci and identify any cumulative or epistatic effects of the

mutations on carotenoid production and qE NPQ protein expression. Once the mutations have been isolated, genomic sequencing can then be performed to identify the genes and pathways responsible for its superior lutein production capacity and dysregulated qE NPQ. It is highly likely that at least one mutation lies within a key regulatory component of photosynthetic stress signalling in EMS-Mut-5, which, if discovered, would be highly valuable to fundamental photosynthesis research given that *C. reinhardtii* is the model organism for green algal photosynthesis, and could perhaps reveal yet more biotechnologically useful targets for enhancing photosynthetic efficiency and photoprotection in *C. reinhardtii* (Vecchi *et al*., 2020).

6.3. Synthetic promoters to expand the range of recombinant protein expression levels from the *C. reinhardtii* nuclear genome

In **Chapter 4**, the promoter regions of genes exhibiting high constitutive expression were analysed *in silico* with the goal of finding DNA elements that can be used to generate high-expression synthetic promoters. The computationally-identified CREs were tested for their ability to induce the expression of a fluorescent protein *in vivo* using flow cytometry, which led to the identification of active elements in *C. reinhardtii* constitutive promoters.

A bottom-up approach towards developing synthetic promoters was applied in **Chapter 5** by identifying and testing individual promoter elements. The next step would be to examine the reduced promoter parts as modules, by combining the CREs to construct novel, 'build-your-own' promoters. This has the potential to generate more robust synthetic promoters, given that CREs can function sub-optimally in isolation and may require other CREs to induce expression (Pilpel *et al*., 2001; Gertz *et al*., 2009). Examples of this phenomena in *C. reinhardtii* include mutually necessary heat shock elements in the Hsp70A enhancer (Strenkert *et al*., 2013), and the 'CCCATGCGA' motif, which is necessary for high expression in some but not other synthetic promoters developed by Scranton *et al*. (2016). Combining motifs could highlight synergistic or antagonistic effects of different CRE combinations, as well as provide insight into positional effects of CRE function with respect to the transcription and translation start sites. As part of this project, an attempt was made to build randomised synthetic promoters containing the strongest 6 motifs, however due to time and funding restrictions, this work was not completed. The ultimate intention was, following promoter testing and analysis by flow cytometry, to use synthetic promoters of differing strengths to drive co-expression of crOR and crPSY, with the goal of increasing and optimising lutein production (See **Chapter 3**).

In other biological systems, synthetic promoters are often first tested transiently (Rushton *et al*., 2002; Brown *et al*., 2014). As a transient expression system does not currently exist in *C. reinhardtii*, synthetic promoters had to be tested following full genomic integration. This puts *C. reinhardtii* at a disadvantage in promoter development, as a large number of transformant strains have to be tested to get a result, especially given the unpredictability of genomic integration. Developing a transient expression system for *C. reinhardtii* could increase the speed at which genetic engineering tools can be developed.

As discussed in **Chapter 5**, using computational techniques alone to identify TFBSs and their cognate TF partners is difficult, at least in algal systems where the understanding of gene regulation is still in its early stages. Increasing the number of characterised CREs and TFs in *C. reinhardtii* could dramatically improve informative promoter design. Although the *C. reinhardtii* genome is predicted to code for ~350 TFs (Pérez-Rodríguez *et al*., 2010), very few TFs and their binding sites have thus far been characterised experimentally (Yoshioka *et al*., 2004; Tsai *et al*., 2014; Kropat *et al*., 2005; Ibáñez-Salazar *et al*., 2014; Salas-Montantes *et al*., 2018; Li *et al*., 2018). Anderson *et al*. (2017) attempted an *in vitro* approach to characterise *C. reinhardtii* TFs using a yeast one-hybrid assay against the promoters of 5 highly expressed genes. Their TF library was then transformed into *C. reinhardtii* to examine how the overexpression of TFs would affect the transcriptome, but due to problems with transformation, expression and cell survival, only one TF, a basic helix-loop-helix protein referred to as TF64, was ultimately characterised. Unfortunately, a common binding motif was not found within the genes regulated by TF64. Their study exemplifies some of the issues with using *C. reinhardtii* as a nuclear genetic engineering system, as well as with TFBS discovery in general. Performing a genome-wide chromatin immunoprecipitation sequencing (ChIP-seq) experiment to identify TFBSs *in vivo* using a *C. reinhardtii* TF library, would be a logical next step; this would identify TF DNA binding regions, providing more refined data for computational motif analysis (Franco-Zorrilla and Solano, 2017). Moreover, a pull-down assay using the pCRE DNA sequences as bait could be performed and followed by proteomic analysis of the bound protein complexes, thus potentially verifying their roles as TFBSs and identifying their cognate TFs (Wierer and Mann, 2016).

Developing a more comprehensive understanding of core promoter elements and structure in *C. reinhardtii* could further facilitate promoter building and optimisation, and lead to more promoter parts for tailored promoter design. In other host organisms, viral promoter parts are often used as strong core promoters (Pauli *et al*., 2004; Brown *et al*., 2014). To date, no known viruses can infect

C. reinhardtii (Weynberg *et al*., 2017), and the strong plant viral promoter CaMV35 drives poor gene expression (Ruecker *et al*., 2008; Díaz-Santos *et al*., 2013). Future core promoter design could be performed *in silico*, similar to the core promoter used in this study (Scranton *et al*., 2016), but with more in-depth and focussed analysis of the core promoter region. Identifying core promoter structural and functional elements in *C. reinhardtii* and testing them in different combinations, positions and copy numbers would be an obvious way forward. The methodology developed in this study could also be applied to find inducible promoter elements, such as those regulated by light or minerals that can be added to the media *e.g.* copper. Inducible motifs could be combined with other promoter elements to create strong inducible promoters.

Finally, synthetic promoters that can function across various species would be extremely useful. Some algal promoter elements display orthogonality and can drive expression other biological systems (Baek *et al*., 2016a; López-Paz *et al*., 2017). TFBSs tend to be highly conserved, so it would be likely that a *C. reinhardtii* promoter would be at least partially functional in other microalgae or even plants. Testing pCREs in another biotechnologically useful alga would be an interesting endeavour, as well as introducing pCREs from other taxa into microalgae.

The combination of computational promoter analysis and *in vivo* testing has augmented our understanding of what makes a strong promoter (Venter, 2007; Scranton *et al*., 2016; Koschmann *et al*., 2012). This study has provided some interesting leads for building optimised *C. reinhardtii* synthetic promoter modules, but more cycles of design-build-test will need to be implemented to optimise promoter motifs to a competitive standard. This has expanded the *C. reinhardtii* promoter repertoire, and potentially provided some insight into gene regulation in algae through the discovery of TFBSs.

6.4. Final remarks

The overarching aims of this thesis were to develop and apply various genetic engineering methods to improve *C. reinhardtii* as a biotechnological host for high-value carotenoid production. Targeted and reverse metabolic engineering approaches were successfully applied to increase lutein and b-carotene production, and novel promoter elements were identified and applied to improve the expression of transgenes from the *C. reinhardtii* nuclear genome and to facilitate the rational design of genetic engineering devices.

Working to replace current industrial processes that are reliant on petrochemicals with photosynthetic biotechnological systems, which have the dual benefit of reducing our dependence

on fossil fuels and of naturally fixing atmospheric carbon, is an increasingly important objective for researchers. The continued development and optimization of algal genome editing methods, as well as the identification of metabolic engineering targets, will, with perseverance, enable this goal to be achieved. The work undertaken in this thesis strives to contribute to the development of green algae as industrial biotechnology hosts.

Appendix A

This appendix contains recipes in the order of appearance in **Chapter 2**, and supplementary materials.

Appendix Table A1: Tris-acetate-phosphate (TAP) medium

Tris base (eg Trizma)		2.42 g
TAP salts solution		25 mL
Phosphate solution		1 mL
Hutner's trace elements		1 mL
(Glacial) acetic acid	CH_3COOH	1 mL

Make up to 1 L solution with dH_2O. Autoclave.

Appendix Table A2: TAP salts solution

Ammonium chloride	NH_4Cl	15 g
Magnesium sulphate heptahydrate	$MgSO_4(H_2O)_7$	4 g
Calcium chloride dihydrate	$CaCl_2(H_20)_2$	2 g

Make up to 1 L solution with dH_2O. Autoclave and store at 4°C.

Appendix Table A3: Phosphate solution for TAP medium

Monopotassium phosphate	K_2HPO_4	28.8 g
Dipotassium phosphate	KH_2PO_4	14.4 g

Make up to 100 mL solution with dH_2O.

Appendix Table A4: Hutner's Trace elements

Sodium EDTA dihydrate (titriplex III)	$Na_2EDTA(H_2O)_2$	5.00 g
Zinc sulphate heptahydrate	$ZnSO_4(H_2O)_7$	2.20 g
Boric acid	H_3BO_3	1.14 g

Magnesium chloride tetrahydrate	$MnCl_2(H_2O)_4$	0.50 g
Iron sulphate heptahydrate	$FeSO_4(H_2O)_7$	0.50 g
Cobalt chloride hexahydrate	$CoCl_2(H_2O)_6$	0.16 g
Copper sulphate pentahydrate	$CuSO_4(H2O)_5$	0.16 g
Hexaammonium molybdate	$(NH_4)_6MoO_3$	0.11 g

Make up to 100 mL solution with dH_2O. Filter sterilise and store at 4°C. Hutner's trace elements for this work were purchased premade from the Chlamydomonas Resource Centre.

Appendix Table A5: 10x TAE buffer

Tris base (eg Trizma)		96.8 g
Glacial acetic acid	CH_3COOH	22.8 mL
EDTA		7.4 g

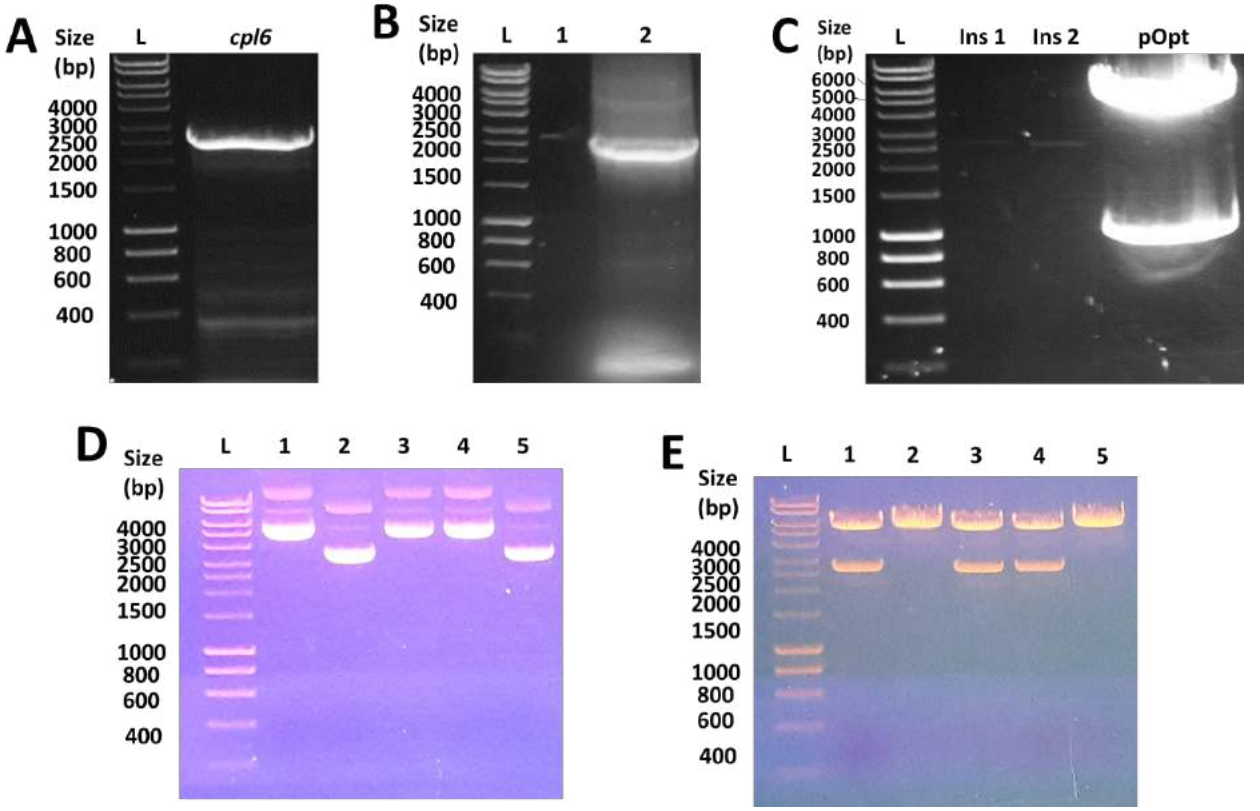
Make up to 2 L solution with dH_2O. Dilute 10x for use in DNA gel electrophoresis.

Appendix Table A6: IUPAC nucleotide base nomenclature system

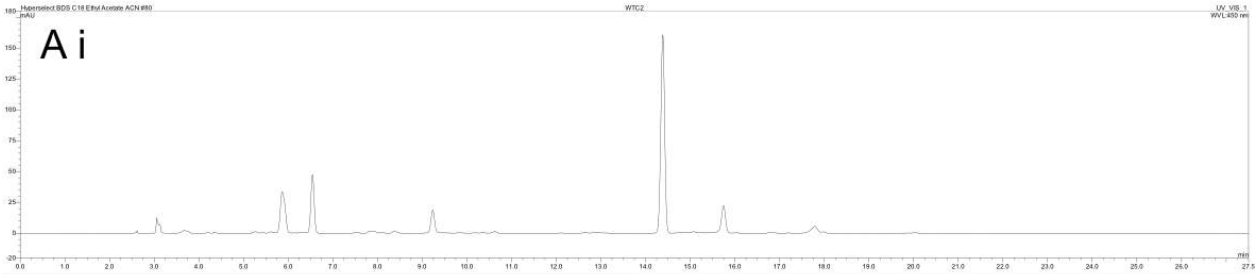
UPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T

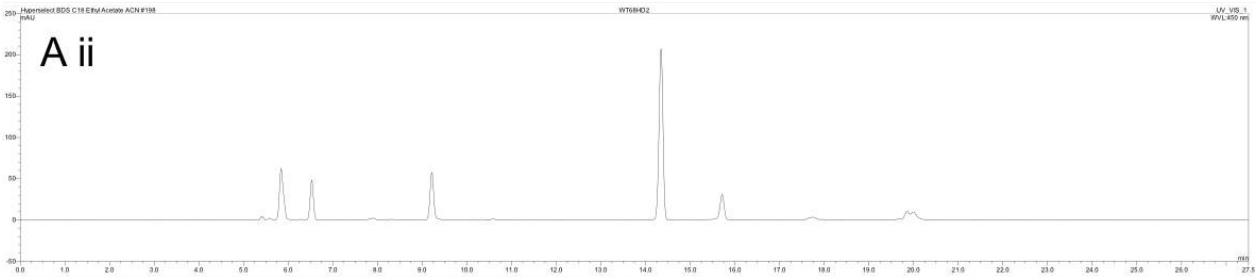
D	A or G or T
H	A or C or T
V	A or C or G
N	Any base or gap

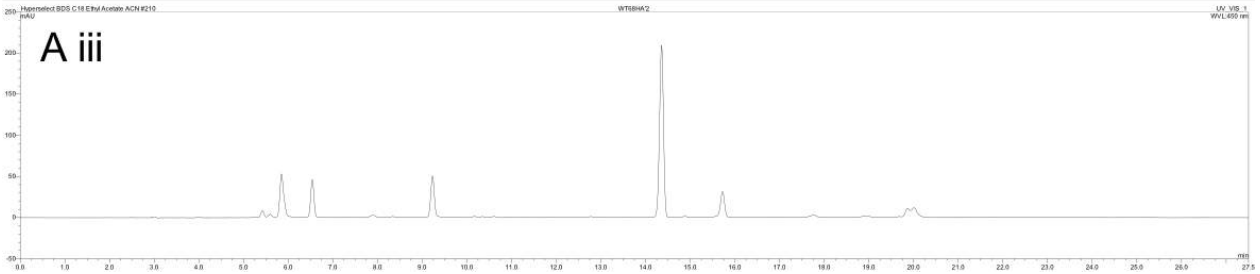
Appendix B: Supplementary material for Chapter 3

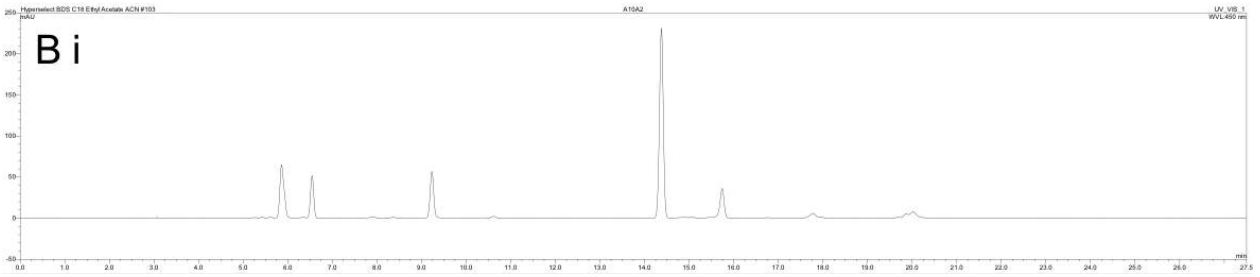


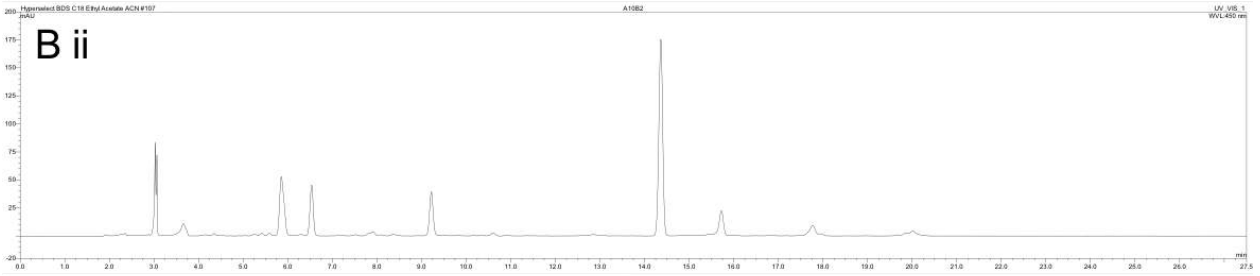
Appendix Figure B1: Gels showing stages of pOpt_crOR vector creation. For all gels, L = 1 kb DNA ladder (**Figure 2.2A**). **A -** Amplification of *cpl6* gene from CC-4533 gDNA. L, 1 kb Ladder; *clp6*, PCR product from amplification of CC-4533 gDNA using primers crOR_F and crOR_R (**Table 2.1**) for *cpl6* gene. Annealing temperature 62°C. *cpl6* expected fragment size = 2557 bp; the fragment at ~2500 bp was excised and gel purified. **B -** Amplification of His_6-*cpl6* fragment using primers crOR_N-His_F and crOR_R (**Table 2.1**). Purified *cpl6* DNA (Lane 1) was used as a template for His_6-*cpl6* (Lane 2). Expected fragment size = 2593 bp. **C -** Digested pOpt_mVenus_Paro vector and insert with NdeI and EcoRV restriction enzymes. Vector backbone (Lane pOpt; larger 5642 bp fragment) was excised and gel extracted for ligation with His_6-*cpl6* fragment (Lane Ins2; 2593 bp fragment). **D -** DNA minipreps of ligated pOpt vector and His_6-*cpl6* insert, extracted from transformed chemically competent *E. coli* DH5α cells. L, Ladder; 1–5, miniprepped DNA of selected colonies from pOpt + His_6-*cpl6* ligation. 1 µL miniprep DNA analysed per lane. **E -** Diagnostic digestion of plasmid DNA shown in **D** with NdeI and EcoRV restriction enzymes.

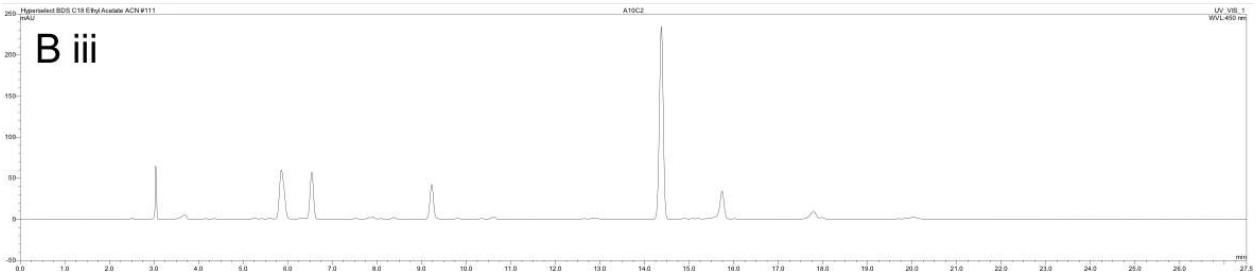

A i

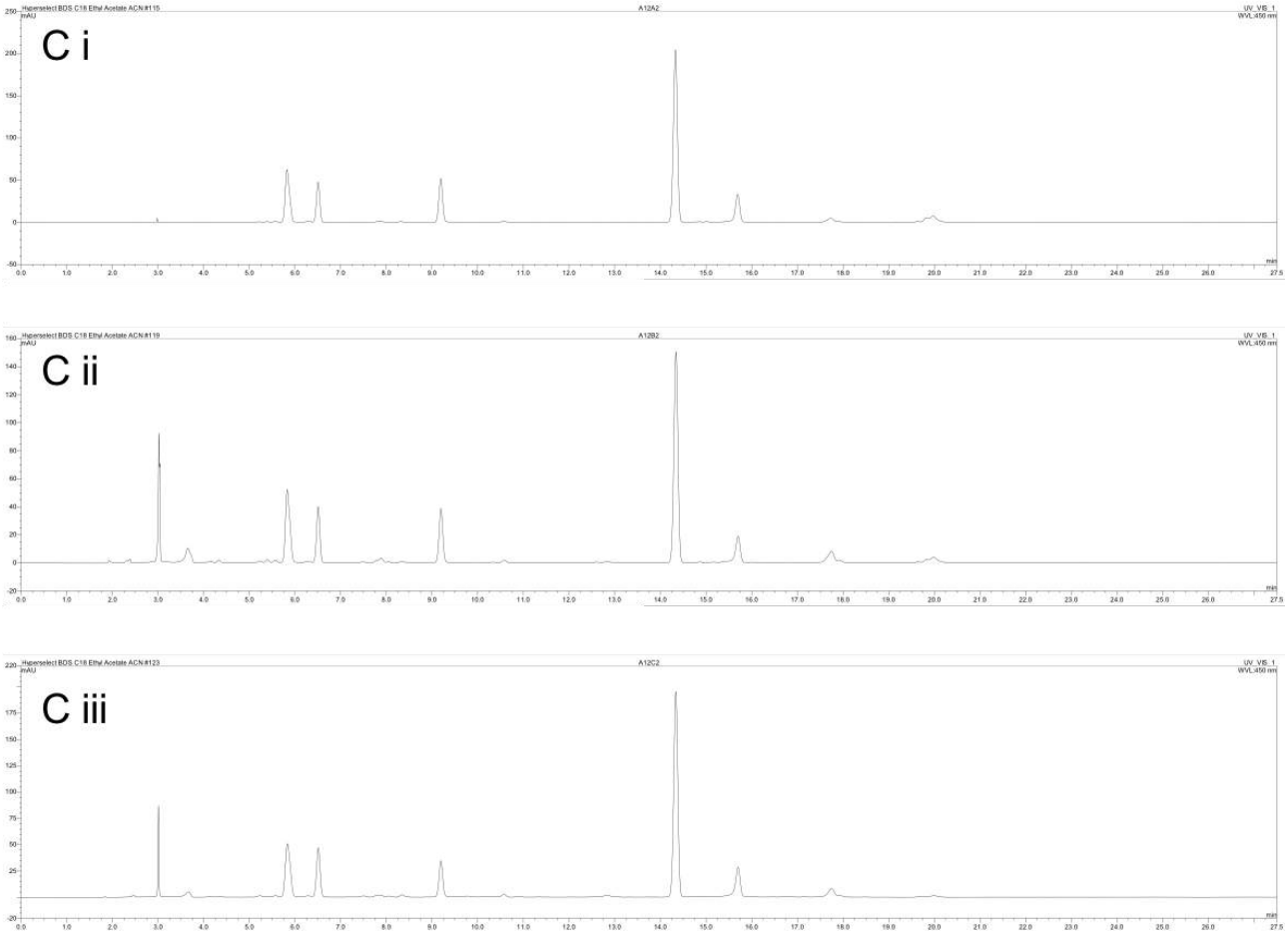

A ii


A iii

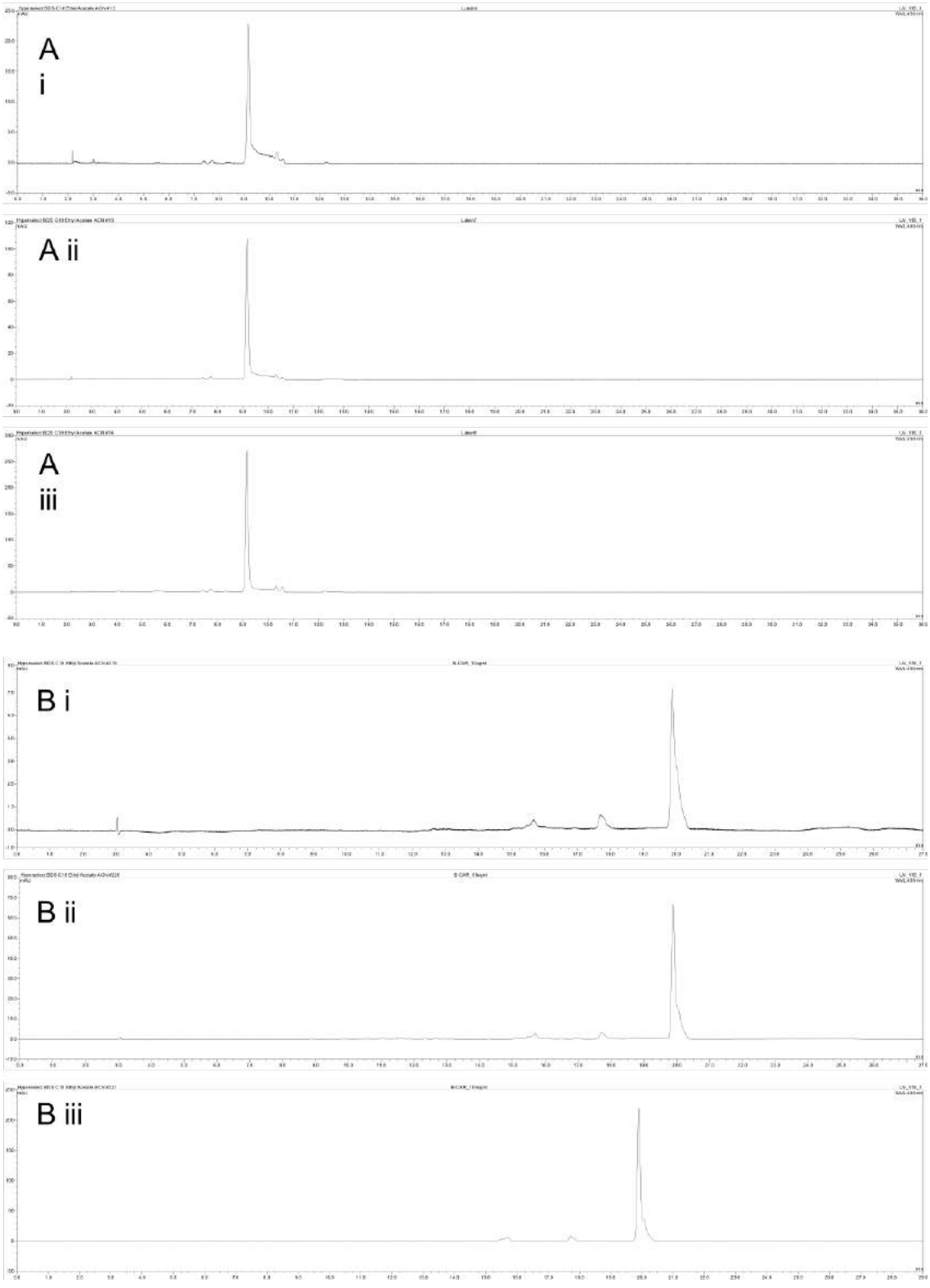

B i


B ii

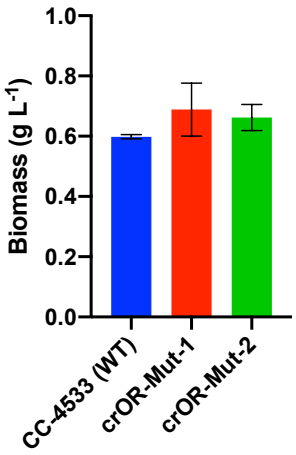

B iii



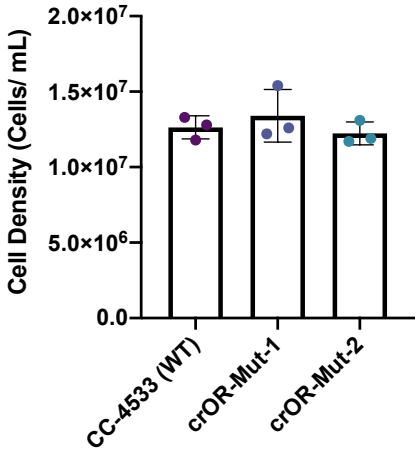
Appendix Figure B2: HPLC chromatograms of separated pigments extracted from strains CC-4533, crOR-Mut-1 and crOR-Mut-2. A – WT pigment separation. **B** – crOR-Mut-1 pigment separation. **C** – crOR-Mut-2 pigment separation. i, ii and iii represent pigment separations for individual biological replicates. X-axis shows retention time (min), y-axis shows absorption at 450 nm.



Appendix Figure B3: HPLC chromatograms of lutein and β-carotene pigment standards. A – lutein standard injected at i, 0.1 µg; ii, 0.3 µg; iii, 1 µg. **B** - β-carotene standard injected at i, 0.1 µg; ii, 0.5 µg; iii, 1 µg.



Appendix Figure B4: Dry Cell Weight (DCW) measurements for CC-4533 and transformant strains crOR-Mut-1 and crOR-Mut-2. N = 2 for CC-4533, n = 3 for crOR-Mut-1 and crOR-Mut-2



Appendix Figure B5: Cell density measurements for CC-4533 and transformant strains crOR-Mut-1 and crOR-Mut-2. Cell counts of cultures extracted for carotenoid analysis. These measurements were used to calculate carotenoids/ cell.

Appendix Table B1: Retention times and peak areas for lutein and β-carotene standards

Carotenoid standard	Concentration (µg mL^{-1})	Mass injected (µg)	Peak Area (mAU)	Retention Time (min)
β-carotene	5	0.05	0.111	19.88
	10	0.1	1.520	19.89
	50	0.5	10.497	19.89
	100	1	31.171	19.90
lutein	10	0.100	2.553	9.17

	30	0.300	12.345	9.17
	100	1.000	31.977	9.16

Appendix Table B2: Raw retention times for pigment elution for CC-4533 and transformant strains crOR-Mut-1 and crOR-Mut-2

		Retention time (min)					
Strain	BR	Neo	Viol	Lut	Chl-*b*	Chl-*a*	B-car
	1[a]	5.87	6.55	9.25	14.40	15.76	19.89
WT	2[a]	5.87	6.55	9.24	14.40	15.76	19.90
	3[a]	5.85	6.54	9.23	14.38	15.75	19.88
	1[b]	6.10	6.80	9.60	14.89	16.24	20.39
crOR1	2[b]	6.10	6.80	9.61	14.90	16.25	20.41
	3[a]	5.85	6.54	9.24	14.38	15.75	19.89
	1[b]	6.12	6.82	9.62	14.90	16.25	20.41
crOR2	2[b]	6.11	6.81	9.61	14.89	16.25	20.39
	3[a]	5.85	6.54	9.24	14.38	15.75	19.89

Raw retention times for each eluted pigment to 2 decimal places. BR = biological replicate. Two 'sets' of retention times were observed between the samples, denoted a and b. In set b, the retention times are shifted later in time towards the right of the spectra, but appear to represent the same pigment profiles (**Appendix Figure B2**). Leakage of the HPLC system was an issue during several of the sample runs, which could explain the set of altered retention times.

Appendix Table B3: Raw values for pigments detected via HPLC for CC-4533 and transformant strains crOR-Mut-1 and crOR-Mut-2

		Peak Area (mAU)					
Strain	BR	Neo	Viol	Lut	Chl-*b*	Chl-*a*	B-car
	1	5.523	4.417	4.050	18.083	3.516	0.094
WT	2	5.378	4.794	3.684	17.290	2.991	0.005
	3	5.275	4.911	4.138	18.114	2.774	1.004
	1	9.285	5.961	6.760	28.114	5.447	0.481
crOR1	2	8.521	5.402	6.433	24.779	4.436	0.391

	3	7.690	5.325	5.136	21.882	3.070	1.138
	1	8.804	5.271	6.819	25.755	4.776	0.698
crOR2	2	9.008	5.466	8.807	30.924	5.558	1.707
	3	7.921	5.475	5.755	24.926	3.703	1.193

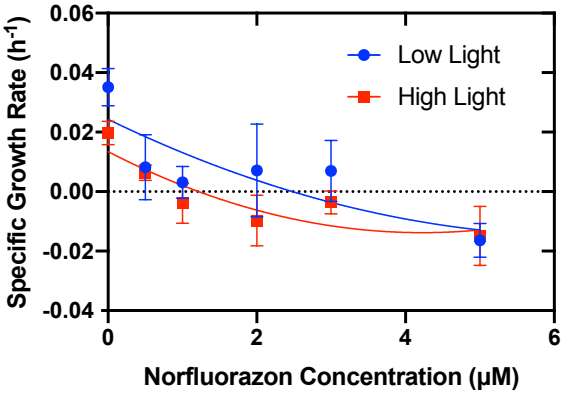
Raw peak areas for each eluted pigment to 4 significant figures. BR = biological replicate.

Appendix Table B4: Peak areas for all pigments detected via HPLC for CC-4533 and transformant strains crOR-Mut-1 and crOR-Mut-2 adjusted to cell number

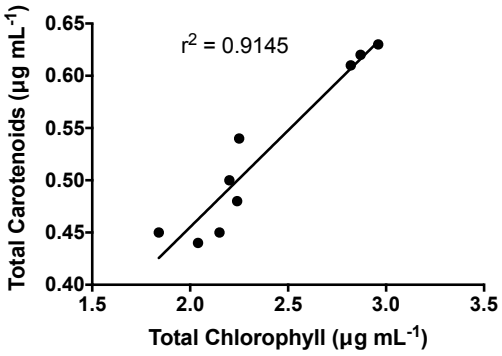
	Peak area / 10^7 cells					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	B-car
CC-4533	4.29 ± 0.38	3.73 ± 0.04	3.14 ± 0.28	14.17 ± 1.05	2.47 ± 0.47	0.28 ± 0.41
crOR-Mut-1	$6.38 \pm 0.38^{**}$	$4.18 \pm 0.27^{*}$	$4.58 \pm 0.48^{*}$	$18.65 \pm 0.95^{**}$	3.20 ± 0.59	0.52 ± 0.36
crOR-Mut-2	$7.02 \pm 0.45^{**}$	$4.42 \pm 0.22^{**}$	$5.79 \pm 0.96^{**}$	$22.18 \pm 1.38^{**}$	$3.81 \pm 0.62^{*}$	0.97 ± 0.36

Peak area calculated as follows: (Peak Area / cell number) x 10^7. Statistically significant differences from the CC-4533 means were calculated using a student's t-test. *$P < 0.05$, **$P < 0.01$. Error bars = SD; n = 3.

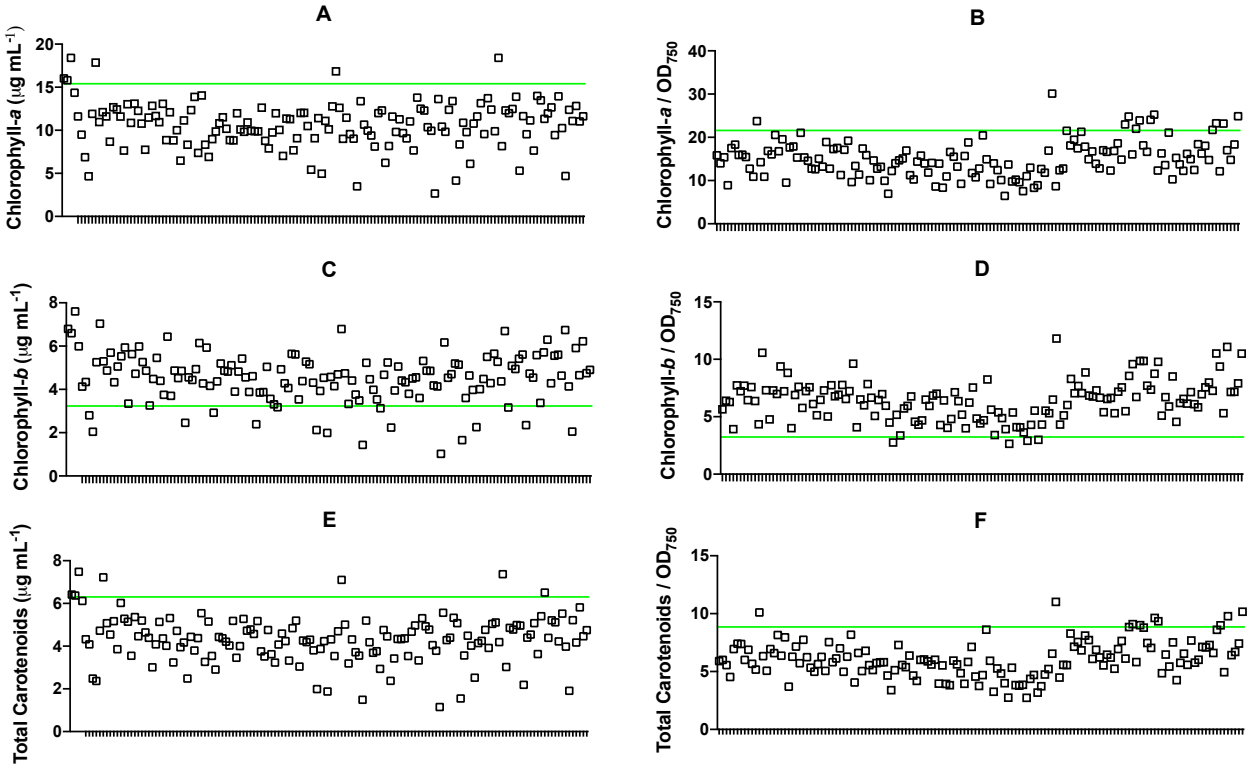
Appendix C: Supplementary material for Chapter 4



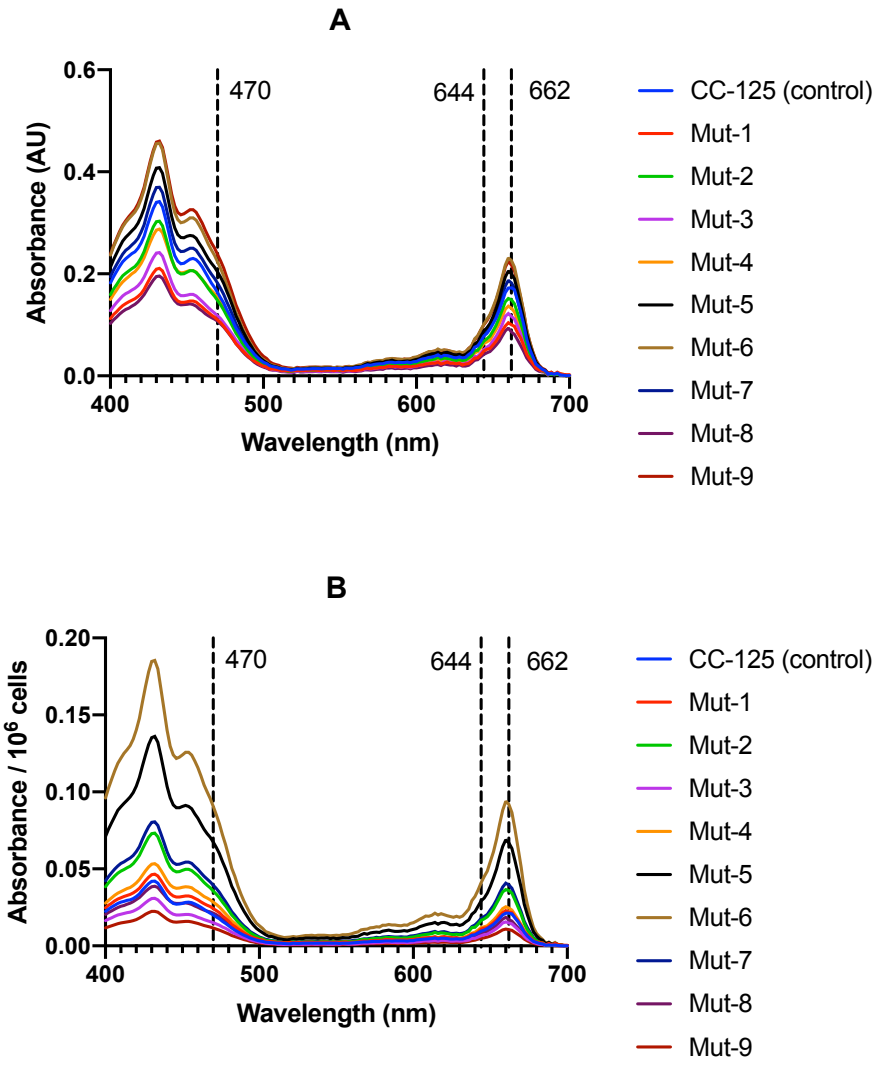
Appendix Figure C1: Specific growth rates of CC-125 grown in increasing levels of norflurazon in low- and high-light conditions. Cultures were grown in 96-well plates in triplicate. Specific growth rate calculated from chlorophyll fluorescence values measured at OD750 using a microplate reader. Means of individual SGRs for each condition plotted with error bars representing SD. Polynomial regression shown for each dataset using following equations: Low light Y = 0.02431 + -0.01212*X + 0.0009359*X^2 (r^2 = 0.5603); High light Y = 0.01346 + -0.01289*X + 0.001526*X^2 (r^2 = 0.6066). SGRs calculated from cell density measurements at 750 nm were significantly lower under high light conditions than in low light (*P* = 0.0141). For high light-grown cultures supplemented with norflurazon, SGRs appear to be lower than those grown in low-light, however there is no significant difference between norflurazon concentrations grown in either light condition; the coefficients of variation for norflurazon concentrations > 0 µM are > 30%, which is likely due to uneven growth conditions within 96-well plates. Despite the lack of statistical significance between high and low light conditions at norflurazon concentrations > 0 µM, high light conditions were still taken forward for selection with norflurazon.



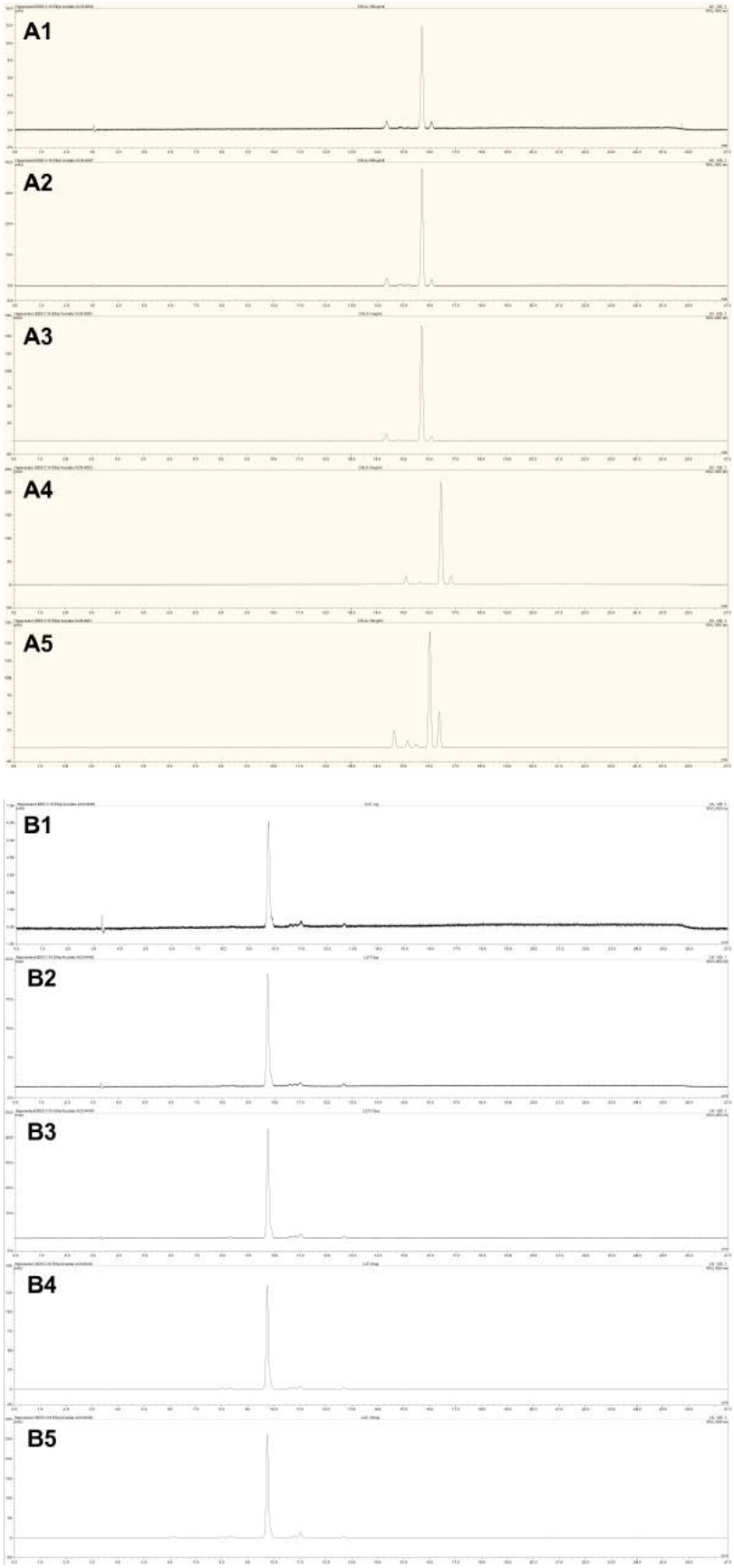
Appendix Figure C2: Correlation between total carotenoid and chlorophyll concentrations. Values obtained from measurements of WT (CC-4533) and pOpt_crOR transformants crOR-Mut-1 and crOR-Mut-2.



Appendix Figure C3: Total pigment contents for 144 EMS mutants. Pigment measurements calculated from pigment extractions of 144 mutants grown on 24-well plates during round 2 of screening. **A** – Chlorophyll-*a* measurements / mL culture. **B** – Chlorophyll-*a* measurements normalised to OD750. **C** – Chlorophyll-*b* measurements / mL culture. **D** – Chlorophyll-*b* measurements normalised to OD750. **E** – Total carotenoid measurements / mL culture. **F** – Total carotenoids normalised to OD750. Each square represents an individual mutant. Green line represents mean value for WT cultures.

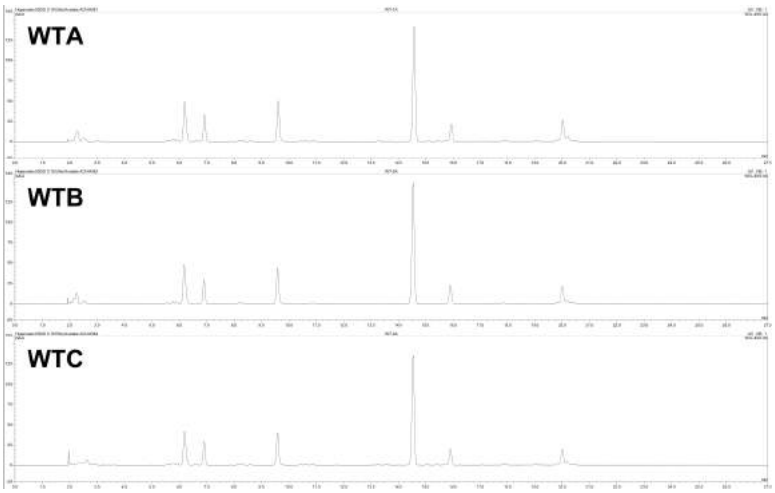


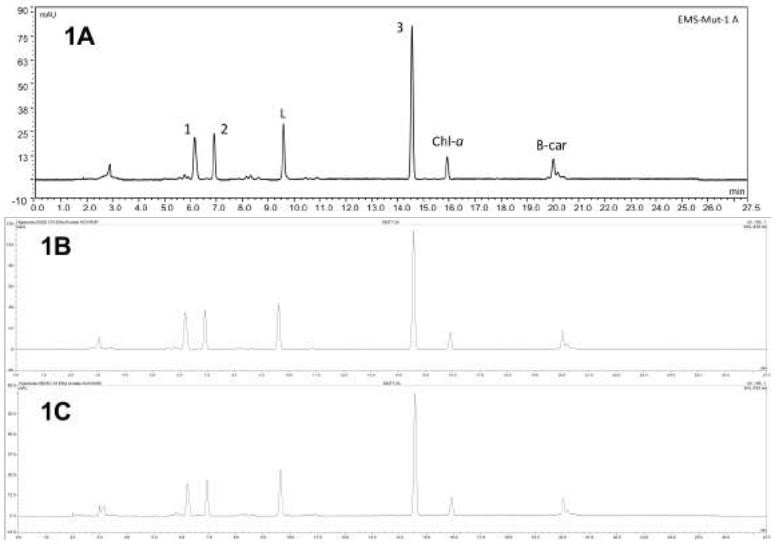
Appendix Figure C4: Wavelength scans for EMS mutant and parental CC-125 strain pigment extractions with 100% acetone. A – Raw absorption measurements. **B** – Absorption normalised to OD_{750} measurements of culture taken prior to extraction. Readings were taken at 2 nm intervals between 400 and 700 nm wavelengths. Wavelengths of peaks used to calculate pigment concentrations highlighted with dotted line.

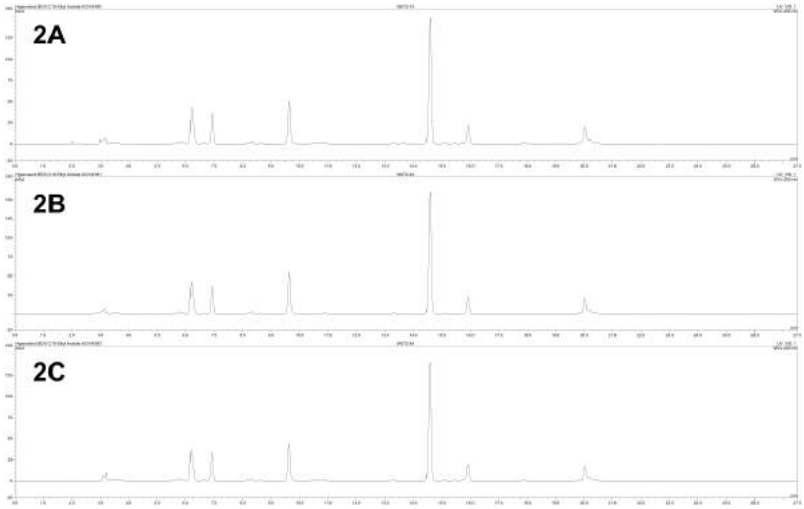


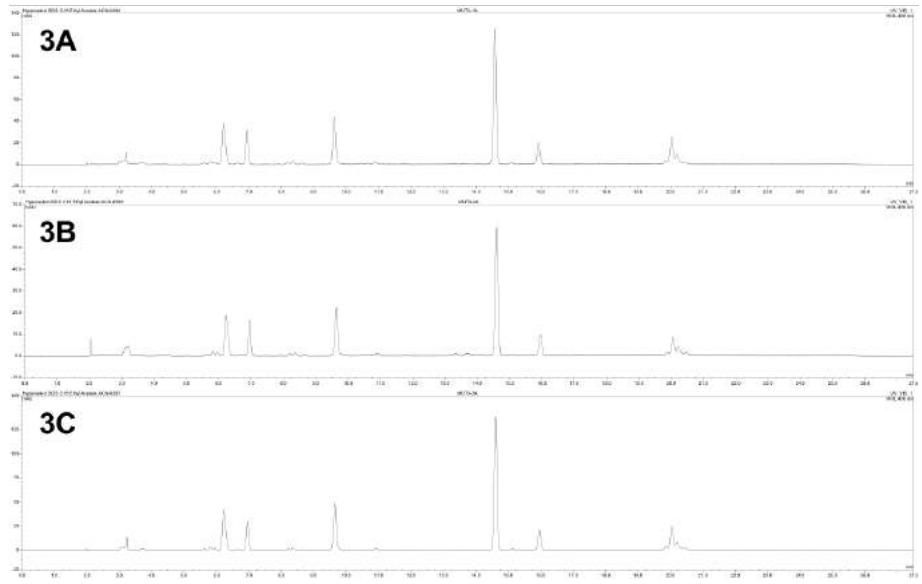
Appendix Figure C5: HPLC chromatograms of pigment standards. A – Chlorophyll-a standard at the following concentrations injected: A1, 100 µg/ mL; A2, 300 µg/ mL; A3, 1 mg/ mL; A4, 3 mg/ mL; A5, 10 mg/ mL. **B** –

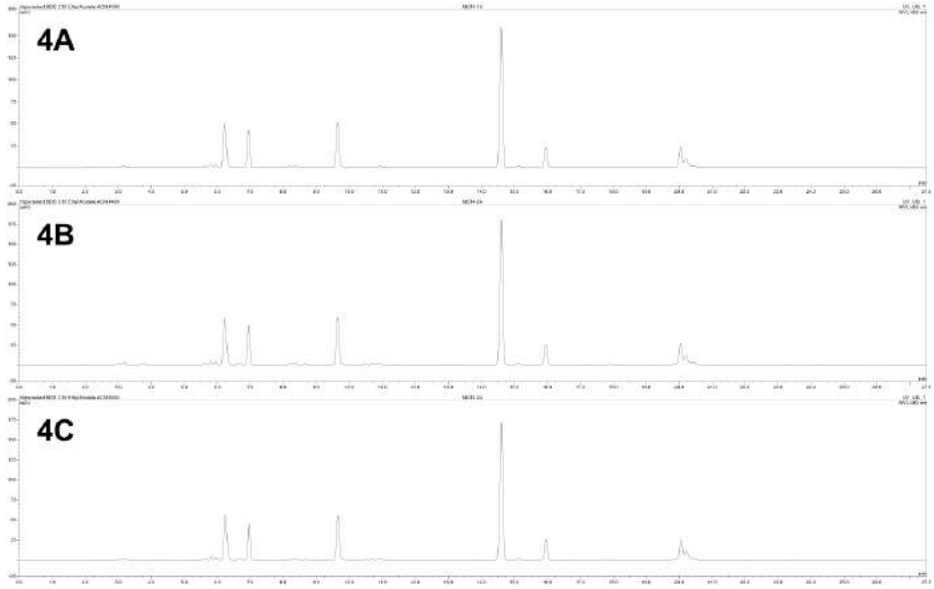
Lutein standard injected at the following concentrations: B1, 1 µg/ mL; B2, 3 µg/ mL; B3, 10 µg/ mL; B4, 30 µg/ mL; B5, 100 µg/ mL.

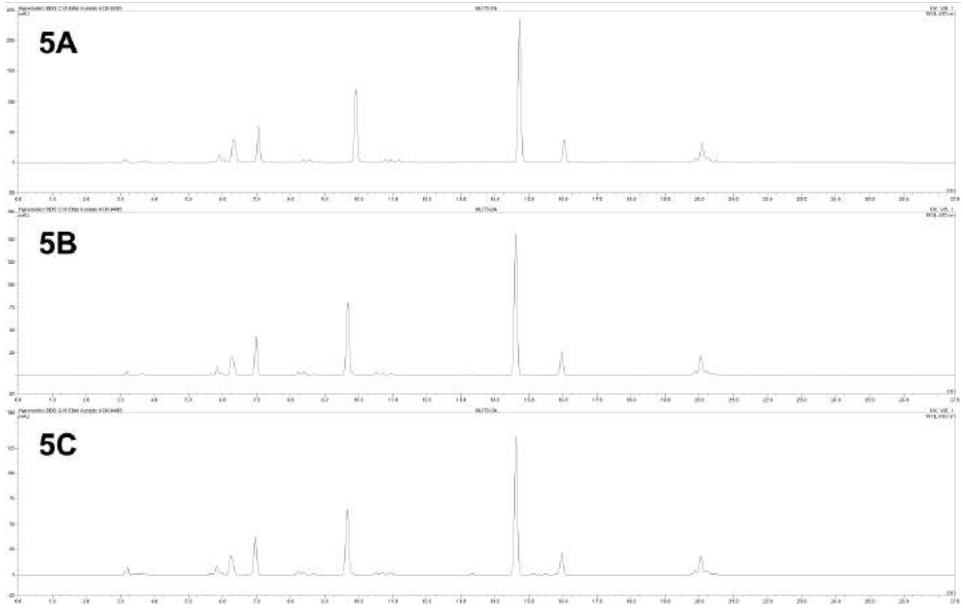


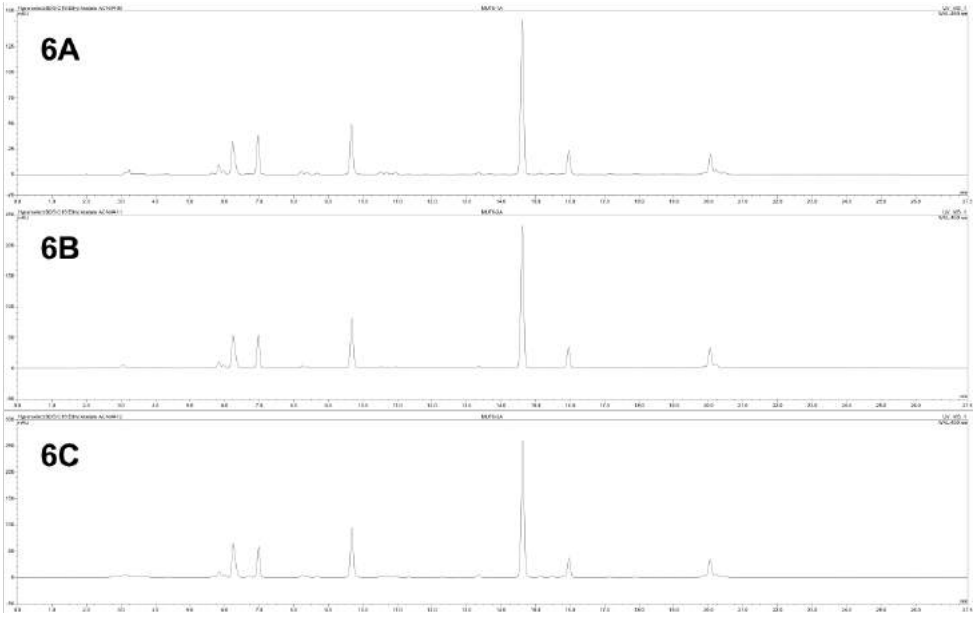


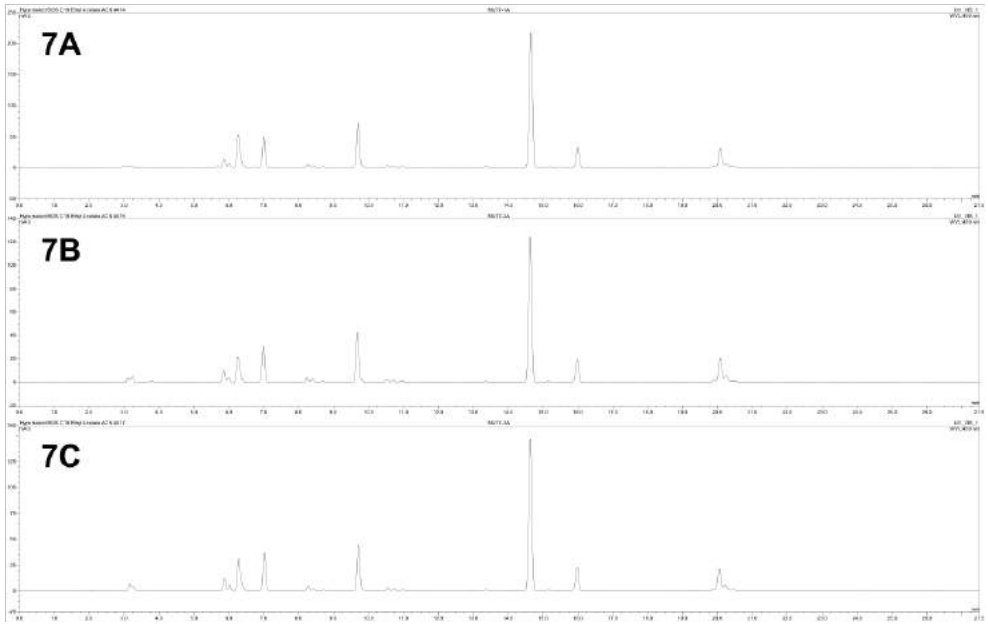


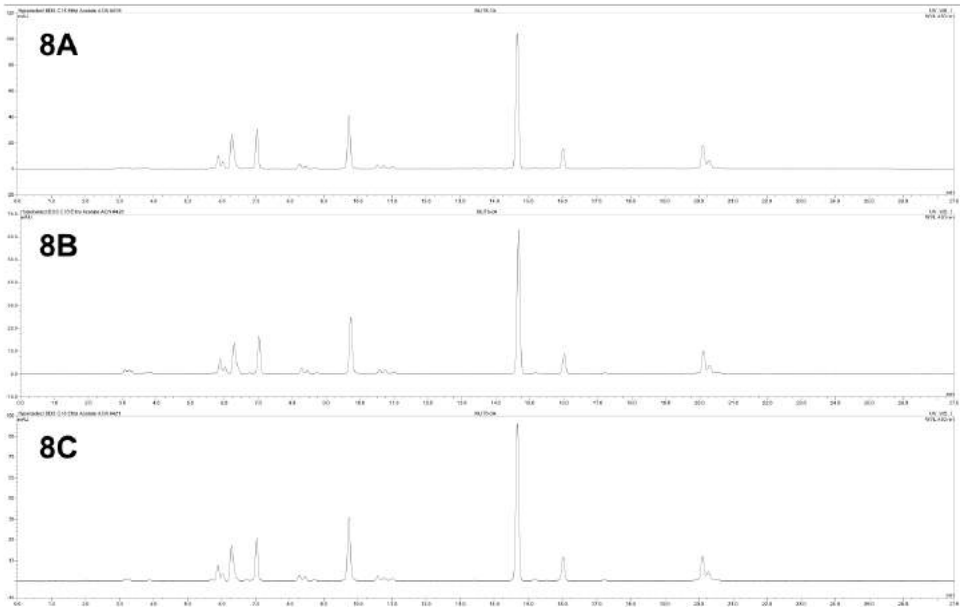

3A
3B
3C

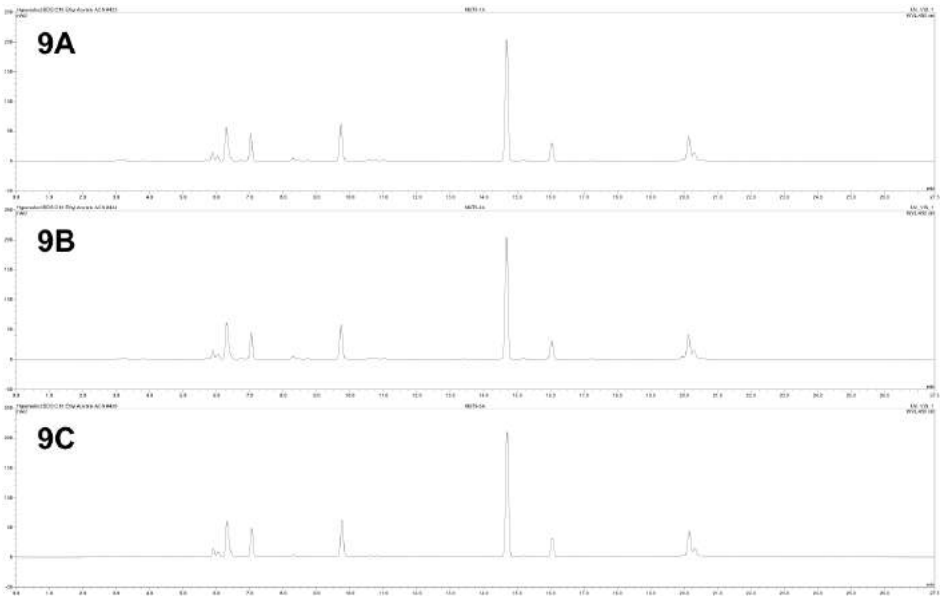

4A
4B
4C


5A
5B
5C

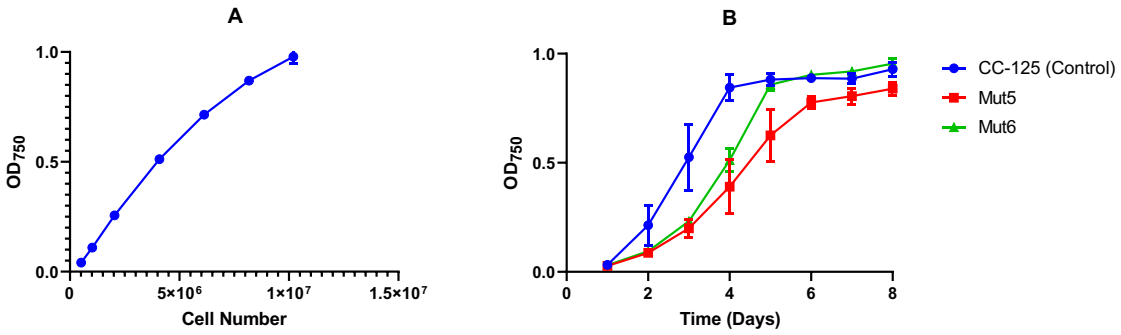

6A
6B
6C


7A
7B
7C

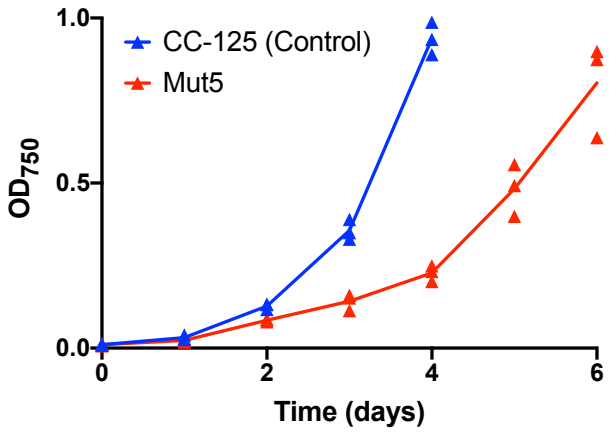

8A
8B
8C



Appendix Figure C6: HPLC chromatograms of pigment extractions for each EMS mutant and WT. WTA, B, C – WT pigment spectra for biological replicates 1–3. **1–9** – Pigment spectra for EMS mutants 1–9. Biological replicates denoted A, B and C.

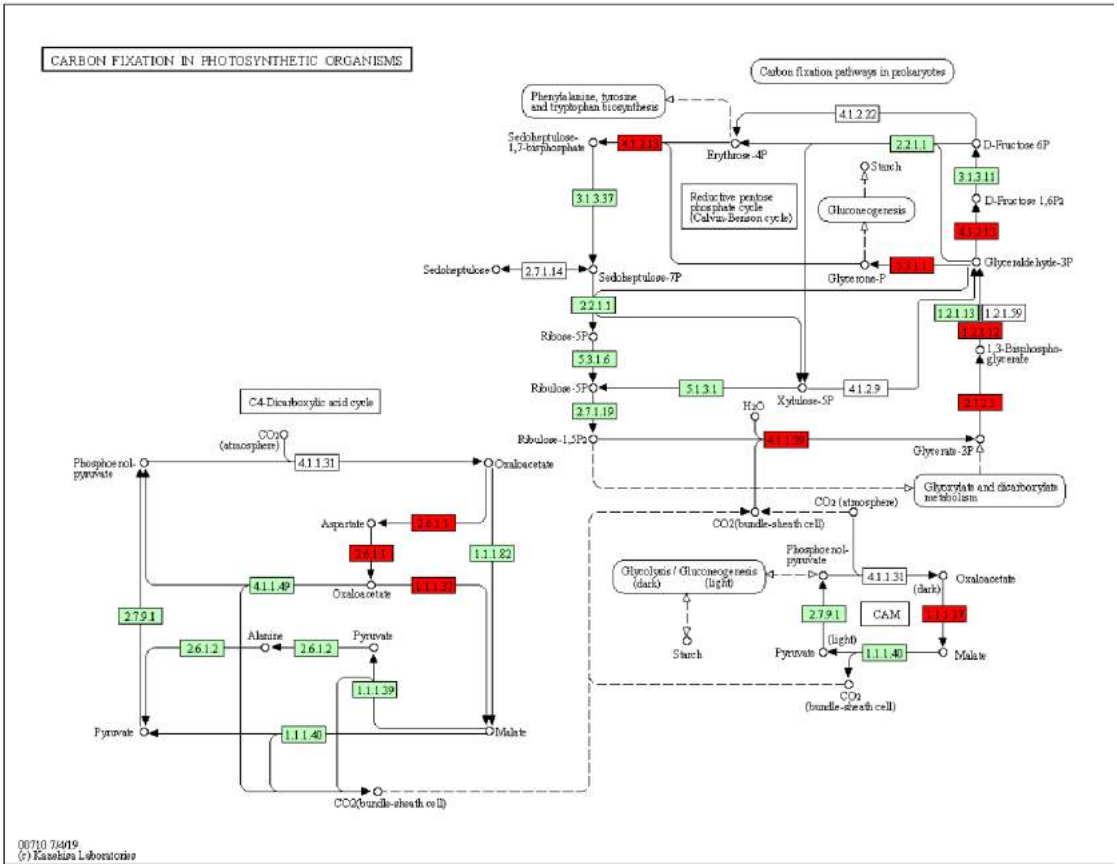


Appendix Figure C7: Calibration curve OD vs cell number. Calibration curve calculated using CC-125 strain using a haemocytometer. Values interpolated using Graphpad Prism.

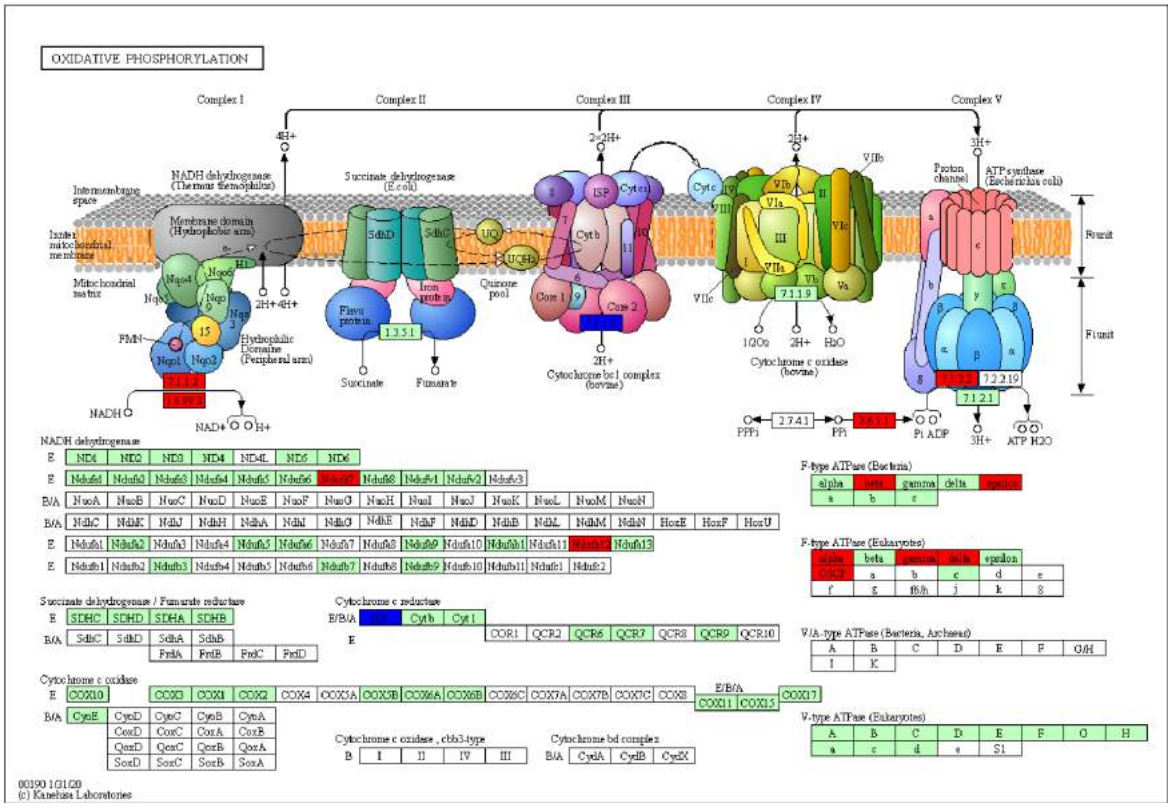


Appendix Figure C8: Growth curves of samples taken for proteomics analysis. No significant difference between final OD750 measurements for CC-125 and EMS-Mut-5 (*P* = 0.2028; *t*-test)

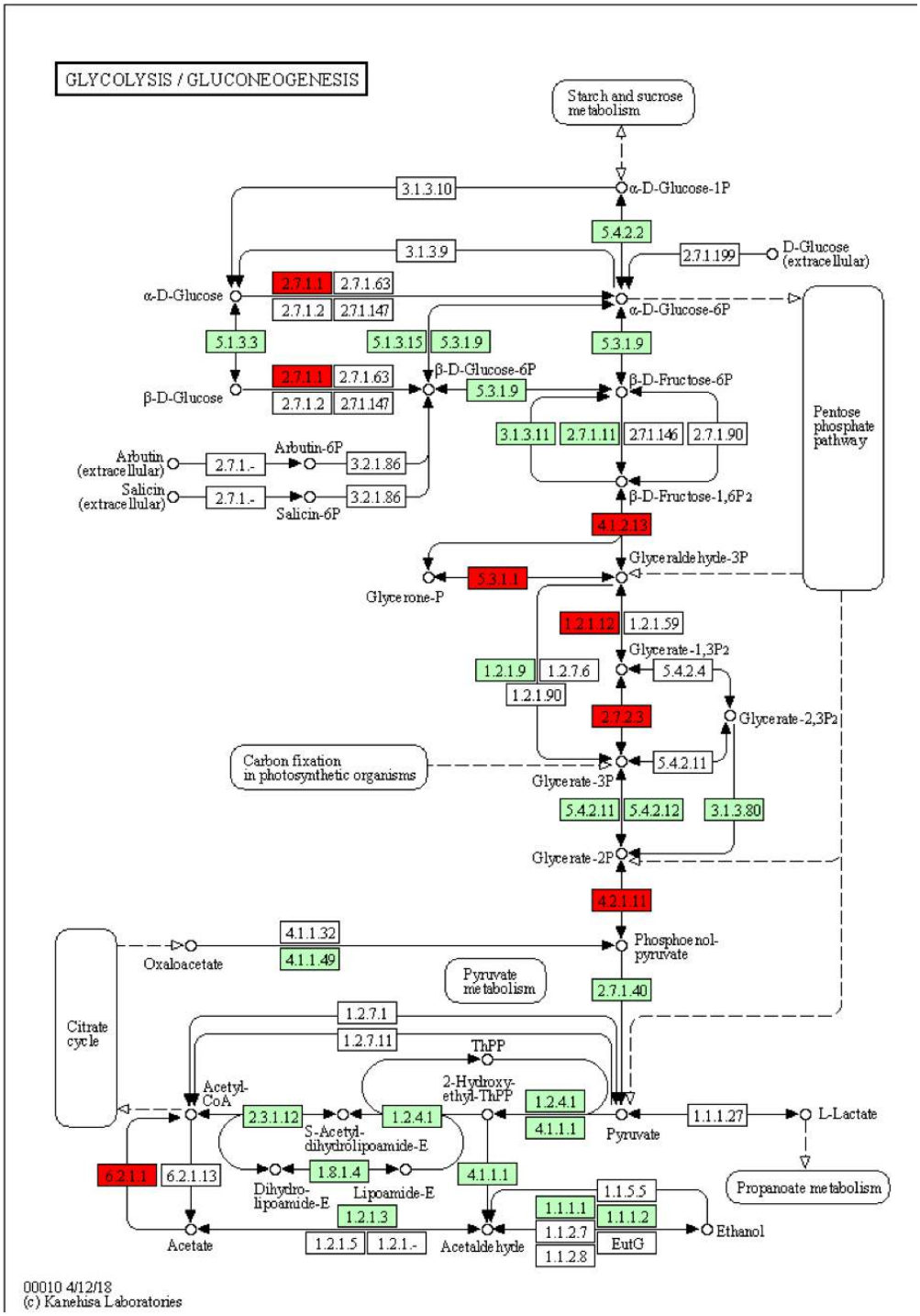
A



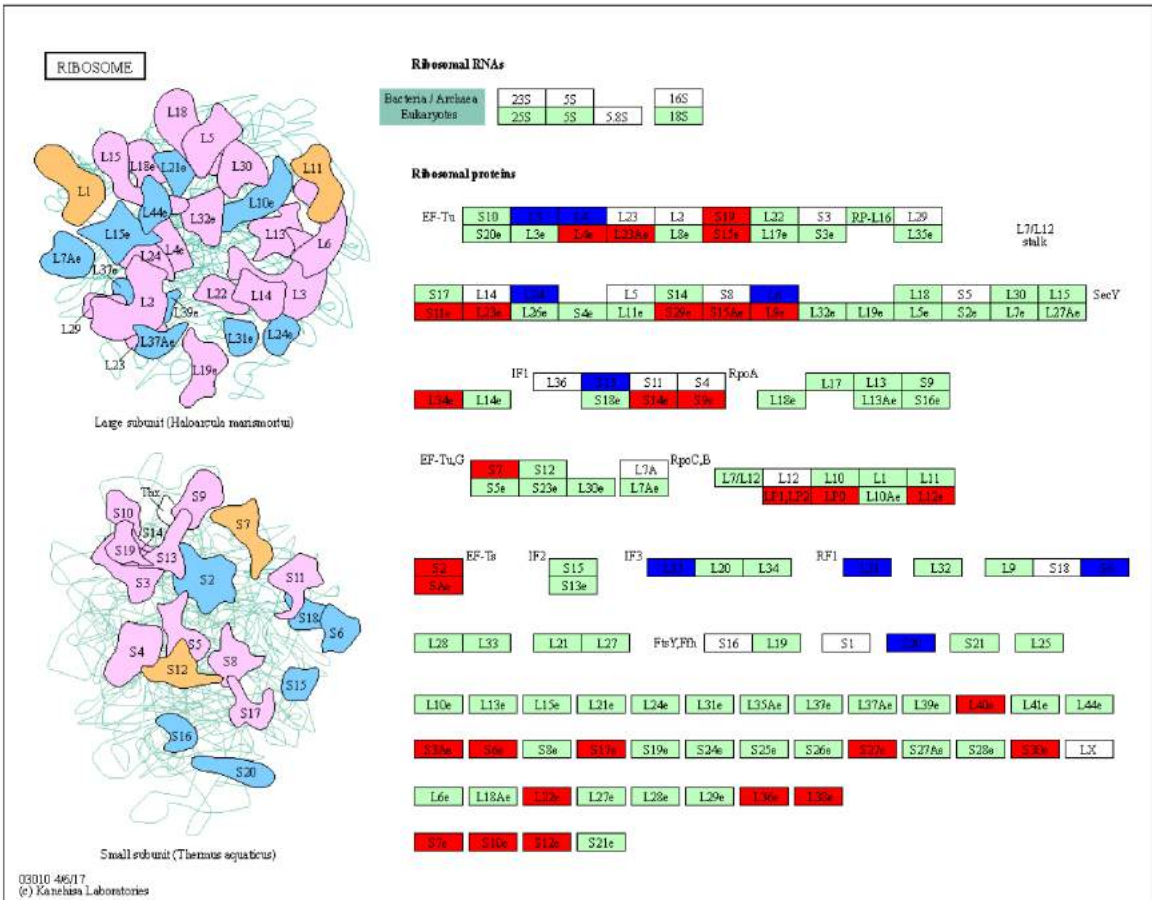
B


OXIDATIVE PHOSPHORYLATION
Complex I
Complex II
Complex III
Complex IV
Complex V
NADH dehydrogenase (Thermus thermophilus)
Succinate dehydrogenase (E.coli)
Proton channel
ATP synthase (Escherichia coli)
Intermembrane space
Inner mitochondrial membrane
Membrane domain (Hydrophobic arm)
Mitochondrial matrix
Hydrophilic Domain (Peripheral arm)
Flavo protein
Iron protein
Quinone pool
Core 1
Core 2
Cytochrome bc1 complex (bovine)
Cytochrome c oxidase (bovine)
NADH
NAD+
Succinate
Fumarate
PPi
Pi ADP
ATP H2O
NADH dehydrogenase
F-type ATPase (Bacteria)
F-type ATPase (Eukaryotes)
Succinate dehydrogenase / Fumarate reductase
Cytochrome c reductase
V/A-type ATPase (Bacteria, Archaea)
Cytochrome c oxidase
Cytochrome c oxidase , cbb3-type
Cytochrome bd complex
V-type ATPase (Eukaryotes)
00190 1G1G9
(c) Kanehisa Laboratories

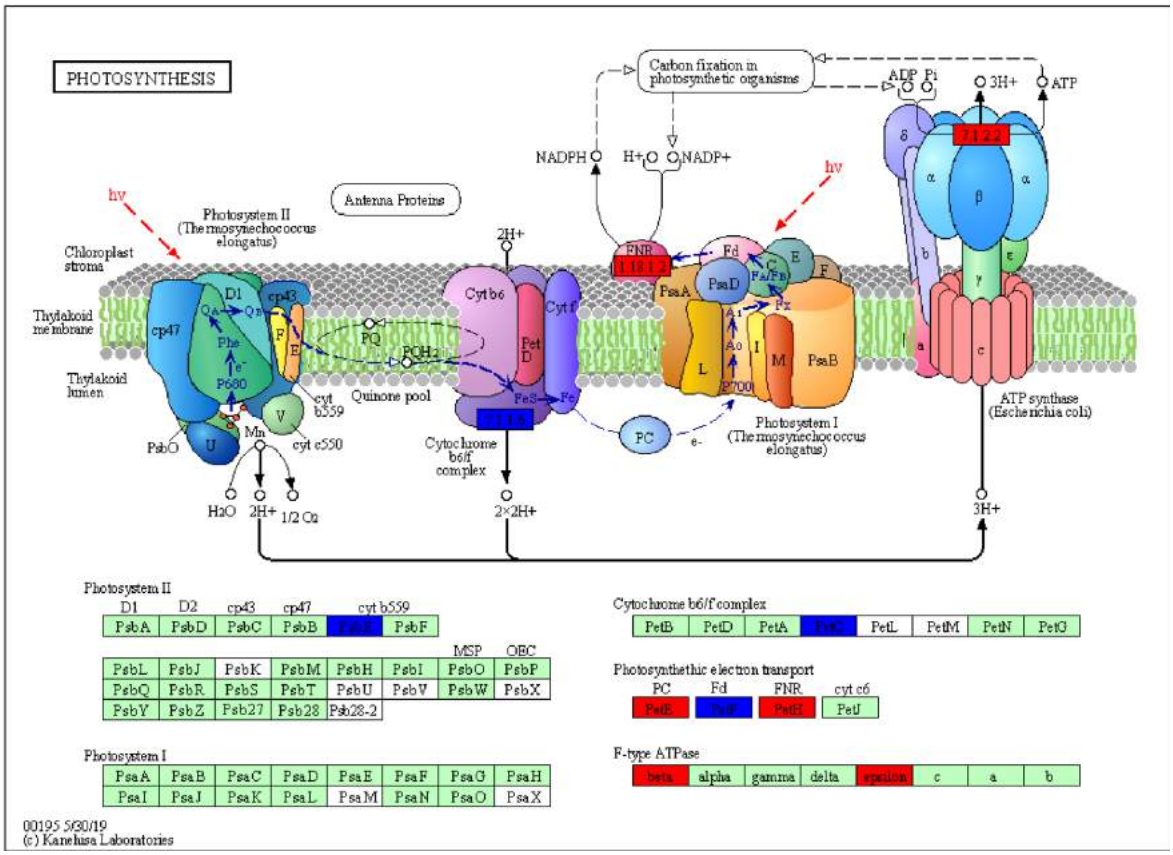
C


GLYCOLYSIS / GLUCONEOGENESIS
Starch and sucrose metabolism
α-D-Glucose-1P
3.1.3.10
5.4.2.2
3.1.3.9
2.7.1.199
D-Glucose (extracellular)
α-D-Glucose
2.7.1.1
2.7.1.63
2.7.1.2
2.7.1.147
α-D-Glucose-6P
5.1.3.3
5.1.3.15
5.3.1.9
β-D-Glucose
β-D-Glucose-6P
β-D-Fructose-6P
Pentose phosphate pathway
3.1.3.11
2.7.1.11
2.7.1.146
2.7.1.90
Arbutin (extracellular)
Arbutin-6P
2.7.1.-
3.2.1.86
Salicin (extracellular)
Salicin-6P
β-D-Fructose-1,6P2
4.1.2.13
Glyceraldehyde-3P
Glycerone-P
5.3.1.1
1.2.1.12
1.2.1.59
Glycerate-1,3P2
1.2.1.9
1.2.7.6
1.2.1.90
5.4.2.4
2.7.2.3
Glycerate-2,3P2
Carbon fixation in photosynthetic organisms
5.4.2.11
Glycerate-3P
5.4.2.11
5.4.2.12
3.1.3.80
Glycerate-2P
4.2.1.11
4.1.1.32
4.1.1.49
Oxaloacetate
Phosphoenol-pyruvate
Pyruvate metabolism
2.7.1.40
Citrate cycle
1.2.7.1
1.2.7.11
ThPP
Acetyl-CoA
2-Hydroxy-ethyl-ThPP
1.2.4.1
4.1.1.1
2.3.1.12
1.2.4.1
S-Acetyl-dihydrolipoamide-E
Pyruvate
1.1.1.27
L-Lactate
6.2.1.1
6.2.1.13
1.8.1.4
4.1.1.1
Dihydro-lipoamide-E
Lipoamide-E
Propanoate metabolism
1.1.5.5
1.1.1.1
1.1.1.2
1.2.1.3
1.1.2.7
EutG
Ethanol
Acetate
1.2.1.5
1.2.1.-
Acetaldehyde
1.1.2.8
00010 4/12/18
(c) Kanehisa Laboratories

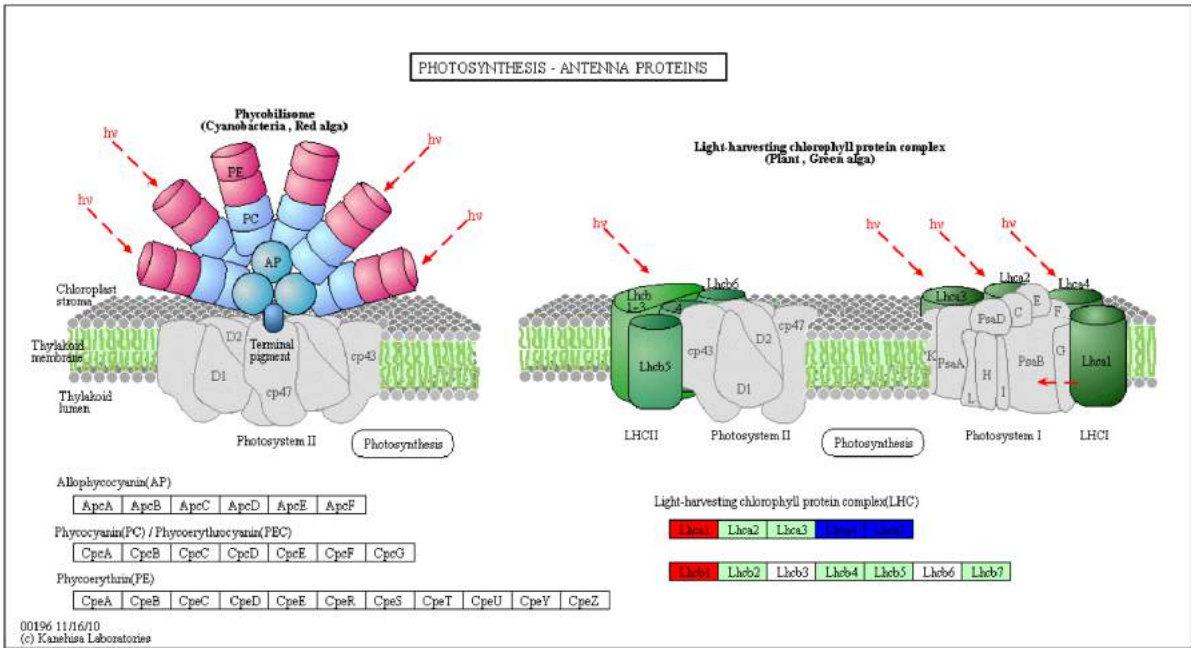
D


RIBOSOME
Large subunit (Haloarcula marismortui)
Small subunit (Thermus aquaticus)
Ribosomal RNAs
Bacteria / Archaea
Eukaryotes
23S
5S
16S
5.8S
18S
Ribosomal proteins
EF-Tu
RP-L16
L7/L12 stalk
SecY
RpoA
RpoC,B
EF-Ts
IF1
IF2
IF3
RF1
FtsY,Ffh
03010 46/17
(c) Kanehisa Laboratories

E


PHOTOSYNTHESIS
Carbon fixation in photosynthetic organisms
ADP Pi
3H+
ATP
NADPH
H+
NADP+
hv
Antenna Proteins
Photosystem II (Thermosynechococcus elongatus)
Chloroplast stroma
Thylakoid membrane
Thylakoid lumen
2H+
Cyt b6
Cyt f
PsaA
PsaD
PsaB
Fd
FNR
PC
Quinone pool
cyt b559
Mn
cyt c550
PsbO
H2O
2H+
1/2 O2
Cytochrome b6/f complex
2×2H+
Photosystem I (Thermosynechococcus elongatus)
ATP synthase (Escherichia coli)
Photosystem II
D1 D2 cp43 cp47 cyt b559
PsbA PsbD PsbC PsbB PsbE PsbF
MSP OEC
PsbL PsbJ PsbK PsbM PsbH PsbI PsbO PsbP
PsbQ PsbR PsbS PsbT PsbU PsbV PsbW PsbX
PsbY PsbZ Psb27 Psb28 Psb28-2
Photosystem I
PsaA PsaB PsaC PsaD PsaE PsaF PsaG PsaH
PsaI PsaJ PsaK PsaL PsaM PsaN PsaO PsaX
Cytochrome b6/f complex
PetB PetD PetA PetC PetL PetM PetN PetG
Photosynthetic electron transport
PC Fd FNR cyt c6
PetE PetF PetH PetJ
F-type ATPase
beta alpha gamma delta epsilon c a b
00195 5/30/19
(c) Kanehisa Laboratories

F



Appendix Figure C9: KEGG pathways of interest showing proteins that are increased (blue) or decreased (red) in EMS-Mut-5 compared to WT. A – Carbon fixation of photosynthetic organisms. **B** – oxidative phosphorylation. **C** – glycolysis/ gluconeogenesis. **D** – ribosomal proteins. **E** – photosynthesis. **F** – photosynthesis antenna proteins.

Appendix Table C1: Results from first round of mutant screening for potential high carotenoid-producing strains

Plate	Initial NF conc. (μM)	Specific growth rate (h^{-1})		Chlorophyll fluorescence (AU)		# hits
		Mean	Mean + 1 SD	Mean	Mean + 1 SD	
A	0.5	0.02974 ± 0.01534	0.04509	80.37 ± 36.97	117.3	17
B	0.5	0.03077 ± 0.01938	0.05015	69.14 ± 37.48	106.6	16
C	0.5	0.03100 ± 0.01505	0.04605	73.31 ± 44.65	118.0	13
D	0.5	0.03096 ± 0.01524	0.04620	65.68 ± 34.22	99.89	17
E	1	-0,01116 ± 0.02812	0.01697	43.74 ± 23.89	67.63	28
F	1	-0.01680 ± 0.03861	0.02181	51.38 ± 31.82	83.19	20
G	1	0.006438 ± 0.02542	0.03186	38.68 ± 29.50	68.18	16
H	1	0.02281 ± 0.01680	0.03960	55.77 ± 44.96	100.7	17

8 separate 96-well plates containing individually picked mutant colonies in individual wells, named A–H. Initial NF conc. = concentration of norflurazon on which picked colonies were first grown before transfer to

1 μM norflurazon. Mean calculated specific growth rate and chlorophyll fluorescence shown ± SD, to 4 significant figures. Specific growth rate calculated between days 1 and 2. Chlorophyll fluorescence measurement taken on day 4. # hits represent number of mutants per plate that passed one or both criteria depicted in **Figure 4.3**.

Appendix Table C2: Total pigment quantities of EMS mutants and CC-125 parental strain

Strain	Chl *a* / cell (pg)	Chl *b* / cell (pg)	Car (pg)	Chl *a* / *b*	Chl / Car
CC-125 (WT)	1.10 ± 0.25	0.53 ± 0.12	0.30 ± 0.06	2.09 ± 0.10*	5.34 ± 0.25
EMS-Mut-1	1.13 ± 0.19	0.63 ± 0.07	0.47 ± 0.13	1.80 ± 0.14	3.96 ± 1.26
EMS-Mut-2	1.86 ± 0.21*	0.91 ± 0.13*	0.55 ± 0.04**	2.04 ± 0.10	5.04 ± 0.33
EMS-Mut-3	0.77 ± 0.26	0.35 ± 0.13	0.24 ± 0.08	2.20 ± 0.11	4.64 ± 0.15*
EMS-Mut-4	1.27 ± 0.26	0.67 ± 0.17	0.46 ± 0.07*	1.89 ± 0.11	4.23 ± 0.35*
EMS-Mut-5	3.46 ± 0.71**	1.54 ± 0.29**	1.09 ± 0.19**	2.24 ± 0.14	4.58 ± 0.20*
EMS-Mut-6	2.70 ± 0.17**	1.27 ± 0.11**	0.81 ± 0.06**	2.14 ± 0.06	4.90 ± 0.03
EMS-Mut-7	2.05 ± 0.59	0.98 ± 0.30	0.62 ± 0.17*	2.10 ± 0.04	4.90 ± 0.07*
EMS-Mut-8	0.87 ± 0.21	0.48 ± 0.12	0.35 ± 0.07	1.81 ± 0.04*	3.91 ± 0.18**
EMS-Mut-9	0.82 ± 0.07	0.41 ± 0.04	0.28 ± 0.03	2.02 ± 0.01	4.36 ± 0.09*

Pigments measured by spectrophotometer at λ 470, 644 and 662 nm, and calculated using equations in *Section 2.6.2*. Figures shown to two decimal places. Values ± SD. Significance calculated using one-way ANOVA (Tukey's test; $P < 0.05$).

Appendix Table C3: Average retention time for detected peaks

	Retention time (T_R)					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	β-car
CC-125	6.18 ± 0.00577	6.91 ± 0.00577	9.61 ± 0.0116	14.56 ± 0.0173	15.91 ± 0.0231	20.01 ± 0.0116
EMS-Mut-1	6.19 ± 0.0100	6.92 ± 0.0100	9.61 ± 0.0100	14.56 ± 0.0058	15.91 ± 0.0058	20.02 ± 0.0058
EMS-Mut-2	6.20 ± 0.0100	6.93 ± 0.0100	9.63 ± 0.00577	14.58 ± 0.00	15.93 ± 0.0058	20.03 ± 0.0058
EMS-Mut-3	6.22 ± 0.00577	6.95 ± 0.00577	9.64 ± 0.0100	14.60 ± 0.0100	15.95 ± 0.0100	20.05 ± 0.0116
EMS-Mut-4	6.23 ± 0.0116	6.96 ± 0.0116	9.65 ± 0.00577	14.61 ± 0.00	15.96 ± 0.0058	20.05 ± 0.0116
EMS-Mut-5	6.27 ± 0.0416	7.00 ± 0.0473	9.74 ± 0.145	14.64 ± 0.0666	15.98 ± 0.0436	20.05 ± 0.0231
EMS-Mut-6	6.24 ± 0.0100	6.98 ± 0.00577	9.68 ± 0.00577	14.62 ± 0.0100	15.96 ± 0.0100	20.05 ± 0.0116
EMS-Mut-7	6.27 ± 0.0153	7.00 ± 0.0153	9.70 ± 0.0208	14.64 ± 0.0153	15.99 ± 0.0116	20.07 ± 0.0173
EMS-Mut-8	6.29 ± 0.00577	7.02 ± 0.00577	9.73 ± 0.00	14.67 ± 0.0058	16.02 ± 0.00	20.12 ± 0.0058
EMS-Mut-9	6.30 ± 0.0153	7.04 ± 0.0153	9.74 ± 0.0153	14.69 ± 0.0100	16.04 ± 0.0100	20.14 ± 0.0116

Average retention time (T_R) ± SD. To 2 decimal places for mean, 3 significant figures for SD. Retention time slippage is apparent across the samples. Samples were injected in triplicate spaced by blanks in the order portrayed in this table. **Appendix Table C4** shows the difference in retention times between each peak for each sample run; the differences are similar across all samples (coefficient of variation < 2). Slippage between the first (CC-125) and last (EMS-Mut-9) samples for each peak was ~0.13 min^{-1}. This suggests that each of the peaks represent the same pigment in each sample. Peak slippage over time is a common phenomenon that

can occur due to sample mixing, aging of the column, evaporation of solvents causing changes in mobile phase composition, and system leakage.

Appendix Table C4: Differences in retention times between peaks detected via HPLC

	Viola - Neo			Lut - Viola			Chl-b - Lut			Chl-a - Chl-b			B-Car - Chl-a		
	BioRep-1	BioRep-2	BioRep-3	BioRep-1	BioRep-2	BioRep-3	BioRep-1	BioRep-2	BioRep-3	BioRep-1	BioRep-2	BioRep-3	BioRep-1	BioRep-2	BioRep-3
WT	0,73	0,73	0,73	2,70	2,69	2,69	4,96	4,95	4,95	1,36	1,35	1,35	4,09	4,10	4,10
Mut1	0,73	0,73	0,73	2,69	2,69	2,69	4,96	4,95	4,95	1,35	1,35	1,35	4,09	4,11	4,10
Mut2	0,73	0,73	0,73	2,69	2,70	2,70	4,95	4,95	4,96	1,36	1,35	1,35	4,09	4,09	4,10
Mut3	0,73	0,73	0,73	2,69	2,69	2,70	4,96	4,96	4,96	1,35	1,35	1,35	4,10	4,10	4,08
Mut4	0,73	0,73	0,73	2,71	2,70	2,70	4,96	4,96	4,95	1,35	1,35	1,36	4,09	4,08	4,10
Mut5	0,74	0,72	0,72	2,86	2,69	2,69	4,81	4,93	4,96	1,31	1,35	1,34	4,04	4,09	4,08
Mut6	0,73	0,73	0,73	2,70	2,70	2,70	4,94	4,94	4,95	1,35	1,34	1,34	4,10	4,09	4,08
Mut7	0,73	0,73	0,73	2,70	2,69	2,70	4,95	4,95	4,92	1,35	1,35	1,34	4,08	4,09	4,07
Mut8	0,74	0,73	0,74	2,70	2,71	2,71	4,95	4,94	4,95	1,34	1,35	1,35	4,11	4,10	4,09
Mut9	0,74	0,74	0,74	2,70	2,70	2,71	4,96	4,95	4,94	1,35	1,35	1,35	4,10	4,09	4,09
Mean	0,731			2,703			4,945			1,348			4,091		
SD	0,004			0,030			0,027			0,008			0,013		
CoV	0,582			1,100			0,544			0,582			0,582		

Differences between peaks were calculated by subtracting the retention time of each peak from the retention time of the following peak. Viola = violaxanthin, neo = neoxanthin, lut = lutein, Chl-a = chlorophyll-a, Chl-b = chlorophyll-b, B-car = b-carotene. CoV = coefficient of variation. The differences between the peaks were similar for each sample despite the drift in retention time between samples.

Appendix Table C5: Average peak area for each pigment detected

A

	Peak Area					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	β-car
CC-125	4.99 ± 0.38	2.64 ± 0.15	4.39 ± 0.49	14.53 ± 0.61	2.11 ± 0.09	2.28 ± 0.47
EMS-Mut-1	3.03 ± 1.01	2.47 ± 0.83	3.31 ± 0.79	9.17 ± 2.15	1.25 ± 0.29	1.37 ± 0.51
EMS-Mut-2	4.67 ± 0.43	3.09 ± 0.22	4.96 ± 0.54	15.37 ± 1.09	2.11 ± 0.15	2.18 ± 0.35
EMS-Mut-3	3.68 ± 1.41	2.25 ± 0.73	3.77 ± 1.43	11.00 ± 4.37	1.66 ± 0.63	2.35 ± 1.33
EMS-Mut-4	6.21 ± 0.46	3.95 ± 0.30	5.39 ± 0.38	17.46 ± 1.02	2.46 ± 0.17	2.58 ± 0.19
EMS-Mut-5	3.39 ± 1.31	4.10 ± 1.01	8.83 ± 3.03	18.00 ± 5.14	2.81 ± 0.82	2.75 ± 1.24
EMS-Mut-6	5.92 ± 1.94	4.48 ± 0.99	7.50 ± 2.39	22.08 ± 5.79	3.11 ± 0.69	3.26 ± 0.95
EMS-Mut-7	4.03 ± 1.75	3.39 ± 0.78	5.33 ± 1.74	16.81 ± 5.19	2.45 ± 0.73	2.67 ± 0.53
EMS-Mut-8	2.38 ± 0.71	2.13 ± 0.63	3.41 ± 0.85	8.99 ± 2.21	1.26 ± 0.39	1.39 ± 0.50
EMS-Mut-9	6.59 ± 0.23	4.06 ± 0.11	6.03 ± 0.25	21.28 ± 0.34	3.15 ± 0.04	4.24 ± 0.72

B

	Peak Area/ 10^6 cells					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	β-car
CC-125	0.61 ± 0.10	0.32 ± 0.04	0.54 ± 0.09	1.78 ± 0.33	0.26 ± 0.05	0.28 ± 0.06
EMS-Mut-1	0.67 ± 0.14	0.54 ± 0.12*	0.73 ± 0.09	2.03 ± 0.25	0.28 ± 0.03	0.30 ± 0.07
EMS-Mut-2	1.12 ± 0.03**	0.75 ± 0.04***	1.19 ± 0.02***	3.71 ± 0.16***	0.51 ± 0.02**	0.52 ± 0.03**
EMS-Mut-3	0.46 ± 0.21	0.28 ± 0.10	0.47 ± 0.22	1.38 ± 0.66	0.21 ± 0.09	0.29 ± 0.17
EMS-Mut-4	1.15 ± 0.11**	0.73 ± 0.07**	1.00 ± 0.09**	3.24 ± 0.25**	0.46 ± 0.04**	0.48 ± 0.04**
EMS-Mut-5	1.09 ± 0.12**	1.37 ± 0.27**	2.87 ± 0.32***	5.95 ± 1.00**	0.93 ± 0.15**	0.87 ± 0.10***

EMS-Mut-6	1.32 ± 0.21*	1.07 ± 0.04***	1.665 ± 0.20**	5.10 ± 0.32**	0.74 ± 0.01***	0.76 ± 0.12**
EMS-Mut-7	0.87 ± 0.37	0.74 ± 0.20*	1.18 ± 0.44	3.67 ± 1.23	0.54 ± 0.18	0.59 ± 0.15*
EMS-Mut-8	0.47 ± 0.09	0.42 ± 0.09	0.68 ± 0.14	1.78 ± 0.35	0.25 ± 0.06	0.27 ± 0.08
EMS-Mut-9	0.55 ± 0.13	0.34 ± 0.09	0.51 ± 0.13	1.79 ± 0.43	0.26 ± 0.06	0.36 ± 0.13

C

	Peak Area/ g DW					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	β-car
CC-125	7.16 ± 0.44	3.79 ± 0.20	6.31 ± 0.64	20.89 ± 0.48	3.04 ± 0.07	3.27 ± 0.64
EMS-Mut-1	6.09 ± 0.78	4.96 ± 0.65	6.75 ± 0.28	18.68 ± 0.62	2.54 ± 0.10	2.75 ± 0.50
EMS-Mut-2	8.16 ± 0.29	5.42 ± 0.33	8.66 ± 0.25	26.90 ± 1.48	3.69 ± 0.18	3.79 ± 0.20
EMS-Mut-3	6.58 ± 2.32	4.02 ± 1.01	6.76 ± 2.35	19.68 ± 7.19	2.96 ± 1.01	4.13 ± 2.19
EMS-Mut-4	10.66 ± 0.47	6.78 ± 0.36	9.25 ± 0.42	29.98 ± 0.87	4.22 ± 0.15	4.43 ± 0.19
EMS-Mut-5	5.61 ± 0.88	6.91 ± 0.15	14.69 ± 1.70	30.20 ± 1.75	4.70 ± 0.30	4.50 ± 1.09
EMS-Mut-6	9.04 ± 1.64	6.93 ± 0.43	11.46 ± 1.96	33.98 ± 3.64	4.81 ± 0.30	5.00 ± 0.73
EMS-Mut-7	6.85 ± 1.58	5.92 ± 0.20	9.20 ± 1.17	29.03 ± 2.97	4.23 ± 0.39	4.68 ± 0.07
EMS-Mut-8	4.88 ± 1.30	4.38 ± 1.20	7.06 ± 1.87	18.55 ± 4.66	2.60 ± 0.79	2.87 ± 0.99
EMS-Mut-9	9.09 ± 0.67	5.63 ± 0.55	8.40 ± 1.13	29.58 ± 3.06	4.38 ± 0.41	5.94 ± 1.46

A = Average peak area obtained for each peak for each strain. **B** = average peak area per 10^6 cells ((Peak area / cell number) * 10^6). **C** = average peak area per g of dry biomass (peak area / DCW). All ± SD, to 2 decimal places

Appendix Table C6: Fold difference from CC-125 of each pigment detected per 10^6 cells

	Fold change from CC-125 control per pigment					
Strain	Neo	Viol	Lut	Chl-*b*	Chl-*a*	β-car
EMS-Mut-1	1.09	1.68	1.37	1.14	1.06	1.08
EMS-Mut-2	1.84	2.32	2.23	2.08	1.96	1.89
EMS-Mut-3	0.75	0.87	0.88	0.77	0.80	1.03
EMS-Mut-4	1.89	2.27	1.87	1.82	1.76	1.73
EMS-Mut-5	1.79	4.25	5.35	3.34	3.57	3.12
EMS-Mut-6	2.16	3.33	3.11	2.86	2.84	2.74
EMS-Mut-7	1.43	2.30	2.20	2.06	2.06	2.11
EMS-Mut-8	0.76	1.30	1.26	1.00	0.96	0.99
EMS-Mut-9	0.90	1.06	0.95	1.00	1.02	1.31

To 2 decimal places.

Appendix Table C7: Concentration of 2-D cleaned-up protein extractions for proteomics analysis

Sample	Biological Replicate	10x dilution (µg/ µL)	Concentration (µg/ µL)
WT	1	2.531	25.31
	2	2.919	29.19
	3	2.958	29.58
EMS-Mut-5	1	2.397	23.97
	2	2.471	24.71
	3	2.603	26.03

Samples diluted in Urea Buffer and measured using Nanodrop 2000 (Thermo Scientific).

Appendix Table C8: Summary of MaxQuant MS/ MS data

Sample	BR	TR	MS	MS/ MS	MS/ MS Submitted	MS/ MS Identified	MS/ MS ID%	Pep seq ID	Peaks	aAMD [ppm]	MSD [ppm]
WT	1	1	26047	61276	70066	12276	17.52	7239	2229638	0.22314	0.35648
		2	25857	61759	70681	12578	17.8	7615	2239905	0.2254	0.3583
	2	1	27686	55773	63317	9460	14.94	5877	2248865	0.202	0.30897
		2	27992	54956	62363	9667	15.5	5842	2302184	0.20202	0.30435
	3	1	26997	58079	65749	10387	15.8	6387	2249174	0.22202	0.36498
		2	27176	57416	64916	10790	16.62	6560	2326210	0.22338	0.36597
Mut-5	1	1	26504	60105	69725	11607	16.65	6663	2107681	0.21758	0.3447
		2	26337	60579	70290	12023	17.1	6918	2116100	0.22139	0.34004
	2	1	28040	54736	63344	10134	16	5727	2232550	0.20762	0.32399
		2	28076	54644	63064	9748	15.46	5499	2343984	0.20694	0.32267
	3	1	30564	46386	51668	6716	13	4160	2505121	0.19447	0.29418
		2	30796	45739	50986	7278	14.27	4346	2520724	0.2076	0.31925
Total			332072	671448	766169	122664	16.01	11626	27422136	0.21432	0.33773

BR = biological replicate; TR = technical replicate; Pep seq ID = peptide sequences identified; aAMD = average absolute mass deviation; MSD = mass standard deviation

Appendix Table C9: Full list of proteins with significantly higher EMS-Mut-5/WT Log_2 protein intensity ratios

Protein name	Description	Log_2 fold change	*P*-value
P93664	Light-harvesting complex stress-related protein 1, chloroplastic (Chlorophyll a-b binding protein LHCSR1)	9.329	0.00E+00
A8HPM5	Photosystem II protein PSBS2 (Protein PHOTOSYSTEM II SUBUNIT S2)	5.774	0.00E+00
A8IWW7	Uncharacterized protein	5.368	0.00E+00
A8HQC2	Uncharacterized protein	5.215	0.00E+00
A0A2K3DL78	NAD(P)-bd_dom domain-containing protein	5.177	0.00E+00
A8I686	Uncharacterized protein	4.765	1.50E-05
A8JDR9	Fasciclin-like protein	4.683	0.00E+00
A0A2K3D1W6	Uncharacterized protein	4.669	1.00E-06
A0A2K3CVL1	TMEM189_B_dmain domain-containing protein	4.669	0.00E+00

A0A2K3DMK2	DLH domain-containing protein	4.569	0.00E+00
A8IYH1	Uncharacterized protein	4.497	0.00E+00
A0A2K3DSJ9	PDZ domain-containing protein	4.486	0.00E+00
A0A2K3CTK5	Uncharacterized protein	4.484	3.00E-06
A0A2K3DCA6	Peptidylprolyl isomerase (EC 5.2.1.8)	4.435	0.00E+00
A0A2K3DQI7	Aldo_ket_red domain-containing protein	4.403	0.00E+00
A0A2K3D5Y3	ADK_lid domain-containing protein	4.383	4.20E-05
P0DO19	Light-harvesting complex stress-related protein 3.2, chloroplastic (Chlorophyll a-b binding protein LHCSR3.2)	4.319	0.00E+00
A0A2K3DUD0	Uncharacterized protein	4.275	0.00E+00
A8IVS3	Uncharacterized protein	4.208	2.60E-05
O48663	Chloroplast w6 desaturase (Omega-6-FAD, chloroplast isoform)	4.165	1.70E-05
A0A2K3D0R0	Uncharacterized protein	4.105	0.00E+00
A0A2K3DAQ7	Uncharacterized protein	4.038	0.00E+00
Q5W9T4	Early light-inducible protein (Lhc-like protein Lhl1)	3.884	4.70E-05
A8HPN4	Uncharacterized protein	3.839	1.85E-02
A0A2K3D3L4	Glutaredoxin domain-containing protein	3.792	0.00E+00
A8J5N6	Predicted protein	3.780	1.00E-06
A0A2K3D0U9	Amino_oxidase domain-containing protein	3.770	0.00E+00
A8J6G0	Thylakoid lumenal protein	3.743	0.00E+00
A0A2K3CN34	#N/A	3.741	0.00E+00
A0A2K3E0P9	AA_TRNA_LIGASE_II domain-containing protein	3.738	0.00E+00
A0A2K3CXI9	Peptide-methionine (R)-S-oxide reductase (EC 1.8.4.12)	3.711	1.16E-03
A0A2K3DG80	bPH_2 domain-containing protein	3.562	0.00E+00
A8J2S3	RNA binding protein	3.532	0.00E+00
A8IZ88	Uncharacterized protein	3.530	3.00E-06
A8HYP8	Uncharacterized protein	3.489	0.00E+00
A0A2K3E2M0	Uncharacterized protein	3.474	9.21E-03
A8IWN6	Uncharacterized protein	3.425	0.00E+00
A8IQW5	Uncharacterized protein	3.350	4.50E-05
A8IVH2	Uncharacterized protein	3.329	3.43E-04
A0A2K3E132	PDZ domain-containing protein	3.316	0.00E+00
A0A2K3CPW9	Uncharacterized protein	3.300	1.10E-05
A0A2K3DBZ5	Aldo_ket_red domain-containing protein	3.282	0.00E+00
A8J3M8	Chloroplast Mn superoxide dismutase (Superoxide dismutase [Mn])	3.277	0.00E+00
A0A2K3D5N5	AB hydrolase-1 domain-containing protein	3.273	1.70E-05
A0A2K3D9Z9	Uncharacterized protein	3.249	1.00E-06
Q8HUH1	Putative 30S ribosomal S2-like protein	3.207	1.90E-02
A8HR79	Uncharacterized protein	3.199	0.00E+00
A8IZY2	Uncharacterized protein	3.189	0.00E+00
A8JCK3	GTP binding protein TypA	3.169	3.90E-04
A0A2K3DIB0	Alpha-amylase (EC 3.2.1.1)	3.123	2.60E-05
A0A2K3DYR0	PAP_fibrillin domain-containing protein	3.089	2.00E-06
A8J230	Predicted protein	3.053	0.00E+00
A8JDK2	Predicted protein	3.014	0.00E+00

A8IAE5	Uncharacterized protein	2.988	0.00E+00
A0A2K3DMQ1	Uncharacterized protein	2.982	0.00E+00
A0A2K3CR99	SE domain-containing protein	2.948	6.20E-05
A8JF72	Saccharopine dehydrogenase-like protein	2.929	1.10E-05
A0A2K3DPB0	NAD(P)-bd_dom domain-containing protein	2.913	1.34E-04
A0A2K3DY03	DUF1995 domain-containing protein	2.902	1.70E-02
A0A2K3DLZ0	SufE domain-containing protein	2.896	1.00E-06
A8HRQ4	Uncharacterized protein	2.887	1.00E-06
A0A2K3DPM9	NAD(P)-bd_dom domain-containing protein	2.886	3.00E-06
A0A2K3DQ98	Uncharacterized protein	2.856	2.10E-05
A0A2K3DHV3	NAD(P)H-hydrate epimerase (EC 5.1.99.6) (NAD(P)HX epimerase)	2.855	2.04E-04
A0A2K3CUT8	Aldo_ket_red domain-containing protein	2.853	1.00E-05
V9P7H6	Peptide-methionine (R)-S-oxide reductase (EC 1.8.4.12) (Fragment)	2.830	1.75E-04
A0A2K3DN15	Uncharacterized protein	2.819	1.19E-03
A8J4Q7	Rieske [2Fe-2S] protein	2.812	0.00E+00
A0A2K3DNQ6	NAD(P)-bd_dom domain-containing protein	2.791	1.70E-05
A0A2K3DXW2	CBM20 domain-containing protein	2.782	1.83E-02
A8IZV5	Uncharacterized protein	2.757	3.50E-04
A0A2K3D072	Uncharacterized protein	2.738	1.00E-06
A0A2K3DWN1	Translation factor GUF1 homolog, chloroplastic (EC 3.6.5.n1) (Elongation factor 4 homolog) (EF-4) (GTPase GUF1 homolog) (Ribosomal back-translocase)	2.728	1.20E-05
A0A2K3DGL4	Uncharacterized protein	2.693	6.24E-04
A0A2K3D2C1	tRNA-synt_1g domain-containing protein	2.661	1.00E-05
A0A2K3CXL6	Polyketide_cyc domain-containing protein	2.650	1.50E-03
A0A2K3E0H7	Uncharacterized protein	2.642	0.00E+00
A8IAJ4	Uncharacterized protein	2.639	8.76E-04
A0A2K3CW82	Uncharacterized protein	2.627	1.73E-04
A0A2K3DH17	Alanine--tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase) (AlaRS)	2.609	3.20E-05
A8JDW2	Predicted protein	2.599	6.00E-06
A0A2K3DNJ3	NAD(P)-bd_dom domain-containing protein	2.591	1.00E-06
A0A2K3D0Z6	Uncharacterized protein	2.583	1.28E-04
A8HZ72	Uncharacterized protein	2.583	1.00E-06
Q84X79	CR008 protein	2.576	0.00E+00
A0A2K3DY62	Uncharacterized protein	2.576	0.00E+00
A0A2K3DNT0	NAD(P)-bd_dom domain-containing protein	2.570	5.50E-05
A0A2K3E5T3	Uncharacterized protein	2.561	0.00E+00
A0A2K3DPD0	NAD(P)-bd_dom domain-containing protein	2.558	3.38E-04
A8HPK7	Translation factor	2.549	2.20E-05
A8J306	Uncharacterized protein CPLD1	2.540	1.00E-06
D3K371	Cytochrome c synthesis 5 protein	2.534	2.74E-04
A8IIN4	Uncharacterized protein	2.520	9.70E-04
A0A2K3E1Q7	Peptidylprolyl isomerase (EC 5.2.1.8)	2.495	0.00E+00
A0A2K3DSZ0	Uncharacterized protein	2.486	2.45E-04
A0A2K3E7M3	Uncharacterized protein	2.483	7.20E-05

A8III5	Uncharacterized protein	2.477	0.00E+00
O24426	Actin-like protein (Actin-related protein)	2.470	4.38E-03
A8I0K9	Uncharacterized protein	2.470	6.91E-04
A8JGX5	Protein arginine N-methyltransferase	2.468	5.40E-03
A0A2K3D9Z5	Aldo_ket_red domain-containing protein	2.445	2.62E-02
A0A2K3D6U5	ADK_lid domain-containing protein	2.443	1.00E-05
A0A2K3DQG7	Uncharacterized protein	2.439	1.99E-02
Q2HZ23	Ferredoxin	2.423	8.30E-05
Q6WEE4	Cyanobacterial-type MPBQ/MSBQ methyltransferase (MPBQ/MSBQ transferase cyanobacterial type) (Predicted protein)	2.399	3.84E-03
A0A2K3CZ18	Uncharacterized protein	2.371	7.48E-04
A0A2K3CWU8	Uncharacterized protein	2.364	7.59E-04
A0A2K3CQM1	Ubiquitin carboxyl-terminal hydrolase (EC 3.4.19.12)	2.353	4.54E-04
A8J2X8	Predicted protein	2.352	0.00E+00
A0A2K3DL88	Uncharacterized protein	2.348	2.97E-04
A0A2K3CVH1	Uncharacterized protein	2.337	1.70E-05
A0A2K3D4W9	Flavodoxin-like domain-containing protein	2.327	1.30E-05
A0A2K3DK43	Methyl-accepting transducer domain-containing protein	2.326	5.65E-03
A0A2K3D023	Uncharacterized protein	2.325	1.30E-05
A0A2K3CUI1	Pyr_redox_2 domain-containing protein	2.318	2.21E-04
A0A2K3CQR4	Uncharacterized protein	2.317	5.30E-03
P26565	50S ribosomal protein L20, chloroplastic	2.306	7.88E-03
A8IVP7	Uncharacterized protein	2.285	2.75E-02
A0A2K3CYZ7	Obg-like ATPase 1	2.251	1.20E-05
A0A2K3DXF8	Uncharacterized protein	2.250	1.05E-02
TRXh2a	Snf1-like protein kinase (EC 2.7.1.-) (Sulfur stress regulator)	2.247	0.00E+00
A0A2K3DHB0	Uncharacterized protein	2.224	5.25E-04
A0A2K3DFA0	AA_TRNA_LIGASE_II domain-containing protein	2.220	4.94E-04
A8J635	Predicted protein	2.212	0.00E+00
A0A2K3E7E9	PAP_fibrillin domain-containing protein	2.193	4.24E-03
A0A2K3CRB5	Uncharacterized protein	2.187	5.00E-06
A0A2K3D3G2	PDZ domain-containing protein	2.187	1.70E-05
A0A2K3DSZ3	Katanin p60 ATPase-containing subunit A-like 2 (Katanin p60 subunit A-like 2) (EC 5.6.1.1) (p60 katanin-like 2)	2.164	0.00E+00
A8IL08	Uncharacterized protein	2.155	0.00E+00
A8I2V7	Uncharacterized protein	2.152	3.38E-04
A0A2K3DNI1	NAD(P)-bd_dom domain-containing protein	2.148	0.00E+00
A8IKN8	Uncharacterized protein	2.124	0.00E+00
A0A2K3DDU7	Protein kinase domain-containing protein	2.121	1.52E-03
A8JF62	SOUL heme-binding protein	2.116	9.00E-05
A8J6C7	Membrane AAA-metalloprotease (EC 3.4.24.-)	2.099	0.00E+00
A0A2K3DS48	Rhodanese domain-containing protein	2.093	1.32E-02
A0A2K3E4U8	Lipase_3 domain-containing protein	2.088	1.30E-05
A0A2K3DU31	S1-like domain-containing protein	2.081	9.10E-03
A0A2K3E5I4	Peptidylprolyl isomerase (EC 5.2.1.8)	2.058	8.96E-03

Q9ATG8	Ferrochelatase (EC 4.99.1.1)	2.050	8.52E-04
A1YSB4	Photosystem-II repair protein	2.040	3.00E-06
A2BCY1	Chloroplast nucleosome assembly protein-like (Nucleosome assembly protein) (Fragment)	2.026	0.00E+00
A0A2K3D954	PDZ domain-containing protein	2.026	1.56E-03
A0A2K3DID5	Alpha-amylase (EC 3.2.1.1)	2.022	7.08E-04
A0A2K3DQ47	Methyltransf_25 domain-containing protein	1.983	1.40E-05
A0A2K3E0Y4	Uncharacterized protein	1.962	3.33E-04
A0A2K3DYQ7	Peptide-methionine (R)-S-oxide reductase (EC 1.8.4.12)	1.959	1.16E-02
A0A2K3D0E7	Protein kinase domain-containing protein	1.956	1.84E-02
A8J3K3	Predicted protein	1.953	2.83E-04
A0A2K3E436	Lon N-terminal domain-containing protein	1.949	8.10E-05
A8I9A1	Uncharacterized protein	1.919	3.40E-05
A0A2K3DZR6	Uncharacterized protein	1.916	3.97E-03
A8JEQ7	Predicted protein	1.899	1.90E-05
A0A2K3D385	Uncharacterized protein	1.892	9.01E-04
A8I368	Uncharacterized protein	1.889	3.34E-03
A0A2K3D7P2	FAD-binding PCMH-type domain-containing protein	1.889	1.00E-06
A0A2K3E6T3	Uncharacterized protein	1.883	2.51E-02
A8JEP1	50S ribosomal protein L35	1.876	1.86E-03
A0A2K3DK39	APH domain-containing protein	1.818	1.43E-03
A0A2K3E5A4	Chalcone-flavonone isomerase family protein	1.811	2.27E-02
A8IKA6	Uncharacterized protein	1.810	5.28E-03
A0A2K3D5G7	Amino_oxidase domain-containing protein	1.809	1.00E-06
A0A2K3E592	CYTOSOL_AP domain-containing protein	1.806	2.00E-06
A8JAY5	Tyrosine--tRNA ligase (EC 6.1.1.1) (Tyrosyl-tRNA synthetase)	1.805	2.09E-02
A0A2K3D3Y6	Aldo_ket_red domain-containing protein	1.786	1.17E-03
A0A2K3E4L0	Uncharacterized protein	1.783	1.00E-06
A0A2K3CSH6	Flavodoxin-like domain-containing protein	1.775	4.42E-03
A8IW09	Uncharacterized protein	1.774	0.00E+00
A8HWS8	Uncharacterized protein	1.771	2.12E-04
A0A2K3E6S5	PDZ domain-containing protein	1.763	0.00E+00
A0A2K3CZS7	Peptidase_M3 domain-containing protein	1.759	6.50E-05
A8JF87	Predicted protein	1.748	1.50E-05
Q8HEB4	Cytochrome b-c1 complex subunit Rieske, mitochondrial (EC 7.1.1.8)	1.737	6.66E-04
A8IS75	Uncharacterized protein	1.720	4.00E-06
A0A2K3D0R8	Photolyase/cryptochrome alpha/beta domain-containing protein	1.716	3.58E-03
Q84XR9	Thioredoxin x	1.706	0.00E+00
A0A2K3E4Y3	C2 domain-containing protein	1.667	5.38E-04
A8JH52	4-alpha-glucanotransferase (EC 2.4.1.25) (Amylomaltase) (Disproportionating enzyme)	1.663	3.16E-02
A0A2K3DXT0	NAD(P)-bd_dom domain-containing protein	1.645	9.97E-04
A8IKE6	Uncharacterized protein	1.644	0.00E+00
A0A2K3DE58	AA_TRNA_LIGASE_II domain-containing protein	1.635	1.19E-02
A0A2K3CT48	Uncharacterized protein	1.624	1.23E-02

A0A2K3CSZ6	Amino_oxidase domain-containing protein	1.623	2.82E-04
A8J2N2	Chlorophyll a-b binding protein, chloroplastic	1.610	0.00E+00
A0A2K3DCL9	Uncharacterized protein	1.598	0.00E+00
A0A2K3DIW8	Glutaredoxin domain-containing protein	1.597	1.70E-02
A8I6B9	Uncharacterized protein	1.589	4.00E-06
A0A2K3E651	M16C_associated domain-containing protein	1.588	0.00E+00
A8IRT4	Uncharacterized protein	1.585	2.00E-06
A0A2K3CRF7	Uncharacterized protein	1.561	1.84E-03
A0A2K3DHT0	Uncharacterized protein	1.560	8.43E-04
A0A2K3D8U2	Uncharacterized protein	1.558	1.00E-06
A0A2K3CUL8	Uncharacterized protein	1.553	7.57E-04
A0A2K3DW05	SPOC domain-containing protein	1.538	6.18E-03
A0A2K3CPN0	Pyr_redox_2 domain-containing protein	1.521	4.00E-06
Q9FYU1	Fe-hydrogenase (EC 1.18.99.1) (Iron hydrogenase) (Iron-hydrogenase HydA1)	1.514	1.08E-02
Q75NZ1	Low-CO2 inducible protein (Low-CO2 inducible protein LCIC)	1.512	0.00E+00
O22448	Glutathione peroxidase	1.509	1.03E-04
A8INE5	Uncharacterized protein	1.500	3.95E-03
A0A2K3E4Y0	Uncharacterized protein	1.494	1.00E-06
A0A2K3DIN0	DUF953 domain-containing protein	1.420	6.55E-03
A8JDP6	Plastid ribosomal protein S13	1.402	1.34E-04
A0A2K3E4J9	Uncharacterized protein	1.401	3.09E-02
A0A2K3CXA5	Uncharacterized protein	1.397	4.36E-02
A8IRU6	Uncharacterized protein	1.388	3.00E-06
Q6J213	Chloroplast phytoene desaturase (Phytoene desaturase) (EC 1.3.-.-) (EC 1.3.99.-)	1.370	0.00E+00
A0A2K3DT83	Katanin p60 ATPase-containing subunit A-like 2 (Katanin p60 subunit A-like 2) (EC 5.6.1.1) (p60 katanin-like 2)	1.368	5.80E-04
A0A2K3DH99	Uncharacterized protein	1.356	7.30E-03
A8I775	Uncharacterized protein	1.337	1.00E-06
A8I5A0	Uncharacterized protein	1.336	0.00E+00
A8HPM8	Uncharacterized protein	1.323	5.40E-05
A8J664	Steroid dehydrogenase	1.320	2.95E-02
A8I7F2	Uncharacterized protein	1.320	2.40E-05
A0A2K3DYK2	AB hydrolase-1 domain-containing protein	1.318	1.40E-02
A8J463	3-dehydroquinate synthase (EC 4.2.3.10) (EC 4.2.3.4)	1.317	2.49E-02
Q0ZAI6	LciB (Low-CO2-inducible protein)	1.315	0.00E+00
A8JGZ8	YCII-related protein	1.291	6.20E-05
A8JAX9	Predicted protein	1.281	3.80E-02
A0A2K3DKY9	Uncharacterized protein	1.247	1.08E-02
A8IEF7	Uncharacterized protein	1.234	5.21E-03
A0A2K3DVC5	Uncharacterized protein	1.229	2.49E-03
A0A2K3D6G1	ADK_lid domain-containing protein	1.209	4.60E-02
A0A2K3D3X9	Protein translocase subunit SecA	1.205	6.00E-06
O20032	30S ribosomal protein S18, chloroplastic	1.200	1.04E-03
A0A2K3D070	Uncharacterized protein	1.198	1.39E-03

A0A2K3D9C0	Peptidase_S9 domain-containing protein	1.197	1.40E-05
A0A2K3DVM0	Uncharacterized protein	1.191	3.00E-03
Q39588	Carbonic anhydrase, alpha type (Intracellular carbonic anhydrase, alpha type)	1.191	5.16E-04
A0A2K3E888	Thioredoxin reductase (EC 1.8.1.9)	1.190	4.00E-05
A0A2K3E408	Uncharacterized protein	1.180	3.82E-02
A8IXU9	Uncharacterized protein	1.171	0.00E+00
A0A2K3CQW3	Uncharacterized protein	1.164	1.00E-06
A8HY43	Uncharacterized protein	1.150	0.00E+00
A8IKD4	Uncharacterized protein	1.136	3.81E-02
A0A2K3E434	Thioredoxin domain-containing protein	1.133	4.20E-02
A0A2K3D1N2	Protein kinase domain-containing protein	1.131	1.92E-03
A0A2K3CZN2	Uncharacterized protein	1.117	3.38E-04
A8JH60	Predicted protein	1.102	0.00E+00
A8IR98	Uncharacterized protein	1.093	1.16E-04
A0A2K3E1H9	Uncharacterized protein	1.093	8.08E-03
A0A2K3DRU9	Uncharacterized protein	1.080	2.20E-05
A0A2K3DW10	RRF domain-containing protein	1.078	1.33E-04
A8IY43	Uncharacterized protein	1.064	3.26E-03
Q84X70	CR084 protein (Predicted protein)	1.059	2.70E-05
A8J9D9	Histone domain-containing protein	1.057	3.79E-03
A0A2K3DSB1	Uncharacterized protein	1.048	5.07E-03
A0A2K3E7W0	TNase-like domain-containing protein	1.045	1.03E-03
A8I1D3	Uncharacterized protein	1.045	4.07E-04
A8J637	Elongation factor Ts, mitochondrial (EF-Ts) (EF-TsMt)	1.036	0.00E+00
P48267	30S ribosomal protein S7, chloroplastic	1.032	1.42E-03
A0A2K3E4N1	FAD-binding FR-type domain-containing protein	1.026	4.00E-06
A8HXM1	Uncharacterized protein	1.020	4.00E-03
A0A2K3DL28	Uncharacterized protein	1.016	2.00E-06
A8IJ60	Uncharacterized protein	1.015	1.44E-02
A8JDN8	Plastid ribosomal protein S16	1.015	1.85E-04
Q66YD0	Chloroplast vesicle-inducing protein in plastids 1 (Vesicle inducing protein in plastids 1)	1.014	2.20E-04
A8I647	Uncharacterized protein	1.011	8.00E-06
A8J311	Predicted protein	0.990	6.01E-04
Q84U22	Plastid ribosomal protein L4 (Ribosomal protein L4)	0.984	4.50E-05
A0A2K3DTR8	Starch synthase, chloroplastic/amyloplastic (EC 2.4.1.-)	0.976	2.87E-03
A0A2K3E242	Peptidylprolyl isomerase (EC 5.2.1.8)	0.975	1.69E-04
A8HNJ8	Plastid ribosomal protein L18	0.971	6.20E-04
A0A2K3D735	Uncharacterized protein	0.953	6.00E-06
A8JGM1	Rhodanese-like Ca-sensing receptor	0.952	0.00E+00
O20029	30S ribosomal protein S9, chloroplastic	0.950	4.82E-04
P93109	Carbonic anhydrase (EC 4.2.1.1) (Carbonate dehydratase)	0.945	1.24E-04
A8IYS5	Uncharacterized protein	0.943	9.54E-03
A8J503	Plastid ribosomal protein L6	0.939	0.00E+00

A0A2K3DNZ6	NAD(P)-bd_dom domain-containing protein	0.930	1.34E-03
A8J8M5	Plastid ribosomal protein S5	0.929	0.00E+00
P11094	50S ribosomal protein L14, chloroplastic	0.926	3.00E-06
A8IUC3	Uncharacterized protein	0.922	2.97E-03
Q8HTL2	50S ribosomal protein L2, chloroplastic	0.921	3.00E-06
A0A2K3CXF2	Ribosomal_L18e/L15P domain-containing protein	0.918	6.10E-05
O20030	Photosystem I assembly protein Ycf4	0.898	5.40E-03
P59775	30S ribosomal protein S8, chloroplastic	0.874	5.90E-05
Q8HTL1	50S ribosomal protein L5, chloroplastic	0.849	1.00E-06
A8JG56	L-ascorbate peroxidase	0.834	2.00E-06
A8ICE4	Uncharacterized protein	0.834	4.10E-05
A8IYJ5	Uncharacterized protein	0.832	1.32E-02
Q08365	30S ribosomal protein S3, chloroplastic (ORF 712)	0.828	3.00E-06
A8HVP7	Uncharacterized protein	0.827	2.00E-06
A0A2K3CXE4	Uncharacterized protein	0.825	9.00E-06
A8IA26	Uncharacterized protein	0.823	3.16E-02
A8IIP7	Uncharacterized protein	0.818	3.01E-02
A0A2K3CSD6	Uncharacterized protein	0.816	1.26E-02
A0A2K3D5T7	ADK_lid domain-containing protein	0.815	0.00E+00
A8I3M4	Uncharacterized protein	0.810	3.40E-05
P48270	30S ribosomal protein S4, chloroplastic	0.808	3.00E-06
P14149	30S ribosomal protein S12, chloroplastic	0.808	1.03E-03
A8JDN4	Plastid ribosomal protein S20	0.803	7.00E-06
A8IW44	Uncharacterized protein	0.800	2.00E-06
A8I8A3	Uncharacterized protein	0.798	2.20E-05
O81954	1-deoxy-D-xylulose-5-phosphate synthase	0.792	5.00E-05
A8J2E9	Chlorophyll a-b binding protein, chloroplastic	0.791	4.80E-05
A0A2K3D7Z5	Uncharacterized protein	0.786	2.30E-05
A8I7V2	Uncharacterized protein	0.783	6.45E-03
Q8HTL3	50S ribosomal protein L23, chloroplastic	0.783	7.00E-06
A8IWQ1	Uncharacterized protein	0.780	1.44E-02
A8IT01	Uncharacterized protein	0.777	0.00E+00
A8ICK6	Uncharacterized protein	0.771	4.89E-04
A0A2K3E5G0	Pept_C1 domain-containing protein	0.766	1.32E-03
A0A2K3D3L7	Glutamine amidotransferase type-2 domain-containing protein	0.764	1.20E-05
A8JE35	Plastid ribosomal protein L3	0.756	5.00E-06
A0A2K3E076	Elongation factor G, chloroplastic (cEF-G)	0.755	9.10E-05
A8I8Z4	Uncharacterized protein	0.754	6.00E-06
A0A2K3D985	AA_TRNA_LIGASE_II domain-containing protein	0.749	2.70E-05
A8HWZ8	Uncharacterized protein	0.745	3.30E-05
A0A2K3DMK3	DLH domain-containing protein	0.743	3.60E-04
Q8W4V3	Serine hydroxymethyltransferase (EC 2.1.2.1)	0.741	2.29E-04
A0A2K3DDD4	Uncharacterized protein	0.739	6.90E-05
Q70DX8	Plastid ribosomal protein S1 (Ribosomal protein S1 homologue)	0.728	1.75E-04

A8IVM9	Uncharacterized protein	0.727	3.40E-05
A8HYV3	Uncharacterized protein	0.717	9.30E-05
A8JCJ9	Predicted protein	0.711	1.55E-02
A0A2K3CQC6	Starch synthase, chloroplastic/amyloplastic (EC 2.4.1.-)	0.697	4.84E-04
A8J6Y3	Predicted protein	0.694	8.00E-06
A0A2K3D0E8	Uncharacterized protein	0.694	1.00E-05
A0A2K3CRX7	Uncharacterized protein	0.674	1.78E-02
A8J5Y7	Plastid ribosomal protein S6	0.662	4.00E-05
A0A2K3CY14	Uncharacterized protein	0.662	2.68E-02
O47027	30S ribosomal protein S2, chloroplastic	0.658	1.80E-05
A8I8A6	Uncharacterized protein	0.648	6.03E-03
A0A2K3E5V6	Uncharacterized protein	0.636	2.45E-04
A8J820	Predicted protein	0.634	6.50E-03
A8HWZ6	Uncharacterized protein	0.625	8.17E-04
A8I9I9	Uncharacterized protein	0.622	5.00E-06
A8J3L3	Peptidylprolyl isomerase (EC 5.2.1.8)	0.615	7.10E-03
A8HPS2	Uncharacterized protein	0.613	3.26E-03
A8JFR4	Ornithine transaminase (EC 2.6.1.13)	0.609	3.56E-02
Q05093	Chlorophyll a-b binding protein, chloroplastic	0.591	3.53E-03
Q945T2	GrpE protein homolog	0.584	3.25E-02
A8J7S1	Predicted protein	0.584	4.07E-03
P59776	30S ribosomal protein S19, chloroplastic	0.580	6.00E-06
A8JC71	Peptidylprolyl isomerase (EC 5.2.1.8)	0.567	4.69E-02
A0A2K3CSJ6	Uncharacterized protein	0.547	4.00E-03
A0A2K3DJF1	Nucleoside diphosphate kinase (EC 2.7.4.6)	0.520	1.20E-03
A8IGH1	Uncharacterized protein	0.518	1.16E-03
A0A2K3DTW7	Uncharacterized protein	0.516	8.08E-03
A0A2K3E2Q4	Uncharacterized protein	0.516	2.43E-02
A8IX41	Uncharacterized protein	0.514	2.47E-02
A0A2K3DN68	NAD(P)-bd_dom domain-containing protein	0.506	3.60E-05
A0A2K3DXG9	Uncharacterized protein	0.505	3.88E-03
Q75VY7	Chlorophyll a-b binding protein, chloroplastic	0.504	5.70E-05
Q6EMK7	38 kDa RNA-binding protein (Chloroplast-targeted RNA-binding protein) (RNA-binding protein RB38)	0.503	2.31E-03
A0A2K3DYJ4	AA_TRNA_LIGASE_II domain-containing protein	0.502	1.84E-02
A0A2K3DPC8	NAD(P)-bd_dom domain-containing protein	0.502	4.84E-02
A8I547	Uncharacterized protein	0.500	8.70E-03
A0A2K3E673	Peptidase_M3 domain-containing protein	0.498	3.03E-02
A8HMQ3	3,8-divinyl protochlorophyllide a 8-vinyl reductase	0.489	1.06E-03
Q75VY8	Chlorophyll a-b binding protein, chloroplastic	0.475	2.09E-04
A0A2K3CWZ6	SOR_SNZ domain-containing protein	0.468	2.51E-02
A0A2K3E0L8	Uncharacterized protein	0.466	2.00E-05
A8HPJ2	NADPH-protochlorophyllide oxidoreductase (EC 1.3.1.33)	0.463	2.64E-03
A8J9Y1	Cytochrome b6-f complex iron-sulfur subunit (EC 7.1.1.6)	0.462	1.75E-04
A0A2K3CTB3	Uncharacterized protein	0.461	1.42E-04

A0A2K3DII9	tRNA-synt_1c domain-containing protein	0.454	6.55E-03
Q75VY6	Chlorophyll a-b binding protein, chloroplastic	0.444	3.44E-02
A0A2K3CS36	DUF1336 domain-containing protein	0.444	3.16E-03
A8IFC8	Uncharacterized protein	0.423	5.02E-03
Q75VZ0	Chlorophyll a-b binding protein, chloroplastic	0.418	9.50E-05
A0A2K3DJF7	PAP_fibrillin domain-containing protein	0.410	2.28E-03
A8J3Z3	50S ribosomal protein L31	0.402	4.06E-02
A8IAW5	Uncharacterized protein	0.398	3.33E-04
P17746	Elongation factor Tu, chloroplastic (EF-Tu)	0.383	2.22E-02
A0A2K3CND7	#N/A	0.381	7.82E-04
Q84Y02	Chlorophyll a-b binding protein, chloroplastic	0.371	2.91E-02
A8HND3	Predicted protein	0.359	6.30E-04
A0A2K3CZB8	AMP-binding domain-containing protein	0.356	1.24E-02
A8JFY9_CHLRE	Serine glyoxylate aminotransferase (EC 2.6.1.45)	0.349	2.92E-02
A8IIK4	Uncharacterized protein	0.345	1.80E-03
A8ICV4	Uncharacterized protein	0.325	8.35E-03
A0A2K3DA85	Uncharacterized protein	0.316	1.93E-02
P48268	Cytochrome b559 subunit alpha (PSII reaction center subunit V)	0.298	2.57E-03
A8IL21	Uncharacterized protein	0.298	3.14E-02
A8I528	Uncharacterized protein	0.266	2.35E-02
A0A2K3D2I6	Lipoxygenase (EC 1.13.11.-)	0.264	6.55E-03
A8IKC8	Uncharacterized protein	0.259	3.82E-02
A8IW39	Uncharacterized protein	0.257	5.34E-03
A0A2K3DDN5	Uncharacterized protein	0.255	2.43E-02
A8JHC9	Citrate synthase	0.248	1.73E-02
A0A2K3DGW0	Uncharacterized protein	0.244	8.17E-03
A8IXV0	Uncharacterized protein	0.236	5.16E-03
A0A2K3DRY2	Acyl-coenzyme A oxidase	0.208	3.57E-02
A8J264	Chlorophyll a-b binding protein, chloroplastic	0.174	2.74E-02
A8J270	Chlorophyll a-b binding protein, chloroplastic	0.174	2.74E-02
P07891	ATP synthase epsilon chain, chloroplastic (ATP synthase F1 sector epsilon subunit) (F-ATPase epsilon subunit)	0.163	1.90E-02
A0A2K3DW88	Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.-)	0.162	4.81E-02
A8HUU9	Uncharacterized protein	0.156	1.24E-03
A0A2K3DAI3	HTH La-type RNA-binding domain-containing protein	0.098	3.35E-03
A0A2K3CZE0	Uncharacterized protein	0.084	1.88E-02
A8JGS4	Acyl-coa-binding protein	0.066	4.77E-02
A8JDL5	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	0.026	4.14E-02

Appendix Table C10: Full list of proteins with significantly lower EMS-Mut-5/WT Log_2 protein intensity ratios

Protein name	Description	Log_2	p-value
A0A2K3CVE9	Peptidylprolyl isomerase (EC 5.2.1.8)	-0.010	9.40E-03
A0A2K3DTW6	Uncharacterized protein	-0.015	1.54E-02
A8IIS8	Uncharacterized protein	-0.093	3.35E-02
A8JH68	Plastocyanin	-0.113	4.31E-02
P06541	ATP synthase subunit beta, chloroplastic (EC 7.1.2.2) (ATP synthase F1 sector subunit beta) (F-ATPase subunit beta)	-0.139	3.51E-02
A0A2K3DYQ5	Uncharacterized protein	-0.155	1.87E-02
A0A2K3CRS9	Uncharacterized protein	-0.183	4.77E-02
A8J7H8	Cytochrome c oxidase 11 kD subunit	-0.184	2.12E-02
A8ILN4	Uncharacterized protein	-0.187	2.25E-02
A8HNE8	Geranylgeranyl reductase (EC 1.3.1.-)	-0.195	1.60E-03
A8JFB1	Porphobilinogen deaminase	-0.196	4.27E-02
Q8GV23	Nucleic acid binding protein (Putative nucleic acid binding protein)	-0.196	8.70E-03
A8J4Z4	Mitochondrial F1F0 ATP synthase associated 45.5 kDa protein	-0.196	3.06E-02
A0A2K3D1P1	Malate dehydrogenase (EC 1.1.1.37)	-0.198	7.24E-03
A8I2V3	Uncharacterized protein	-0.200	2.30E-02
A0A2K3CPR8	Uncharacterized protein	-0.212	6.31E-03
A0A2K3DD77	DUF3707 domain-containing protein	-0.215	4.35E-02
A8J5F7_CHLRE	6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44)	-0.216	3.30E-03
A8HX89	Uncharacterized protein	-0.221	3.92E-02
A8J493	Ribosomal protein S15	-0.226	4.13E-03
A0A2K3E5S1	Uncharacterized protein	-0.227	3.82E-02
A8JDL8	Predicted protein	-0.231	2.82E-02
A0A2K3E3Q0	Malate dehydrogenase (EC 1.1.1.37)	-0.234	1.38E-03
A8IJY7	Uncharacterized protein	-0.235	1.29E-02
A8HNC0	Diaminopimelate decarboxylase (EC 4.1.1.20)	-0.237	3.00E-02
A8J0Q8	Thioredoxin-related protein CITRX	-0.240	1.47E-04
A8J6Y8	Ferredoxin--NADP reductase, chloroplastic (FNR) (EC 1.18.1.2)	-0.244	3.15E-03
A0A2K3DZ59	Uncharacterized protein	-0.249	2.23E-02
A0A2K3CVN0	Uncharacterized protein	-0.249	4.23E-04
A8JC04	Phosphoglycerate kinase (EC 2.7.2.3)	-0.251	4.77E-02
A0A2K3DEF1	Guanylate kinase-like domain-containing protein	-0.255	6.11E-03
A8IMZ5	Uncharacterized protein	-0.256	2.69E-02
A0A2K3D7U3	3-oxoacyl-[acyl-carrier-protein] synthase	-0.257	6.36E-03
Q8LK22	Cytochrome c oxidase subunit	-0.263	2.54E-02
A0A220IUF1	#N/A	-0.266	2.72E-02
A8HTX7	Uncharacterized protein	-0.270	1.33E-02
O48949	Protein disulfide-isomerase (EC 5.3.4.1)	-0.270	3.57E-02
A0A2K3E3Y9	Alanine--tRNA ligase (EC 6.1.1.7) (Alanyl-tRNA synthetase) (AlaRS)	-0.277	6.55E-03
A8I5N5	Uncharacterized protein	-0.284	8.90E-03
A8IN92	Uncharacterized protein	-0.294	3.72E-03
A0A2K3CZF3	Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose pyrophosphorylase)	-0.299	1.54E-02
A8JDV9	F1F0 ATP synthase gamma subunit	-0.304	1.11E-02

Q6SA05	Rubisco activase	-0.305	1.22E-03
A0A2K3DY10	FBPase domain-containing protein	-0.306	4.28E-03
A8IKI9	Uncharacterized protein	-0.306	5.97E-03
A0A2K3CVB0	Uncharacterized protein	-0.307	2.15E-03
A8J1C1	Ubiquitin-activating enzyme E1 (EC 6.3.2.19)	-0.309	1.77E-02
A8IUU3	Uncharacterized protein	-0.314	2.59E-02
A0A2K3D1B0	B5 domain-containing protein	-0.317	8.72E-03
A8IZZ4	Uncharacterized protein	-0.326	1.80E-03
Q5S7Y5	Chloroplast triosephosphate isomerase (Triose phosphate isomerase)	-0.328	1.87E-02
A8J5P7	Ubiquinol:cytochrome c oxidoreductase 50 kDa core 1 subunit	-0.329	6.41E-03
Q6Y682	38 kDa ribosome-associated protein (Chloroplast stem-loop-binding protein)	-0.334	4.34E-02
A8ISZ1	Uncharacterized protein	-0.338	8.55E-03
A0A2K3DD83	Uncharacterized protein	-0.341	5.42E-03
A8JCE9	Mitochondrial F1F0 ATP synthase associated 36.3 kDa protein	-0.348	4.00E-03
A8J9X1	Mitochondrial F1F0 ATP synthase, delta subunit (EC 3.6.3.14)	-0.349	2.45E-02
Q96550	ATP synthase subunit alpha	-0.358	1.68E-02
A8IFZ9	Uncharacterized protein	-0.359	4.37E-02
A8JH45	Dynein light chain	-0.362	3.06E-02
A8J3U9	Mitochondrial ATP synthase subunit 5, OSCP subunit (EC 3.6.3.14)	-0.362	1.02E-03
A0A2K3DGK5	Alpha-amylase (EC 3.2.1.1)	-0.364	5.44E-03
A0A2K3DU16	PPM-type phosphatase domain-containing protein	-0.364	8.65E-03
A0A2K3DGA4	Uncharacterized protein	-0.365	3.69E-03
Q84X74_CHLRE	CR057 protein (Mitochondrial phosphate carrier 1)	-0.365	3.09E-02
A8J3F8	Predicted protein	-0.365	4.64E-02
A8JH98	Enolase (EC 4.2.1.11)	-0.376	1.21E-03
A0A2K3DMK8	Uncharacterized protein	-0.377	1.10E-02
A8IAT4	Uncharacterized protein	-0.380	3.91E-03
A8J6R7	Predicted protein	-0.380	1.60E-02
A8J506	Argininosuccinate synthase (EC 6.3.4.5)	-0.382	1.39E-03
A8HQT1	Uncharacterized protein	-0.382	2.49E-02
A0A2K3DZ72	PfkB domain-containing protein	-0.388	3.10E-02
A8JE07	Ribosomal protein S15a	-0.389	1.01E-03
A0A2K3E6Y0	Glucosamine_iso domain-containing protein	-0.396	1.00E-02
A8JG73	Flagellar associated protein	-0.408	1.85E-04
A8IMN5	Uncharacterized protein	-0.412	1.53E-02
A8IKQ0	Uncharacterized protein	-0.412	2.18E-03
A2PZC2	UDP-Glucose:protein transglucosylase (UDP-glucose protein: protein trans glycosylase)	-0.414	3.11E-02
A0A2K3D1L8	CoA carboxyltransferase N-terminal domain-containing protein	-0.415	1.11E-02
Q6QAY3	Mitochondrial F1F0 ATP synthase associated 19.5 kDa protein (Mitochondrial NADH:ubiquinone oxidoreductase 19 kDa subunit) (EC 1.6.5.3) (EC 1.6.99.3)	-0.416	1.56E-03
A8JEU4	Heat shock protein 70A (EC 3.6.1.3)	-0.422	3.23E-03
A0A2K3CZL0	40S ribosomal protein S3a	-0.423	6.60E-05
A0A2K3DA10	Methylenetetrahydrofolate reductase (EC 1.5.1.20)	-0.425	8.21E-04

A0A2K3DS60	Branched-chain-amino-acid aminotransferase (EC 2.6.1.42)	-0.426	8.80E-05
A8J537	Catalase (EC 1.11.1.6)	-0.427	6.21E-03
A8JCP5	NAD(P) transhydrogenase	-0.427	1.33E-04
A0A2K3D8R9	FAD-binding FR-type domain-containing protein	-0.428	4.68E-04
A8HP06	Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial (EC 1.3.5.1)	-0.436	2.46E-03
A0A2K3E272	Pyruvate dehydrogenase E1 component subunit alpha (EC 1.2.4.1)	-0.439	1.02E-02
A8IVP1	Uncharacterized protein	-0.441	1.09E-02
Q8LPD9	Phototropin (EC 2.7.11.1) (Blue light receptor PHOT)	-0.442	1.02E-04
A8HP84	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)	-0.448	6.55E-03
A0A2K3DAF7	Uncharacterized protein	-0.448	1.54E-02
A0A2K3E8D8	Uncharacterized protein	-0.455	2.68E-03
A8HZZ1	Uncharacterized protein	-0.456	5.16E-04
A0A2K3DFI1	Uroporphyrinogen-III synthase (EC 4.2.1.75)	-0.457	1.01E-02
A0A2K3DGE7	Biotin carboxylase (EC 6.3.4.14) (EC 6.4.1.2)	-0.459	1.20E-04
A0A2K3CX84	Uncharacterized protein	-0.462	3.11E-02
A0A2K3DKV5	Uncharacterized protein	-0.467	1.32E-04
A0A2K3DCF4	Acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	-0.478	3.77E-02
A0A2K3DM34	AIG1-type G domain-containing protein	-0.484	3.06E-02
A8HVA3	Uncharacterized protein	-0.486	2.60E-05
A8HYN3	Uncharacterized protein	-0.487	9.54E-03
A8IZS7	Uncharacterized protein	-0.491	4.35E-02
A8IQU3	Uncharacterized protein	-0.492	4.94E-04
A8J6J6	Acetyl-CoA acyltransferase (EC 2.3.1.16)	-0.496	4.00E-03
O49822	Ascorbate peroxidase (EC 1.11.1.11)	-0.497	3.81E-03
A8J8P4	Ribosomal protein L34	-0.499	2.99E-04
A0A2K3DKU5	Uncharacterized protein	-0.501	1.75E-04
A8JGF8	Ribosomal protein S9, component of cytosolic 80S ribosome and 40S small subunit	-0.501	2.20E-05
A8IR81	Uncharacterized protein	-0.503	5.00E-06
A0A2K3DLZ7	Uncharacterized protein	-0.504	3.61E-02
A0A2K3DQS2	Uncharacterized protein	-0.506	1.44E-04
A0A2K3CXI7	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit (Oligosaccharyl transferase 48 kDa subunit)	-0.507	1.94E-02
A8J1G8	40S ribosomal protein S6	-0.507	1.15E-04
A0A2K3CSB8	Arginine biosynthesis bifunctional protein ArgJ, chloroplastic [Cleaved into: Arginine biosynthesis bifunctional protein ArgJ alpha chain; Arginine biosynthesis bifunctional protein ArgJ beta chain] [Includes: Glutamate N-acetyltransferase (GAT) (EC 2.3.1.35) (Ornithine acetyltransferase) (OATase) (Ornithine transacetylase); Amino-acid acetyltransferase (EC 2.3.1.1) (N-acetylglutamate synthase) (AGS)]	-0.508	1.04E-02
A8I495	Uncharacterized protein	-0.508	1.04E-02
A8I017	Uncharacterized protein	-0.510	2.82E-04
A8HVU5_CHLRE	Uncharacterized protein	-0.512	2.08E-02
A8IIP9	Uncharacterized protein	-0.512	2.97E-02

P00877	Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit) (EC 4.1.1.39)	-0.516	2.48E-04
A0A2K3DAR1	Aamy domain-containing protein	-0.517	1.42E-04
A0A2K3D598	Uncharacterized protein	-0.520	3.40E-05
A0A2K3DRN1	Coatomer subunit gamma	-0.524	9.09E-04
A8IUV7	Uncharacterized protein	-0.526	1.40E-05
A8HMC0	Calreticulin	-0.527	3.51E-03
A8IYP4	Uncharacterized protein	-0.533	6.33E-03
A8J355	Cystathionine gamma-synthase	-0.534	1.34E-04
Q9LLL6	Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose pyrophosphorylase)	-0.535	3.60E-04
A0A2K3DG74	Uncharacterized protein	-0.536	4.80E-02
A0A2K3DYL5	Pyruvate dehydrogenase E1 component subunit beta (EC 1.2.4.1)	-0.536	1.20E-03
A0A2K3DDZ0	Uncharacterized protein	-0.536	1.77E-02
A8JGJ6	Mg protoporphyrin IX S-adenosyl methionine O-methyl transferase (EC 2.1.1.11)	-0.538	2.25E-02
A8IMP6	Uncharacterized protein	-0.545	1.23E-04
A8JI94	Ribosomal protein L22	-0.546	4.00E-05
A8JGY8	Sarcaleumin-like protein	-0.548	1.07E-02
A0A2K3E5K9	Carboxypeptidase (EC 3.4.16.-)	-0.551	3.70E-02
A0A2K3D9Z4	40S ribosomal protein SA	-0.551	5.28E-03
A0A2K3D7C4	Ribosomal_L28e domain-containing protein	-0.553	5.00E-06
A0A2K3E3Z0	Isocitrate dehydrogenase [NAD] subunit, mitochondrial	-0.555	2.96E-04
A8IQE3	Uncharacterized protein	-0.557	1.10E-05
Q6UKY5	Acyl carrier protein	-0.557	7.70E-05
A8J5Z0	60S acidic ribosomal protein P0	-0.558	2.20E-05
A0A2K3E4N7	Uncharacterized protein	-0.561	1.40E-05
A8HP90	Ribosomal protein L6	-0.566	5.16E-04
A0A2K3E794	S5 DRBM domain-containing protein	-0.566	2.23E-03
I2FKQ9	Mitochondrial chaperonin 60	-0.572	7.60E-05
A0A2K3DNX1	NAD(P)-bd_dom domain-containing protein	-0.573	1.51E-02
A0A2K3D2H8	Uncharacterized protein	-0.574	2.82E-04
A8HX04	Uncharacterized protein	-0.576	3.56E-04
A8JE98	Heterogeneous nuclear ribonucleoprotein	-0.577	1.09E-04
A0A2K3DQA6	Ribosomal_S7 domain-containing protein	-0.578	6.20E-05
A0A2K3D4U8	eIF2B_5 domain-containing protein	-0.582	4.00E-03
A0A2K3DD98	RanBD1 domain-containing protein	-0.583	1.13E-02
A0A2K3DQA0	Uncharacterized protein	-0.583	6.41E-03
A8IQC1	Uncharacterized protein	-0.584	4.60E-05
A8J173	Aspartate semialdehyde dehydrogenase	-0.585	1.75E-04
A0A2K3DZI3	Uncharacterized protein	-0.587	1.92E-02
A0A2K3E7A5	Polyadenylate-binding protein (PABP)	-0.591	7.12E-04
A0A2K3DQI2	Uncharacterized protein	-0.592	2.40E-03
D5LAW4	522875p	-0.593	2.99E-03
A8HVQ1	Uncharacterized protein	-0.594	1.10E-05
A8IWB5	Uncharacterized protein	-0.596	3.77E-02

A8IXE0	Uncharacterized protein	-0.596	9.00E-06
A8HNX3	Ribosomal protein L35	-0.596	4.10E-05
P00873	Ribulose bisphosphate carboxylase small chain 1, chloroplastic (RuBisCO small subunit 1) (EC 4.1.1.39)	-0.598	9.83E-04
A8J0I0	Ribosomal protein L4	-0.601	1.44E-04
A0A2K3DMS1	Ribosome assembly factor mrt4	-0.602	2.99E-02
A8JGI9	40S ribosomal protein S7	-0.604	1.00E-06
A8JAV1	Actin	-0.605	4.00E-05
A8I4T2	Uncharacterized protein	-0.606	4.70E-05
A0A2K3DTG4	Coatomer subunit beta (Beta-coat protein)	-0.608	4.04E-03
A8HX38	Uncharacterized protein	-0.610	3.00E-06
A9XPA7	Intraflagellar transport protein 144	-0.610	1.56E-03
A0A2K3CQ27	Nuclear pore protein	-0.611	1.03E-03
A0A2K3D4X3	Uncharacterized protein	-0.611	1.50E-05
A8J5B8	Predicted protein	-0.612	4.00E-06
A8IH03	Uncharacterized protein	-0.612	3.00E-06
A8J646	Acetyl-CoA carboxylase	-0.613	5.26E-03
A8IVZ9	Uncharacterized protein	-0.614	2.10E-03
A0A2K3DCJ3	Uncharacterized protein	-0.614	2.10E-05
A8JHX9	Elongation factor 2 (EC 3.6.5.3)	-0.617	3.98E-04
A8ICT4	Uncharacterized protein	-0.620	1.15E-02
A0A2K3E3Z9	Glycerol-3-phosphate acyltransferase, chloroplastic (GPAT) (EC 2.3.1.15)	-0.620	3.43E-02
P93106	Malate dehydrogenase (NAD-dependent malate dehydrogenase) (EC 1.1.1.37)	-0.622	3.38E-04
A8I2T0	Uncharacterized protein	-0.623	1.80E-05
A8HVP2	Uncharacterized protein	-0.623	5.00E-06
A8HN02	Triose phosphate translocator	-0.624	2.20E-05
A8J6A7	Adenylylphosphosulfate reductase	-0.628	6.00E-06
A8ID84	Uncharacterized protein	-0.630	5.20E-05
A8IYC7	Uncharacterized protein	-0.631	2.91E-04
A0A2K3D9Q8	Uncharacterized protein	-0.637	1.84E-02
A8J1B6	Adenine nucleotide translocator	-0.638	7.04E-03
Q6UP31	NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	-0.638	3.03E-02
A8HSU7	Uncharacterized protein	-0.640	1.51E-03
A8I0I1	Uncharacterized protein	-0.643	3.30E-05
A0A2K3DWR0	Uncharacterized protein	-0.644	2.00E-02
A8HY08	Uncharacterized protein	-0.644	1.13E-03
A8HN50	60S ribosomal protein L18a	-0.644	2.20E-05
A0A2K3E2C0	Uncharacterized protein	-0.645	4.47E-03
A8HQ81	Uncharacterized protein	-0.646	4.00E-05
A8HRZ9	Uncharacterized protein	-0.647	7.95E-04
A0A2K3E6L7	Uncharacterized protein	-0.648	9.10E-05
A0A2K3CZ51	Eukaryotic translation initiation factor 3 subunit A (eIF3a) (Eukaryotic translation initiation factor 3 subunit 10)	-0.649	1.29E-02
A8I0H0	Uncharacterized protein	-0.650	1.80E-03
A8J9W0	Ribosomal protein L23	-0.650	6.10E-05

A8IZK3	Uncharacterized protein	-0.651	1.00E-05
A0A2K3E2C7	Uncharacterized protein	-0.654	1.80E-05
A8JCY4	Fructose-bisphosphate aldolase (EC 4.1.2.13)	-0.655	1.82E-02
A8JAW4	Predicted protein	-0.656	1.40E-05
A8J6F2	Mitochondrial F1F0 ATP synthase associated 14.3 kDa protein	-0.656	3.86E-04
A0A2K3CZA0	Adenylosuccinate lyase (ASL) (EC 4.3.2.2) (Adenylosuccinase)	-0.657	2.57E-02
A0A2K3D9L7	Uncharacterized protein	-0.658	1.10E-05
A0A2K3DNH7	NAD(P)-bd_dom domain-containing protein	-0.659	2.41E-02
A8JF47	Mitochondrial carrier protein	-0.660	6.71E-04
A0A2K3DSI4	Uncharacterized protein	-0.661	7.24E-03
A8JIN6	Histone H2B	-0.661	3.88E-04
A0A2K3CV04	40S ribosomal protein S26	-0.666	2.30E-05
A8IAA3	Uncharacterized protein	-0.668	4.42E-03
A0A2K3D577	GST C-terminal domain-containing protein	-0.676	3.69E-02
A0A2K3D4Q4	Ribosomal_L7Ae domain-containing protein	-0.677	8.00E-06
A8J4Q3	Ribosomal protein S10	-0.677	1.05E-03
A0A2K3E7P1	Uncharacterized protein	-0.678	1.27E-03
A8JCT1	Histone H2B	-0.678	5.25E-04
A8J768	Ribosomal protein S14	-0.683	1.60E-05
A8IRX5	Uncharacterized protein	-0.684	1.23E-04
A8IJR6	Uncharacterized protein	-0.691	4.01E-04
A0A2K3CTK0	Ribosomal_S10 domain-containing protein	-0.691	1.40E-05
A8HVK4	Uncharacterized protein	-0.691	1.30E-04
A8IF08	Uncharacterized protein	-0.692	1.10E-05
A8HP55	Ribosomal protein L5	-0.695	2.70E-05
Q8HTL6	DNA-directed RNA polymerase subunit beta N-terminal section (EC 2.7.7.6) (PEP) (Plastid-encoded RNA polymerase subunit beta N-terminal section) (RNA polymerase subunit beta N-terminal section)	-0.698	8.21E-04
O22547	Acetolactate synthase (EC 2.2.1.6)	-0.700	8.84E-03
A8HN42	Small rab-related GTPase	-0.702	3.91E-02
A0A2K3DLS9	Uncharacterized protein	-0.703	2.98E-02
A0A2K3DRT1	Uncharacterized protein	-0.706	7.47E-04
Q8LKK4	Protofilament ribbon protein of flagellar microtubules (RIB72 protein) (p72)	-0.706	2.28E-02
A8IDP6	Uncharacterized protein	-0.709	1.23E-04
A8JHW7	Eukaryotic translation initiation factor 3 subunit K (eIF3k) (eIF-3 p25)	-0.722	1.52E-03
A0A2K3E0J7	Uncharacterized protein	-0.724	1.00E-06
A0A2K3DNY1	NAD(P)-bd_dom domain-containing protein	-0.725	4.00E-06
A8IWL4	Uncharacterized protein	-0.731	3.39E-03
A0A2K3E5E9	Terpene cyclase/mutase family member (EC 5.4.99.-)	-0.731	1.60E-05
A8IXG3	Uncharacterized protein	-0.735	3.40E-05
A0A2K3E513	Corrinoid adenosyltransferase (EC 2.5.1.17)	-0.735	3.33E-04
A8J0R4	Acidic ribosomal protein P2	-0.735	5.20E-05
A8IKZ2	Uncharacterized protein	-0.735	1.74E-03
A8IJ34	Uncharacterized protein	-0.736	1.00E-05
A0A2K3DS62	Uncharacterized protein	-0.737	8.40E-05

A8J8Y6	Predicted protein	-0.739	5.91E-03
A8I8Z1	Uncharacterized protein	-0.739	7.86E-03
A8J597	Ribosomal protein L12	-0.741	7.00E-06
A8IIL1	Uncharacterized protein	-0.746	1.00E-05
A8JHU2	60S ribosomal protein L36	-0.746	6.23E-03
Q763T6	UDP-sulfoquinovose synthase	-0.747	3.41E-04
A0A2K3DFI2	Uncharacterized protein	-0.749	5.34E-03
A0A2K3DLX7	S-adenosylmethionine synthase (EC 2.5.1.6)	-0.755	1.11E-04
A0A2K3E650	Ribosome biogenesis regulatory protein	-0.756	1.65E-04
A0A2K3DQY7	Uncharacterized protein	-0.756	2.40E-02
A0A2K3CWF7	NAC-A/B domain-containing protein	-0.757	1.20E-05
A8J8Y1	Receptor of activated protein kinase C 1	-0.758	7.10E-05
A8JDP4	Ribosomal protein L9	-0.758	1.30E-05
A8JD64	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	-0.760	2.00E-06
A8IYK1	Uncharacterized protein	-0.760	5.50E-05
A8JC40	Centrin	-0.762	4.61E-04
A0A2K3CYM5	Aldo_ket_red domain-containing protein	-0.762	2.40E-02
Q9S7V1	Coproporphyrinogen III oxidase	-0.763	9.35E-04
A8J146	Guanosine nucleotide diphosphate dissociation inhibitor	-0.764	2.35E-03
A0A2K3CWD5	PBP_domain domain-containing protein	-0.765	7.10E-03
A0A2K3CPI0	Glutathione synthetase (GSH-S) (EC 6.3.2.3)	-0.767	2.70E-02
A8ICT1	Uncharacterized protein	-0.767	2.00E-06
A8IXQ5	Uncharacterized protein	-0.768	1.00E-05
A0A2K3DG23	Uncharacterized protein	-0.770	8.74E-04
A8IQ05_CHLRE	Uncharacterized protein	-0.775	2.82E-04
A8J9M0	Histone domain-containing protein	-0.776	4.40E-05
A0A2K3E189	Ribosomal protein L15	-0.784	7.62E-04
A8J9S9	UDP-glucose 4-epimerase (EC 5.1.3.-)	-0.786	5.20E-05
A8I980	Uncharacterized protein	-0.786	1.30E-05
A8J914	UDP-glucose 6-dehydrogenase (EC 1.1.1.22)	-0.788	4.10E-05
A8JHC3	Ribosomal protein S11	-0.789	5.60E-05
A8I175	Uncharacterized protein	-0.791	7.85E-03
A8IZ36	Uncharacterized protein	-0.794	8.00E-06
A8JI07	Dual function alcohol dehydrogenase / acetaldehyde dehydrogenase	-0.797	1.00E-05
A8J2L0	Chlorophyll a-b binding protein, chloroplastic	-0.799	3.72E-03
A0A2K3DJU4	Uncharacterized protein	-0.801	4.53E-03
A0A2K3CNK7	Uncharacterized protein	-0.803	1.39E-03
A8I0Y2	Uncharacterized protein	-0.805	9.35E-04
A0A2K3CNQ0	Eukaryotic translation initiation factor 3 subunit L (eIF3l)	-0.807	2.22E-03
A0A2K3CQH8	Uncharacterized protein	-0.810	4.98E-03
A8JBQ5	Predicted protein	-0.814	1.83E-02
A0A2K3DB47	Uncharacterized protein	-0.815	1.07E-03
Q6DN05	Betaine lipid synthase (Diacylglyceryl-N,N,N-trimethylhomoserine synthesis protein)	-0.818	4.30E-02
A8J513	Nucleosome assembly protein	-0.827	2.97E-04

A8JFZ2	Predicted protein	-0.828	1.66E-02
A8IPQ9	Uncharacterized protein	-0.830	2.82E-04
A0A2K3D193	Uncharacterized protein	-0.837	1.41E-04
A8IT25	Uncharacterized protein	-0.837	1.07E-02
A0A2K3D8G0	Uncharacterized protein	-0.839	1.32E-04
A0A2K3DSL4	Glutamine amidotransferase type-2 domain-containing protein	-0.843	6.95E-03
A0A2K3D2K9	Uncharacterized protein	-0.845	1.50E-05
A0A2K3DSL2	Uncharacterized protein	-0.845	1.00E-06
A0A2K3E7I5	Aconitate hydratase, mitochondrial (Aconitase) (EC 4.2.1.-)	-0.847	1.70E-05
A8II42	Uncharacterized protein	-0.849	3.95E-03
A0A2K3DTD5	Eukaryotic translation initiation factor 3 subunit C (eIF3c) (Eukaryotic translation initiation factor 3 subunit 8) (eIF3 p110)	-0.851	2.20E-05
A8ITS8	Uncharacterized protein	-0.854	9.79E-04
A0A2K3DQH9	Serine hydroxymethyltransferase (EC 2.1.2.1)	-0.856	2.00E-06
A8IHB3	Uncharacterized protein	-0.856	4.84E-02
A0A2K3DD97	Uncharacterized protein	-0.857	8.85E-03
A8J576	40S ribosomal protein S27	-0.862	2.75E-02
A8J1X0	Uncoupling protein	-0.862	4.77E-02
D5LAU0	SPS1p	-0.862	3.84E-03
A8IV37	Uncharacterized protein	-0.863	9.12E-04
A8HS59	Uncharacterized protein	-0.864	3.36E-02
A0A2K3D9S0	Uncharacterized protein	-0.867	3.37E-04
A8J087	Vasa intronic gene	-0.867	9.60E-05
A0A2K3E6E9	T-complex protein 1 subunit eta (TCP-1-eta) (CCT-eta)	-0.870	3.10E-05
A0A2K3DX78	PUA domain-containing protein	-0.872	1.33E-02
A0A2K3D552	Ribosomal_L2_C domain-containing protein	-0.875	5.21E-04
A8J7C8	Lysine--tRNA ligase (EC 6.1.1.6) (Lysyl-tRNA synthetase)	-0.876	2.10E-05
Q540H1	Tubulin alpha chain	-0.878	1.50E-05
A8JIE5	Ribosomal protein S29	-0.885	1.10E-02
A0A2K3CS97	Uncharacterized protein	-0.894	1.16E-02
A8I531	Uncharacterized protein	-0.895	1.04E-03
Q9ZSM9	Snf1-like protein kinase (EC 2.7.1.-) (Sulfur stress regulator)	-0.901	4.88E-03
B1B601	Parkin-co-regulated gene product	-0.902	4.92E-02
A0A2K3DS55	Uncharacterized protein	-0.904	1.53E-02
A8IXZ0	Uncharacterized protein	-0.908	2.90E-05
A0A2K3E8I5	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	-0.910	5.10E-03
A8IA18	Uncharacterized protein	-0.922	1.21E-04
A8IAN1	Uncharacterized protein	-0.925	1.00E-06
Q84X71_CHLRE	ANK_REP_REGION domain-containing protein	-0.925	1.90E-02
A0A2K3DYI3	Uncharacterized protein	-0.926	6.60E-04
A8I5R9	Uncharacterized protein	-0.934	1.80E-02
A8JCC6	Acidic ribosomal protein P1	-0.943	3.96E-03
A8J091	Hybrid-cluster protein	-0.943	1.88E-02
A0A2K3D581	Transket_pyr domain-containing protein	-0.944	1.00E-06
A0A2K3D802	Uncharacterized protein	-0.946	7.65E-03

A8IX19	Uncharacterized protein	-0.949	2.73E-04
A0A2K3E625	Uncharacterized protein	-0.951	9.66E-03
A0A2K3CQ32	Protein kinase domain-containing protein	-0.952	9.68E-04
A0A2K3DVJ7	Citrate synthase (EC 2.3.3.16)	-0.958	2.00E-06
A8JAP7	Inosine-5'-monophosphate dehydrogenase (IMP dehydrogenase) (IMPD) (IMPDH) (EC 1.1.1.205)	-0.961	5.00E-06
A8JH37	Cobalamin-independent methionine synthase (EC 2.1.1.14)	-0.961	1.00E-06
A0A2K3DN75	NAD(P)-bd_dom domain-containing protein	-0.969	2.52E-03
A8JGS8	Coatomer subunit beta'	-0.977	5.28E-03
A8J1M5	Flagellar outer dynein arm heavy chain beta	-0.980	3.08E-02
A8IGY1	Uncharacterized protein	-0.984	0.00E+00
A8IXR5	Uncharacterized protein	-0.991	1.49E-04
A0A2K3CR77	Endo/exonuclease/phosphatase domain-containing protein	-0.993	4.39E-04
A0A2K3CR90	AIG1-type G domain-containing protein	-0.997	5.27E-03
A2PZC3	UDP-glucose 6-dehydrogenase (EC 1.1.1.22)	-0.999	4.00E-06
A0A2K3CR83	Uncharacterized protein	-1.001	2.09E-02
A0A2K3DT88	Katanin p60 ATPase-containing subunit A-like 2 (Katanin p60 subunit A-like 2) (EC 5.6.1.1) (p60 katanin-like 2)	-1.004	2.54E-04
A8JAG1	Predicted protein	-1.004	2.90E-04
A8J129	Aspartate aminotransferase (EC 2.6.1.1)	-1.007	1.16E-04
A0A2K3DY96	Uncharacterized protein	-1.007	5.26E-04
A0A140CTI1	#N/A	-1.009	6.84E-04
A0A2K3DF88	CTP synthase (EC 6.3.4.2) (UTP--ammonia ligase)	-1.010	2.13E-03
A8J0N1	Chloroplast lumenal protein	-1.010	3.66E-03
A0A2K3E652	Uncharacterized protein	-1.017	4.40E-02
A8IRV6	Uncharacterized protein	-1.020	2.60E-05
A8JFR9	Acetyl-coenzyme A synthetase (EC 6.2.1.1)	-1.021	1.40E-05
A0A2K3DTT8	Isocitrate dehydrogenase [NADP] (EC 1.1.1.42)	-1.021	0.00E+00
A0A2K3CSD7	Uncharacterized protein	-1.022	1.67E-03
A8I403	Uncharacterized protein	-1.023	2.04E-04
Q6X898	Malate synthase (EC 2.3.3.9)	-1.029	4.90E-05
A0A2K3CNY6	Uncharacterized protein	-1.031	4.70E-05
A0A2K3CVJ1	Methylthioribose-1-phosphate isomerase (M1Pi) (MTR-1-P isomerase) (EC 5.3.1.23) (S-methyl-5-thioribose-1-phosphate isomerase) (Translation initiation factor eIF-2B subunit alpha/beta/delta-like protein)	-1.034	3.14E-02
A8ISV8	Uncharacterized protein	-1.041	1.67E-03
A8JA22	Predicted protein	-1.041	4.66E-02
A0A2K3D6V9	ADK_lid domain-containing protein	-1.043	1.20E-05
A0A2K3D5B6	Uncharacterized protein	-1.049	3.34E-03
A8HRS8	Uncharacterized protein	-1.052	3.74E-02
A8J9T0	40S ribosomal protein S12	-1.060	6.00E-06
A8JFW5	Centriole proteome protein	-1.063	7.88E-03
A0A2K3D5Q3	ADK_lid domain-containing protein	-1.073	3.86E-04
A0A2K3E4D4	Uncharacterized protein	-1.074	1.74E-03
A8J239	Ribosomal protein L23a	-1.084	1.00E-06
A0A2K3E4Q3	Uncharacterized protein	-1.085	1.27E-03

A8JE83	Translocon-associated protein alpha subunit	-1.086	3.95E-03
A0A2K3DTA6	Katanin p60 ATPase-containing subunit A-like 2 (Katanin p60 subunit A-like 2) (EC 5.6.1.1) (p60 katanin-like 2)	-1.094	5.00E-03
A0A2K3D4M6	Uncharacterized protein	-1.096	1.80E-02
A0A2K3DCQ1	AAA domain-containing protein	-1.107	6.84E-04
A0A2K3D261	S1 motif domain-containing protein	-1.110	4.92E-02
A8JBX6	Nascent polypeptide-associated complex subunit beta	-1.113	4.23E-04
A8I6R4	Uncharacterized protein	-1.115	8.19E-04
Q84UB2	GMP synthetase	-1.120	1.60E-05
A8IGE2	Uncharacterized protein	-1.122	4.79E-03
A0A2K3D212	TPR_REGION domain-containing protein	-1.125	1.07E-02
A8HSH7	Uncharacterized protein	-1.126	1.78E-02
A8J7J2	T-complex protein, epsilon subunit	-1.132	5.74E-03
A0A2K3DMI3	Pyruvate carboxyltransferase domain-containing protein	-1.133	5.00E-06
A8IZS5	Uncharacterized protein	-1.145	4.00E-05
A8IVR6	Uncharacterized protein	-1.153	7.89E-03
A0A2K3CSB1	MlaD domain-containing protein	-1.155	2.08E-03
A8JGU7	Hexokinase	-1.164	3.17E-03
A0A2K3D2U0	Uncharacterized protein	-1.169	1.73E-03
A0A2K3D4W3	Uncharacterized protein	-1.176	5.08E-03
A0A2K3DKM3	Phosphodiesterase (EC 3.1.4.-)	-1.179	6.29E-04
A0A2K3D149	Uncharacterized protein	-1.185	1.11E-03
A8IXZ2	Uncharacterized protein	-1.185	7.58E-04
A8IJZ3	Uncharacterized protein	-1.190	3.63E-04
A0A2K3DV76	AAA domain-containing protein	-1.190	3.00E-06
A0A2K3E798	Uncharacterized protein	-1.193	1.23E-03
A0A2K3E6U0	Apple domain-containing protein	-1.197	1.70E-05
A0A2K3DBP3	ACT domain-containing protein	-1.202	3.53E-02
A8J1X8	Aspartyl-tRNA synthetase (EC 6.1.1.12)	-1.207	2.07E-03
A8HZ87	Uncharacterized protein	-1.210	1.26E-02
A0A2K3E0C1	Uncharacterized protein	-1.211	0.00E+00
A8JCQ8	Acetyl CoA synthetase (EC 6.2.1.1)	-1.217	4.82E-04
A8J6Q7	Phospho-2-dehydro-3-deoxyheptonate aldolase (EC 2.5.1.54)	-1.219	2.00E-06
A0A2K3E0D8	Uncharacterized protein	-1.223	3.00E-06
A8JG07	Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4)	-1.224	9.30E-05
A0A2K3DWS8	Uncharacterized protein	-1.226	6.71E-03
A0A2K3D4R1	MI domain-containing protein	-1.237	2.14E-02
A0A2K3DQE1	Uncharacterized protein	-1.238	3.73E-04
A8ITJ3	Uncharacterized protein	-1.244	2.66E-02
A8J8V5	26S proteasome regulatory subunit	-1.249	6.04E-03
A8J4N7	Protein phosphatase 2C	-1.262	3.35E-02
Q6V9B0	NADH:ubiquinone oxidoreductase subunit 10 (EC 1.6.5.3)	-1.267	4.21E-02
A0A2K3E3K1	Uncharacterized protein	-1.274	1.07E-03
A8INH7	Uncharacterized protein	-1.282	2.71E-02
A0A2K3DCP5	Uncharacterized protein	-1.283	4.84E-02

Q5CB51	Gamma-tocopherol methyltransferase (EC 2.1.1.95)	-1.283	5.00E-06
A0A2K3CUK8	Uncharacterized protein	-1.287	2.00E-06
Q27YU5	Radial spoke protein 9	-1.289	1.23E-02
A0A2K3D977	Glycerol-3-phosphate dehydrogenase [NAD(+)] (EC 1.1.1.8)	-1.291	4.47E-02
A0A2K3DG12	Uncharacterized protein	-1.291	9.70E-03
A8HYA9	Uncharacterized protein	-1.295	5.00E-06
A0A2K3CST8	Uncharacterized protein	-1.299	1.29E-03
A0A2K3DU93	V-type proton ATPase subunit a	-1.308	4.47E-02
A8J363	Matrix metalloproteinase-like protein	-1.319	4.55E-03
A0A2K3CVP8	Uncharacterized protein	-1.322	2.40E-05
A8IR41	Uncharacterized protein	-1.332	6.00E-05
A8IP37	Uncharacterized protein	-1.334	1.32E-02
A0A2K3DPW5	Tr-type G domain-containing protein	-1.334	6.05E-04
A0A2K3E0G2	Uncharacterized protein	-1.336	3.32E-03
A0A2K3DVA7	Uncharacterized protein	-1.344	3.84E-02
Q6PSL4	Fe-hydrogenase assembly protein (Hydrogenase assembly factor)	-1.349	1.17E-03
A8IYG8	Uncharacterized protein	-1.367	5.18E-03
A0A2K3DFF7	Uncharacterized protein	-1.372	3.15E-02
A8J6H7	Predicted protein	-1.373	4.82E-04
A0A2K3CUQ9	Importin N-terminal domain-containing protein	-1.376	4.00E-03
A0A2K3CQ54	Peptidase_M11 domain-containing protein	-1.376	4.00E-06
A0A2K3DMS8	Uncharacterized protein	-1.378	3.00E-05
A8IA86	Uncharacterized protein	-1.381	1.00E-06
A0A2K3E702	Importin N-terminal domain-containing protein	-1.387	7.62E-04
A8JGK1	Ribosomal protein S17	-1.404	6.93E-04
A0A2K3CQL4	Uncharacterized protein	-1.407	2.54E-03
A0A2K3DMH6	Aspartate aminotransferase (EC 2.6.1.1)	-1.410	2.33E-04
A0A2K3DJ22	Uncharacterized protein	-1.413	9.50E-05
A8IKQ6	Uncharacterized protein	-1.415	3.38E-04
A0A2K3DSI2	Uncharacterized protein	-1.430	2.20E-05
A8I0I4	Uncharacterized protein	-1.431	1.87E-04
A0A2K3CWI8	Uncharacterized protein	-1.435	2.87E-02
A8J073	Mago nashi-like protein	-1.456	4.18E-02
A8IBU6	Uncharacterized protein	-1.459	4.70E-05
A8J0A0	Predicted protein	-1.465	3.81E-03
A0A2K3DD34	Uncharacterized protein	-1.473	1.73E-02
A8IE32	Uncharacterized protein	-1.475	6.30E-04
A0A2K3CVI3	Uncharacterized protein	-1.482	4.85E-02
A8JCR1	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B	-1.501	1.44E-02
A8I208	Uncharacterized protein	-1.527	1.60E-05
A0A2K3CR25	Uncharacterized protein	-1.534	7.79E-03
A8JGX0	Spermidine synthase	-1.556	4.12E-04
A0A2K3DCG7	Methyltranfer_dom domain-containing protein	-1.569	5.40E-03
A0A2K3DLP9	Bifunctional lysine-specific demethylase and histidyl-hydroxylase (EC 1.14.11.-)	-1.572	4.00E-06

A8ITZ0	Uncharacterized protein	-1.574	2.00E-05
A0A2K3CYQ5	Uncharacterized protein	-1.577	1.09E-04
Q6V505	NADH:ubiquinone oxidoreductase 17.8 kDa subunit (Putative NADH:ubiquinone oxidoreductase 17.8 kDa subunit)	-1.580	1.43E-04
A8JF80	Casein kinase 2 alpha chain, CK2A	-1.583	2.13E-03
A0A2K3CQS9	26S proteasome non-ATPase regulatory subunit 1 homolog	-1.593	4.46E-02
A8IY95	Uncharacterized protein	-1.603	4.28E-03
A0A2K3DBE2	Uncharacterized protein	-1.603	0.00E+00
A8IFK0	Uncharacterized protein	-1.607	0.00E+00
A8I263	Uncharacterized protein	-1.611	1.02E-02
A8IRT6	Uncharacterized protein	-1.613	1.07E-02
A8HUE0	Uncharacterized protein	-1.616	3.88E-04
A8IWT7	Uncharacterized protein	-1.620	5.44E-03
A8IJG5	Uncharacterized protein	-1.638	1.82E-04
A0A2K3DDM4	Calcium-transporting ATPase (EC 7.2.2.10)	-1.654	4.91E-02
A8HN52	Glutaredoxin, CGFS type	-1.662	2.40E-03
A8J814	Prefoldin-related KE2-like protein	-1.665	1.02E-03
Q39568	G-strand telomere binding protein 1 (Gbp1p)	-1.666	9.00E-06
A0A2K3E417	AA_TRNA_LIGASE_II domain-containing protein	-1.670	2.06E-02
A8IHF4	Uncharacterized protein	-1.671	3.00E-06
A8I4H4	Uncharacterized protein	-1.689	3.56E-04
A8J8L3	Cytochrome b5 protein	-1.693	2.00E-06
A0A2K3DAC3	Uncharacterized protein	-1.696	2.73E-02
A8HQ77	Uncharacterized protein	-1.705	7.48E-04
Q66YD3	Chloroplast DnaJ-like protein (Chloroplast DnaJ-like protein 1)	-1.715	2.60E-02
A8I3V3	Uncharacterized protein	-1.716	4.00E-03
A8IJQ4	Uncharacterized protein	-1.718	5.00E-06
A0A2K3DPW8	Peptidyl-prolyl cis-trans isomerase (PPIase) (EC 5.2.1.8)	-1.724	1.94E-02
Q6QAY4	Mitochondrial NADH:ubiquinone oxidoreductase 23 kDa subunit (EC 1.6.5.3) (EC 1.6.99.3) (NADH:ubiquinone oxidoreductase B14.7 subunit)	-1.727	6.33E-03
A0A2K3DJN8	CaMKII_AD domain-containing protein	-1.730	4.70E-02
A0A2K3CSZ9	Uncharacterized protein	-1.731	9.14E-04
A0A2K3CNF5	ARL2_Bind_BART domain-containing protein	-1.732	2.20E-04
A0A2K3E4Y8	DJ-1_PfpI domain-containing protein	-1.734	2.12E-04
A0A2K3DDD0	Protein kinase domain-containing protein	-1.736	8.30E-05
Q8GUQ9	60S ribosomal protein L38 (Ribosomal protein L38)	-1.764	3.14E-03
A0A2K3DWC4	Fe2OG dioxygenase domain-containing protein	-1.764	2.00E-06
A8IWJ8_CHLRE	Uncharacterized protein	-1.768	1.02E-02
A8HQ72	Uncharacterized protein	-1.773	4.50E-05
A0A2K3D6X8	ADK_lid domain-containing protein	-1.784	6.30E-05
A8IZV1	Uncharacterized protein	-1.792	6.26E-03
A0A2K3CTM9	Uncharacterized protein	-1.795	6.05E-04
A0A2K3CZF8	H/ACA ribonucleoprotein complex subunit	-1.811	8.00E-06
A8J2Q0	Predicted protein	-1.813	9.37E-03
A0A2K3D8J9	WPP domain-containing protein	-1.829	1.60E-05

A8I1E4	Uncharacterized protein	-1.844	2.18E-03
A0A2K3E2S0	Uncharacterized protein	-1.846	2.44E-04
A8HQ21	Uncharacterized protein	-1.859	1.00E-05
A0A2K3DPM5	NAD(P)-bd_dom domain-containing protein	-1.869	8.50E-03
A8J841_CHLRE	Hydroxymethylpyrimidine phosphate synthase	-1.913	3.50E-05
A0A2K3DWN5	Uncharacterized protein	-1.921	1.00E-06
A8HYQ6	Uncharacterized protein	-1.923	7.58E-04
A0A2K3CV63	LRRcap domain-containing protein	-1.929	4.49E-03
A8HUK0	Uncharacterized protein	-1.930	4.03E-04
A0A2K3CZ88	Adenylosuccinate lyase (ASL) (EC 4.3.2.2) (Adenylosuccinase)	-1.935	4.00E-06
A8JF66	40S ribosomal protein S30	-1.940	5.80E-04
A0A2K3D8I4	Uncharacterized protein	-1.940	1.17E-02
A0A2K3DH22	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2 (Ribophorin-2)	-1.948	4.14E-02
A8J7I9	Mitochondrial ubiquinol-cytochrome c oxidoreductase subunit 8 (EC 1.10.2.2) (Ubiquinol:cytochrome c oxidoreductase 9 kDa subunit)	-1.983	4.20E-02
A0A2K3E259	Uncharacterized protein	-1.990	0.00E+00
A7UCH9	Uncharacterized protein	-1.992	1.89E-03
Q9XGU3	Serine/threonine-protein phosphatase (EC 3.1.3.16)	-1.993	2.20E-05
A0A2K3DE56	Lipase_3 domain-containing protein	-1.994	1.30E-05
A0A2K3DKE9	Uncharacterized protein	-2.001	2.87E-03
A0A2K3DCX4	Uncharacterized protein	-2.008	9.37E-03
A0A2K3DE67	Serine/threonine-protein phosphatase (EC 3.1.3.16)	-2.029	0.00E+00
A0A2K3DTB8	Uncharacterized protein	-2.044	3.00E-06
A0A2K3DAI9	Nop domain-containing protein	-2.066	0.00E+00
A8IKD6	Uncharacterized protein	-2.068	1.17E-02
A8JHR9_CHLRE	Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.-)	-2.107	1.00E-06
Q5VLJ9	Aquaporin, glycerol transport activity (Putative aquaporin)	-2.150	4.30E-04
A0A2K3D9T6	Uncharacterized protein	-2.159	3.34E-03
A0A2K3DCH9	Uncharacterized protein	-2.188	6.50E-05
A8HXD3	Uncharacterized protein	-2.207	0.00E+00
A0A2K3D0X3	Kinesin motor domain-containing protein	-2.216	3.72E-02
A8JBB4	Glutathione S-transferase	-2.243	1.75E-03
A8J0N7_CHLRE	Phosphoenolpyruvate carboxykinase, splice variant (EC 4.1.1.49)	-2.252	4.00E-06
Q8VZZ5	Eukaryotic release factor 1	-2.255	0.00E+00
A8J906	Predicted protein	-2.258	1.84E-03
A0A2K3CX88	Sm domain-containing protein	-2.276	0.00E+00
A8IRB7	Uncharacterized protein	-2.295	1.27E-03
A8HRZ0	Uncharacterized protein	-2.303	0.00E+00
A0A2K3DT03	Katanin p60 ATPase-containing subunit A-like 2 (Katanin p60 subunit A-like 2) (EC 5.6.1.1) (p60 katanin-like 2)	-2.326	1.90E-05
A8J133	Soluble inorganic pyrophosphatase	-2.333	0.00E+00
A0A2K3DNK6	NAD(P)-bd_dom domain-containing protein	-2.336	1.00E-06
A0A2K3CY13	Uncharacterized protein	-2.392	0.00E+00
A8HXE1	Uncharacterized protein	-2.423	4.23E-04
A0A2K3DLB7	DUF667 domain-containing protein	-2.466	1.14E-03

A0A2K3D2I5	Protein kinase domain-containing protein	-2.558	0.00E+00
A8JHJ5	Predicted protein	-2.558	1.10E-05
A8JEG6	Predicted protein	-2.558	3.03E-04
A0A2K3DLU5	Ribonucloprotein	-2.585	4.51E-04
A8IQG4	Uncharacterized protein	-2.638	6.55E-03
A8IP53	Uncharacterized protein	-2.766	0.00E+00
A0A2K3DZM1	Uncharacterized protein	-2.790	1.00E-06
A8J3F0	High mobility group protein	-2.943	1.75E-04
A0A2K3D4E7	Nop domain-containing protein	-3.012	1.00E-06
A8J244	Isocitrate lyase (EC 4.1.3.1)	-3.244	0.00E+00
A0A2K3CP17	Uncharacterized protein	-3.258	2.88E-03
A8JGK5	Eukaryotic translation initiation factor 3 subunit E (eIF3e) (Eukaryotic translation initiation factor 3 subunit 6)	-3.289	4.00E-06
A0A2K3CP19	Uncharacterized protein	-3.455	0.00E+00

Appendix Table C11: Enriched biological process GO terms for proteins of increased abundance in EMS-Mut-5

GO Biological Process	GO ID	Proteins	Fold Enrichment	*P*-value
photosystem II repair	GO:0010206	6	40.69	1.12E-07
response to high light intensity	GO:0009644	6	35.61	1.93E-07
nonphotochemical quenching	GO:0010196	4	31.65	3.39E-05
regulation of photosynthesis, light reaction	GO:0042548	3	23.74	6.67E-04
photosystem II assembly	GO:0010207	5	23.74	9.21E-06
photosynthesis, light harvesting in photosystem I	GO:0009768	10	19.78	1.04E-09
cellular response to superoxide	GO:0071451	3	17.8	1.27E-03
protein-chromophore linkage	GO:0018298	10	15.83	6.04E-09
thylakoid membrane organization	GO:0010027	5	15.83	4.37E-05
tRNA aminoacylation for protein translation	GO:0006418	14	15.46	6.82E-12
dicarboxylic acid biosynthetic process	GO:0043650	4	14.61	3.43E-04
hydrogen peroxide catabolic process	GO:0042744	4	14.61	3.43E-04
terpenoid metabolic process	GO:0006721	5	10.79	2.02E-04
regulation of translational fidelity	GO:0006450	4	10	1.16E-03

GO terms with > 10-fold enrichment shown. Only most specific subclass of each cluster shown.

Appendix Table C12: Enriched biological process GO terms for proteins of decreased abundance in EMS-Mut-5

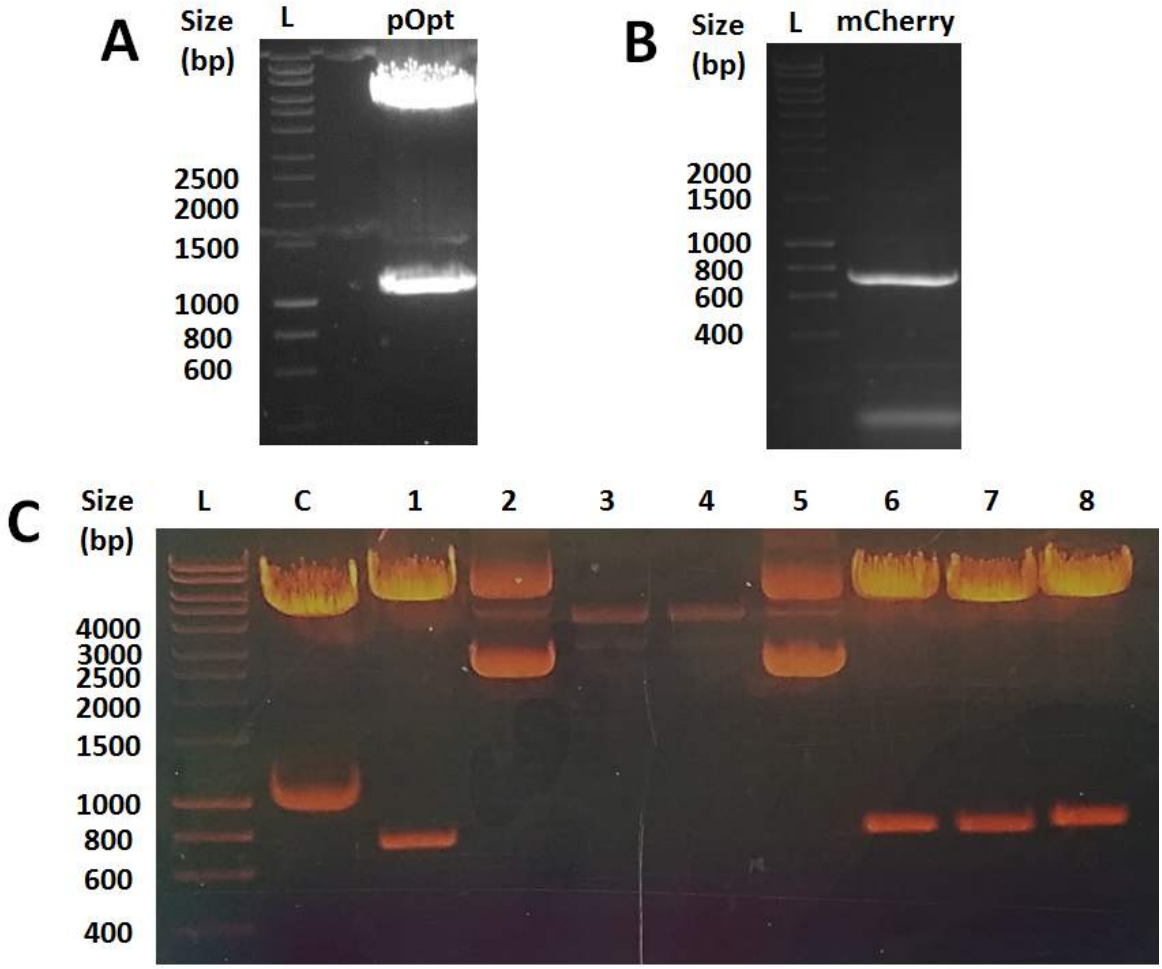
GO biological process complete	GO ID	Proteins	Fold Enrichment	*P*-value
valine biosynthetic process	GO:0009099	3	31.79	5.27E-04

acetyl-CoA biosynthetic process from acetate	GO:0019427	3	31.79	5.27E-04
S-adenosylmethionine metabolic process	GO:0046500	3	23.85	9.01E-04
formation of cytoplasmic translation initiation complex	GO:0001732	9	23.85	4.53E-09
gluconeogenesis	GO:0006094	7	22.26	3.52E-07
isoleucine biosynthetic process	GO:0009097	4	21.2	1.55E-04
fructose 1,6-bisphosphate metabolic process	GO:0030388	5	19.87	2.72E-05
positive regulation of RNA polymerase II transcriptional preinitiation complex assembly	GO:0045899	3	19.08	1.41E-03
tricarboxylic acid metabolic process	GO:0072350	3	19.08	1.41E-03
acetate transmembrane transport	GO:0035433	3	19.08	1.41E-03
tricarboxylic acid cycle	GO:0006099	14	18.55	2.52E-12
succinate transmembrane transport	GO:0071422	3	15.9	2.06E-03
galactose catabolic process via UDP-galactose	GO:0033499	3	15.9	2.06E-03
methionine biosynthetic process	GO:0009086	5	14.45	8.56E-05
arginine biosynthetic process	GO:0006526	4	14.13	4.91E-04
glycogen biosynthetic process	GO:0005978	3	13.63	2.88E-03
ribosomal small subunit assembly	GO:0000028	8	13.39	9.46E-07
glycolytic process	GO:0006096	9	13.01	2.37E-07
ribosomal subunit export from nucleus	GO:0000054	4	12.72	6.71E-04
glutamine metabolic process	GO:0006541	6	11.92	3.85E-05
one-carbon metabolic process	GO:0006730	4	11.56	8.93E-04
pentose-phosphate shunt	GO:0006098	4	11.56	8.93E-04
purine nucleoside bisphosphate biosynthetic process	GO:0034033	6	11.22	5.07E-05
ribosomal large subunit assembly	GO:0000027	9	10.6	9.64E-07

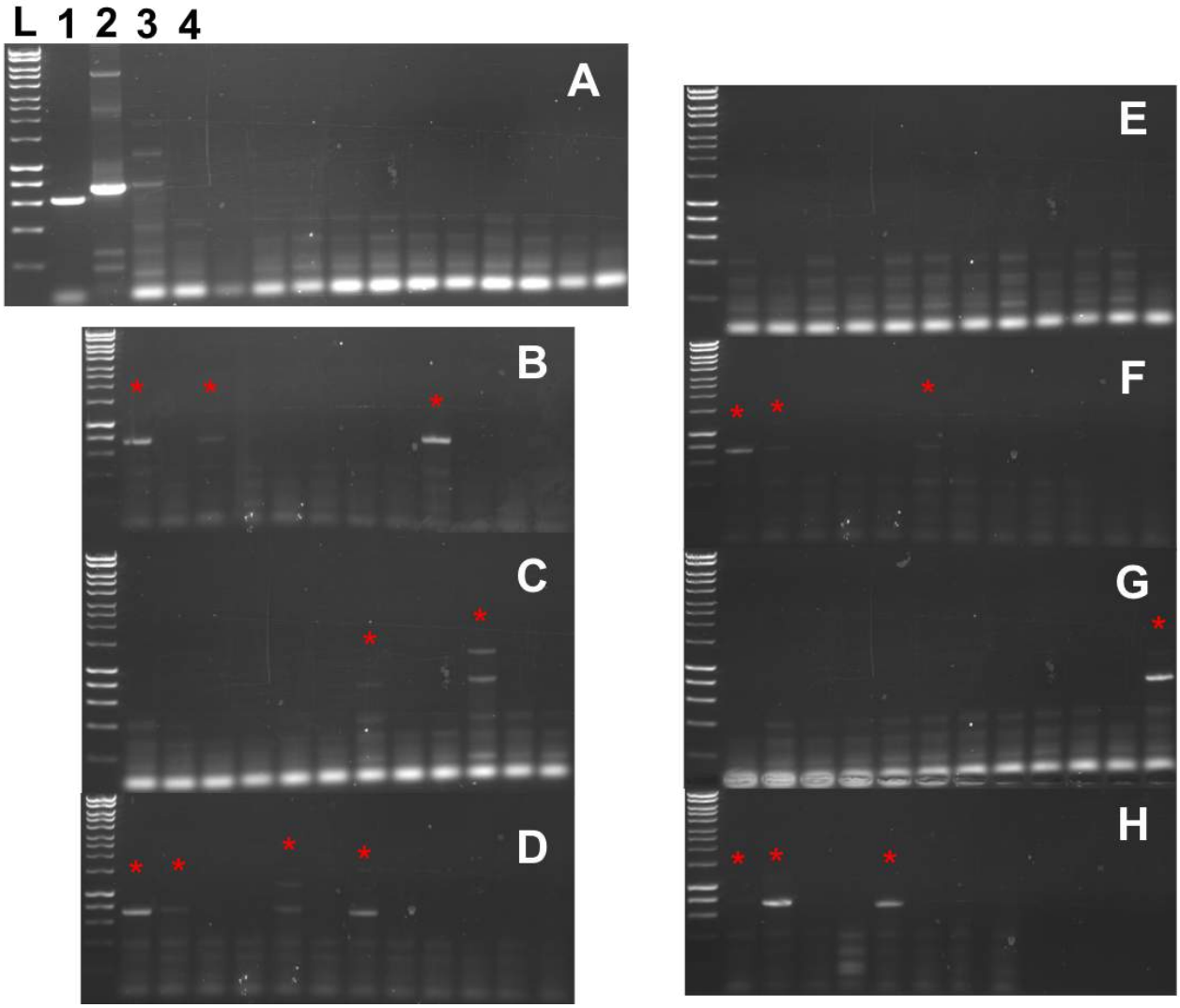
GO terms with > 10-fold enrichment shown. Only most specific subclass of each cluster shown.

Appendix D: Supplementary material for Chapter 5

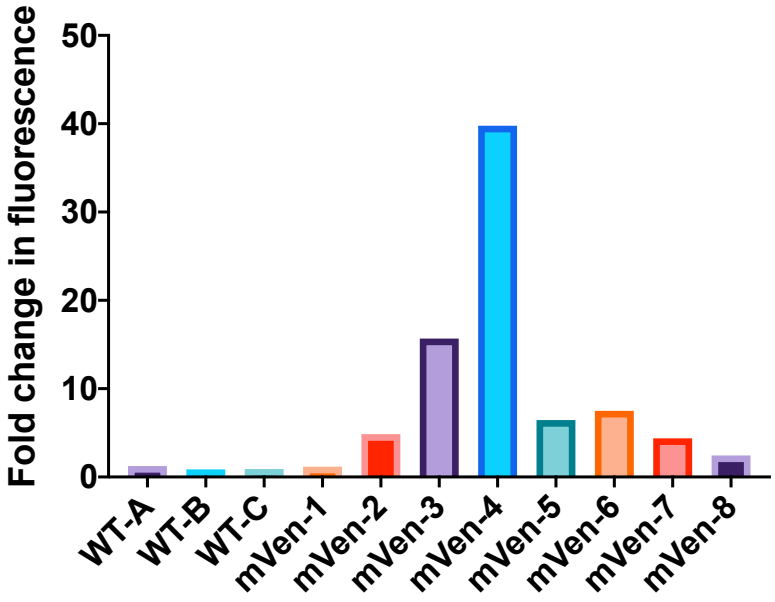
Appendix D Figures



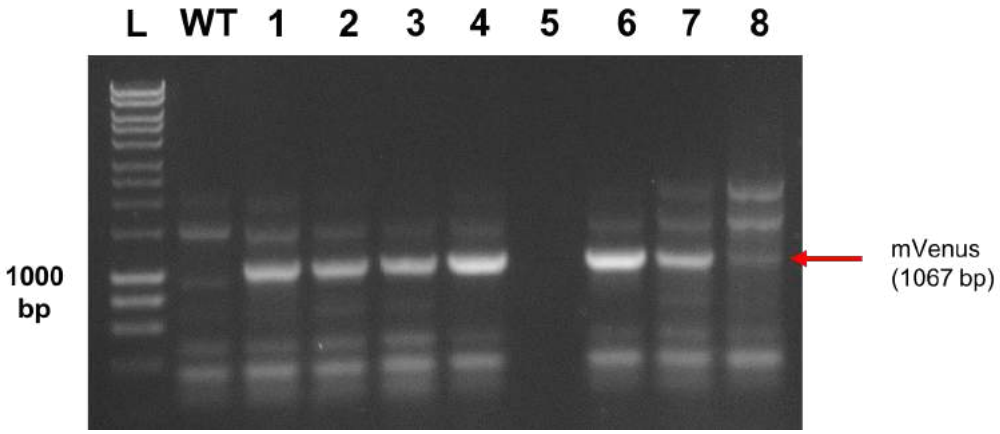
Appendix Figure D1: Gels showing stages in construction of pOpt_mCherry. A – L, Ladder; pOpt, pOpt_mVenus_Paro digested with NdeI and EcoRI. The heavier DNA band in the pOpt lane (5604 bp) shows the cut vector backbone, and the lighter band at 1067 bp shows the cut mVenus insert. The vector backbone was excised and purified for ligation with digested mCherry fragment (**B**) to form the pOpt_mCherry vector (**Figure 5.7**). **B –** L, ladder; mCherry, 733 bp PCR product of mCherry amplified from plasmid pUC_mCherry using primers mCherry_NdeI_F and mCherry_EcoRI_R (**Table 2.2**), annealing temperature 59°C. This fragment was cleaned and digested with NdeI and EcoRI, then ligated with the cut vector backbone depicted in **A**. Ligations were transformed into *E. coli* DH5α cells using ampicillin for plasmid selection. Plasmid DNA was isolated from individual colonies. **C** - Diagnostic digestion of pOpt_mCherry minipreps using restriction enzymes NdeI and EcoRI. L, Ladder; C, pOpt_mVenus_Paro control; 1–8, digested minipreps of selected colonies. Lanes 1, 6, 7 and 8 show inserts 600–800 bp in size, which correspond to the mCherry insert (733 bp), and these vectors were sequenced.



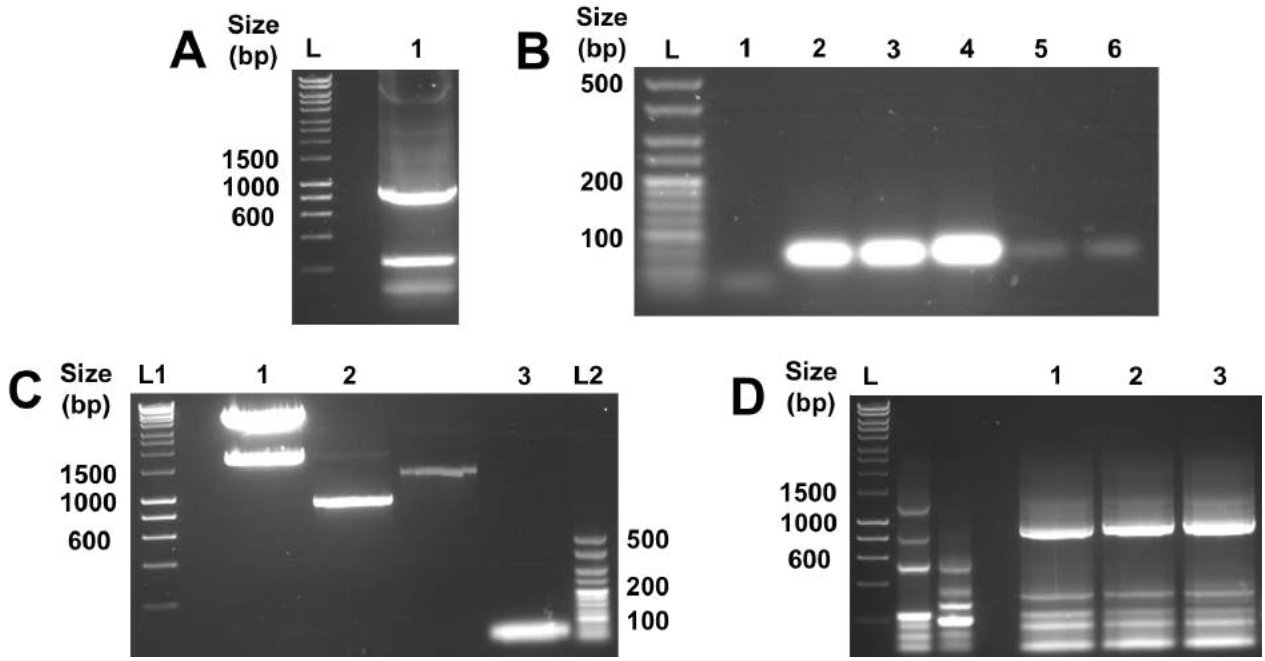
Appendix Figure D2: Agarose gels showing PCR screen of 92 paromomycin-resistant pOpt_mCherry-transformed colonies for integration of mCherry gene. Lane L, 1 kbp ladder; lane 1, *mid* gene DNA positive control; lane 2, pOpt_mCherry vector positive control; lane 3, previously confirmed mCherry-positive clone; lane 4, WT negative control. Gels A–H each contain 1 kbp DNA ladder in their first lane; all other lanes (besides lanes 1–4 on plate A) contain amplified DNA from the mCherry PCR screen. mCherry fragment (733 bp) was amplified using primers mCherry_NdeI_F and mCherry_EcoRI_R (**Table 2.2**); positive clones showing a band ~700 bp are signified using a red asterisk. The *mid* gene, a 622 bp amplicon that determines mating type in *C. reinhardtii* (Ferris and Goodenough, 1997), was amplified from a randomly selected clone to ensure gDNA was sufficiently extracted for PCR in lane A1 using primers Mid_mt_F and Mid_mt_R (**Table 2.2**).



Appendix Figure D3: mVenus measurements of pOpt_mVenus_Paro transformant strains grown in 6 well plates. Eight surviving transformant strains that initially exhibited > 2.5-fold higher mVenus fluorescence than WT when grown on 96-well plates following transformation with pOpt_mVenus_Paro (**Figure 5.10**). mVenus measurements normalised to chlorophyll fluorescence. Fold change relative to WT measurements. mVenus measured using fluorescent plate reader at Ex500/Em550 nm.

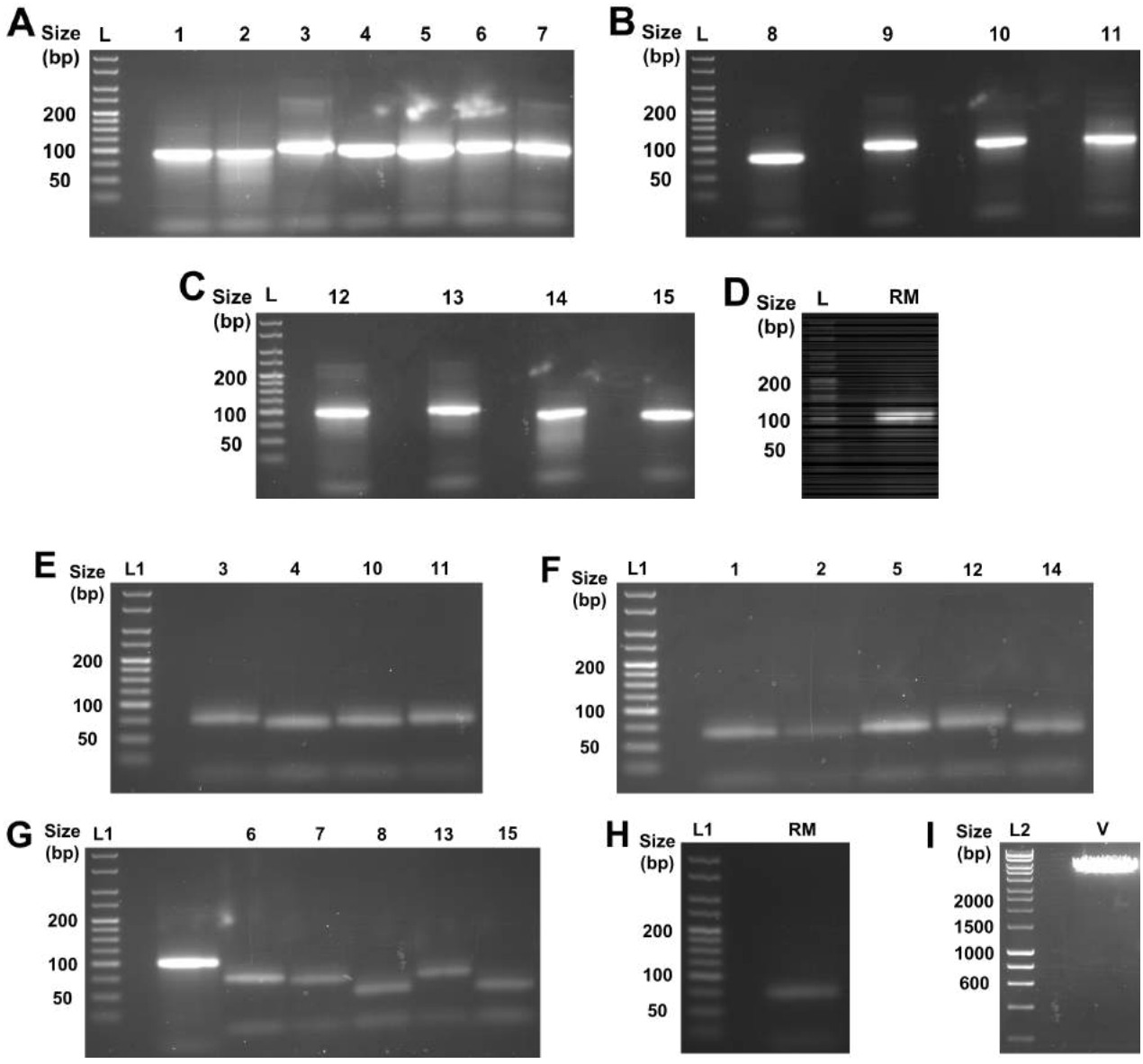


Appendix Figure D4: PCR screening of selected mVenus-fluorescing transformants for integration of mVenus gene. L, 1 kbp DNA ladder; **1–8**, mVenus amplification product of selected clones; **WT**, wild-type. mVenus DNA was amplified using primers mVen_Screen_F and mVen_Screen_R (**Table 2.2**), annealing temperature 61°C. 1% agarose gel.



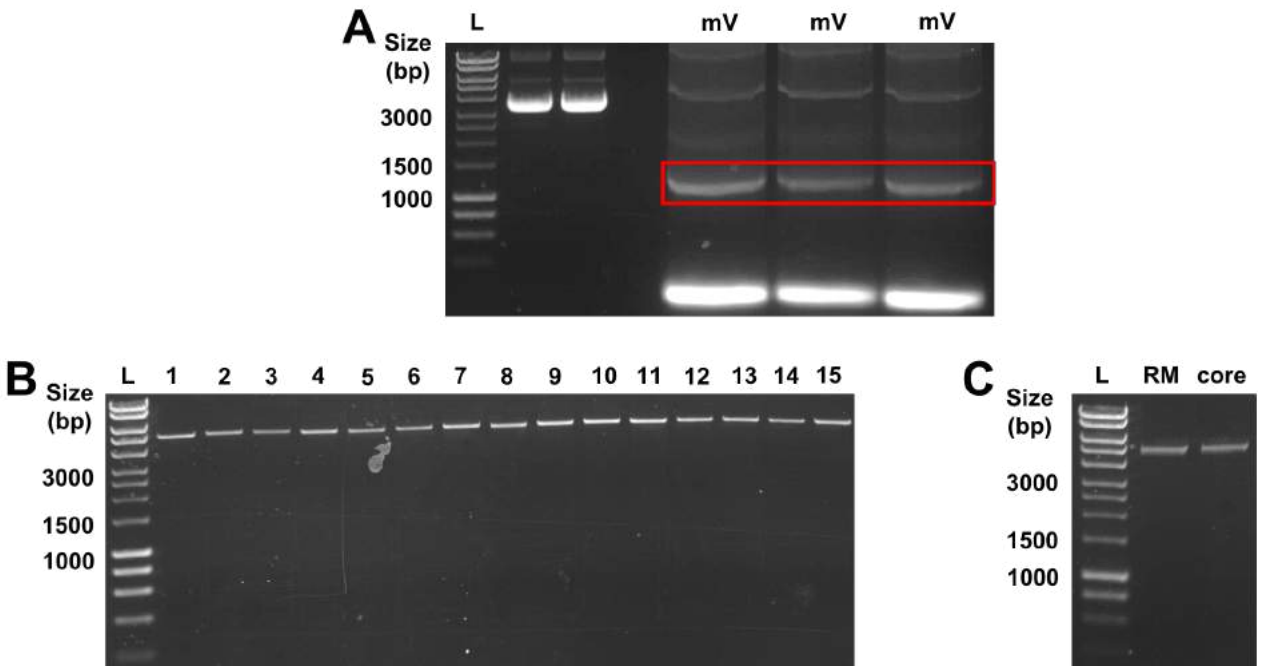
Appendix Figure D5: Construction of the pOpt_Core_mCherry vector.

A - 1% agarose gel showing PCR amplification of iRbcS2_mCherry fragment using primers iRbcS2_Amp_F and mCherry_EcoRI_R (Annealing temp, 60°C; Extension time, 1 min 30 s). L, 1 kbp ladder; 1, 888 bp iRbcS2_mCherry fragment. **B** – 2% agarose gel showing PCR amplification of pCore fragment using primers pCore_Amp_F and pCore_Amp_R (Annealing temp, 60 °C; Extension time, 1 min 15 s). L, 25 kb ladder; 1, ssDNA pCore template (50 bp); 2–4, PCR amplification of ssDNA fragment (70 bp); 5–6, gel extracted pCORE fragment (70 bp). **C** – 1% agarose gel showing fragment digestion for pOpt_Core_mCherry building. L1, 1 kbp ladder; 1, pOpt_mCherry cut with XbaI + EcoRI; 2, iRbcS2_mCherry cut with ClaI + EcoRI; 3, pCore cut with ClaI + XbaI; L2, 25 kbp ladder. **D** – 1% agarose gel showing PCR amplification of ligation mixture with primers pCore_Amp_F and mCherry_EcoRI_R. Expected fragment size 944 bp. (Annealing temp, 60°C; Extension time, 1 min 40 s). L, 1 kbp ladder; 1–3, pCore_iRbcS2_mCherry fragment. All primers listed in **Table 2.2**.



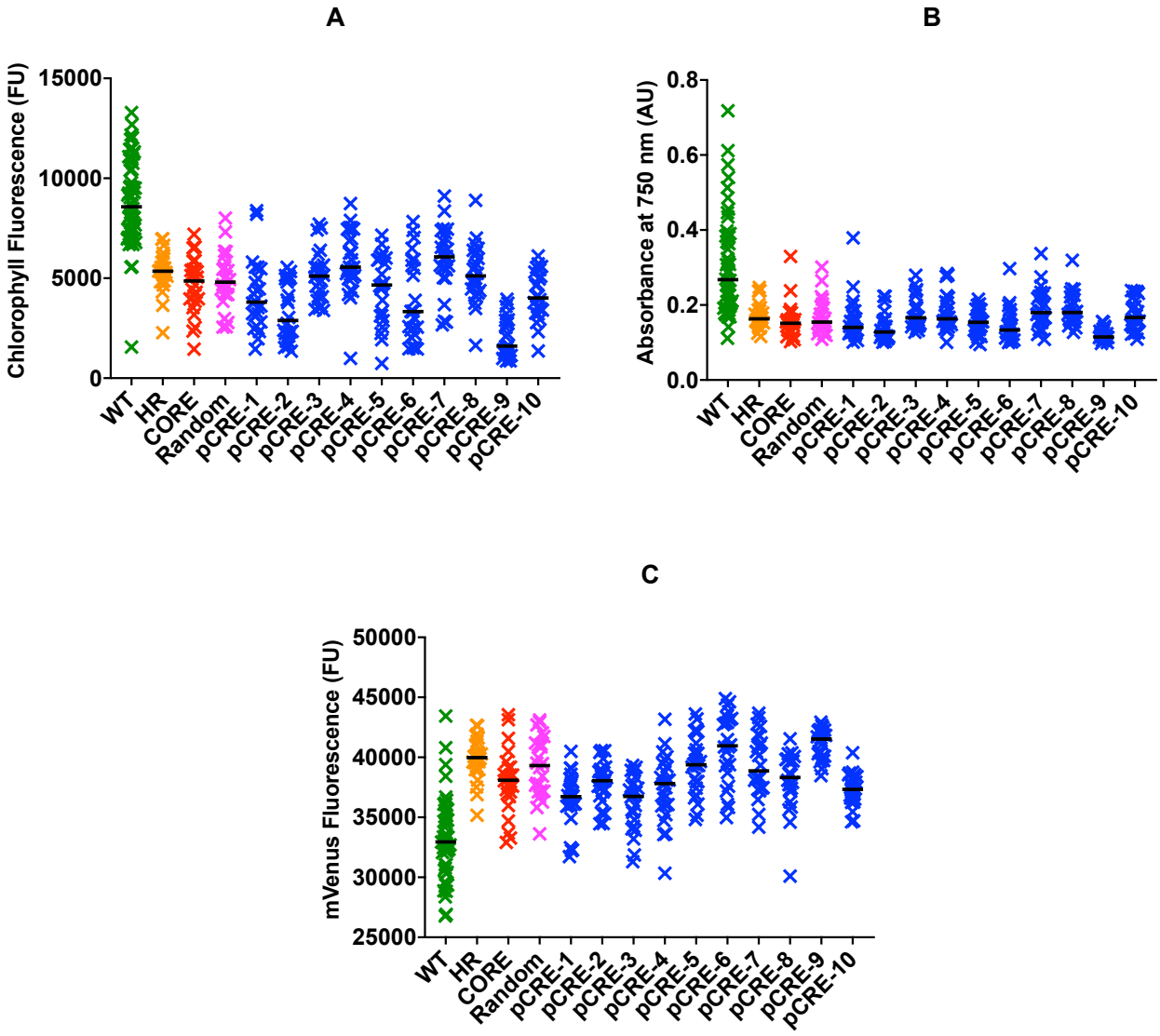
Appendix Figure D6: Gels showing construction of pCRE vector suite.

A–D - 3% agarose gels showing amplified ssDNA fragments listed in **Table 2.3**, gels labelled **A–D.** Lane **L** for each gel – 25 bp DNA ladder (**Figure 2.2B**). 1–15 represent pCRE-1 to pCRE-15. RM – pCRE-RM. ssDNA templates (**Table 2.3**) were amplified using primers pCRE_Amp_F and pCRE_Amp_R (**Table 2.2**), annealing temperature 62°C. **E–H** – 3% agarose gels showing amplified pCRE fragments (**A–D**) following digestion with SacI and XbaI restriction enzymes. **L1** – 25 bp DNA ladder (**Figure 2.2B**). Lanes are numbered according to their respective pCRE (pCRE-1 to pCRE-15) and RM (pCRE-RM). **I** – 1% agarose gel showing pOpt_Core_mCherry vector digested with SacI and XbaI restriction enzymes (**V**). **L2** – 1 kbp DNA ladder (**Figure 2.2B**).



Appendix Figure D7: Gels showing construction of pCRE_mVenus vector suite.

A – The mVenus gene plus its iRbcS2 intron (iRbcS2-mVenus) was PCR amplified from pOpt_mVenus_Paro for insertion into pCRE_mCherry vectors (**Appendix Table D9**) using primers iRbcS2_Amp_F and mVenus_EcoRI_R, annealing temperature 60°C. **L**, 1 kbp DNA ladder (**Figure 2.2A**); mV, amplified vector fragments. The bands encompassed within the red rectangle (1260 bp) were excised and purified, then digested with ClaI and EcoRI for insertion into pCRE vectors (gel not shown). **B** and **C** – 1% agarose gels showing gel extracted pCRE_mCherry vector backbones that were digested with ClaI and EcoRI to remove the iRbcS2-mCherry fragment for insertion of iRbcS2-mVenus. **L**, 1 kbp DNA ladder (**Figure 2.2A**); **1–15** and **RM**, backbones of digested pCRE_mCherry vectors 1-15 and pCRE-RM_mCherry (**Appendix Table D9**); **core**, pOpt_core_mCherry backbone minus the iRbcS2-mCherry insert. Gel extractions of the vector backbones were ligated with the iRbcS2-mVenus insert to produce the pCRE_mVenus vectors tested in **Chapter 5** and listed in **Table 2.5**.



Appendix Figure D8: Raw plate reader measurements for WT and strains transformed with test vectors for measuring pCREs 1–10, AR-1 and core promoters.

A – Raw chlorophyll fluorescence measurements (Ex440/ 9, Em680/ 20). **B** – OD750 measurements. **C** – Raw mVenus fluorescence measurements before normalisation to chlorophyll fluorescence (Ex515/ 9, Em529/ 20). FU, Fluorescence units. AU, absorbance units. Black line shows median value per population.

Appendix D Tables

Appendix Table D1: Top 300 constitutively expressed genes from Mettler *et al.* (2014)

Gene Name	Description	PFAM ID
Cre12.g548950	Chlorophyll a/b binding protein of LHCII	PF00504
Cre01.g066917	Chlorophyll a/b binding protein of LHCII	PF00504
Cre12.g512600	Ribosomal protein L18, component of cytosolic 80S ribosome and 60S large subunit	0
Cre04.g232104	Light-harvesting complex II chlorophyll a/b binding protein M3	PF00504

Cre06.g278135	Ribosomal protein L21, component of cytosolic 80S ribosome and 60S large subunit	PF01157
Cre06.g263450	(M=2) K03231 - elongation factor 1-alpha	PF00009
Cre11.g480150	Ribosomal protein S14, component of cytosolic 80S ribosome and 40S small subunit	PF00411
Cre14.g626700	Ferredoxin	PF00111
Cre12.g550850	Oxygen-evolving enhancer protein 2 of photosystem II	PF01789
Cre06.g250200	S-adenosylmethionine synthetase	PF02773
Cre10.g456200	Ribosomal protein S24, component of cytosolic 80S ribosome and 40S small subunit	PF01282
Cre16.g666301	Ribosomal protein S30, component of cytosolic 80S ribosome and 40S small subunit	PF04758
Cre04.g211800	Ribosomal protein L23, component of cytosolic 80S ribosome and 60S large subunit	PF00238
Cre14.g617900	Ribosomal protein L35, component of cytosolic 80S ribosome and 60S large subunit	PF00831
Cre08.g358556	Ribosomal protein S29, component of cytosolic 80S ribosome and 40S small subunit	PF00253
Cre06.g285250	Chloropyll a/b binding protein of LHCII type I, chloroplast precursor	PF00504
Cre06.g272800	Ribosomal protein S8, component of cytosolic 80S ribosome and 40S small subunit	PF01201
Cre07.g325746	Ribosomal protein L38, component of cytosolic 80S ribosome and 60S large subunit	PF01781
Cre06.g283950	Chlorophyll a/b binding protein of LHCII	PF00504
Cre09.g396213	Oxygen-evolving enhancer protein 1 of photosystem II	PF01716
Cre14.g630100	Ribosomal protein L13, component of cytosolic 80S ribosome and 60S large subunit	PF01294
Cre12.g560950	Photosystem I reaction center subunit V	PF01241
Cre02.g102250	Ribosomal protein S3, component of cytosolic 80S ribosome and 40S small subunit	PF00189
Cre10.g420350	Photosystem I 8.1 kDa reaction center subunit IV	PF02427
Cre02.g120100	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 1, chloroplast precursor	PF00101
Cre02.g120150	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2	PF00101
Cre12.g548400	Light-harvesting protein of photosystem II	PF00504
Cre17.g720250	Chlorophyll a/b binding protein of photosystem II	PF00504
Cre12.g484050	Ribosomal protein L36, component of cytosolic 80S ribosome and 60S large subunit	PF01158
Cre05.g238332	Photosystem I reaction center subunit II, 20 kDa	PF02531
Cre16.g661050	Ribosomal protein, L34e superfamily, component of cytosolic 80S ribosome and 60S large subunit	PF01199
Cre08.g382500	Ribosomal protein S25, component of cytosolic 80S ribosome and 40S small subunit	PF03297
Cre02.g075700	Ribosomal protein L19, component of cytosolic 80S ribosome and 60S large subunit	PF01280
Cre06.g278213	Light-harvesting protein of photosystem I	PF00504
Cre07.g330250	Subunit H of photosystem I	PF03244
Cre12.g494050	Ribosomal protein L9, component of cytosolic 80S ribosome and 60S large subunit	PF00347
Cre02.g143050	Acidic ribosomal protein P2	PF00428
Cre06.g310700	Ribosomal protein L36a, component of cytosolic 80S ribosome and 60S large subunit	PF00935
Cre09.g402219	Low-CO2-inducible protein	0
Cre06.g283050	Light-harvesting protein of photosystem I	PF00504
Cre07.g357850	Ribosomal protein L22, component of cytosolic 80S ribosome and 60S large subunit	PF01776
Cre06.g273600	Ribosomal protein S27a, component of cytosolic 80S ribosome and 40S small subunit	PF01599
Cre06.g296350	(M=3) PTHR12606 - SENTRIN/SUMO-SPECIFIC PROTEASE	PF02902
Cre12.g498900	Ribosomal protein S7, component of cytosolic 80S ribosome and 40S small subunit	PF01251
Cre03.g207050	Ribosomal protein L29, component of cytosolic 80S ribosome and 60S large subunit	PF01779
Cre16.g687900	Light-harvesting protein of photosystem I	PF00504
Cre06.g282500	Ribosomal protein L23a, component of cytosolic 80S ribosome and 60S large subunit	PF03939
Cre12.g508750	Light-harvesting protein of photosystem I	PF00504
Cre17.g732000	Mitochondrial F1F0 ATP synthase subunit 9, isoform A	PF00137
Cre03.g182551	Pre-apoplastocyanin	PF13473
Cre06.g272650	Light-harvesting protein of photosystem I	PF00504
Cre08.g365400	Plastid ribosomal protein L31	PF01197
Cre12.g494750	Plastid ribosomal protein S20	PF01649
Cre01.g2717058	Chlorophyll a/b binding protein of LHCII	PF00504
Cre09.g412100	Photosystem I reaction center subunit III	PF02507
Cre10.g420750	Ribosomal protein L30, component of cytosolic 80S ribosome and 60S large subunit	PF01248
Cre16.g682300	Ribosomal protein S26, component of cytosolic 80S ribosome and 40S small subunit	PF01283
Cre02.g115200	Ribosomal protein L27a, component of cytosolic 80S ribosome and 60S large subunit	PF00828

Cre06.g257150	Ribosomal protein L37a, component of cytosolic 80S ribosome and 60S large subunit	PF01780
Cre08.g359750	Ribosomal protein S9, component of cytosolic 80S ribosome and 40S small subunit	PF01479
Cre12.g546150	Cytochrome b6f complex PetM subunit	PF08041
Cre12.g529400	Ribosomal protein S27e isoform 1, component of 80S ribosome and 40S small subunit	PF01667
Cre10.g425900	Light-harvesting protein of photosystem I	PF00504
Cre11.g475250	Ubiquinol:cytochrome c oxidoreductase 7 kDa subunit	PF05365
Cre12.g530650	Glutamine synthetase	PF03951
Cre06.g284250	Chlorophyll a/b binding protein of LHCII	PF00504
Cre12.g497300	Rhodanese-like Ca-sensing receptor	PF00581
Cre06.g259900	Chloroplast ATP synthase gamma chain	PF00231
Cre03.g204250	S-Adenosyl homocysteine hydrolase	PF05221
Cre17.g698000	Mitochondrial F1F0 ATP synthase, beta subunit	PF00006
Cre13.g568900	Ribosomal protein L17, component of cytosolic 80S ribosome and 60S large subunit	PF00237
Cre17.g724300	Photosystem I reaction center subunit psaK	PF01241
Cre02.g082750	4.1 kDa photosystem II subunit	PF06596
Cre02.g082500	Photosystem I reaction center subunit N, chloroplastic	PF05479
Cre18.g744400	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre12.g529650	Ribosomal protein S27e isoform 1, component of 80S ribosome and 40S small subunit	PF01667
Cre08.g360900	Ribosomal protein S15, component of cytosolic 80S ribosome and 40S small subunit	PF00203
Cre12.g528750	Ribosomal protein L12, component of cytosolic 80S ribosome and 60S large subunit	PF03946
Cre10.g459250	Ribosomal protein L35a, component of cytosolic 80S ribosome and 60S large subunit	PF01247
Cre08.g380250	Small protein associating with GAPDH and PRK	PF02672
Cre11.g481450	CF0 ATP synthase subunit II precursor	PF00430
Cre11.g476750	Ferredoxin-NADP reductase	PF00175
Cre02.g080200	Transketolase	PF00456
Cre08.g372450	Oxygen evolving enhancer protein 3	PF05757
Cre07.g340350	Mitochondrial F1F0 ATP synthase associated 60.6 kDa protein	0
Cre12.g537800	Ribosomal protein L7, component of cytosolic 80S ribosome and 60S large subunit	PF08079
Cre13.g577100	Acyl-carrier protein	PF00550
Cre03.g165100	Photosystem I reaction centre, subunit VIII	PF00796
Cre10.g417700	Ribosomal protein L3, component of cytosolic 80S ribosome and 60S large subunit	PF00297
Cre09.g402300	Mitochondrial F1F0 ATP synthase associated 10.0 kDa protein	0
Cre03.g203450	Ribosomal protein S21, component of cytosolic 80S ribosome and 40S small subunit	PF01249
Cre12.g483950	NAD-dependent malate dehydrogenase, mitochondrial	PF02866
Cre06.g298650	(M=1) K03257 - translation initiation factor 4A	PF00270
Cre03.g203850	ATP-sulfurylase	PF01747
Cre02.g114600	2-cys peroxiredoxin	PF10417
Cre02.g091100	Ribosomal protein L15, component of cytosolic 80S ribosome and 60S large subunit	PF00827
Cre13.g598750	Epsilon subunit of COP-I complex	PF04733
Cre07.g331900	Ribosomal protein S13, component of cytosolic 80S ribosome and 40S small subunit	PF08069
Cre05.g234550	Fructose-1,6-bisphosphate aldolase	PF00274
Cre16.g650550	Flagellar Associated Protein, nucleoside diphosphate kinase-like	PF00334
Cre17.g701650	Ribosomal protein L27, component of cytosolic 80S ribosome and 60S large subunit	PF01777
Cre01.g027000	Ribosomal protein L11, component of cytosolic 80S ribosome and 60S large subunit	PF00673
Cre10.g430400	Ribosomal protein L37, component of cytosolic 80S ribosome and 60S large subunit	PF01907
Cre41.g786600	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre09.g386650	ADP/ATP carrier protein, mitochondrial	PF00153
Cre17.g738300	Acidic ribosomal protein P1	PF00428
Cre02.g079800	Mitochondrial F1F0 ATP synthase associated 13.3 kDa protein	0
Cre06.g258800	Hydroxyproline-rich glycoprotein component of the outer cell wall	0
Cre02.g101350	Ribosomal protein L10a, component of cytosolic 80S ribosome and 60S large subunit	PF00687
Cre10.g434750	Acetohydroxy acid isomeroreductase	PF01450
Cre12.g513200	Enolase	PF00113
Cre06.g298100	Translation initiation protein	PF01253

Cre07.g349350	0	0
Cre07.g338050	Mitochondrial F1F0 ATP synthase associated 36.3 kDa protein	0
Cre12.g514500	Ribosomal protein S11, component of cytosolic 80S ribosome and 40S small subunit	PF16205
Cre06.g289550	Ribosomal protein L32, component of cytosolic 80S ribosome and 60S large subunit	PF01655
Cre12.g510400	Putative rubredoxin-like protein	0
Cre01.g039250	Ribosomal protein S2, component of cytosolic 80S ribosome and 40S small subunit	PF00333
Cre12.g504200	Ribosomal protein S23, component of cytosolic 80S ribosome and 40S small subunit	PF00164
Cre03.g179800	Low-CO2-inducible membrane protein	PF07466
Cre12.g510050	Copper target 1 protein	PF02915
Cre10.g440400	0	0
Cre04.g229300	Rubisco activase	PF00004
Cre01.g052100	Plastid ribosomal protein L18	PF00861
Cre03.g154350	Cytochrome c oxidase subunit II, protein IIa of split subunit	PF02790
Cre09.g400650	Ribosomal protein S6, component of cytosolic 80S ribosome and 40S small subunit	PF01092
Cre16.g660851	0	0
Cre13.g568650	Ribosomal protein S3a, component of cytosolic 80S ribosome and 40S small subunit	PF01015
Cre22.g765228	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre06.g272950	Ribosomal protein S18, component of cytosolic 80S ribosome and 40S small subunit	PF00416
Cre06.g262700	Ubiquinol:cytochrome c oxidoreductase 14 kDa subunit, mitochondrial	PF02271
Cre09.g415550	Mitochondrial F1F0 ATP synthase associated 45.5 kDa protein	0
Cre08.g372950	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	PF02401
Cre02.g088900	Plastid ribosomal protein L1	PF00687
Cre12.g516200	Elongation Factor 2	PF00009
Cre01.g049500	Cytochrome c oxidase subunit II, protein IIb of split subunit	PF00116
Cre06.g257601	2-cys peroxiredoxin, chloroplastic	PF08534
Cre02.g126450	Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit 2	PF00101
Cre12.g493950	Plastid ribosomal protein S13	PF00416
Cre03.g158000	Glutamate-1-semialdehyde aminotransferase	PF00202
Cre09.g388200	Ribosomal protein L10, component of cytosolic 80S ribosome and 60S large subunit	PF00252
Cre12.g535850	Glycine cleavage system, P protein	PF02347
Cre13.g596450	Epsilon subunit of COP-I complex	PF04733
Cre07.g325500	Magnesium chelatase subunit H	PF02514
Cre13.g604650	(M=1) PTHR10804:SF11 - PROLIFERATION-ASSOCIATED PROTEIN 2G4	PF00557
Cre17.g721300	Mitochondrial F1F0 ATP synthase associated 14.3 kDa protein	0
Cre03.g187450	Ribose-5-phosphate isomerase	PF06026
Cre13.g567950	ADP-glucose pyrophosphorylase large subunit 1	PF00483
Cre22.g763250	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre01.g051900	Rieske iron-sulfur protein of mitochondrial ubiquinol-cytochrome c reductase	PF00355
Cre10.g452100	Ycf32-related polyprotein of photosystem II	PF06298
Cre16.g675550	Peptidyl-prolyl cis-trans isomerase, FKBP-type	PF00254
Cre16.g680000	Mitochondrial ATP synthase subunit 5, OSCP subunit	PF00213
Cre17.g713200	Chloroplast oxoglutarate-malate translocator	PF00939
Cre12.g537450	Cytochrome c oxidase 12 kDa subunit	0
Cre06.g264350	Plastid ribosomal protein L13	PF00572
Cre01.g010900	Glyceraldehyde-3-Phosphate Dehydrogenase	PF02800
Cre01.g016900	0	PF14347
Cre12.g509750	Mitochondrial processing peptidase alpha subunit	PF00675
Cre09.g409150	Ubiquinol:cytochrome c oxidoreductase 10 kDa subunit	0
Cre09.g409150	Ubiquinol:cytochrome c oxidoreductase 10 kDa subunit	0
Cre06.g290950	Ribosomal protein S5, component of cytosolic 80S ribosome and 40S small subunit	PF00177
Cre01.g018800	Mitochondrial F1F0 ATP synthase subunit 6	PF00119
Cre06.g307500	Low-CO2 inducible protein	0
Cre03.g188250	ADP-glucose pyrophosphorylase small subunit	PF00483
Cre01.g007051	(M=1) PF01020 - Ribosomal L40e family	PF01020

Cre06.g250100	Heat shock protein 70B	PF00012
Cre17.g715250	Acetyl-CoA biotin carboxyl carrier	PF00364
Cre12.g557050	Predicted protein	PF02325
Cre16.g650100	Subunit of the chloroplast cytochrome b6f complex	PF03742
Cre12.g510650	Fructose-1,6-bisphosphatase	PF00316
Cre09.g416050	Argininosuccinate synthase	PF00764
Cre12.g514050	Ferredoxin-dependent glutamate synthase	PF00310
Cre10.g420700	Mitochondrial F1F0 ATP synthase, epsilon subunit	PF04627
Cre12.g520600	Plastid ribosomal protein S6	PF01250
Cre03.g181150	Dynein arm light chain 8, LC8	PF01221
Cre17.g720050	FtsH-like membrane ATPase/metalloprotease	PF01434
Cre01.g026450	Serine/arginine-rich pre-mRNA splicing factor	PF00076
Cre16.g663900	Porphobilinogen deaminase	PF01379
Cre12.g496000	Plastid ribosomal protein S20	PF01649
Cre12.g486300	Photosystem I reaction center subunit XI	PF02605
Cre13.g581600	Mitochondrial F1F0 ATP synthase associated 31.2 kDa protein	0
Cre05.g233950	0	PF14159
Cre12.g519200	Adenylylphosphosulfate reductase	PF01507
Cre03.g157700	Cytochrome c oxidase 11 kD subunit	0
Cre12.g517150	Adenylylphosphosulfate reductase	PF01507
Cre06.g308250	Ribosomal protein S4, component of cytosolic 80S ribosome and 40S small subunit	PF00900
Cre01.g011000	Ribosomal protein L6, component of cytosolic 80S ribosome and 60S large subunit	PF01159
Cre19.g751700	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre05.g241950	(M=2) PTHR11743 - VOLTAGE-DEPENDENT ANION-SELECTIVE CHANNEL	PF01459
Cre13.g599400	Epsilon subunit of COP-I complex	PF04733
Cre05.g237450	Plastid-specific ribosomal protein 1	PF02482
Cre12.g558900	Cytochrome b6f complex subunit V	0
Cre15.g635850	Mitochondrial F1F0 ATP synthase, gamma subunit	PF00231
Cre18.g745500	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre01.g047750	Ribosomal protein L18a, component of cytosolic 80S ribosome and 60S large subunit	PF01775
Cre03.g159500	Ornithine decarboxylase 1	0
Cre06.g257450	Vacuolar ATP synthase subunit C proteolipid	PF00137
Cre18.g743700	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre01.g016500	Dihydrolipoamide dehydrogenase	0
Cre13.g592450	Epsilon subunit of COP-I complex	PF04733
Cre03.g191050	Ran-like small GTPase	PF00025
Cre01.g063200	Plastid ribosomal protein L18	PF00861
Cre10.g422600	NADH:ubiquinone oxidoreductase 51 kDa subunit	PF01512
Cre06.g253350	Glycine cleavage system, H-protein	PF01597
Cre06.g265800	Plastid ribosomal protein L28	PF00830
Cre01.g050550	(M=1) PF09360 - Iron-binding zinc finger CDGSH type	PF09360
Cre10.g432800	Ribosomal protein Sa, component of cytosolic 80S ribosome and 40S small subunit	PF00318
Cre09.g411100	Ribosomal protein S10, component of cytosolic 80S ribosome and 40S small subunit	PF03501
Cre03.g182150	Predicted protein	PF04536
Cre13.g573400	0	0
Cre12.g483850	0	0
Cre01.g004500	Isopropylmalate dehydratase, large subunit	PF00330
Cre12.g498250	Ribosomal protein S17, component of cytosolic 80S ribosome and 40S small subunit	PF00833
Cre08.g368050	(M=1) PTHR10815 - METHYLATED-DNA--PROTEIN-CYSTEINE METHYLTRANSFERASE	PF01035
Cre01.g040000	Ribosomal protein L26, component of cytosolic 80S ribosome and 60S large subunit	PF16906
Cre10.g459750	NADH:ubiquinone oxidoreductase 8 kDa subunit	PF15879
Cre14.g621450	Ribosomal protein L5, component of cytosolic 80S ribosome and 60S large subunit	PF00861
Cre03.g206600	Acetohydroxyacid dehydratase	PF00920
Cre08.g362450	Alpha-amylase	PF00128

Cre03.g199900	Eukaryotic translation initiation factor 4E	PF01652
Cre12.g494750	Plastid ribosomal protein S20	PF01649
Cre31.g780600	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre01.g037850	Acetyl-CoA biotin carboxyl carrier	PF00364
Cre12.g530300	Peptidyl-prolyl cis-trans isomerase, FKBP-type	PF00254
Cre12.g559250	14-3-3 protein	PF00244
Cre12.g542250	Beta tubulin 1	PF00091
Cre06.g304350	Mitochondrial cytochrome c oxidase subunit	PF02297
Cre06.g269450	Eukaryotic translation initiation factor 3, subunit G	PF00076
Cre03.g156950	Ubiquinol:cytochrome c oxidoreductase 9 kDa subunit	PF10890
Cre01.g019250	Putative dTDP-glucose 4-6-dehydratase	PF16363
Cre03.g160300	(M=1) PTHR15601 - STRESS ASSOCIATED ENDOPLASMIC RETICULUM PROTEIN (SERP1/RAMP4)	PF06624
Cre16.g659700	0	0
Cre01.g044800	Pyruvate-formate lyase	PF02901
Cre07.g356400	Centriole proteome protein; GMP phosphodiesterase, delta subunit	PF05351
Cre07.g340200	Thylakoid transmembrane protein involved in cyclic electron flow	0
Cre11.g468950	Ubiquinol:cytochrome c oxidoreductase 7 kDa subunit	PF05365
Cre06.g261000	10 kDa photosystem II polypeptide	PF04725
Cre02.g097400	Eukaryotic initiation factor, eIF-5A	PF01287
Cre12.g528000	(M=2) KOG1764 - 5'-AMP-activated protein kinase, gamma subunit	PF00571
Cre02.g111450	Rhodanese-like protein	PF00581
Cre12.g534800	Glycine cleavage system, P protein	PF02347
Cre09.g410700	NADP-dependent malate dehydrogenase, chloroplastic	PF00056
Cre13.g596800	Epsilon subunit of COP-I complex	PF04733
Cre03.g189400	Seryl-tRNA(Sec) synthetase	PF00587
Cre19.g757350	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre13.g571150	NADH:ubiquinone oxidoreductase 10 kDa subunit	0
Cre06.g257500	14-3-3 protein	PF00244
Cre10.g458550	(M=1) PF02575 - YbaB/EbfC DNA-binding family	PF02575
Cre16.g686900	(M=1) KOG3410 - Conserved alpha-helical protein	PF08555
Cre09.g410600	(M=2) PTHR13832//PTHR13832:SF140 - PROTEIN PHOSPHATASE 2C // SUBFAMILY NOT NAMED	PF00481
Cre05.g233800	(M=1) K01880 - glycyl-tRNA synthetase	PF03129
Cre02.g085450	Coproporphyrinogen III oxidase	PF01218
Cre01.g024350	(M=2) PTHR10766:SF14 - TRANSMEMBRANE 9 SUPERFAMILY PROTEIN MEMBER 1	PF02990
Cre01.g013700	0	PF01459
Cre11.g468450	Centrin present in monomeric inner arm dyneins b, e, and g	PF13499
Cre17.g734200	L,L-diaminopimelate aminotransferase	PF00155
Cre13.g562850	Thylakoid formation protein	PF11264
Cre06.g251000	0	0
Cre06.g299000	Plastid ribosomal protein L21	PF00829
Cre03.g195400	Ran-like small GTPase	PF00025
Cre05.g236150	(M=4) 2.7.7.2 - FAD synthetase.	PF13419
Cre13.g573351	(M=1) K02960 - small subunit ribosomal protein S16e	PF00380
Cre16.g677450	(M=2) PTHR11122//PTHR11122:SF9 - APOSPORY-ASSOCIATED PROTEIN C-RELATED // SUBFAMILY NOT NAMED	PF01263
Cre17.g738050	Flagellar membrane protein, paralog of AGG2	PF04749
Cre10.g423500	Heme oxygenase	PF01126
Cre02.g103550	Eukaryotic translation initiation factor 1A, eIF-1A	PF01176
Cre03.g151200	Predicted protein	0
Cre09.g396100	Cell wall protein pherophorin-C15	PF12499
Cre03.g152150	0	0
Cre05.g243800	Predicted protein	PF13326
Cre09.g415700	Carbonic anhydrase 3	PF00194
Cre10.g452800	Low-CO2-inducible protein	0
Cre01.g005050	(M=2) K06816 - golgi apparatus protein 1	PF00839

Cre16.g694850	N-acetylglutamate synthase	PF01960
Cre12.g510450	Ribosomal protein S28, component of cytosolic 80S ribosome and 40S small subunit	PF01200
Cre12.g494350	(M=1) PTHR10766:SF1 - TRANSMEMBRANE 9 SUPERFAMILY PROTEIN	PF02990
Cre09.g396300	Protoporphyrinogen oxidase	PF13450
Cre01.g064400	Plastid ribosomal protein L18	PF00861
Cre02.g096150	Mn superoxide dismutase	PF02777
Cre17.g741850	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre10.g438550	TatA-like sec-independent protein translocator subunit	PF02416
Cre26.g773800	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre13.g576450	0	0
Cre18.g749750	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre38.g785050	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre03.g169400	UDP-D-glucuronic acid decarboxylase	PF01073
Cre12.g537050	0	0
Cre07.g329650	(M=1) PF06041 - Bacterial protein of unknown function (DUF924)	PF06041
Cre16.g651550	Mitochondrial transcription termination factor	PF02536
Cre05.g232600	(M=48) PF05548 - Gametolysin peptidase M11	PF05548
Cre05.g247600	Ubiquitin-conjugating enzyme E2	PF00179
Cre19.g751950	Heterogeneous nuclear ribonucleoprotein	PF14259
Cre10.g451900	Threonine synthase	PF00291
Cre02.g079900	Mitochondrial F1F0 ATP synthase associated 13.3 kDa protein	0
Cre15.g643550	Pyridoxin biosynthesis protein	PF01680
Cre16.g691850	Cytochrome c oxidase subunit	0
Cre07.g343100	Mitochondrial F1F0 ATP synthase associated 60.6 kDa protein	0
Cre06.g269050	(M=1) PTHR10366//PTHR10366:SF122 - NAD DEPENDENT EPIMERASE/DEHYDRATASE // SUBFAMILY NOT NAMED	PF13460
Cre12.g522450	COP-II coat subunit	PF07304

Gene list taken from open access microarray dataset (Mettler *et al*., 2014), and descriptions acquired from Phytozome Biomart.

Appendix Table D2: Motifs discovered in top 267 expressed genes using Weeder2.0

Motif name	Motif forward	Motif reverse
Weeder_1	TCCCTCTTTC	GAAAGAGGGA
Weeder_2	GCCCCATGCA	TGCATGGGGC
Weeder_3	TGCATGGGCC	GGCCCATGCA
Weeder_4	GAGAAAGAGA	TCTCTTTCTC
Weeder_5	TCTCTTTC	GAAAGAGA
Weeder_6	GCCCCATT	AATGGGGC
Weeder_7	GGGGTACT	AGTACCCC
Weeder_8	CGTACGGC	GCCGTACG
Weeder_9	CGAGAGGT	ACCTCTCG
Weeder_10	GACCCATG	CATGGGTC
Weeder_11	CGAGGGAC	GTCCCTCG
Weeder_12	CTCTCG	CGAGAG
Weeder_13	CTCTTT	AAAGAG
Weeder_14	AATGGG	CCCATT
Weeder_15	GTACCC	GGGTAC

Weeder_16	GGGGTC	GACCCC
Weeder_17	GTACGG	CCGTAC
Weeder_18	GTCTCT	AGAGAC
Weeder_19	AAGGGG	CCCCTT
Weeder_20	CATGGG	CCCATG
Weeder_21	AGAGCC	GGCTCT
Weeder_22	CTCGGT	ACCGAG
Weeder_23	GTAGGC	GCCTAC
Weeder_24	CTACAC	GTGTAG
Weeder_25	GTAGCC	GGCTAC

Appendix Table D3: Results from HOMER *de novo* motif discovery

Motif name	Motif logo	*P*-value	# occurrences within top 267 promoters
Homer_1		1.00E-19	29
Homer_2		1.00E-12	64
Homer_3		1.00E-11	66
Homer_4		1.00E-11	21
Homer_5		1.00E-11	27
Homer_6		1.00E-10	73
Homer_7		1.00E-10	52
Homer_8		1.00E-10	27
Homer_9		1.00E-09	55
Homer_10		1.00E-09	25

Homer_11		1.00E-09	25
Homer_12		1.00E-09	16
Homer_13		1.00E-08	30
Homer_14		1.00E-08	9
Homer_15		1.00E-08	10
Homer_16		1.00E-08	13
Homer_17		1.00E-08	20
Homer_18		1.00E-08	46
Homer_19		1.00E-08	61
Homer_20		1.00E-08	8
Homer_21		1.00E-08	9
Homer_22		1.00E-08	34
Homer_23		1.00E-08	34
Homer_24		1.00E-08	28
Homer_25		1.00E-08	14
Homer_26		1.00E-07	23

Homer_27		1.00E-07	7
Homer_28		1.00E-07	80
Homer_29		1.00E-07	56
Homer_30		1.00E-07	17
Homer_31		1.00E-06	7
Homer_32		1.00E-06	41
Homer_33		1.00E-05	11
Homer_34		1.00E-05	24
Homer_35		1.00E-05	20
Homer_36		1.00E-04	36
Homer_37		1.00E-00	5

Motif logos generated in HOMER program.

Appendix Table D4: DREME Results

Motif name	Motif logo	p-value	E-value
DREME_1		6.20E-11	2.70E-06
DREME_2		5.70E-08	2.40E-03

DREME_3		7.90E-08	3.30E-03
DREME_4		3.60E-07	1.50E-02
DREME_5		3.70E-07	1.60E-02
DREME_6		5.00E-07	2.10E-02
DREME_7		3.40E-16	1.50E-11
DREME_8		3.70E-13	1.60E-08
DREME_9		4.00E-09	1.70E-04
DREME_10		5.80E-09	2.50E-04
DREME_11		1.30E-08	5.40E-04
DREME_12		6.40E-08	2.60E-03
DREME_13		2.40E-07	9.90E-03
DREME_14		5.60E-07	2.30E-02

DREME program run with 1000 bp cut into 100 bp fragments. Logos created in MEME suite.

Appendix Table D5: Full list of clusters found in RSAT matrix clustering program

Motif Name	Motif F	Motif R	# motifs merged	Motifs merged list

cluster_1	TGCCGTACGA	TCGTACGGCA	6	DREME_13, Homer_7, Weeder_17, Weeder_8, DREME_3, DREME_7
cluster_2*	GCCCCATKCAGG	CCTGMATGGGGC	10	Homer_6, Homer_3, DREME_8, Weeder_2, Weeder_3, Weeder_14, Weeder_20, Homer_2, Weeder_10, Weeder_6
cluster_3	CGAGAGVC	GBCTCTCG	5	Weeder_18, Weeder_21, Weeder_11, Weeder_12, Weeder_9
cluster_4*	GHGAAAGARRGAGA	TCTCYYTCTTTCDC	9	DREME_10, DREME_2, Homer_29, Homer_1, Weeder_1, Weeder_13, DREME_4, Weeder_4, Weeder_5
cluster_5	CCTSGCC	GGCSAGG	2	DREME_12, DREME_5
cluster_6	SRGTMCCCC	GGGGKACYS	5	Homer_36, Weeder_16, Homer_28, Weeder_15, Weeder_7
cluster_7	CTCCAGGKTA	TAMCCTGGAG	2	DREME_6, Homer_10
cluster_8	TGTAGSCAGG	CCTGSCTACA	3	Homer_35, Weeder_23, Weeder_25
cluster_9*	TRTGYAGG	CCTRCAYA	4	DREME_14, DREME_1, DREME_11, Weeder_24
cluster_10*	CTCGGT	ACCGAG	1	Weeder_22
cluster_11	CRGTWCSGTGTG	CACACSGWACYG	2	Homer_21, Homer_34
cluster_12*	CCMTCKCGMSCVA	TBGSKCGMGAKGG	3	Homer_18, Homer_16, Homer_4
cluster_13*	GTATGCHTGCTG	CAGCADGCATAC	2	Homer_21, Homer_34
cluster_14	CCMTCKCGMSCVA	TBGSKCGMGAKGG	3	Homer_18, Homer_16, Homer_4
cluster_15	ACGCGGGGTA	TACCCCGCGT	1	Homer_13
cluster_16	AACCASGGYTAG	CTARCCSTGGTT	1	Homer_31
cluster_17*	GTCCACCTGG	CCAGGTGGAC	1	Homer_30
cluster_18	SATSSACCAGGW	WCCTGGTSSATS	1	Homer_8
cluster_19	GCCCTYCCAAGG	CCTTGGRAGGGC	2	DREME_9, Homer_9
cluster_20*	CGAGCGTTTTCT	AGAAAACGCTCG	1	Homer_20
cluster_21	KCYARCGYKC	GMRCGYTRGM	1	Homer_32
cluster_22	TGTGGTMTTTGC	GCAAAKACCACA	1	Homer_14
cluster_23	GTGGTGGTGGTG	CACCACCACCAC	1	Homer_37
cluster_24	AAGCGGCACG	CGTGCCGCTT	1	Homer_19
cluster_25	ACTCTGCAAG	CTTGCAGAGT	1	Homer_22
cluster_26	TYAYGGGWCC	GGWCCCRTRA	1	Homer_11
cluster_27	CGGGCGGGTCAG	CTGACCCGCCCG	1	Homer_25
cluster_28	CACTKACTGC	GCAGTMAGTG	1	Homer_23
cluster_29	GMAGCSCATGTC	GACATGSGCTKC	1	Homer_5
cluster_30	AKGGRTTCWT	AWGAAYCCMT	1	Homer_24
cluster_31	TAGTTCMGSA	TSCKGAACTA	1	Homer_26
cluster_32	GCCTTGGCCCCT	AGGGGCCAAGGC	1	Homer_33
cluster_33	CTTTTACGTC	GACGTAAAAG	1	Homer_17
cluster_34	CCAAATGCCGTG	CACGGCATTTGG	1	Homer_15

cluster_35	MAKS	SMTK	1	Weeder_19

Clusters found by all 3 discovery programs highlighted with an asterisk. IUPAC nucleotide nomenclature is listed in **Appendix Table A6.**

Appendix Table D6: DNA Cloning fragment generated for pOpt_mCherry

Fragment	Sequence 5' → 3'	Properties
mCherry fragment for insertion into pOpt_mVenus_Paro	TCACATATGATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCA AGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCA CGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCAC CCAGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCC TGGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAA GCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCT TCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCG TGACCCAGGACTCCTCCTTGCAGGACGGCGAGTTCATCTACAAGGTGAA GCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAG ACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGC GCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGC CACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGC AGCTGCCCGGCGCCTACAACGTCAACATCAAGTTGGACATCACCTCCCAC AACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGC CACTCCACCGGCGGCATGGACGAGCTGTACAAGTAATAAGAATTCGTA	*Nde*I, *EcoR*V 2593 bp

Restriction sites are underlined, and supplementary bases incorporated for the binding of restriction enzymes are highlighted in grey.

Appendix Table D7: DNA Cloning fragments generated for pOpt_Core_mCherry

Fragment	Sequence 5' → 3'	Properties
iRbcS2_mCherry	ACGTATCGATGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGC GTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGG CTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTC GCCGTTTCCATTTGCAGGATGCATATGATGGTGAGCAAGGGCGAG GAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTG CACATGGAGGGCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGC GAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGCCAAGCT GAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCT GTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCC GCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCA AGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACC GTGACCCAGGACTCCTCCTTGCAGGACGGCGAGTTCATCTACAAGG TGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCA GAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCC CGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGC TGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACA AGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACA TCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGA ACAGTACGAACGCGCCGAGGGCCGCCACTCCACCGGCGGCATGGA CGAGCTGTACAAGTAATAAGAATTCGTA	mCherry fluorescent reporter gene with iRbcS2 intron upstream ClaI, NdeI, EcoRI 888 bp
pCore_iRbcS2_mCherry	ACGTTCTAGAAAGCCGAGCGAGCCCGCTGCAGGTTAGTCTTTCTTT TAGCGTGTGCCCACATCGATGTGAGTCGACGAGCAAGCCCGGCGG ATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTG	Fragment generated by PCR of ligation mixture of pCore,

	TCGACGAAGGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCT AACCCTGCGTCGCCGTTTCCATTTGCAGGATGCATATGATGGTGAG CAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCG CTTCAAGGTGCACATGGAGGGCTCCGTGAACGGCCACGAGTTCGA GATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGA CCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCT GGGACATCCTGTCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGT GAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCC GAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGG CGTGGTGACCGTGACCCAGGACTCCTCCTTGCAGGACGGCGAGTT CATCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGC CCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCTCCGAG CGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCA GAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCA AGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCT ACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTA CACCATCGTGGAACAGTACGAACGCGCCGAGGGCCGCCACTCCAC CGGCGGCATGGACGAGCTGTACAAGTAATAAGAATTCGTA	iRbcS2_mCherry and cut pOpt_mCherry using primers pCore_Amp_F and mCherry_EcoRI_R XbaI, ClaI, NdeI, EcoRI 944 bp

Restriction sites underlined and listed in properties from 5' to 3'. Supplementary bases incorporated for the binding of restriction enzymes are highlighted in grey. DNA fragment sections highlighted as follows: pCore, red; iRbcS2, blue; mCherry, purple.

Appendix Table D8: mCherry plasmids for pCRE testing

Plasmid	Properties	Source/ Reference
pOpt_Core_mCherry	pOpt vector with mCherry reporter gene with Hsp70A/RbcS2 promoter replaced with core promoter and added restriction sites	This work
pCRE-1_mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-1 proximal promoter	This work
pCRE-2_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-2 proximal promoter	This work
pCRE-3_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-3 proximal promoter	This work
pCRE-4_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-4 proximal promoter	This work
pCRE-5_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-5 proximal promoter	This work
pCRE-6_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-6 proximal promoter	This work
pCRE-7_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-7 proximal promoter	This work

pCRE-8_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-8 proximal promoter	This work; linearised with BsaI
pCRE-9_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-9 proximal promoter	This work
pCRE-10_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-10 proximal promoter	This work
pCRE-11_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-11 proximal promoter	This work
pCRE-12_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-12 proximal promoter	This work
pCRE-13_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-13 proximal promoter	This work
pCRE-RM_ mCherry	pOpt vector backbone with mCherry reporter driven by Core promoter and pCRE-RM proximal promoter	This work

Appendix Table D9: DNA Cloning fragments generated for pCRE_mVenus vectors

Fragment	Sequence	Properties
iRbcS2_mVenus	ACGTATCGATGTGAGTCGACGAGCAAGCCCGGCGGATCAGGCAGCG TGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAAGGCTT TTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCG TTTCCATTTGCAGGATGCATATGAGATCTGACGTCATCGAGGGCAGG GTGAGCAAGGGCGAGGAGCTGTTCACCGGCGTGGTGCCCATCCTGG TGGAGCTGGACGGCGACGTGAACGGCCACAAGTTCAGCGTGAGCGG CGAGGGCGAGGGCGACGCCACCTACGGCAAGCTGACCCTGAAGCTG ATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTGGTGAC CACCCTGGGCTACGGCCTGCAGTGCTTCGCCCGCTACCCCGACCACAT GAAGCAGCACGACTTCTTCAAGAGCGCCATGCCCGAGGGCTACGTGC AGGAGCGCACCATCTTCTTCAAGGACGACGGTGAGCTTGCGGGGTTG CGAGCAACACTCCAGCAACGAACAGTGCCCAAGTCAGGAATCTGCAG TCAGCCTGGGCTTTCGGCGGCTTTTTCTTGGGCAAACAGCTTGCACTC ATGCCAGCGCGGCTTGTCCAGCCTCACTTGAGCTTTCCAGCTGCTACC AGCCGGGCTATACGACAGCGACAGAGCCATAGCGTGGAATCACTTAT TTGGGTTGCCGAAGTAGCGGTCGGAGCGTGAGTTCTTGGTCAAGCCG CCCCTTATCCGGTTCCTGTCCGTGTCTTTGTCCCTCGTTCACCCTTCGC GGCACCCTTCATCCCCTTGCTTGCAGGTAACTACAAGACCCGCGCCGA GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAG GGCATCGACTTCAAGGAGGACGGCAACATCCTGGGCCACAAGCTGG AGTACAACTACAACAGCCACAACGTGTACATCACCGCCGACAAGCAG AAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA CGGCGGCGTGCAGCTGGCCGACCACTACCAGCAGAACACCCCCATCG GCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCTACCAG AGCAAGCTGAGCAAGGACCCCAACGAGAAGCGCGACCACATGGTGC TGCTGGAGTTCGTGACCGCCGCCGGCATCACCCTGGGCATGGACGAG CTGTACAAGATCGAGGGCAGGGATATCGAATTCACGT	ClaI; NdeI; EcoRI 1260 bp

Restriction sites underlined and listed in order of appearance (5’ to 3’). Supplementary bases incorporated for the binding of restriction enzymes are highlighted in grey. DNA fragment sections highlighted as follows: iRbcS2, blue; mVenus, green. Fragment generated by PCR of pOpt_mVenus_Paro with primers iRbcS2_Amp_F and mVenus_EcoRI_R.