University of Southampton

Faculty of Natural and Environmental Sciences
School of Ocean and Earth Science

Improving Photosynthetic Conversion Efficiency in Marine Microalgae

by

Staffan Andreas Johansson

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

May 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
School of Ocean and Earth Science

Doctor of Philosophy

IMPROVING THE PHOTOSYNTHETIC CONVERSION EFFICIENCY IN MARINE MICROALGAE

by Staffan Andreas Johansson

Marine photosynthetic microalgae have great potential in biotechnology. They have huge genetic diversity and naturally make an array of metabolites that are precursors in high value products such as fuels and pharmaceuticals. They do not compete with agriculture for land or fresh water and can be used to reduce industrial carbon-emissions. In order to realize this potential however much work needs to be done to overcome the engineering challenges of growing microalgae on large scales and developing the genetic tools required to increase the yield and diversity of products synthesized by cells. Irrespective of the type of microalgal species selected for growth, the efficiency at which light energy is converted into product, the 'photosynthetic conversion efficiency' sets a fundamental limitation on the potential yield. In natural systems as much as 90% of absorbed light energy is re-emitted as heat or fluorescence, representing a major loss in overall efficiency. In this thesis a high-throughput pipeline using random mutagenesis and live single cell sorting has been used to isolate two cell-lines of the eukaryotic microalgae *Dunaliella tertiolecta* with reduced chlorophyll content (termed lca1 and lca2). As there is no published genome for *D. tertiolecta* and the species is difficult to transform, this approach represents a feasible method to develop improved cell-lines from any microalgal species. These cell lines are characterized physiologically and shown to increase the maximum rate of chlorophyll normalized photosynthesis P_{max} by 289 (lca1) and 131% (lca2) respectively. The molecular basis of these random mutations characterized by transcriptomics using next-generation sequencing approaches, helps define the differences in regulation in light-harvesting and photosynthesis between the lca1, lca2 and wild-type. The approaches applied in this thesis therefore show how microalgal strains with poor genetic characterization can rapidly be selected for biotechnological applications, and providing new gene targets and valuable insights into the fundamental mechanisms of photosynthesis.

Contents

List of Figures

List of Tables

Listings

Declaration of Authorship

I, Staffan Andreas Johansson, declare that this thesis titled, 'Improving Photosynthetic Conversion Efficiency in Marine Microalgae' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Acknowledgements

I would like to thank my supervisors, Thomas S. Bibby, C. Mark Moore, Matthew J. Terry, but also my previous supervisor Richard J. Edwards for all the help and the good advice that you have given me. I would especially like to thank you for the substantial feedback and all the comments you have provided during the writing of this thesis. I would like to thank Nicola Pratt and John Gittins for help with practical matters in the laboratory and good company. I would also like to thank Patrick Stephenson for all the work he did previously as part of this project. This work would not have been possible without him.

Big thanks to Ben Lucker at MSU in Lansing, Michigan, for having the patience to run all the growth experiments, and David E. Kramer who allowed us to use his laboratories setup of photobioreactors.

On a more personal note I would like to thank my wife's parents for supporting us in all ways possible, including financially during the last year. Thank you for having us over for two months and all the good food you provided. On the same note I would like to thank my mother and grandmother, who also allowed us to stay at their place while I finished writing up this thesis.

Thank you to all the people in the lab, and especially Despo and Andy, you are the best. Thank you to Christos and Iordanis for being great friends and flatmates and the best suppliers of coffee at NOCS.

Most of all I would like to thank my wife, Yue Chang, for helping me during this whole period. I couldn't have done it without you!

Thank you.

Abbreviations

PSII	**P**hoto **S**ystem **II**
PSI	**P**hoto **S**ystem **I**
LHCII	**L**ight **H**arvesting **C**omplex **II**
LHCI	**L**ight **H**arvesting **C**omplex **I**
OEC	**O**xygen **E**volving **C**omplex
NPQ	**N**on **P**hotochemical **Q**uenching
FRRf	**F**ast **R**epetition **R**ate **f**luoroemeter
PAM	**P**ulse **A**mplitude **M**odulation
MWM	**M**olecular **W**eight **M**arker
PCR	**P**olymerase **C**hain **R**eaction
qPCR	**Q**uantitative **R**eal **T**ime **PCR**
MDS	**M**ulti **D**imensional **S**caling
BCV	**B**iological **C**oefficient of **V**ariation
SDE	**S**ignificantly **D**ifferentially **E**xpressed
PBR	**P**hoto **B**io **R**eactor

Chapter 1

Introduction

1.1 Photosynthesis

The process of photosynthesis, meaning "synthesis with light" in which light energy is stored by an organism in the form of organic compounds, is the major underlying mechanism supplying energy for life on earth and supports an estimated 600 to 1000 Gt of biomass (Falkowski et al., 2000). The light energy is typically derived from electromagnetic radiation in the 400 to 700 nm interval commonly known as PAR "photosynthetically active radiation", although certain organisms use light energy shifted towards lower wavelengths down to 760 nm in the near-infrared range (Chen and Blankenship, 2011). The active range that can be utilized is broadly delimited by the energy contained within the light quanta. Electromagnetic radiation of a higher wavelength in the ultraviolet range has enough energy to cause the formation of reactive oxygen species and results in UV-inhibition (White and Jahnke, 2002), while energy with wavelengths longer than 740 nm in the far infra-red range carries too little energy to excite pigment molecules (Zhu et al., 2008). The generalized equation for photosynthesis is:

$$\mathrm{CO_2} + 2\,\mathrm{H_2A} + \text{Light energy} \longrightarrow [\mathrm{CH_2O}] + 2\,\mathrm{A} + \mathrm{H_2O} \tag{1.1}$$

In which A is the electron donor, which in the majority of present day species is H_2O, however, alternative electron donors such as H_2S and AsO_3^{3-} (arsenite) are also in use in some bacterial species (Kulp et al., 2008). The electron is then transferred through a chain of protein complexes before being used to reduce NADP+ (nicotinamide adenine dinucleotide phosphate) to NADPH, providing reducing power for the cell as needed and in the process creating a proton gradient across the thylakoid membrane.

This potential in the form of a proton motive force is used to convert ADP (adenosine diphosphate) into ATP (adenosine triphosphate) by attaching an additional phosphate group to the molecule (Bolton, 1977). These molecules of ATP and NADPH store the light energy and allow the cell to allocate the energy as needed for downstream processes. The electrons that feed into this process are sourced from the splitting of water molecules in oxygenic species, which account for the vast majority of all photosynthesis (Barber et al., 1997). The light energy as shown in equation 1.1 is stored in organic compounds, which in turn provides the building blocks for life. The highly simplified picture drawn by the equation masks the complex steps that enables the process. In reality, the capture of energy is spatially and sometimes temporally separated from the fixation of inorganic atmospheric carbon into complex carbon molecule. While spatially separated, both processes takes place in the cellular compartment provided by the chloroplast in eukaryotic cells.

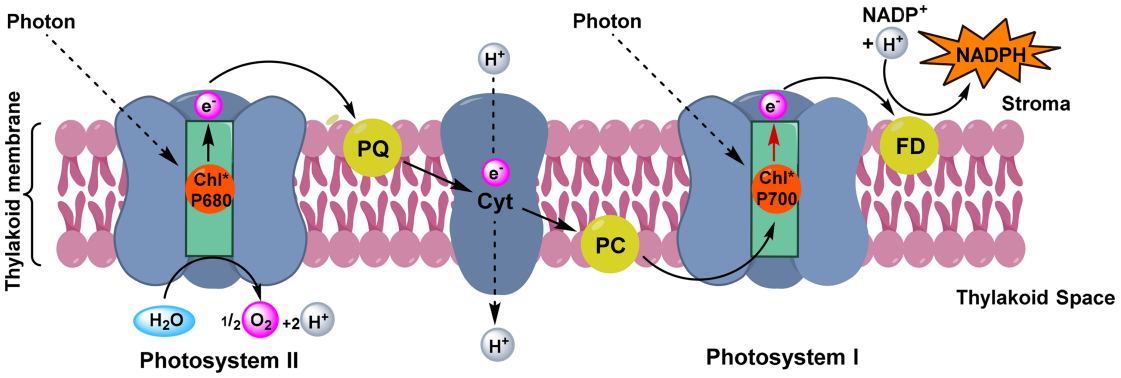


FIGURE 1.1: The complexes directly involved in the electron transport chain in oxygenic photosynthesis, showing the excited reaction centre chlorophyll of PSII (Chl*/P680), plastoquinone (PQ), cytochrome complex (Cyt), plastocyanin (PC), the excited reaction centre chlorophyll of PSI (Chl*/P700) and ferrodoxin (FD) Figure from Senge et al. (2014).

These complexes (see Figure 1.1) are not directly attached to each other but are instead linked through carrier molecules that transfer electrons between the complexes in the case of photosystem II (PSII), Cyt b_6/f and photosystem I (PSI). The electron transfer chain formed in oxygenic photosynthesis together with the increase in oxidation-reduction potential of the electrons passing through the chain after each subsequent excitation evening in PSII and PSI is usually referred to as the z-scheme (Senge et al., 2014).

The first part of the process, the capture of energy is commonly referred to as the light reaction since it needs light to proceed. The key protein complexes and reactions of this process are all membrane bound and located in either the thylakoid membrane of the chloroplast in eukaryotic cells or in the homologue membranes of prokaryote

cells. The system consists of two photosystems with associated light harvesting complexes, responsible for the capture of light, an intermediary complex, Cyt b_6/f, responsible for creating a proton gradient and finally a pump, ATP-synthase (ATPase) that uses the gradient to store energy in the form of ATP. The system operates in a series of connected processes starting with PSII that contains a special pair of chlorophyll molecules in the centre (P_{680}) with specific properties due to molecular interactions with the surrounding matrix. When a quanta of light energy is captured by surrounding pigment molecules it is transfered in the form of an exciton until it reaches the special pair, also known as the reaction centre (Pessarakli, 2001). The energy excites an electron in the special pigment pair, which is donated to an acceptor consisting of a pheophytin molecule. This redox reaction results in the creation of a $P680^+$ and a pheophytin$^-$ pair. The electron is very quickly (10^{-10}s) passed on to another donor in the form of a bound quinone (Q_A), before subsequent transfer to plastoquinone (Q_A). The electrons lost in the process are replaced by electrons scavenged from the splitting of water and supplied by the oxygen evolving complex (OEC) situated just below the special pair. The oxidised $P680^+$ supplies the reducing power via an intermediary in the form of a tyrosine side chain situated on the D1 protein that forms part of the core of the photosystem and interconnects the OEC and the special pigment pair (Nelson and Yocum, 2006). The reaction can be summarized as:

$$2\,H_2O \longrightarrow O_2 + 4\,H^+ + 4\,e^- \tag{1.2}$$

To balance the equation shown in 1.2, four electrons need to be extracted, however, the reaction centre only provides enough reducing power to sufficiently reduce one electron at a time, which necessitates the need for a system that can temporarily store reducing power until enough energy is provided for the full reaction. This is provided by four manganese ions that form the core of the OEC, together with interactions from surrounding Ca ions. Each cycle, or S state, results in the reduction of one of the manganese ions, followed by release of oxygen after four cycles (Ferreira et al., 2004). Since the OEC is situated on the lumenal side of the thylakoid membrane the process does not only supply electrons for the photosystem but the release of the protons also helps to establish and maintain a proton gradient across the thylakoid membrane.

Plastoquinone, a membrane-soluble benso-quinone, transfers 2 electrons (e^-) from PSII to the next complex in the electron transfer chain, the Cyt b_6/f complex which is responsible for providing the majority (up to two thirds) of the protons that drive the synthesis of ATP molecules (Hasan et al., 2013). The complex also plays an important part in balancing the ratio of NADPH and ATP through the action of cyclic

electron flow from PSI via ferrodoxin. Electrons from PSII are passed on to an iron-sulfure centre (Rieske Fe-S protein) that passes on the electrons via cytochrome f to a second carrier in the form of plastocyanin (Green and Parson, 2003). In the process, two protons extracted from the stromal side of the membrane and carried via plastoquinone (known as a plastoquinol in this reduced state) are released on the lumenal side. This is followed by a second cycle in which another plastoquinol binds to the complex and releases another electron to plastocyanin via the iron-sulfur centre, followed by regeneration of a plastoquinone through uptake of two more protons from the stromal side. In total, four protons are moved from the stroma and released into the lumen. During cyclic electron transfer from PSI, ferrodoxin passes on its electrons to plastoquinone (Barber et al., 1997).

Plastocyanin containing a copper-centre binds and transfers one electron to PSI, which similarly to PSII contains a reaction centre in the form of a special pair of pigments $P700^{+}$. Upon excitation of the pair an electron is donated to another chlorophyll molecule (A_0). Again the charge is transferred to another acceptor in the form of a phylloquinone ((A_1)), followed by transfer through three iron-sulfur centres (4Fe-4S) (F_x, F_A and F_B), before the last electron donor in the chain (F_x) reduces ferredoxin (Chitnis, 2001). The oxidised $P700^{+}$ is reduced by transfer of an electron from plastocyanin.

Life on earth depends on building blocks that use carbon as the backbone; the second part of photosynthesis, the carbon reactions (also known as the Calvin cycle), involves the incorporation of carbon from inorganic carbon dioxide into more complex organic molecules in the form of carbohydrates. This process occurs outside of the thylakoid membranes and is initiated by the enzyme D-ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCo) that catalyses the carboxylation of Ribulose 1,5-bisphosphate (RuBP), a five carbon compound, into a six carbon intermediary that is rapidly converted into two molecules of 3-Phosphoglycerate (PGA) (see Figure 1.2).

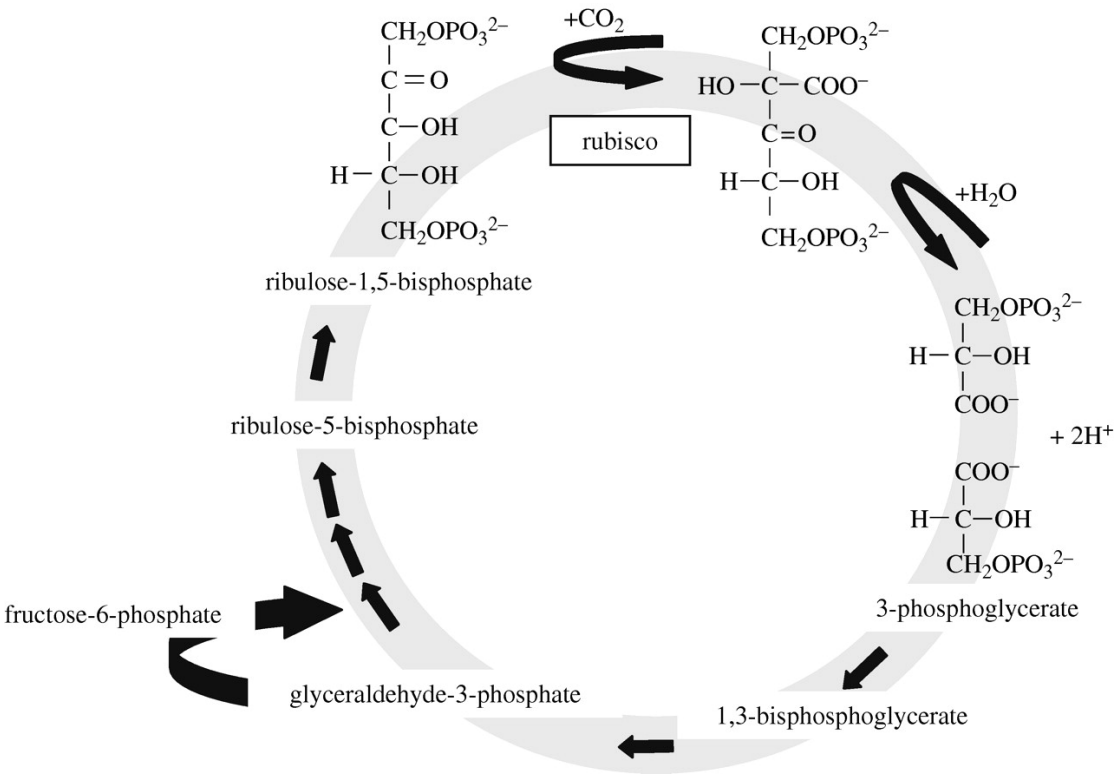


FIGURE 1.2: Simplified depiction of the key events in the carbon reactions with the action of RuBisCo (ribulose-1,5-bisphosphate carboxylase/oxygenase). Figure from (Rothschild, 2008).

The stored energy in the form of ATP is used to convert PGA to Glyceraldehyde 3-phosphate (G3P) and resulting in the release of NADP+, ADP and phosphate ions that are cycled back into the light dependent reaction. The majority (5 out of 6 molecules) of the G3P is then used to regenerate the RuBP that, together with CO_2, forms the substrate for the action of RuBisCo. The remaining G3P, together with remaining NADPH and ATP is the net gain from photosynthesis. The allocation of G3P by the cell is a dynamic process with more G3P being allocated to the re-generation of RuBisCo under low light when the total pool is small (Blankenship, 2002), compared to under higher light, when more G3P can be diverted for starch production or export from the chloroplast.

1.1.1 The photosynthetic apparatus

The photosynthetic apparatus in oxygenic species consists of four major protein complexes, photosystem II (PSII) and its associated light harvesting antenna (LHCII),

cytochrome b/6-f (Cyt b_6/f), photosystem I (PSI) and antenna (LHCI), and ATP-synthase (ATPase) that are all primarily situated within the thylakoid membrane in eukaryotic species or in the inner membrane in prokaryotic organisms.

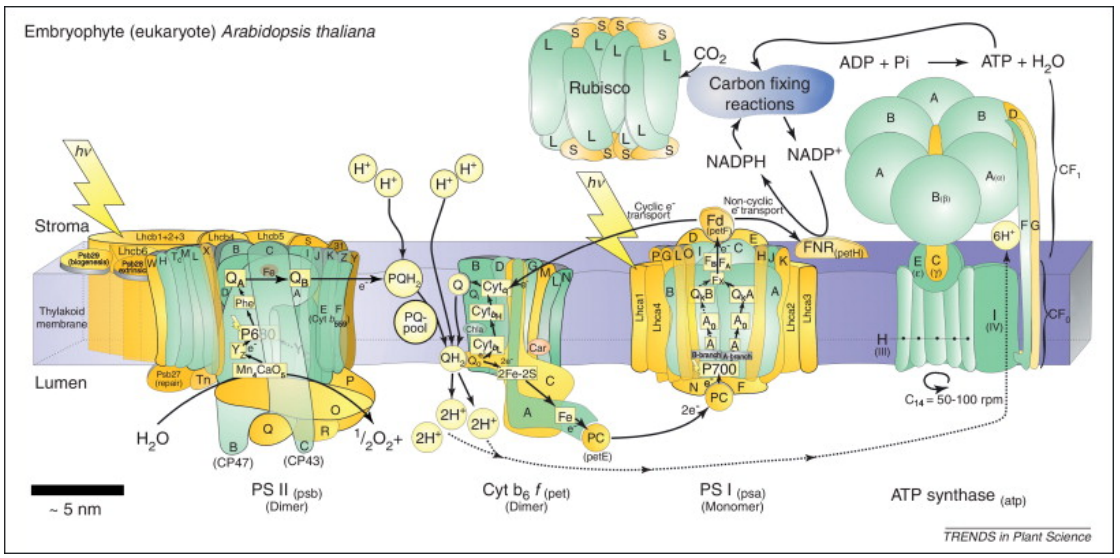


FIGURE 1.3: Overview of the photosynthetic apparatus. Subunits encoded by the nucleus are shown in yellow while subunits encoded by the plastid are shown in green. Figure from (Allen et al., 2011).

Photosystem II is a protein complex consisting of 25 or more subunits, including the core photosystem as well as the oxygen evolving complex that splits water. Two main subunits, D1 and D2, comprise the core of PSII, while three subunits (PsbO, PsbP and PsbQ) forms the OEC. Two core antenna proteins, CP43 and CP47, are also closely associated with PSII, providing a minimal intrinsic antenna. PSII is complemented by an outer antenna in the LHCII in eukaryotes or phycobilisomes in cyanobacteria and red algae. Structural analysis of photosystem I shows that it have a core consisting of 11 to 12 subunits in cyanobacteria and 14 subunits in eukaryotes, which together with LHCI supports around 200 chlorophyll molecules, in addition to 3 FE-S clusters and 2 phylloquinones (Ben-Shem et al., 2003, Chitnis, 2001). Similar to PSII, photosystem I has a core of two homologue proteins (PsaA and PsaB) (Chitnis, 2001) with PsaC, PsaO, PsaH and PsaE situated on the stromal side and providing the docking site for ferredoxin, while PsaN forms the plastocyanin docking site on the lumenal side in eukaryotes. Nearly all the co-factors with the exception of 2 of the iron-sulfur clusters that are bound by PsaC and some of the chlorophyll, are co-ordinated and bound by PsaA and PsaB (Chitnis, 2001). The remaining subunits are situated within the thylakoid and are involved in influencing the co-factors, binding the remaining chlorophyll molecules or providing binding sites for other proteins that associate with PSI (Chitnis, 2001, Fromme et al., 2001).

The proteins of LHCII and LHCI are related and form a family consisting of at least 6 members in LHCII (Lhcb1 to Lhcb6) and 4 members in LHCI (Lhca1 to Lhca4). These apoproteins, together with associated pigments, form 4 subcomplexes in LHCII (LHCIIa through c) and LHCIa and b in LHCI. Lhcb1 and Lhcb2 are the most common and form trimers, while Lhcb4-6 are present as monomers (Grossman et al., 1995).

1.1.2 Pigments and light-harvesting strategies

The majority of the pigments in photosynthesising organisms are maintained in light harvesting complexes, specialized pigment complexes organized on a protein backbone that fine tune the performance of the pigments through molecular interactions and allow for spatial organisation. Light harvesting complexes are cheap to produce and maintain compared to the core photosystem and are used to collect energy of many different wavelengths and shuttle it to the reaction centre (Croce and van Amerongen, 2014).

Table 1.1: Pigments associated with light harvesting complexes and photosystems. The peak absorption maxima for each pigment is shown with its associated pigment family and the groups of organisms that utilizes it. Modified from (Grossman et al., 1995).

Pigment family	Pigment	Absorption maxima (nm)	Phylogenetic group
Chlorophyll	Chlorophyll a	662, 430	All groups
	Chlorophyll b	644, 430	Land plants, Chlorophyta, Euglenophyto, Prochlorophyta
	Chlorophyll c	628, 578, 444	Chromophytic algae, Cryptophyta, Pyrrophyta, Prochlorophyta
Carotenoids	Beta-Carotene	449, 477	All groups
	Lutein	443, 472	Land plants, Rhodophyta
	Zeaxanthin	451, 479	Land plants
	Fucoxanthin	438, 451, 469	Chromophytic algae
	Peridinin	475	Pyrrophyta
Phycobilins	Phycoerythrobilin	535-567	Cyanophyta, Rhodophyta, Cryptophyta
	Phycocyanobilin	590-670	Cyanophyta, Rhodophyta, Cryptophyta
	Phycourobilin	498	Cyanophyta, Rhodophyta

While chlorophylls are found in all groups of photosynthetic organisms, they are complemented by phycobilins in cyanobacteria (Cyanophyta) and red algae (Rhodophyta),

which allow them to utilize a wider range of the available light, especially in the green band.

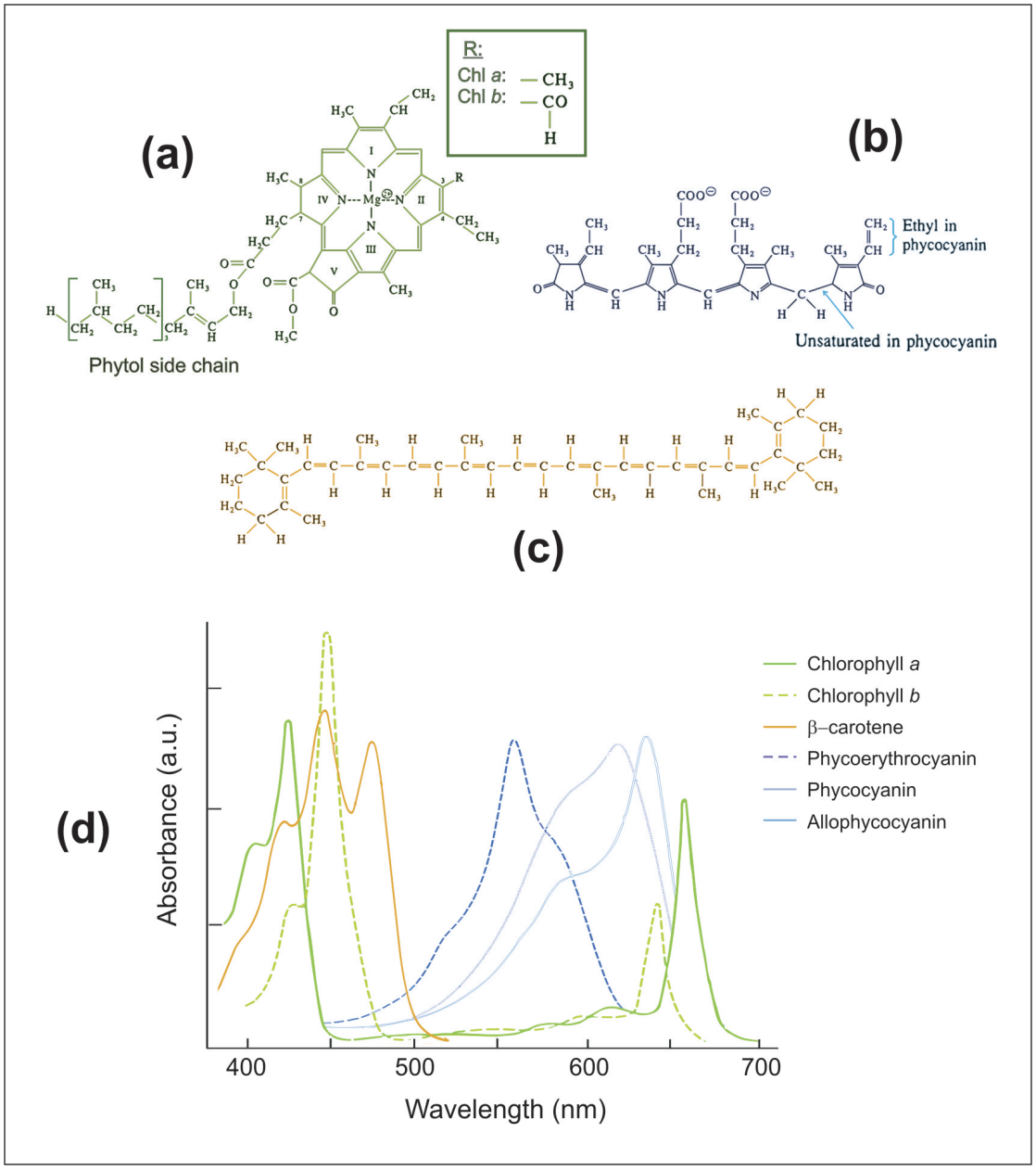


FIGURE 1.4: Photosynthetic light harvesting pigments. (a) Structure of chlorophylls a and b. (b) The structure of allophycocyanin and phycocyanin. (c) The structure of β-carotene. (d) The absorbance spectra of several light harvesting pigments. Adapted from (Hall and Rao, 1994)

Phycobilins are linear tetrapyrroles that are covalently attached to the apoprotein backbone of the LHC matrix, while chlorophylls are cyclic tetrapyrroles that are non-covalently and easily detached from the apoproteins (Grossman et al., 1995). The phycobilins are maintained in phycobilisomes, LHC's peripherally located and attached

to core photosystems on the stromal side of the thylakoid membrane, in contrast to the LHCII and LHCI that are integral to the thylakoid membrane. In addition, the chlorophyll a/b binding (CAB) proteins forming the protein backbone in the LHCs bind carotenoid components (including lutein) that help dissipate excess light energy as part of the xanthophyll cycle (discussed in section 1.1.4) (Horton et al., 1996). Together these pigments cover the whole light spectra (see Figure 1.4) including the so called 'green gap' in the 500 to 600 nm range (Croce and van Amerongen, 2014).

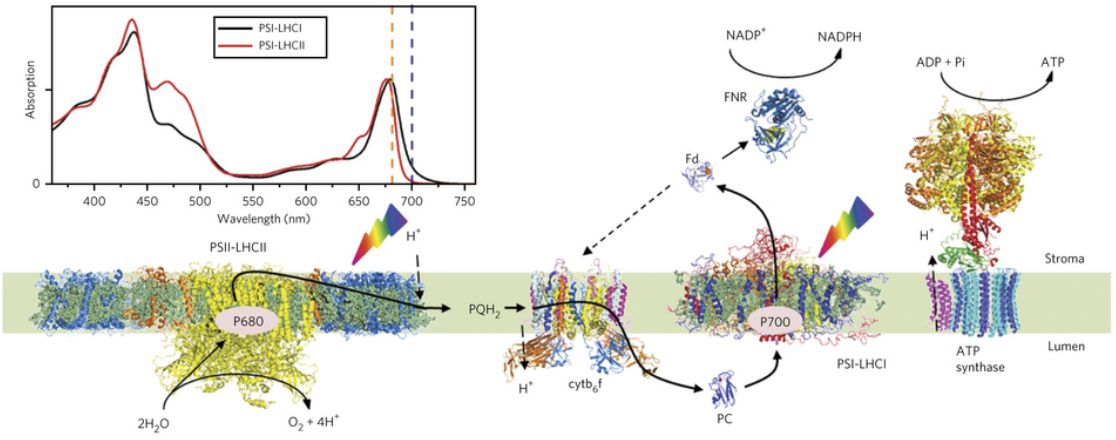


FIGURE 1.5: Schematic representation of the key components in the light reaction in green algae and higher plants. Energy is absorbed by the light harvesting complex of PSII (PSII-LHCII) and channelled to the reaction centre of PSII (P680), which source electrons from the splitting of water using the oxygen evolving complex (below P680). Reduced electrons are transferred to the cytochrome complex ($cytb_6f$) using plastoquinole (PQH_2) and used to increase the electrochemical potential through transfer of protons (H^+) from the stroma to the lumen. Plastocyanin (PC) another electron carrier transfer the partially reduced electrons to the reaction centre of PSI (P700), where following a second excitation event involving the light harvesting complex of PSI (PSI-LHCI), the electron is ultimately donated to ferrodoxin (Fd) and used to reduce $NADP^+$ to NADPH through the action of ferredoxin- NADP-reductase (FNR) or alternatively, through cyclic electron flow returned to $cytb_6f$. The electrochemical potential established is used to drive the synthesis of ATP using the ATP synthase complex. The absorption spectra of the LHC of each photosystem is shown, together with the absorption wavelength of each reaction centre. Figure from (Croce and van Amerongen, 2014)

Figure 1.5 shows a representation of the key components of the photosynthetic apparatus with each photosystem and its peripheral light harvesting antenna within the thylakoid membrane, as found in higher plants and green algae.

1.1.3 The evolution of photosynthesis

The exact evolution of the key processes of photosynthesis is naturally hard to decipher from historical evidence. Out of three domains of life, photosynthesis is present

in bacteria and eukaryotes but not archeans (Björn and Govindjee, 2008). It is also commonly accepted that the chloroplasts that are present in eukaryotes originate from bacteria via symbiosis with a cyanobacteria. This strongly suggests that photosynthesis originally evolved in a bacterial ancestor. Current knowledge suggest that anoxygenic photosynthesis operating using reductans that were present in large amounts in the early atmosphere has operated since at least 3.4 Ga years ago in the early Archaen epoch (Björn and Govindjee, 2008). Bacteria carrying out anoxygenic photosynthesis can be dividied into two groups, one including the green sulphur bacteria and heliobacteria and another one containing green nonsulfur and purple bacteria. Both of these groups utilize only a single photosystem, with the first groups resembling PSI and the second PSII but without the OEC (Björn and Govindjee, 2008, Rothschild, 2008). The carbon reactions of the process likely developed around this time or even earlier with evidence of carbon fixation from as early as 3.8 Ga years ago, with RuBisCo largely influencing the greenhouse effect of the atmosphere from around 2.9 Ga years to present (Nisbet and Nisbet, 2008) with definite large scale oxygenic photosynthesis (i.e. the great oxidation event) occurring approximately 2.3 to 2.2 billion Ga years ago, resulting in the oxygenation of our atmosphere (see Figure 1.6). However, it is likely that oxygenic photosynthesis had already operated for a considerable time (300-500 Myr) resulting in a lower but stable oxygen level (Goldblatt et al., 2006, Planavsky et al., 2014).

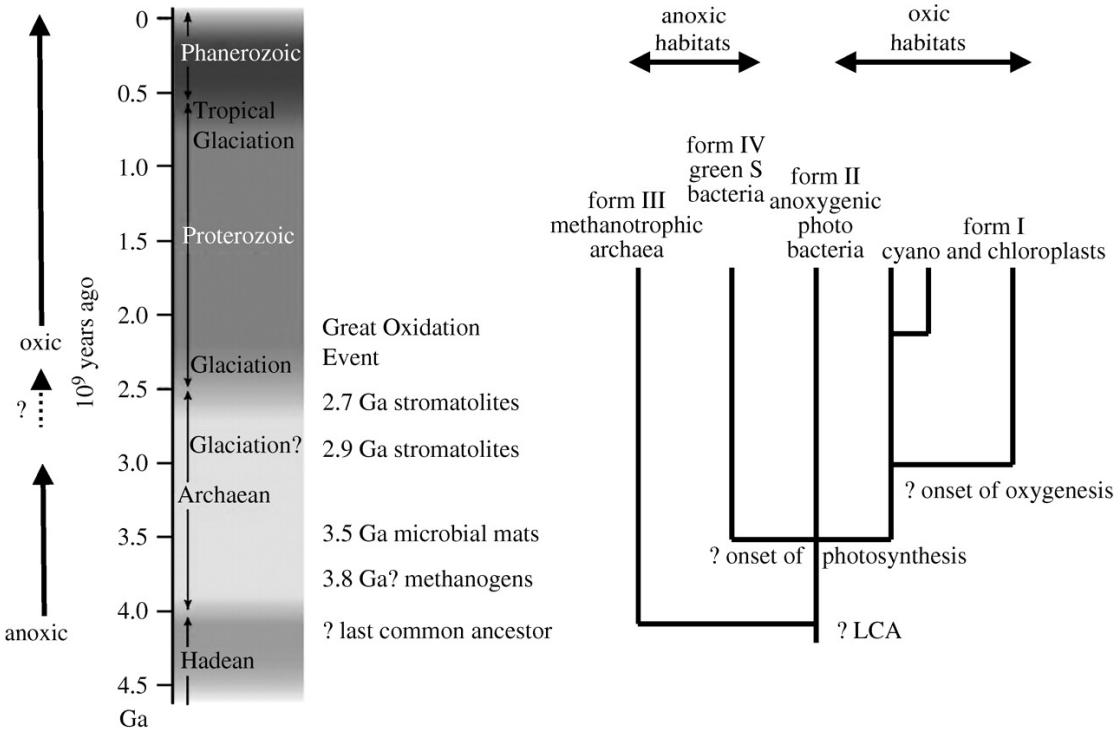


FIGURE 1.6: The key events in the evolution of photosynthesis as shown on a geological time-scale and phylogenetic time scale. Figure from (Nisbet and Nisbet, 2008).

Whether the early oxidation of the atmosphere followed the evolution of cyanobacteria as has been proposed is currently still unclear (Rasmussen et al., 2008), however, since oxygenic photosynthesis is only performed by this group of bacteria and the descendent chloroplasts in eukaryotes, large scale oxygenation must be linked to the development of cyanobacteria. Oxygenic photosynthesis differs from early anoxygenic photosynthesis in that it uses two photosystems operating in a series and use water as the electron donor (Hill, 1937, Hill and Bendall, 1960).

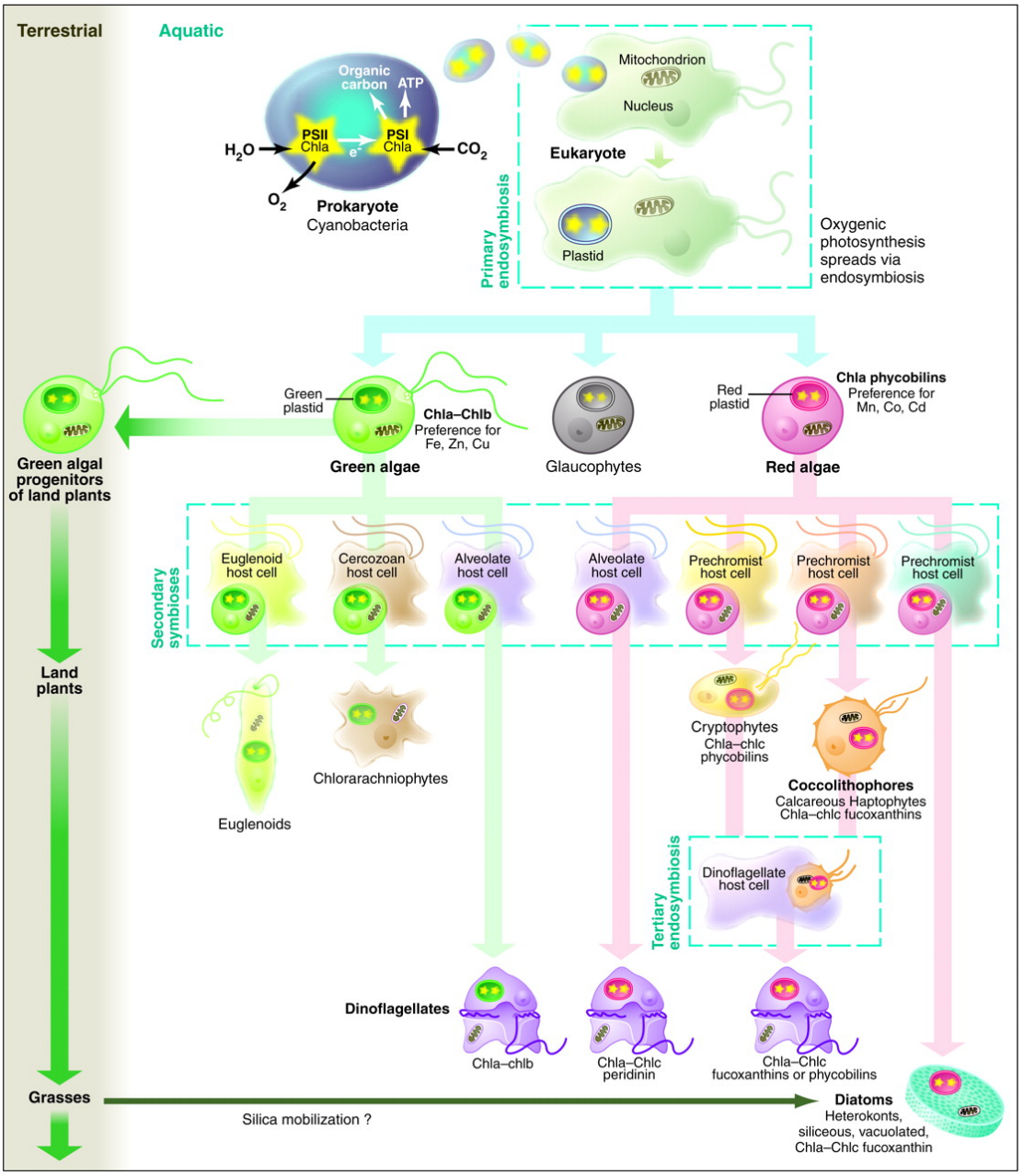


FIGURE 1.7: Primary endosymbioses of an ancestral cyanobacterial cells are thought to have given rise to three groups, green algae, glaucophytes and red algae. Secondary endosymbioses of cells from these groups after further diversification resulted in the euglenoids, the chlorarachniophytes and a small group of 'green' dinoflagellates, while red algae gave rise to cryptophytes and coccolithophores, that together through tertiary endosymbioses gave rise to another group of dinoflagellates. Figure from (Falkowski et al., 2004).

The primary endosymbioses of a cyanobacterial ancestral cell by a eukaryotic host, which in turn already contained a mitochondrion (resulting from an even earlier endosymbiosis event that may have been the origin of the eukaryotes (van der Giezen, 2011)), resulted in the first eukaryotic photoautotrophs. This early ancestor gave rise

to two primary lineages, the green lineage including the green algae (chlorophyte and charophyte) with the land plants (embryophytes) emerging from the charophytes, as well as the red lineage with red algae (rhodophytes) (see Figure 1.7). Further diversification through secondary and even tertiary endosymbiosis gave rise to the the microalgal groups (dinoflagellates, diatoms, coccolithophorids) that currently dominates the worlds oceans. These three groups all utilize plastids from the red lineage (Falkowski et al., 2004).

1.1.4 Photoacclimation

Photosynthesising organisms are either operating under light limiting or light saturating conditions depending on the quantity (light intensity) and quality (spectral distribution) of the available light at a given time. The phenotypic response to these conditions is termed photoacclimation (Falkowski and LaRoche, 1991). All algal groups are capable of photoacclimation but the range and the type of response differ between groups (Falkowski and LaRoche, 1991). Photoacclimation also occurs in plants but involves different mechanisms. Chlorophytes (i.e green algae) tend to respond to changes in light condition by changing the size of the chlorophyll antenna as well as the number and the activity of photosystems (Neidhardt et al., 1998). The chlorophyll antenna typically increases under low light and decreases under high light, while the number of photosystems similarly decreases under high light and increases under low light (Smith et al., 1990, Webb and Melis, 1995). The photoacclimation response is far more pronounced in microalgae compared to plants (Green and Parson, 2003). This is necessary since plants can adjust to changes in light using a number of mechanisms that are lacking in microalgae, including changing the density and size of the the leaf area and the thickness of individual leafs. Chlorophytes also have a stronger ability to vary their chlorophyll content in response to changing light conditions compared to dinoflagellates and diatoms (Falkowski, 1980), with a light-shade response more similar to the one seen in terrestrial plants, perhaps as an adaptation to high light conditions (Falkowski and Owens, 1980a).

Typically the chlorophyll a:b ratio increases in cells responding to high light and vice versa decreases in cells responding to low light (Smith et al., 1990, Webb and Melis, 1995). While chlorophyll a is associated with both the core antenna and the LHC of both photosystems, chlorophyll b is only associated with the nuclear encoded chlorophyll a/b binding proteins (CAB) that form the peripheral light harvesting antenna (Green, 1988, Grossman et al., 1995, Smith et al., 1990). In addition, the majority of chlorophyll b is found within the LHC II (Webb and Melis, 1995). The ratio is therefore to a degree reflecting the size of the connected peripheral antenna in the form of

the LHC, which explains why the ratio tends to increase under high light with a proportionally smaller LHC that holds less chlorophyll b, and decrease under low light when the proportional size of the LHC to the core antenna increases. Changes can also occur in a response to the spectral composition of light, one example is the complementary chromatic adaptation found in aquatic and soil cyanobacterial species resulting from changes in the ratio of two phycobilin pigments (Kehoe and Gutu, 2006, Marsac, 1977).

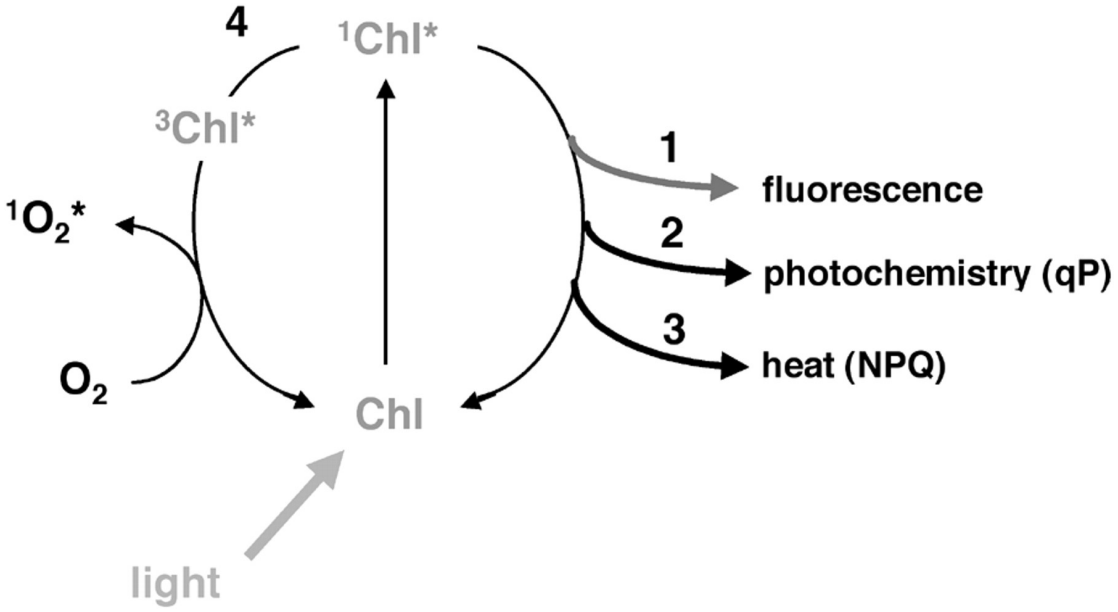


FIGURE 1.8: Energy dissipation after singlet (^{1}Chl) chlorophyll excitation, resulting in fluorescence light emission, quenching through photochemistry (qP) or as thermal emission (NPQ). Triplet excited chlorophyll (^{3}Chl) can result in the formation of reactive oxygen in the form of ($^1O_2$). Figure from (Müller et al., 2001).

In addition to remodelling of the size and pigment composition of the antenna systems, a number of short and medium term responses are deployed by cells in response to saturating light conditions or to balance the amount of light reaching the photosystems. Sensing of the light conditions is either direct using light sensitive receptors or indirect by, for example, sensing changes in the pH of the thylakoid lumen or the presence of reactive oxygen species (ROS), both increase under increasing light (Li et al., 2009). Imbalances in PSII versus PSI energy increases the reduced plastoquinone pool results in a kinase mediated phosphorylation of peripheral LHC units, allowing them to transfer from PSII to PSI. This process, known as a state-transition, helps to equilibrate and re-balance the amount of harvested light energy reaching each pool of photosystems (Niyogi, 1999).

Absorbed energy resulting in singlet excited chlorophyll (^{1}Chl) can be quenched either photochemically (qP) or non-photochemically (NPQ) via thermal dissipation using a fast component (qE) and a slower component (qI) or re-emitted at a longer wavelength as fluorescent light (Müller et al., 2001). State-transitions that result in the quenching of fluorescent light is sometimes also included in NPQ and termed qT (Ihnken et al., 2011). The fast component (qE) is triggered within seconds (or minutes) of a low pH and results in the interconversion of xanthophyll pigments within the LHC coupled with activation of the synthesis of new pigments. These changes trigger conformational changes in PSII and switch the photosystem into a quenched state with a lowered fluorescent yield but increased thermal dissipation (Gilmore, 1997). The slower component, commonly referred to as photoinhibition (qI), visible hours after a change in light, is likely a composite of different effects such as changes in the size of the xanthophyll pool (Müller et al., 2001). In addition, triplet excited chlorophyll (^{3}Chl) can result in the formation of reactive oxygen species, unwanted products that can cause photoinduced damage and a reduced photosynthetic efficiency (Niyogi, 1999). To measure and characterize the status of photoacclimation, photosynthesis-irradiance (PE) response curves are often utilized. The status of photosynthesis in terms of light limitation or light saturating at a specific irradiance can both be parameterized and compared to other organisms under the same conditions.

1.1.5 Photosynthesis in marine environments

Microalgae are the primary producers in the worlds oceans and are responsible for more than 40% of the total global net primary production (Imhoff et al. 2004; Falkowski 1998). This is even more impressive considering that they make up less than 1% of earths total biomass (Finazzi et al., 2010).

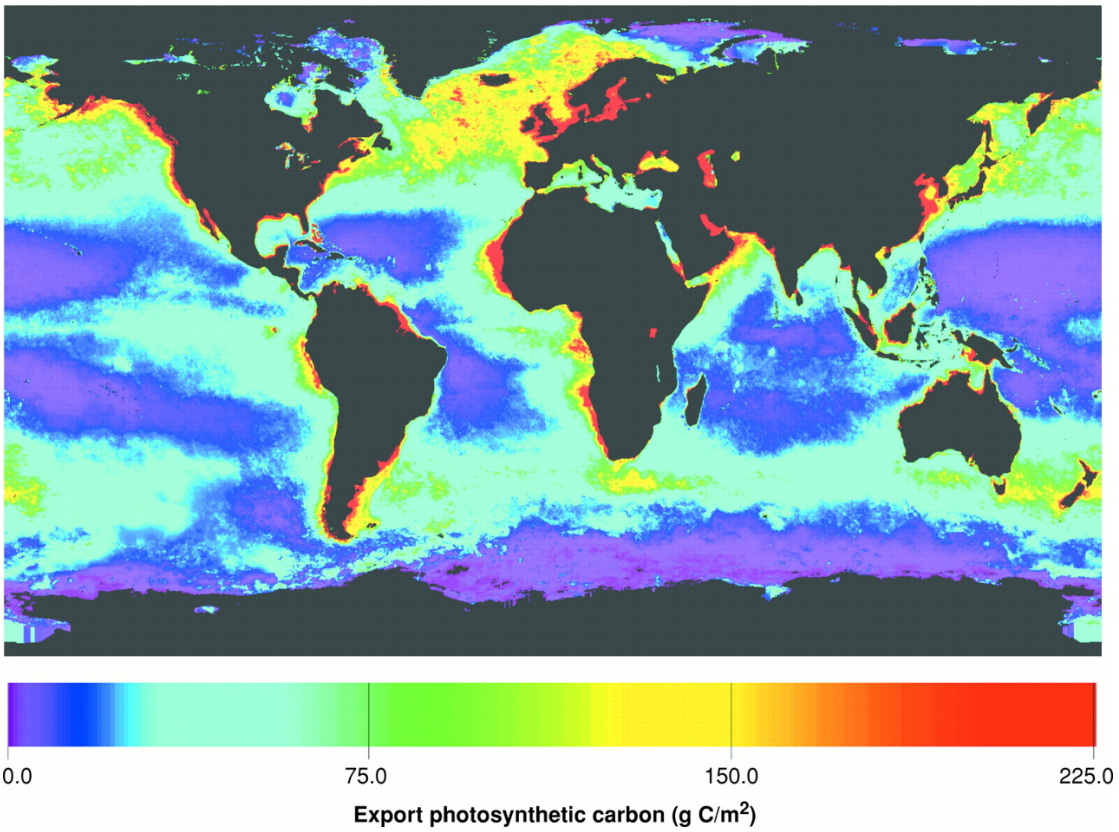


FIGURE 1.9: The estimated mean carbon export from photosynthesis in earths oceans (g C/m^2). Figure from (Falkowski, 1998).

The enormous contribution of carbon by oceanic microalgae (45 gigaton per year) via carbon fixation are central to not only the oceans biogeochemical cycles and the microbial loop, which in turn helps to feed the rest of the oceans food-chain, but also our atmospheric level of CO_2 (Falkowski, 1998, Fenchel, 2008).

The sheer number of algal species, conservatively estimated from 70 000 up to several magnitudes higher, depending on classification, together with the large variety of niches in which they are found, means that there is a tremendous genetic potential within the group (Guiry 2012). Key to understanding the diversity and potential of microalgae is the availability of sequenced genomes. While low (estimated to around 27 in 2010) the numbers are increasing and the results indicate that both the genetic diversity and the amount of species present in the ocean has likely been underestimated (Finazzi et al., 2010).

1.2 Photosynthetic species in biotechnology

Derivatives from organic material originating from photosynthetic processes still provide humanity with the majority of energy needs in the form of various hydrocarbons such as coal, natural gas and petroleum products. The latter also provide the main building blocks for further synthesis of various products such as pharmaceuticals, plastics and various organic compounds used in chemical processes.

Petrochemical products and other fossilized fuels represent stored output, both potential energy and carbon compounds, from photosynthetic processes (Vinayak et al., 2015). While the overall reserves of these fuels are still considerable, it is clear that current use is not sustainable, both from the perspective of available products but also and more importantly from an environmental perspective, since the use of these products releases large amounts of greenhouse gases in the form of CO_2, methane as well as acidification and small particles (Clarens et al., 2010, Eisentraut, 2010). Instead of using fossilized fuels that are primarily derived from microalgae it would be far more prudent to farm, harvest and process organic oils and biomass from grown microalgae (Hallman, 2007, Mostafa, 2012).

TABLE 1.2: Valuable compounds and oil content from a selection of marine microalgae. Table modified from Vinayak et al. (2015)

Phylum	Species	Oil content (% d.w.)	High value compounds
Chlorophyta	*Tetraselmis suecica*	15-32	Carotenoids, chlorophyll, tocopherol, lipids
Chlorophyta	*Ankistrodesmus sp.*	28-40	Mycosporine-like amino acids, polysaccharides
Chlorophyta	*Dunaliella salina*	10	Carotenoid, β carotene, mycosporine-like amino acids, sporopollenin
Chlorophyta	*Dunaliella tertiolecta*	36-42	Carotenoid, β carotene, mycosporine-like amino acids
Chlorophyta	*Neochloris oleoabundans*	35-65	Fatty acids, starch
Chlorophyta	*Botryococcus braunii*	29-75	Isobotryococcene, botryococcene, triterpenes
Chlorophyta	*Chlorella vulgaris*	58	Neutral lipids
Chlorophyta	*Chlorella emersonii*	34	Neutral lipids
Chlorophyta	*Chlorella protothecoides*	15-55	Eicosapentaenoic acid (EPA), ascorbic acid
Chlorophyta	*Chlorella minutissima*	57	C16- and C18-lipids
Heterokontophyta	*Nitzschia laevi*	28-69	EPA
Heterokontophyta	Thalassiosira	21-31	Glycosylglycerides, neutral lipids, TAG
Heterokontophyta	*Schizochytrium limacinum*	50-77	Docosahexaenoic acid (DHA)
Myzozoa	*Crypthecodinium cohni*	20	DHA, Starch
Ochrophyta	*Cyclotella sp.*	42	Neutral lipids
Ochrophyta	*Nannochloropsis sp.*	46-68	EPA, TAG, omega-3, LC-PUFA

Historically microalgae have been used by a number of cultures as a source of nutrients. Various cultures have used cyanobacterial species as a food-source for thousands of years (Spolaore et al., 2006). However, modern day research starting in 1950 led to initial commercial culturing of microalgae a decade later with a primary focus on supplying high value pigments and nutritional supplements from genus such as *Spirulina*, *Chlorella* and *Dunaliella* (Borowitzka, 1999). The production in 2006 of an estimated 5000 tons of dry biomass accounted for a turnover market value of £800 million before processing (Spolaore et al., 2006). Table 1.2 shows the oil content and valuable

compounds from a selection of photosynthetic microalgae that are interesting from a commercial perspective.

The large diversity present in the different groups of microalgae, has direct implications for the selection of candidates for specific applications. The presence of a cell wall improves the ability of the cells to withstand turbulent mixing, such as in a photobioreactor (which may result in reduced repeatability during experiments as shown during recent trials by de Mooij et al. (2015)), while it may prove easier to extract metabolites from species without a cell wall, such as in *Dunaliella salina*, from which β-carotene has successfully been extracted using an organic solvent from intact cells during continuous growth (Hejazi et al., 2004). See Table 1.3 for more details.

Table 1.3: The cell properties of the major groups of marine phytoplankton. Table modified from Beer et al. (2014)

Group	General	Motility	Cell wall	Storage compounds
Cyanobacteria	unicellular, filaments, some colonial	some gliding	peptidoglycan and lipopolysaccharide	cyanophycean starch
Dinoflagellates	mostly unicellular	flagellate	naked or covered in theca (cellulosic plates)	starch, some lipids
Green algae	unicellular, some filamentous or colonial	flagellate, non-flagellate	glycoprotein coat (flagellate cells), cellulose wall (non-flagellate), some naked	starch, lipids
Prasinophyceae	unicellular	motility	organic scales, some naked	starch
Pelagophyceae	unicellular	flagellate	cellulose	chrysolaminarin
Prymnesiophyceae	unicellular	flagellate and haptonema	some scales (calcite in coccolithiphores)	chrysolaminarin
Diatoms	unicellular, filamentous, small colonies	some gliding, some flagellate	silica (frustule)	chrysolaminarin, lipids

The storage compounds used by the different groups differ to a large extent, chrysolaminarin (a water soluable β-1,3-polyglucan) is used by the prymnesiophyceae (including the coccolithiophores) and the heterokonts, including diatoms and pelagopyceae, while dinoflagellates and green algae use starch for carbohydrate storage (Pulz and Gross, 2004, Radakovits et al., 2010). This is complemented by lipid storage in the green algae, dinoflagellates and diatoms and many other oleaginous algae (high oil content algae) typically in the form of triacyl glyceroles (TAG) (Hu et al., 2008), which through esterification can be converted to biofuel (Schenk et al., 2008). See Table 1.2 for the oil content of some of the commercially valuable species of marine microalgae.

1.2.1 Algal biofuels

The raw materials used for the production of biofuels have evolved over time. This has led to a practical classification based on generations. First generation biofuels are derived from edible feedstock; second generation biofuels involve various organic waste byproducts as well as dedicated energy crops grown on land that can not sustain food crops (Moore, 2008). The third and current generation mainly includes microalgae, which do not compete for agricultural land. The limited amount of available land that can sustain the growth of crops suitable for human consumption virtually eliminates the use of first generation biofuels. The second limitation is due to plants low efficiency in converting sun light to biomass, which in some cases is a magnitude lower compared to algal species (Melis, 2009). Using organic byproducts for the generation of biofuels is sound but again limited by the amount of organic material that can be produced. These limitations lead to the concept of using microalgae as the basis for biofuel production. Main advantages include their high solar conversion efficiency as well as their very high biomass turnover rate, estimated to be as high as once a week for the worlds oceans on average (Falkowski, 1998, Melis, 2009). They can also be grown either in ponds on non-arable land, estimated to cover as much as a quarter of all land or directly in the ocean in floating enclosures (Larkum et al., 2011). Since many of the species used are of marine origin there is a much lower dependency on fresh water for the production.

1.2.2 Engineering challenges

Microalgae are typically either grown in open systems such as shallow ponds or raceways or in closed systems in the form of bioreactors of various shapes. Raceways are typically operated in a continuous mode with fresh culture inoculated together with

nutrients in front of a paddle wheel that creates circulation within the system in combination with harvesting of mature algae before the paddle-wheel. Closed systems for algal growth consist of variations of photobioreactor (PBR) setups coupled with extended constructions to optimize the amount of solar radiation that reaches the algal culture. Similarly to raceways these systems use inlets and outlets. Both systems have several advantages and drawbacks. Open systems are comparatively cheap to build and maintain but sensitive to contaminants. Closed systems are considerably more expensive to maintain but allow for precise control of the internal environment and the sealed constructions minimize the risk of introduction of foreign microorganisms (Demirbas, 2011).

1.2.3 Genetic modification challenges

The energy crisis in the 1970s triggered renewed and substantial efforts by the US Department of Energy (US-DOE) to find both suitable species and methods for genetic modifications in aquatic plants, seaweed or microalgae. Initial efforts targeting waste water treatment in combination with energy production was soon focused on large scale technologies from microalgae that, factoring in the high oil content compared to the other sources, would have the potential to replace fossil fuels. This Aquatic Species Program (ASP) as it was known ran from 1978 to 1996 and contained at the peak a library of over 3000 species of algae. Many of the species and the methods used today were either developed or improved during this program, however, regardless of the substantial headway, the funding for the program was terminated in 1996 having spent a total of $25 million. The main reason was simply the understanding at the time that with current technology the price of biofuels could not compete with fossil fuels even after inclusion of a $50 carbon tax per ton of fuel. When summing up the results and providing directions for future research, the authors of the report advice laboratory should focus on naturally effective species while taking advantage of recent advantages in plant biotechnology and maximising the photosynthetic efficiency (Sheehan et al., 1998).

The techniques for genetic engineering can be broadly divided into targeted or random approaches. Directed approaches targeting specific genes rely on the availability of tools for random transformation as well as enough genetic information to identify the target. Direct manipulations of target genes have been complicated by a lack of sequenced genomes or low degrees of homologous versus random integration of inserts during nuclear transformation and diploid life stages in some algal groups including diatoms (Cerutti et al., 2011, Parker et al., 2008). The technology for direct genetic

manipulation of algal species is still being developed and is currently only available for some 30 species (Radakovits et al., 2010).

1.2.4 Random mutagenesis

The concept of using random mutations to improve strains of microalgal species for specific purposes has previously been shown to be an effective tool. Alonso et al. used ultraviolet radiation in 1996 to improve the Eicosapentaenoic acid (EPA) content of cells in the marine diatom *Phaeodactylum tricornutum* by as much as 44% (Alonso et al., 1996). Chaturvedi et al. (Chaturvedi and Fujita, 2006, Chaturvedi et al., 2004) used chemical mutation to improve the EPA content in the eustigmatophyte *Nannochloropsis oculata*, while Mendoza et al. working with *Dunaliella salina* cells suggested the use of flow cytometry for the screening and selection of mutant cells (Mendoza et al., 2008). More recently, Thai et al. used chemical mutations coupled with fluorescence-activated cell sorting (FACS) and flow cytometry (FCM) to successfully increase the palmitoleic fatty acid content of *N. oculata* cells by 30% (Thai et al., 2012). While that study was the first use (to the authors knowledge) of this approach in marine microalgae, chemical mutation in conjunction with FACS was used as early as 1992 to select cells overproducing the antibiotic Gramicidin S in the bacteria *Bacillus brevis* (Azuma et al., 1992), and more than 30 years ago in the yeast *Kluyveromyces lactis*, to select mutants deficient in mannoprotein biosynthesis related enzymes (Douglas and Ballou, 1980). While these studies showed that random mutagenesis can be used successfully to improve the outcome for a broad number of specific applications in a wide number of taxonomic groups, the most important aspect from the perspective of opening up more microalgal species for biotechnological applications, is that in contrast to direct manipulations this technique does not rely on the availability of published genomes.

1.3 Improving the photosynthetic efficiency

While specific processes in photosynthesis can operate at near 100% efficiency, the overall conversion rate in terms of converting light energy into useful carbon based compounds is often below <1%. The calculated maximum conversion efficiencies for captured light for C3 and C4 plants at 4.6 and 6.0% sets the upper limit for any improvement that operates within the current boundaries of photosynthesis (Zhu et al., 2008). The peak measured capacity in terrestrial plants rarely exceeds 3-4% and dropping by a further percent over a growing season, while higher numbers approaching

the upper theoretical limit have been been observed in microalgae under highly controlled conditions (Ort and Melis, 2011). A number of possible improvements have been suggested, including increasing the PAR range by utilizing pigments covering a wider spectra, which at 400-700 nm covers only 43% of available energy in sunlight at ground level or 48% if including light up to 740 nm (Kruse et al., 2005, Zhu et al., 2008) as well as reducing energy loss during carbon fixation by minimizing the ratio of oxygenation to carboxylation, either by incorporating C_4 systems or by the use of other carbon concentrating mechanisms (CCM) that lower the amount of photorespiration (ranging from 10 to 28% in marine microalgae) (Birmingham et al., 1982, Murchie et al., 2009). Perhaps the most obvious target for reducing energy losses lies in minimizing the overabsorption of light by individual cells, which can account for up to 80% of the harvested light (Polle et al., 2002). This excess energy, which is mainly dissipated using NPQ mechanisms or as fluorescent light, could improve the photosynthetic efficiency of the system as a whole by as much as 300% (Murchie et al., 2009, Stephenson et al., 2011).

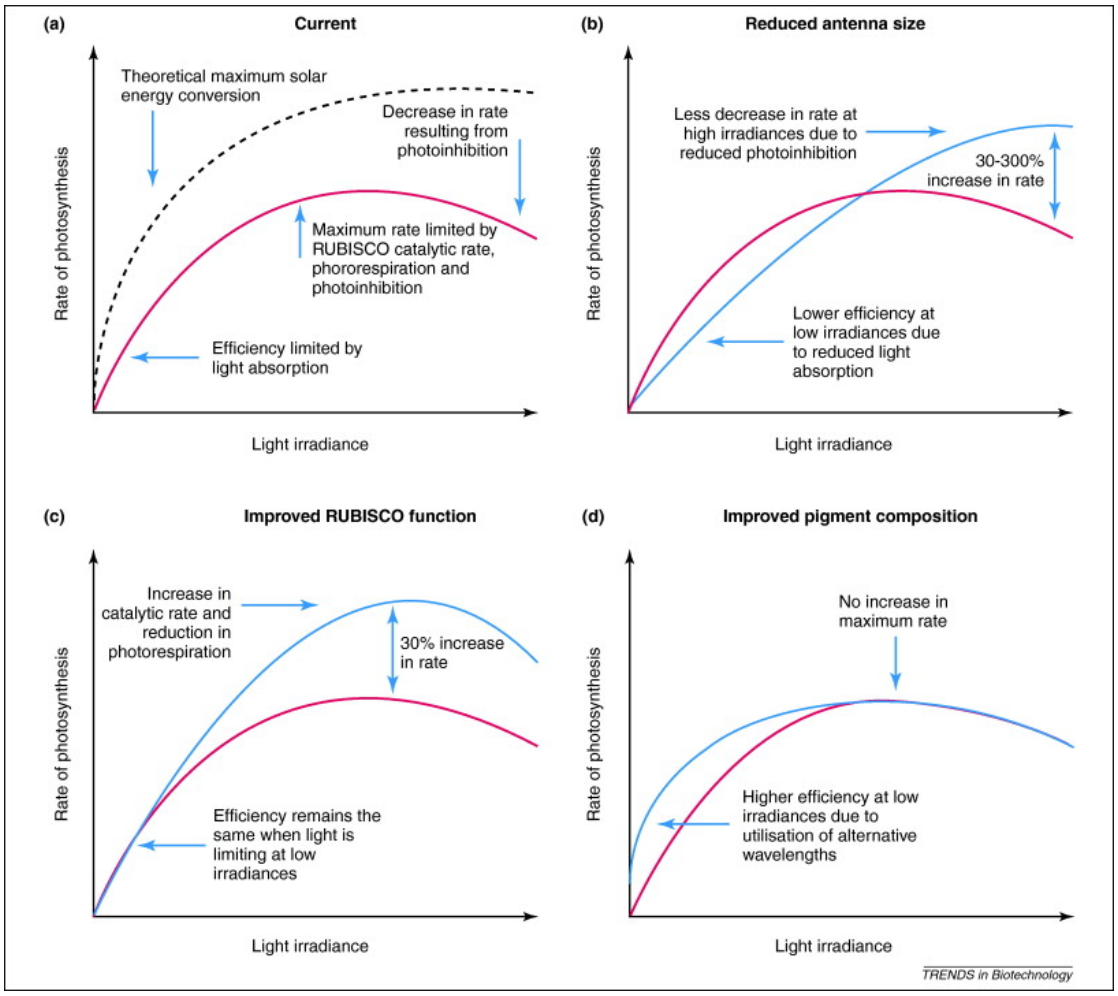


FIGURE 1.10: Photosynthesis irradiance (P/I) curves showing the potential for improvement by comparing the current performance (red line) of photosynthetic organisms with the estimated theoretical potential (dashed line). A) The combined maximum improvement in photosynthetic conversion efficiency, from B) a reduction in antenna size, C) an improved catalytic rate and reduced photorespiration in RUBISCO, D) improvements in the pigment composition. Figure from (Stephenson et al., 2011).

Table 1.10 shows possible improvements in the rate of photosynthesis in relation to the light intensity. Improvements in the function of RuBisCo through further advances in the use of directed protein evolution in the last decade (Mueller-Cajar and Whitney, 2008) have increased the chances of finding an improved version of the enzyme with either a higher catalytic rate or a higher specificity for carbon dioxide versus oxygen (thereby lowering the amount of photorespiration). Recently, it was shown that a *Synechocystis* RuBisCo mutant developed using this method did show an actual improvement in the photosynthetic efficiency of 55% per RuBisCo, although the production of the enzyme was impaired to a certain extent (Durão et al., 2015). These

results also highlight the need for careful considerations of the source to sink relationship within individual cells. Studies have shown that even a reaction such as photoinhibition, generally considered to be deleterious to the cell, can be useful under specific scenarios, including stress reduction and as a source of metabolites (Wingler et al., 2000). This suggests that the incorporation of systems biology that traces the allocation of energy and carbon compounds from source to sink may be essential to successfully improve the yield of biomass, through improvements in the photosynthetic efficiency.

1.4 Aims of thesis

In this thesis we characterize two novel mutant cell-lines from the green algae *Dunaliella tertiolecta* previously generated by random mutagenesis and single cell sorting using flow cytometry. The mutant cell lines were selected for a reduced chlorophyll content and maintained cell size.

The aim of this thesis was to determine if this pipeline had selected cells according to the selection criteria, to assess the possible improvements in the photosynthetic conversion efficiency, resulting from the reduction in chlorophyll content, and to determine if these possible improvements had resulted in realisable improvements in the yield of biomass during growth under simulated real-world conditions.

Chapter 1 provides an introduction to photosynthesis in marine phytoplankton and justifies the need for improvements in the photosynthetic conversion efficiency.

Chapter 2 summarizes the techniques and methods used during this thesis, and also provides a summary of the development of the pipeline used for the selection of these mutant cell lines.

Chapter 3 details the characterisation of the photophysiology in the lca strains compared to the wild-type under different light levels, focusing on differences in the photoacclimation response in terms of the size of the light harvesting antenna, its pigment composition and the ratio of photosystems. This is complemented by measurements of the electron transport rate and the oxygen evolving capacity to help relate these changes to any possible improvements in the photosynthetic conversion efficiency. The objective of these measurements served to establish, if present, the existence of actual significant differences in physiology between the lca strains and the wild-type.

Chapter 4 outlines the specific difficulties involved in analysing multi-factorial transcriptome data from a species that lack a published genome. It provides an overview

of the rationale behind the selection of the statistical packages and the parameters used during the analysis in relation to the quality of the generated data. The main objective of this chapter was to develop an analysis pipeline that allowed for the mapping of short reads from SAGE transcriptome libraries to longer reads from RNA-seq libraries to enable reliable annotation of gene function.

Chapter 5 presents the results from the transcriptome analysis. Enrichment analysis is used extensively to identify changes to specific cellular components or functions based on gene ontology, followed by a presentation of the identified changes in gene expression in regulatory and photosynthesis related pathways. The objective of this chapter was to identify a link between the observed photophysiology and measured gene expression if present.

Chapter 6 synthesises the results from the characterisation of the photophysiology and the molecular gene expression analysis, and relates the outcome to measurements of the productivity. In addition, specific observations of interest is presented together with recommendations for future studies based on the results.

Chapter 2

Material and Methods

2.1 Description of *Dunaliella tertiolecta*

Samples of *Dunaliella tertiolecta* were originally ordered from the NCMA culture collection (formerly known as CCMP) under the name *Dunaliella tertiolecta* CCMP364. This strain was isolated from the Oslo fjord in Norway as early as 1928. The salinity in the fjord varies depending on the depth and location, but generally the brackish surface layers have a salinity below 16 ppt while the deeper layers reaches the typical salinity of oceanic seawater (Alve and Nagy, 1986). The genus of *Dunaliella* was first described in 1905 by Teodoresco in an article describing the holotype species *Dunaliella salina* which is highly halotolerant, as suggested by its name, and also known for its ability to accumulate very high concentrations of β-carotene (Ramos et al., 2011, Teodoresco, 1905). The ability to produce large amounts of protective pigments is shared with other members of the genus including *D. Bardwelli* and *D. tertiolecta* and is currently exploited for commercial reasons in *D. salina* (Barredo, 2012, Ben-Amotz and Avron, 1983, Pisal and Lele, 2005). Halotolerance and the ability to adjust the inner osmotic pressure by producing large amounts of glycerol and the lack of an outer rigid cell wall are other common denominators for the genus (Goyal, 2007a,b, Oren, 2005).

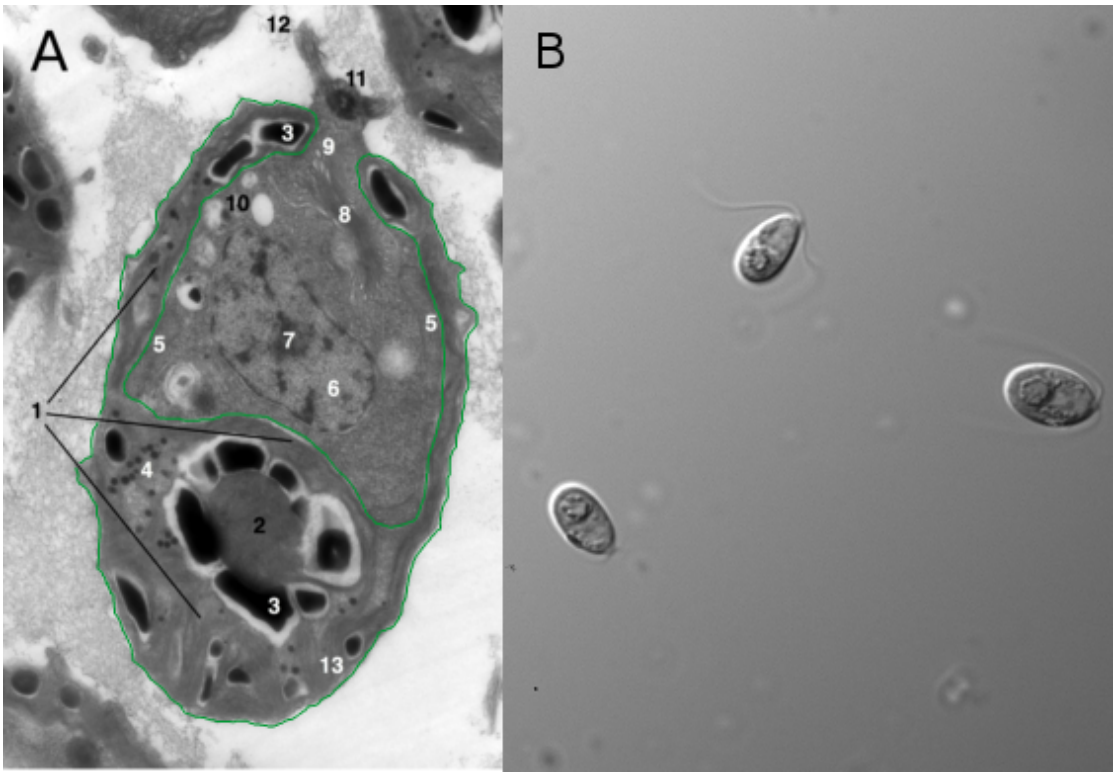


FIGURE 2.1: a) Electron micrograph image of *D. tertiolecta*. Showing a single large cup-shaped chloroplast (1) which has been outlined in green, at the basal end with stacked thylakoid membranes (13) throughout the organelle. The chloroplast contains a single oval shaped pyrenoid (2) surrounded by an amylosphere of starch grains (3) and numerous small plastidic droplets containing carotenoid pigments. The inner upper part of the cell contains the nucleus (6) with nucleolus (7), multiple mitochondria (5), the Golgi apparatus with partially secreted vesicles (9) and droplets most likely containing tri-acyl glycerid (TAG) lipids (10). The most apical point of the cell has a protrusion that supports the insertion points for a pair of flagella (12). These are damaged but normally stretch for approximately one body length past the basal (posterior) end and propel the cell in the opposite direction (i.e. movement of cell the is towards the top of the image). b) Light microscope image (630x magnification) of *D. tertiolecta* cells clearly showing the flagella as well as the single pyrenoid within the cells.

2.1.1 *D. tertiolecta* cell morphology

The cells of *D. tertiolecta* are slightly elongated with an oval shape and a typical size of 6 to 8 μm (see Figure 2.1 b). The cell surface is smooth and even in healthy cells due to the lack of an outer cell wall. The colour is light green due to a single cup shaped chloroplast that envelopes large parts of the inner cell membrane and fills up the majority of the internal space in the distal part of the cell (see Figure 2.1 b). Within the chloroplast the dominating feature is the presence of a single large pyrenoid, a highly localised region in the stromal space containing the majority of the cells ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) enzyme used for the fixation of inorganic carbon. The pyrenoid is normally clearly visible due to the accumulation of

starch sheets, forming what is known as an amylosphere, surrounding the outer surface (Borowitzka and Siva, 2007). Also identifiable are small and numerous granules or droplets containing carotenoid pigments (Borowitzka and Siva, 2007). Outside of the chloroplast the main features are the cell nucleus, and the Golgi apparatus which is involved in the distribution of various synthesized compounds including droplets containing oil (Andersson and Dörmann, 2008). The cell uses twin flagella for locomotion and these are attached to the cell at the apical end on top of a small protrusion that supports their insertion. The flagella are fairly long, stretching well past the full length of the cell.

2.2 Microbiological techniques

2.2.1 Culturing and maintenance of *D. tertiolecta* cells

Cultures of *D. tertiolecta* were continuously maintained under sterile conditions in liquid f/2 - Si media (see Table 2.1 for the medium composition) as well as on f/2 - Si agar plates (Guillard, 1975, Guillard and Ryther, 1962). The media were prepared by filtering seawater through Whatman Grade GF/A Glass microfiber (Fisher Scientific UK Ltd, Loughborough, UK) filter disks followed by autoclaving (121°C for 40 min) and cooling before addition of the media. Agar was prepared by adding 1-2% (w/v) of agar to the filtered sea water followed by re-autoclaving, cooling and addition of media. Sub-culturing was performed every second week for liquid cultures and every second month for plated cultures by transferring 50 µL of culture to 40 mL of fresh liquid media or by transferring cells from old plates to new plates by streaking using single sterile plastic loops. These stock cultures were maintained under 100 - 200 µmol photons m^{-2} s^{-1} of white light under a 12 hr light, 12 hr dark cycle. All handling of cultures was performed in a flow hood to ensure sterility. Cultures for experiments were inoculated using the same number of cells (25 thousand cells per mL of media) to minimize variance. The cultures were shaken and repositioned daily to aerate the bottles while at the same time minimizing differences in growth rate due to variations in light intensity within the light cabinet.

TABLE 2.1: Chemical composition (mg/L) and molar concentration (M) of the f/2 - Si media used for the culturing and maintenance of *D. tertiolecta cells.*

Compound	Amount (mg/L)	Molar composition (M)
$NaNO_3$	75	8.82 x 10^{-4}
$NaH_2PO_4 \cdot H_2O$	5	3.62 x 10^{-5}
$FeCl_3 \cdot 6H_2O$	3.2	1.18 x 10^{-5}
$Na_2EDTA \cdot 2H_2O$	4.4	1.18 x 10^{-5}
$CuSO_4 \cdot 5H_2O$	0.0098	3.93 x 10^{-8}
$Na_2MoO_4 \cdot 2H_2O$	0.0063	2.60 x 10^{-8}
$ZnSO_4 \cdot 7H_2O$	0.022	7.65 x 10^{-8}
$CoCl_2 \cdot 6H_2O$	0.01	4.20 x 10^{-8}
$MnCl_2 \cdot 4H_2O$	0.18	9.10 x 10^{-7}
Cobalamine (Vitamin B_{12})	0.0005	3.69 x 10^{-10}
Biotin (Vitamin H)	0.00005	2.05 x 10^{-9}
Thiamine HCl (Vitamin B_1)	0.1	2.96 x 10^{-7}

f/2 is a general media enriched in nitrogen, phosphate, various trace metals and vitamins, that supports a large range of marine algae. The media is rather rich in composition and works best for coastal species. The original recipe includes silica but this element is normally omitted when not working with diatom species since it enhances the formation of precipitates.

2.2.2 Cell numbers and growth rates

Cell numbers were counted using a Coulter Counter (Multisizer, Beckman Coulter) using 100 to 500 µL of sample together with 9.5 to 9.9 mL of diluent, depending on the cell concentration, for a final sample volume of 10 mL in total. From this diluted sample a total of three technical repeats using 500 µL of sample (volumetric) were counted using a 70 µm aperture tube. The sizing threshold was set at 1.8 µm with a noise level of 1.6 µm. The current was set to 800 µA with a gain of 4. Filtered 3% NaCl solution prepared using deionized water was used as a diluent and electrolyte. The aperture was flushed before and between measurements. In addition to the total cell counts the mean size was also measured based on cells within the range selected for the total counts.

The specific growth rate (µ) was calculated by fitting a linear regression to the natural log of the measured cell count versus the time for the exponential phase (with the

slope corresponding to the growth rate) following Andersen (2005). The doubling time was calculated as the natural log of 2 divided by the specific growth rate (μ).

2.2.3 Continuous growth experiments

Estimating the biomass yield under continuous growth with replicable results is notoriously hard and for this reason a collaboration was initiated with David. E. Kramers group at the University of Michigan (MSU, Lansing, Michigan, USA) who has an established program and laboratory setup specifically designed to allow for replicated growth experiments in photobioreactors. To better estimate the biomass yield under real-world conditions they use so called environmental photobioreactors (ePBRs) that can mimic real-world conditions to a certain extent, the light source is from the top of culture and the PBR can be programmed to follow specific light and temperature conditions. The setup which is highly controlled and replicated using a matrix of reactors connected to a single controlling computer enables realistic comparisons of the productivity of different strains (see Figure 2.2).

FIGURE 2.2: Environmental photobioreactors with cultures set up in a matrix allowing for replication of experimental conditions between strains. The light source is situated at the top of the culture cylinder creating a light gradient within the culture with decreasing light from top to bottom.

During the experiment the ePBRs were run in a turbidostat mode with injection of fresh medium when the optical density reached a pre-specified threshold. This threshold, based on a optical density designed to match a specific chlorophyll content resulting from the combined chlorophyll per cell and the cell density needs to be carefully

selected to result in optimal light penetration. The productivity of the system is finally calculated from the measured ash-free dry weight (AFDW) in relation to the number of dilutions. The key parameters for the setup are shown in Table 2.2.

TABLE 2.2: Experimental conditions used during the first and second productivity experiments in environmental photobioreactors.

Experiment	First	Second
Sinusoidal light curve	Yes	No*
Peak light intensity (µmol photons m{-2} s{-1})	2000	2000
Day length (hr)	14	14
Temperature control	No	No
Measure turbidity every (min)	10	10
OD Target		2 µg Chl/mL
CO_2 injection	5% CO_2 in 95% Air, 1 min per hour	5% CO_2, 95% Air, 1 min per hour
Stirrer (rpm)	200	200

2.3 Determining pigment composition

2.3.1 Chlorophyll a measurements

Chlorophyll a (Chl a) content was routinely measured by filtering 1.0 mL of liquid culture using a hand held pump on to MF 100 microfibre filters (Fisher Scientific UK Ltd, Loughborough, UK). The filtered content was re-suspended in 4 mL of 90% acetone in plastic tubes under low light and then immediately covered in aluminium foil. The extracted dissolved chlorophyll was kept in the cold over night (4°C) and analysed on a Turner Designs TD-700 fluorometer (Turner Designs, Sunnyvale, CA, USA) using the methods of Welschmeyer (1994). Measurements of solid chlorophyll standards were used to re-calibrate the results.

2.3.2 High performance liquid chromatography

The high performance liquid chromatography was performed according to Gibb et al. (1997). Samples from 1.0 mL culture filtered through MF 100 microfibre filters (Fisher Scientific UK Ltd, Loughborough, UK) were transferred to 4 mL of 90% acetone, in dark glass vials, then sonicated for 30 s using a Ultrasonics W-380 sonicator with a C3 probe (Heat Systems Ultrasonics) set at 30% of maximum strength,

and then stored in the dark at 4°C until use. Approximately 1.0 mL of the prepared sample was filtered and loaded onto a SpectraSystem AS3000 autosampler (chilled to 4°C) together with a chlorophyll a standard and a mix of pigment standard. The samples were mixed 1:1 (v/v) with 1M ammonium acetate before injection into a SpectraSystem P2000 HPLC system coupled with a SCM1000 de-gasser (Thermo Fisher Scientific, San Jose, CA, USA). Separation was performed using a C-8 column using a linear gradient of solvents A (70% methanol and 30% 1 M ammonium acetate) and B (100% methanol) at a flow rate of 1.0 mL per min and programmed as previously described in Barlow et al. (1997). Absorption between 366 to 798 nm was measured on a UV-Vis photodiode array detector (UV6000LP, Thermo Fisher Scientific, San Jose, CA, USA) at 2 nm intervals. Peaks were identified using absorbance at 440 nm based on retention times in combination with spectral libraries as needed. The amounts of individual pigments were calculated from the integrated peak area using the ChromQuest software suite (version 4.0, Thermo Fisher Scientific, San Jose, CA, USA).

2.4 Fluorescence based methods

2.4.1 Fast repetition ratio fluorometry (FRRf)

The maximum photochemical efficiency (F_v/F_m) and absorption cross-section of photosystem II (σPSII) was routinely measured using a Chelsea Scientific Instruments MkII FastTracka Fast Repetition Rate fluorometer (FRRf) (Chelsea Technologies Group, West Moseley, Surrey, U.K.). The first parameter is defined as the difference in fluorescence between a dark-adapted state (F_o) and the maximum fluorescence (F_m); this difference (F_v) is normalized using (F_m) to get the final ratio (F_v/F_m (Kolber and Falkowski, 1993) while the second parameter (σ_{PSII}) is the dark-adapted functional cross section.

Table 2.3: The settings used for measurements using the MkII FastTracka Fast Repetition Rate fluorometer.

Parameter	Setting
Sequences per acquisition	24
Sequence interval (ms)	100
Acquisition pitch (s)	3
Flashlets in sat. phase	150
Flashlet duration in sat. phase (µs)	1
Flashlets in rel. phase	25
Flashlet duration in rel. phase (µs)	84
Saturation phase duration (ms)	0.3
Relaxation phase duration (ms)	2.1
LED set	50

Samples were dark adapted for a minimum of 20 min, after which the concentrations were adjusted by diluting 1 mL of cell culture in 1 mL of f/2 -Si medium kept at room temperature before loading in the dark chamber of the instrument. Single turnover (ST) measurements were used for all measurements and using the settings in Table 2.3. The gain of the photomultiplier (PMT) was adjusted during measurements to keep the maximum fluorescence (F_m) centred around 10000±2000 (dimensionless and specific for the instrument).

Table 2.4: The fluorescence parameters with descriptions and units used for the fast repetition rate fluorometry measurements.

Parameter	Definition	Units
F_o	Minimum fluorescence in darkness	Dimensionless
F_m	Maximum fluorescence in darkness	Dimensionless
F'_o	Minimum fluorescence under actinic light	Dimensionless
F'_m	Maximum fluorescence under actinic light	Dimensionless
F_v/F_m	Maximum photochemical efficiency (in darkness)	Dimensionless
F'_q/F'_m	Maximum photochemical efficiency under actinic light	Dimensionless
f'	Proportion of functional PSII reaction centres ($F'_q/F'_m/0.65$)	Dimensionless
σ_{PSII}	Functional absorption cross-section (in darkness)	$Å^2$ (photons)$^{-1}$
σ'_{PSII}	Functional absorption cross-section under actinic light	$Å^2$ (photons)$^{-1}$
E	Irradiance under actinic light	µmol photons m^{-2} s^{-1}

In addition, the photosynthetic electron transport rate of PSII (P^f) under different light levels was estimated by running rapid light curves (RLC) under increasing light intensities from 0 to 2000 μmol photons m^{-2} s^{-1} and calculated using the equation (2.1) below following Fujiki et al. (2007) and based on Kolber and Falkowski (1993) and Gorbunov et al. (2001).

$$P^f = E \times \sigma'_{PSII} \times f' \tag{2.1}$$

The maximum photochemical efficiency parameter (F'_q/F'_m) under actinic light is derived from the difference in the maximum fluorescence yield when all PSII reaction centres are closed (F'_m) and the fluorescence (F') as measured under actinic light divided by F'_m (Butler, 1978). See Table 2.4 for additional definitions and units.

2.4.2 Pulse amplitude modulated (PAM) fluorometry

The absolute size of the functional cross section of PSI (σPSI) was estimated using pulse amplitude modulated fluorometry (PAM) as described in Ryan-Keogh et al. (2012).

Cells were grown under medium light (100 μmol photons m^{-2} s^{-1}) and harvested during the exponential growth phase on day four. The total sample volume used ranged from 40 to 120 mL of liquid culture depending on the cell concentration. The majority of the supernatant was removed using centrifugation for 10 min at 4,700 x g to bring down the total volume to 2 mL and concentrate the sample, which helps to increase the signal to noise ratio. Measurements were taken at three different concentrations, ranging from neat (1:1) to 1:2 and 1:4 dilutions at three different light intensities, 5, 10 and 20 mmol photons m^{-2} s^{-1} for a total of nine measurements per sample. The optical density of the cultures at 635 nm were measured on a Varian Cary 50 Scan spectrophotometer (Agilent Technologies UK Ltd., Wokingham, Berkshire, U.K.) and used to correct for the absorption of the saturating pulse measured using the PAM at the same wavelength. The results were analysed in MATLAB (Mathworks, Cambridge, UK) following Zipfel and Owens (1991).

2.5 Rate of oxygen evolution

Cells were grown under medium light (100 μmol photons m^{-2} s^{-1}) and harvested during the exponential growth phase on day four. The cell concentration for each single

sample was adjusted to 1 million cells per 2 mL of media. For each measurement a fresh sample was used. In addition, 100 µl of sodium bicarbonate (0.5 M solution, pH 7.4) was added to each sample (Polle et al., 2003) to minimize the possibility of carbon availability serving as a bottleneck. The change in dissolved oxygen concentration was measured at a range of 10 different light intensities from 25 to 2000 µmol photons m^{-2} s^{-1} (0, 25, 50, 100, 250, 500, 750, 1000, 1250, 1500, 2000 µmol photons m^{-2} s^{-1}). Base line respiration was measured in the dark by covering the electrode chamber in aluminium foil. Before each measurement the signal from electrode was allowed to stabilize for 3 min, followed by continuous measurements over a 2 min interval. The signal over this interval was then integrated using the Hansatech software suite. The rate of oxygen evolution per mole of chlorophyll was calculated using the molecular weight of chlorophyll a (893.49 g mol^{-1}) and the measured amount in each sample. After plotting the rate of photosynthesis (P) as measured by the rate of oxygen evolution as a function of light intensity (I), a hyperbolic tangent function (Falkowski and Raven, 1997) (equation 2.2) was curve fitted using the least-squares solver algorithm in Microsoft Excel 2010 for each replicate to give the initial slope (α) and the maximum oxygen evolution rate P_{max}, from which the light intensity at which saturation of photosynthetic oxygen production occurs I_k can be calculated (P_{max}/α) (Huesemann et al., 2009). R_{dark} is the rate of dark respiration in absence of light.

$$P = P_{max} \times tan\ h\left(\frac{\alpha \times I}{P_{max}}\right) + R_{dark} \tag{2.2}$$

2.6 Immunoblotting of proteins

2.6.1 Protein extraction

Cells were grown under medium light and harvested during the exponential growth phase on day four or five. Each sample were prepared from between 5 to 35 mL of liquid culture. The supernatant was removed using centrifugation at 4,700 x g at 20°C and the pellet containing the cells was then re-suspended in 1 mL of sample buffer and placed on ice. The sample buffer contained 2% (w/v) lithium dodecyl sulphate (LDS), 10 % (v/v) glycerol, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 141 mM Tris Base, 106 mM Tris-HCL and PefaBloc SC (AEBSF) (Diatchuk et al., 1997). The whole cell protein content was then extracted using a Ultrasonics W-380 sonicator with a C3 probe (Heat Systems Ultrasonics) set at 30% of maximum strength. The samples were flash frozen in liquid nitrogen, then sonicated for approximately 30

s while allowed to partially defrost. This process was repeated 4 times per sample after which the extract was again placed on ice. The samples were then centrifuged and the resulting supernatant containing the extracted protein was removed and stored at -20°C.

2.6.2 Determination of protein concentration

The protein concentration of each sample was measured using a Pierce BCA protein assay kit (Fisher Scientific UK Ltd, Loughborough, UK) using bovine serum albumin (BSA) standard. A set of standards ranging from 25 to 2000 µg/mL were prepared from a 2 mg/mL ampoule using sample buffer as the diluent. Following preparations according to the manufacturer protocol, 25 µL of each sample, standard and blank were set up in duplicate on a 96 well microplate. The plate was covered in plastic film and incubated at 37°C for 30 min. The absorbance was measured at 562 nm on a plate reader using three technical repeats. After blank subtraction of samples and standards, a standard curve was prepared. A four-parameter (quadratic) curve was fitted to the standards and used to estimate the concentration in the samples.

2.6.3 Electrophoresis, blotting and quantification of protein

In preparation for gel loading, the sample volume was adjusted to correspond to a total protein load of 2.5 to 5 µg depending on the protein assayed, corresponding to roughly 1.5 to 10 µL of protein extract. 1.5 µL of dithiothreitol (DTT) was then added, after which the total sample volume was adjusted to a final volume of 15 µL by adding blue loading buffer. The samples were then heated at 70°C for 3 min followed by a brief centrifugation to collect the sample before loading on the gel.

TABLE 2.5: The electrophoresis and membrane transfer run-time in minutes used during the quantification of relative amounts of PsaB, PsaC and RbcL proteins using Western blotting.

Complex	Target protein	Electrophoresis run-time (min)	Membrane transfer run-time (min)
PSII	PsaB	35	57
PSI	PsaC	32	50
RuBisCo	RbcL	35	70

The proteins were separated using electrophoresis at 200V for 32 to 35 min depending on the protein on 4-12% acrylamide mini-gels (NU-PAGE Bis-Tris gels, Invitrogen)

in a MES running buffer (Invitrogen) using the XCell *SureLock* mini-cell system (Invitrogen). The gel was then removed from the unit and covered by a polyvinylidene difluoride (PVDF) blotting membrane that had been pre-soaked in 100% methanol for 1 min followed by submerging in 1x transfer buffer, sandwiched by a single cellulose filter paper on each side before loading on a XCell II blot module. Depending on the size of the proteins the transfer was allowed to run for 50 to 70 min (see Table 2.5). Following completed blotting, the membranes were submerged in blocking solution (either 2% (w/v) ECL advance blocking agent or 2% (w/v) non-fat dry milk powder) and left overnight at 4°C to minimize non-specific binding.

TABLE 2.6: The amount of loaded (total) protein in µg and the dilution factor of primery and secondary antibody used during the quantification of relative amounts of PsaB, PsaC and RbcL proteins using Western blotting.

Protein	Protein load (ug)	Primary antibody dilution	Secondary antibody dilution
PsaB	5	1:20000	1:20000
PsaC	2.5	1:1000	1:5000
RbcL	5	1:10000	1:10000

The primary and secondary antibodies were then diluted to the final strength (see Table 2.6 for concentrations) in blocking solution. The primary and secondary antibodies for each protein were diluted and suspended in blocking solution. The blocking solution was then poured off and replaced by the solution containing the primary antibody and placed on a rocker for 1 hr. The membrane was then rinsed in TBS-T twice, followed by a wash cycle consisting of washing in fresh TBS-T and leaving on the rocker for 10 min, this process was repeated three times while leaving on the rocker for 5 min for each cycle. After discarding the last TSB-T the secondary antibody was added and incubated for 1 hr. The wash cycle was then repeated before the final quantification.

The membrane was then placed on plastic film and covered by a 1:1 Lumigen (ECL Advance) A and B solution (GE Healthcare UK Limited., Little Chalfont, UK) and incubated in the dark for 5 min. The chemiluminescence was detected using a VersaDoc 5000 molecular imager using the Quantity One software suite (Bio-Rad, Hercules, California, USA) with typical exposure times of the final images ranging from 1 to 5 s. Densitometric post-analysis of images using the Image lab software (Bio-Rad, Hercules, California, US) was used to quantify the final protein concentration in samples and standards.

2.7 Methods for molecular and genetical analysis

2.7.1 Agarose gel electrophoresis

Nucleotide fragments and extracts were routinely visualized using agarose gel electrophoresis. The standard recipe used 25 mL of 1x TAE (tris-acetate-EDTA) buffer with 0.25 g of molecular grade agar for 1.0% gels, 0.375g for 1.5% gels or 0.75g for 3% gels and 5 µL of ethidium bromide (10 mg/mL). Gels were covered with 1x TAE used as the running buffer with an additional 5 µL of ethidium bromide (10 mg/mL). For 1.0% gels 5 µL of HyperLadder I (Bioline, London, UK) was used as the molecular weight marker (MWM) or alternatively HyperLadder II for 1.5% gels or HyperLadder V for 3.0% gels. 5 µL of sample was prepared by mixing 5 µL of sample with 1 µL of loading dye and loaded per well. Electrophoresis was run at 60V for 45 min or adjusted as needed, followed by visualization using a UV-transilluminator.

2.7.2 RNA extraction and cDNA synthesis

The original RNA used for the SuperSAGE transcriptome as well as for the first qPCR verification phase was extracted using the cetyltrimethylammonium bromide (CTAB) method according to Pearson et al. (2006). Cells from 35 mL of culture were spun down using centrifugation at 3,500 g for 10 min into a pellet, followed by disruption in a Tissue Lyser II bead beater (Qiagen, Valencia, CA, USA) using 250 µL of beads at 20 RPM for 5 min with 1.5 mL of CTAB buffer (100 mM Tris-HCL (pH 8), 20 mM EDTA, 1.4 M NaCl, 2% CTAB (w/v), 2% PVP (w/v)) added per sample. Following incubation at 60°C for 30 min, 1.5 mL of chloroform:isoamyl alcohol (24:1 v/v) was added, the sample was vortexed and centrifuged at 12,000 x g for 20 min at 20°C after which the aqueous phase was transferred to a new tube. The RNA was precipitated with 0.33 volumes of 10M LiCl for 30 min at -20°C and then collected by centrifugation at 14,000 x g for 20 min at 4°C. The pellet was re-suspended and the RNA precipitated by adding 2x vol isopropanol and 0.1 vol 7.5M ammonium acetate for 1 hr at -20°C. The RNA was collected by centrifugation at 14,000 x g at 4°C. The resulting pellet was washed with 70% EtOH and centrifuged at 14,000 x g for 10 min at 20°C, followed by drying at 20°C for 10-20 min and resuspension in 50 µRNAse-free water. The extracted RNA was then stored at -80°C.

The extracted RNA was transcribed into cDNA using the ImProm-II Reverse Transcription System (Promega Corporation, Madison, WI, USA) using 4 µL ImProm-II 5X Reaction buffer, 2.4 µL $MgCl_2$, 1 µL dNTP mix (10 mM each dNTP), 1 µL recombinant RNasin ribonuclease inhibitor, 1 µL Improm-II reverse transcriptase, 1 µg of

RNA template, 0.5 µg of oligo(dt) primers and 9.1 µL nuclease-free water. The process involved annealing for 25°C for 5 min, first strand extension for 60 min at 42°C followed by heat-inactivation at 70°C for 15 min according to manufacturer instructions.

For the second qPCR-based verification of expression for selected genes an alternative method using a spin column based kit (RNeasy Plant Mini Kit, QIAGEN) according to the protocol with some modifications. For each sample, culture containing approximately 10 million cells were centrifuged for 10 min at 4,700 x g in 15 mL falcon tubes. After removing the majority of the supernatant the remaining sample was again centrifuged for 10 min at 8,000 x g in 2 mL plastic tubes. The remaining supernatant was carefully removed followed by adding 50 µL of lysis buffer (RLT) with added beta-mercaptoethanol (BME). The pellet was then resuspended by vortexing before being flash-frozen in liquid nitrogen. After removal of the tube the sample was ground up using a plastic pestle for 2 min while allowing for partial re-thawing of the sample. 400 µL of RLT with BME (for a total of 450 µL) was then added to the sample, followed by brief vortexing. The sample was processed according to the manufacturers protocol with the exception of two added alcohol washing steps using 500 µL of 96-100% alcohol post loading on the spin column, which helped improve the purity of the eluted sample. The sample was eluted into 30 µL RNase-free water, followed by reloading of the eluate on the spin column and a second elution to improve the yield. The sample purity (260/280 nm and 260/230 nm) and yield (ng/µL) was measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific, NanoDrop products, Wilmington, DE, USA) by loading 1 µL of sample. While the on-column DNase treatment generally removed the majority of the genomic DNA (gDNA), faint bands were observed in some samples. For this reason, the on-column DNase treatment was discarded in favour of post-treatment using Turbo DNase (Thermo Fisher Scientific Inc., Waltham, MA, USA) following the manufacturers, protocol which effectively removed the last traces of gDNA.

Complementary DNA (cDNA) was synthesized using a QuantiTect Reverse Transcription Kit (QIAGEN, Hilden, Germany) according to the supplied protocol using 100 ng of extracted RNA for each reaction. The amount and purity of the extracted RNA was measured using a NanoDrop (as previously described), while the integrity was assessed using agarose gel electrophoresis by loading 4 µL of extract together with 1.0 µL loading dye on 1% gels. Hyperladder II (Bioline) was used as a MWM and ethidium bromide (10 mg/mL) was used to stain the RNA. Results were visualized under ultraviolet light.

2.7.3 Primer design and selected primers

Primers from the second primer set were designed using primer-blast, available online through the NCBI portal (http://www.ncbi.nlm.nih.gov). The software generates primer-pairs using the Primer3 primer design tool (Koressaar and Remm, 2007, Untergasser et al., 2012) but with an added optional BLAST search to minimize non-specific amplification from similar gene targets (Ye et al., 2012).

Table 2.7: The relevant settings used for primer design using primer-blast.

Parameter	Setting
Minimum PCR product size	60
Maximum PCR product size	150
Minimum primer size	15
Optimum primer size	20
Maximum primer size	25
Minimum Tm	59
Optimum Tm	60
Maximum Tm	61
Maximum Tm difference	1
Repeat filter	AUTO
Low complexity filter	Yes

Primers in the second primer set (see Table 2.9) were designed against selected expressed sequence tags (ESTs) using the settings shown in Table 2.7. Primers in the first primer set, and the target genes they were designed for are shown in Table 2.8). Primers were designed with an optimal target melting temperature (T_m) of 60±1°C. The maximum difference in (T_m) between the primers within the pair was set to 1°C The size of the resulting PCR product was set between 60 to 150 bp with an optimal primer size of 20±5 bp. Lyophilised primers were reconstituted in Tris-EDTA prepared using Tris-HCl (final conc. 10 µM) and EDTA (final conc. 0.1 µM) and adjusted to pH 8.0, prepared using DEPC (diethylpyrocarbonate) treated ddH_2 water and kept at -20°C.

2.7.4 PCR amplification of DNA

Amplification of nucleotide sequences using the polymerase chain reaction (PCR) was used as a standard method for checking the quality of primers, the presence of genomic DNA (gDNA) from RNA extractions and for negative and positive controls of

TABLE 2.8: The first set of primers for the verification of the SAGE transcriptome using qPCR. The nucleotide sequence for the forward and reverse primers are shown together with the target gene and the representative sequence tags (EST) matching the gene within the transcriptome.

Primer	Orientation	Nucleotide sequence (5'-3')	Target gene	Representative EST
Cab1	Forward	CATGCTGGCTGGTAGCATTA	Chlorophyll a/b-binding protein (Cab1)	Unknown
	Reverse	CCAGGCTTCACCGATAAAAA		
CSP	Forward	ATGATGTCACGCTTTTCACG	Chloroplast stem-loop-binding protein	isotig16520 isotig16521
	Reverse	TCTTGCTCTTTGCATCCTCA		
FSA	Forward	CGTGCTCCGTGAGATGTTTA	Fructose-6-P aldolase	isotig18730
	Reverse	GCAAGGCAACTTGACAACAA		
Lhc2.2	Forward	TGCTGAAATGCAGAAGATGG	Major light-harvesting chlorophyll a/b protein 2.2	isotig29353
	Reverse	AAGGCATGAGGGAAGGAAAT		
LhcII-3	Forward	TCTCGTTGCCCAGGTAGTTC	Major light-harvesting chlorophyll a/b protein 3	GHEXJPO01EA51P isotig29353
	Reverse	TCTCAGGCATCTCGTGACAG		
Lhca9	Forward	CTTCGGCAAGAACATCACCT	Light-harvesting protein of photosystem I (LHCA9)*	GHEXJPO01EVZYF GHEXJPO02JNZT
	Reverse	TCCAGCTCATGCAAGCATAC		
PHOA	Forward	CGTGGACGAGGTCTTCATCT	Starch phosphorylase (PHOA)	isotig16254
	Reverse	AGTACTCAGCATCGCCGAAC		
Rpi	Forward	TCCTGATGGTGTCAAGCTCA	Ribose-5-phosphate isomerase	isotig09905 isotig09904
	Reverse	AAAGGCAAGAGGGAGAGGA		
S11	Forward	GGCAAGAGTGGGCACCGCTT	S11 ribosomal protein	Control
	Reverse	CCTGGGGGATGACACGCAGC		

TABLE 2.9: The second set of primers for the verification of the SAGE transcriptome using qPCR. The nucleotide sequence for the forward and reverse primers are shown together with the target gene and the representative sequence tags (EST) matching the gene within the transcriptome.

Primer	Orientation	Nucleotide sequence (5'-3')	Target gene	Representative EST
T4	Forward	AAGGTCGTTTGATGTGGGCT	Unknown	isotig19583
	Reverse	CTGTCCGGTCCTGTCTTGTC		
T5	Forward	AGCAGCAGCACTGTTGGTAA	Unknown	isotig28831
	Reverse	CACACAGAGGGCAAACCAGA		
T7	Forward	AGGTTGTGCTGATGGGTCTG	Major light-harvesting chlorophyll a/b protein 2.1	isotig29353
	Reverse	TTCAGCTCAGCGAAGGTGTC		
T10	Forward	TAGTGAGCGTGGTCCTTGTG	Unknown	isotig31036
	Reverse	GCACAAACTCAAGAGCAGGG		
18S	Forward	CCTGCGGCTTAATTTGACTC	18S rRNA gene	Control
	Reverse	ACCGGAATCAACCTGACAAG		
rbcL	Forward	TACGGACGTGCTGTTTACGA	RuBisCo large subunit (rbcL)	Unknown
	Reverse	CGGTCTCTCCAACGCATGAA		
Iso1	Forward	ATGAGGGCTTGTGGTCTTGG	Unknown	isotig13632
	Reverse	AGCCCATCACGGAGAACTTG		
Iso2	Forward	GTGGACCACAGTGATGAGGG	Unknown	isotig27972
	Reverse	ACGTGCTGCTGTATGACTGG		
mimt	Forward	GGGAGCAGCTTGTGGTATGT	Mitochondrial inner membrane translocase	GGWEYJS01B2GV6
	Reverse	CACCCAAGCCACCAGTATGT		
S40	Forward	GTACAAGTACGGCCTGCTGA	40s ribosomal protein	isotig32977
	Reverse	CAACATATCGTTGGCACGCA		

DNA and cDNA as needed. The standard PCR protocol used 10 µL of MyTaq (2X) polymerase, 1.0 µL of template, 0.4 µL of each reverse and forward primer (10 µM) and 8.2 µL of ddH_2O for a total of 20 µL per well. The PCR cycle consisted of hot-start activation for 1 min at 95°C, followed by 30 ampification cycles of 10 s denaturation at 95°C, 10 s of annealing at 60°C (depending of the calculated primer T_m) and 10 seconds of elongation at 7°C, and a single extension step for 2 min at 72°C. Finished amplifications were held at 4°C before visualisation on agarose gels or further storage at -20°C.

2.7.5 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was carried out during two different periods using slightly different set-ups and instruments. The first part of the verification of the transcriptome using qPCR by Patrick Stephenson was carried out on a Opticon qPCR system (Bio-Rad, Hercules, California, USA) using Precision 2x real-time PCR Master Mix with SYBR green (Primerdesign Ltd, Southampton, United Kingdom) using non-skirted, low profile 96 well plates and clear 8-strip caps from Starlab (Starlab International GmbH, Hamburg, Germany) using the following thermal profile, initial denaturation at 95°C for 10 min followed by 40 cycles of denaturation for 15 s at 95°C and combined annealing/elongation for 1 min at 60°C. This was followed by a final extension for 10 min at 72°C. Dissociation curves were recorded from 60 to 95°C using a ramp speed of 2°C s^{-1}. Samples were analysed using duplicate technical repeats together with triplicate biological repeats for a total of six measurements per strain and light level. The samples were analysed using the Livak ($\Delta\Delta$CT) method using the 16S rRNA gene as the reference and the wild type low light sample as the calibrator (Livak and Schmittgen, 2001).

The second part of the verification phase was carried out on a Mx3005P qPCR System using Brilliant III Ultra-Fast SYBR Green QPCR Master Mix, clear non-skirted 96 well plates and strip caps (Agilent Technologies, Santa Clara CA, USA) using the following thermal profile: initial denaturation at 95°C for 3 min followed by 40 cycles of denaturation for 15 s at 95°C and combined annealing/elongation for 20 s at 60°C. Dissociation curves were recorded from 55 to 95°C using a ramp speed of 1°C s^{-1}. Samples were analysed using triplicate technical and biological repeats for a total of nine measurements per strain and light level. No template controls (NTC) were run in triplicate on each plate in combination with an inter-plate calibrator (IPC) consisting of a mix of wild-type cDNA from the three light levels and primers for the *lhca9* gene. The best combination of reference genes based on the calculated stability value was

chosen using the Normfinder Excel add-on package (Dheda et al., 2004). The samples were analysed using a modified version of the example calculations for Normfinder (available at http://moma.dk/normfinder-software).

2.7.6 SAGE library preparation

RNA was extracted using the cetyltrimethylammonium bromide (CTAB) method according to Pearson et al. (2006) using 35 mL of samples from cultures of wild type, lca1 and lca2 grown under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}). The cells were disrupted on a Tissue Lyser II bead beater (Qiagen, Valencia, CA, USA) using 250 mL of beads at 20 RPM for 5 min with 1.5 mL of CTAB buffer (100 mM Tris-HCL (pH 8), 20 mM EDTA, 1.4 M NaCl, 2% CTAB (w/v), 2% PVP (w/v)) added per sample. Following incubation at 60°C for 30 min, 1.5 mL of chloroform:isoamyl alcohol (24:1 v/v) was added, the sample was vortexed and centrifuged at 12,000 x g for 20 min at 20°C after which the aqueous phase was transferred to a new tube. The RNA was precipitated with 0.33 volumes of 10 M LiCl for 30 min at -20°C and then collected by centrifugation at 14,000 x g for 20 min at 4°C. The pellet was re-suspended and the RNA precipitated by adding 2x volumes of isopropanol and 0.1 volumes of 7.5 M ammonium acetate for 1 hr at -20°C. The RNA was collected by centrifugation at 14,000 x g at 4°C. The resulting pellet was washed with 70% EtOH and centrifuged at 14,000 x g for 10 min at 20°C, followed by drying at 20°C for 10-20 min and resuspension in 50 mL RNAse-free water. The extracted RNA was transcribed into cDNA using the ImProm-II Reverse Transcription System (Promega Corporation, Madison, WI, USA) using 4 mL ImProm-II 5X Reaction buffer, 2.4 mL $MgCl_2$, 1 mL dNTP mix (10 mM each dNTP), 1 mL recombinant RNasin ribonuclease inhibitor, 1 mL Improm-II reverse transcriptase, 1 mg of RNA template, 0.5 mg of oligo(dt) primers and 9.1 mL nuclease-free water. The process involved annealing for 25°C for 5 min, first strand extension for 60 min at 42°C followed by heat-inactivation at 70°C for 15 min according to manufacturer instructions. Sage libraries were prepared and sequenced at Iwate Biotechnology Research Centre, Kitakami, Iwate, Japan, following the SuperSAGE method (Matsumura et al. 2006) using a more recent protocol (HT-SuperSAGE) as described in (Matsumura et al. 2011). The cDNA was digested using the NLAIII restriction enzyme, ligated with adapter sequences and further cleaved using EcoP15I linked to a second set of adapter sequences, including four bp index sequences. Pooled samples were sequenced on an Illumina Genome Analyzer II.

2.8 Mutagenesis and flow cytometry

While not part of the work conducted for this thesis the random mutants of *D. tertiolecta* were generated, screened and isolated using the following methods. Five ml of cells (2.5 x 10^6 cell/mL) grown in f/2 - Si medium under 1000 μmol photons m^{-2} s^{-1} of light from the late exponential phase were subjected to treatment by the chemical ethane methanesulfonate (EMS) at five different concentrations (0.01, 0.025, 0.05, 0.1 and 0.15 % v/v). The cultures were vortexed briefly and left in the dark for 2 hr. The EMS was then deactivated using 2 mL of 10% (w/v) sodium thiosulfate. Following centrifugation at 1,200 x g for 5 min the supernatant was removed and the cells were re-suspended in 20 mL of fresh media. The cells were left to grow for 3 days under constant light at 100 μmol photons m^{-2} s^{-1} (24 hr light period). The survival rate was estimated by counting surviving cells using flow cytometry. Based on the survival ratio, the final concentration of EMS used was adjusted to 0.01% (v/v) in an attempt to minimize the number of point mutations generated in each cell. Surviving cells were then grown for 14 days to pre-select for fast-growing cell-lines before cell sorting.

2.8.1 Cell sorting using flow cytometry

A MoFlo XDP high speed flow cytometry cell sorter equipped with a blue laser (488 nm, Coherent Sapphire, Coherent) and coupled with a MoFlo XDP CyClone arm for direct plate sorting (Beckman Coulter) was used for single cell fluorescence-activated cell sorting (FACS) of mutagenized cell lines.

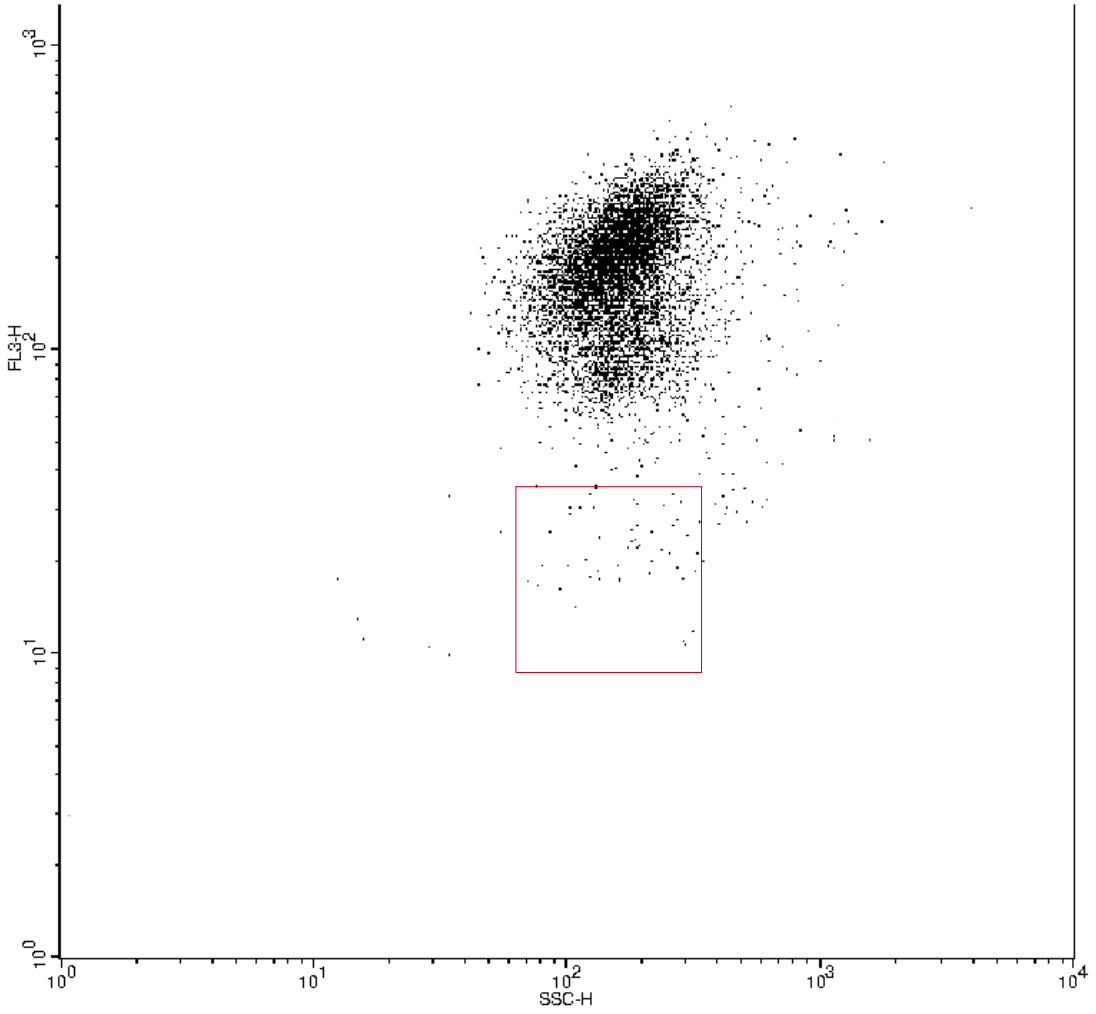


FIGURE 2.3: Selection of low chlorophyll mutagenised *D. tertiolecta* cells using flow cytometry. Each sorting event represents a single cell and is marked by a black dot. The rectangular outline in red shows the selection window gated by the chlorophyll fluorescence as measured by the FL3-H channel and the estimated cell size as measured by the SSC-H side scatter channel.

Single cells were screened based upon side-scatter (emission at the excitation wavelength of 488 nm) and chlorophyll auto-florescence (emission a 650-700 using the red channel at 670 nm (± 30 nm bandpassfilter)). Selection gates were established enabling sorting of cells with the lowest relative chlorophyll fluorescence and maintained WT cell size (see Figure 2.3). Cells were sorted into 96 well plates containing f/2 - Si growth media (amended with catalase to prevent the accumulation of reactive oxygen species) using the software BD FACSDIVA software version 4.1 (BD Biosciences, Ermbodegem, Belgium). Sorted cells were placed in a culture chamber with the growth conditions described above. Viable cell lines in which pigmentation was visible after 14

days were inoculated into larger 40 mL culture volumes and then screened for photophysiological parameters (using the techniques described earlier in this chapter). Sorting was performed at a pressure of 60 psi using a 70 μm nozzle, a nozzle amplitude of ∼12.5 V and a a drop-drive frequency of ∼95,000 Hz, with daily alignment corrections as needed using 3.0 μm flow-check yellow-green microspheres (Polysciences, Germany). The sheath fluid used during the experiments was saline solution (0.1 % w/v in MilliQ water) at a pressure of 20 p.s.i.

2.9 Cryopreservation

2.9.1 Methanol based cryopreservation

Cells of *D. tertiolecta* grown under 100 μmol photons m^{-2} s^{-1} for 7 days were counted and adjusted into volumes containing 20 million cells each depending on the cell concentration. The adjusted cultures were then centrifuged for 1 min at 4,700 x g to pellet the cells. The supernatant was removed and the pellet resuspended in 10 mL of fresh f/2 - Si medium. 10% methanol was prepared by adding 1.0 mL of sterile-filtered (0.22 μm) methanol (analytical grade) to 9 mL of f/2 - Si medium. For each sample 0.5 mL of 10% methanol and 0.5 mL of cell culture (adjusted to 1 million cells) was added to 2 mL cryo-tubes. The tubes were left to incubate for 40 min in the dark at 20°C stored within a plastic container partially filled with isopropanol (MrFrosty, Nalgene). The container was then moved to a -80°C freezer for 4 hr, after which the tubes were finally transferred to a liquid nitrogen storage unit and kept in the vapour phase at -170°C. Thawing of the cryopreserved cells was done by placing the tubes in a covered water bath kept at 20°C for 5 min. The tubes were then centrifuged for 2 min at 5,000 x g to pellet the cells. The supernatant was discarded and replaced by 1 mL of fresh f/2 - Si medium, followed by resuspension of the cells. The tubes were kept covered in aluminium foil for an additional 24 hr while being stored in a light cabinet, to minimize photoinduced damage, after which the cells were transferred to a larger volume of media and cultured according to the standard protocol.

2.9.2 Encapsulation in calcium alginate beads

Cells of *D. tertiolecta* were cryopreserved by encapsulation in calcium alginate beads following the instructions of Tanniou et al. (2012). Cultures were harvested in the exponential phase and samples corresponding to 1 million cells per mL were centrifuged to concentrate the cells. Following removal of the supernatant, 1 mL of 4% sodium alginate was added to the pellet and the cells resuspended by vortexing. Using a pipette,

drops of the solution were added to plastic plates filled with 100 mM calcium chloride solution, which solidifies the drops into semi-solid beads upon contact. The beads were then washed in fresh f/2 - Si media before transfer to a 0.7 M sucrose solution prepared with f/2 - Si media to dehydrate the beads. After 24 hr the beads were removed from the solution and air dried in a flow hood in partially covered petri dishes for 2 hours followed by freezing down to -80°C using a plastic container partially filled with isopropanol (MrFrosty, Nalgene). After 4-5 hr the beads stored in 2 mL cryotubes were moved to a liquid nitrogen storage unit and kept in the vapor phase at -170°C. Beads were defrosted and cells were revived by dissolving the beads in 20 mM sodium citrate under gentle stirring together with gentle manual mixing using a plastic loop as needed. The solution was then centrifuged and the pellet re-dissolved in fresh f/2 - Si media followed by culturing of the cells.

2.10 Screening and selection of mutant cells

The development of the selection process and protocol presented below, including the selection of candidate strains, was carried out before the start of this thesis and the results have been included for clarity. The process of selecting cells with the desired properties from a larger pool of mutated cells was performed in sequential steps and is schematically summarized in Figure 2.4.

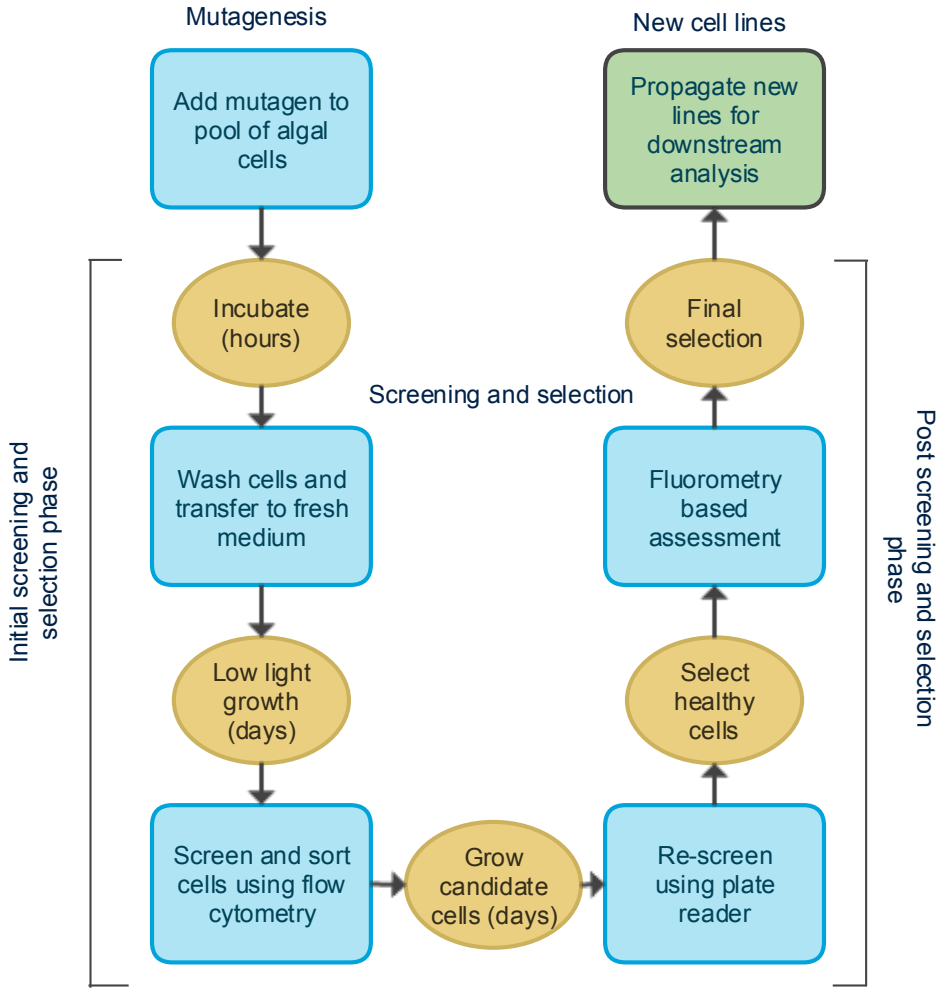


FIGURE 2.4: Flow chart depicting the process of generating mutant cell lines by adding a mutagen to algal cells, After incubation, the cells are washed and transferred to new media and grown under low light. The mutated cells are then screened and sorted using flow cytometry. Cells are then once again incubated and re-screened using a plate reader. Healthy cells are further assessed using fluorometry based measurements before a final selection. Selected cells representing new cell lines are finally propagated for further downstream analysis.

The process starts with the generation of a pool of mutated cells. This was accomplished by adding the chemical ethyl methanesulfonate (EMS) to a cell culture. This chemical agent alkylates guanine (G) residues forming O6-ethylguanine, which pairs with thymine (T) instead of cytosine (C). Following action by the cells normal DNA repair machinery the alkylated guanine residue can be replaced with adenine (A) resulting in a C/G to T/A substitution (Kim et al., 2006). Even though other mutations are possible, they are not commonly found. Greene et al. found at least 99.5% of the changes to be G/C to A/T in a large-scale screening study involving 192 genes and 1906 mutations in *Arabidopsis thaliana* (Greene et al., 2003). They also found the

mutations to be randomly distributed throughout the genome with a mutation frequency of roughly 1 mutation per 300 k basepairs using a EMS concentration ranging from 20 mM to 45 mM (Till et al., 2003). The toxicity to the cells are exponentially additive in relation to the EMS concentration and this is directly reflected in the survival rate of the cells, as seen in Figure 2.5.

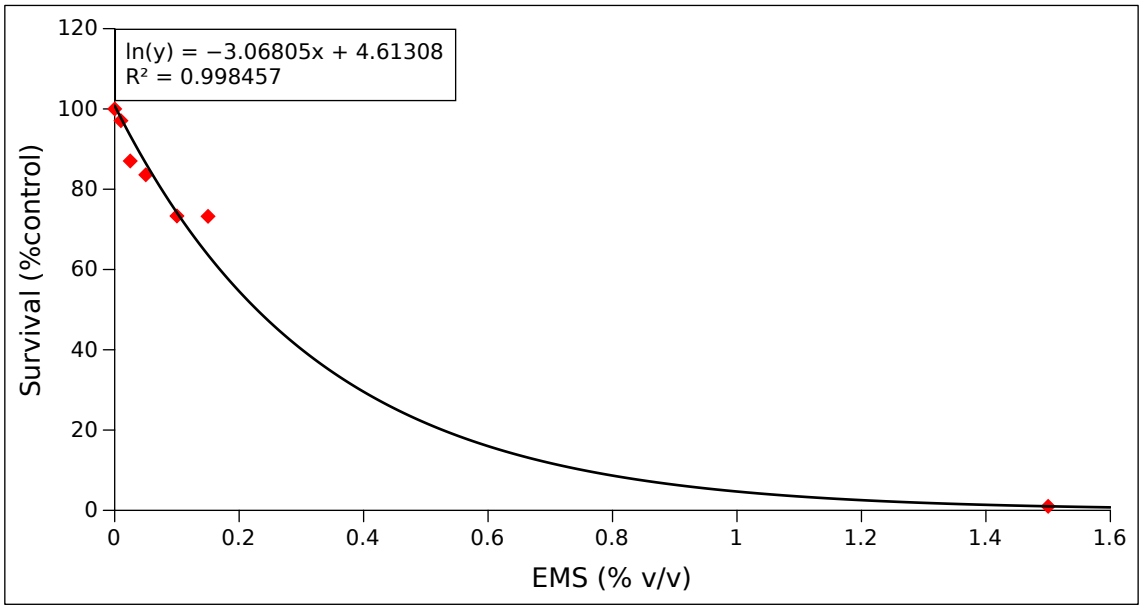


FIGURE 2.5: The actual cell survival rate (red diamonds) in percent compared to controls after chemical mutagenesis using ethyl methanesulfonate (EMS), as determined by counting surviving cells post treatment and under five different concentrations (0.01, 0.025, 0.05, 0.1 and 0.15% v/v) of the chemical. The measured survival rate plotted together with a exponential regression (black line) and its equation and regression coefficient (R^2) to show the expected survival rate under varying concentrations of EMS. Data from P. Stephenson.

In the end the lowest tested concentration of EMS was selected for the generation of mutated cells. This EMS concentration of 0.01% (~0.8 mM) resulted in a survival rate of just above 97%. The rate dropped rather sharply by a further 10 percent at a 0.25% concentration before stabilizing around 1.5% at a survival rate of around 73%. For comparison, Jin et al. used a concentration of 200 mM that resulted in a survival rate of only 5% (Jin et al., 2003), while Chaturvedi and Fujita (2006) used a 100 mM concentration that resulted in a far better 55-60% survival rate.

TABLE 2.10: Cell survival rate (cells survived) and the survival rate in percent (survival rate (%)) after sorting of a specific number of mutated cells (cells sorted) using flow cytometry by the sorting vessel (flasks, tubes or plates) used and the volume (media volume) and if catalase had been added to the media (catalase added).

Sorting vessel	Catalase added	Media volume	Cells sorted	Cells survived	Survival rate (%)
Flasks	No	30 ml	62	47	75.8
Tubes	Yes	500 µl	80	72	90.0
Plates	No	100 µl	480	334	69.6
	Yes	100 µl	2880	2291	79.5

The next phase of the process involved the sorting and selection of cells with certain characteristics from the pool of mutated cells. This phase was optimized for the highest survival rate together with ease of use by testing a small selection of sorting vessels as well as with or without the addition of the enzyme catalase to minimize the negative effect of reactive oxygen species (ROS) produced by the chloroplast (Pérez-Pérez et al., 2012). This enzyme catalyses the breakdown of hydrogen peroxide into hydrogen and oxygen, protecting the cells from oxidative stress (Gill and Tuteja, 2010). The final combination selected consisted of sorting into 96 well plates with catalase added at a concentration of 100 U/mL and maintained under low light. This combination gave an average cell survival rate of 80% two weeks after sorting. The highest survival rate was seen in Eppendorf tubes with a 90% survival rate, while the lowest rate with a 70% survival was seen in 96 well plates without any added catalase (see Table 2.10).

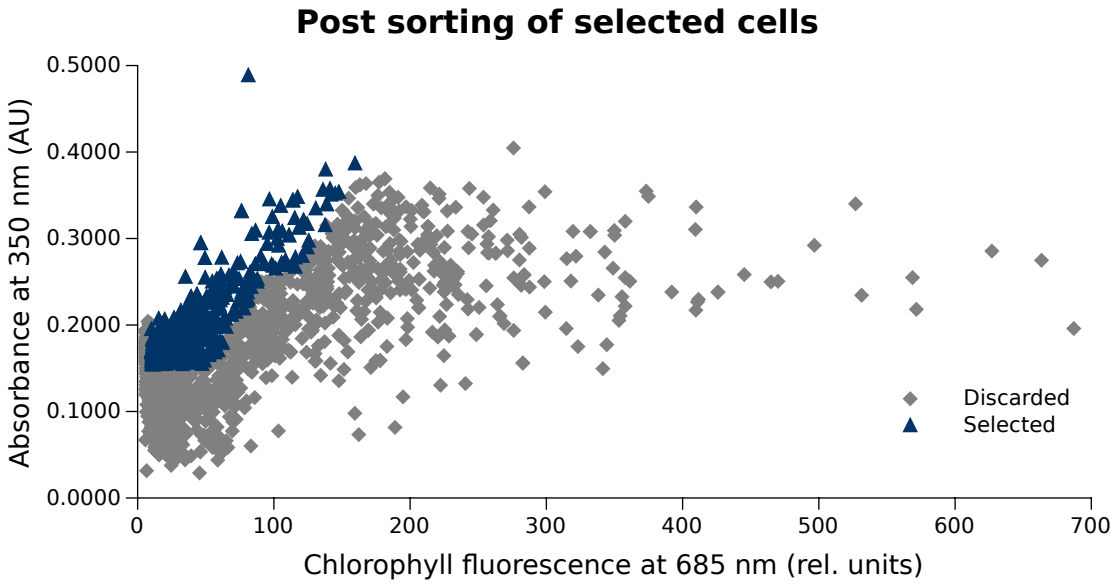


FIGURE 2.6: Post-sorting of low chlorophyll cells using a plate reader 2 weeks after the original selection. The cells were analysed for relative size using the absorbance at 350 nm and the chlorophyll fluorescence at 685 nm as a measure of total chlorophyll content. Selected cells are shown using blue triangles while cells not passing the selection are shown using grey diamonds.

Surviving cells were first subjected to post-screening using a plate reader to again select for cells with a low chlorophyll fluorescence and normal cell size (see Figure 2.6). Cells meeting the criteria were grown in larger volumes (40 mL) for final analysis using FRRf. The functional cross-section (σPSII) together with the maximum photochemical efficiency (F_v/F_m) of PSII were measured and compared to the wild type cells.

TABLE 2.11: The top ten cell lines with most reduced photosystem II antenna size (σPSII) compared to WT selected from mutated cells. The photosynthetic efficiency, in terms of (F_v/F_m), and the difference of σPSII and (F_v/F_m) compared to WT in percent are also shown. Data was collected from cells in the mid exponential phase grown under medium light.

Strain	**σPSII**	**% of WT**	**Fv/Fm**	**% of WT**
#1	0.87	78.3	0.51	102.4
#2	0.87	78.3	0.48	95.6
#3	0.87	78.3	0.54	107.8
#4	0.86	77.4	0.54	108
#5	0.86	77.4	0.48	96
#6	0.83	74.7	0.52	104.6
#7 (lca2)	0.74	66.6	0.52	104
#8	0.72	64.8	0.56	112
#9	0.68	61.2	0.51	102
#10 (lca1)	0.63	56.7	0.53	106

Two cell lines with smaller σPSII and equal or higher (F_v/F_m) compared to the wild type (see Table 2.11) were finally selected for further analysis. The smaller functional cross section of PSII (σPSII) in conjunction with a retained or improved photosynthetic efficiency (as measured by (F_v/F_m)) of PSII of these selected lines have the potential to improve the photosynthetic efficiency of the system. These two mutants, named lca1 and lca2 (for low chlorophyll antenna), are the two strains analysed in more detail in this thesis.

Chapter 3

Photophysiological characterization of two low chlorophyll antenna cell lines from *D. tertiolecta*

3.1 Introduction

The species *Dunaliella tertiolecta* was selected as a candidate for this study. Selection required a marine eukaryotic species lacking a published genome and without an established transformation protocol to evaluate the extent to which a poorly characterized strain could be improved using the non-directed approaches that form the basis of this thesis. It is noted that many studies have improved the photosynthetic efficiency of model cell algal cells lines grown routinely in culture. From the genus of the algae family Dunaliellaceae, *Dunaliella* sp. are motile, unicellular, rod to ovoid shaped green algae (Chlorophyceae), which are common in marine waters. The organisms are relatively simple to cultivate and do not clump or form chains and, in the case of *D. tertiolecta*, have an oil yield of about 37% (organic basis) and a fast growth rate (specific growth rate (μ) of up to 1.71 day^{-1}) (Fujiki et al., 2007, Minowa et al., 1995). These features make it of interest for development as a source of biofuel.

This chapter describes the physiological characterisation of two low chlorophyll antenna cell lines (lca) from *D. tertiolecta* previously selected using the random mutagenesis protocol as described in Chapter 2. The earlier selection process of these mutant cell lines confirmed a stable low chlorophyll phenotype under the specific experimental light conditions (100 μmol photons m^{-2} s^{-1}). However to extend our understanding of how the random mutations impacted the general physiology of the cells, further and more extensive experiments under different light regimes were needed. Initially these experiments focused on establishing basic growth parameters of wild type and lca strains over time, followed by more extensive experiments focusing on specific parameters designed to define the photophysiology of the strains. Fluorescent-based techniques including pulse amplitude modulated (PAM) and fast repetition rate fluorometry (FRRf) were used extensively to measure various photophysiological parameters, including the efficiency of photosynthesis as well as the effective size of each photosystem. Complemented by measurements of the rate of oxygen evolution these techniques allowed us to estimate and compare the maximum rate of photosynthesis in each strain. They also give an insight into how these strains cope with and adjust to potentially damaging light intensities. By combining these data from high performance liquid chromatography (HPLC) and fluorometry-based measurements of the pigment composition within each cell with Western blot-based measurements of the distribution of reaction centres in photosystem II and I, its possible to draw well informed conclusions regarding how each strain adjust the size and chlorophyll distribution within the antenna as well as the stoichiometric distribution of photosystems in regard to changing light conditions. The results from these experiments also allows for a direct comparison of these low chlorophyll strain phenotypes and their distinct photophysiology with the corresponding results from a full gene expression analysis presented in later chapters (Chapters 4 and 5) and how these results compare to previously described cell lines with similar properties reported in the literature.

3.2 Results

The growth characteristics of the lca1 and lca2 strains compared to the wild type were established by culturing the strains under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}) for extended periods until the cultures (based on cell counts) had reached a stationary phase while continuously monitoring a number of basic growth parameters. Typical measurements taken on a daily basis included counting the number of cells per volume, measuring chlorophyll a content and fluorescence-based measurements of photosystem II related parameters, including the functional cross section (σPSII) and the photosynthetic efficiency F_v/F_m.

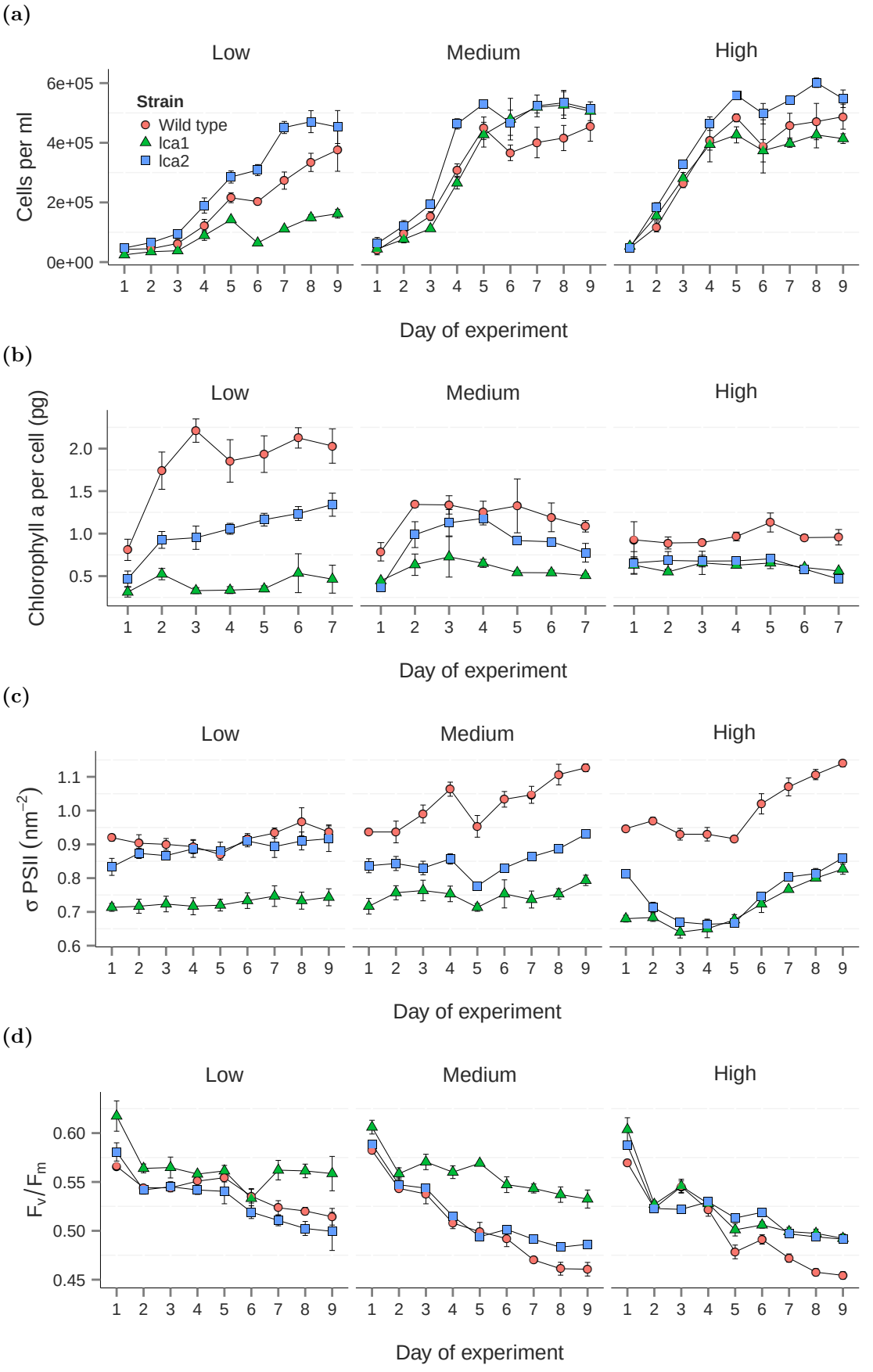


FIGURE 3.1: Photophysiology and growth parameters of wild type (red) and mutant cell lines lca1 (green) and lca2 (blue) measured over 7 or 9 days of growth. Cells were grown under low, medium and high light (20, 100 and 1200 µmol photons m^{-2} s^{-1}) with a) cell density, b) chlorophyll a per cell, c) photosynthetic physiology σPSII and d) (F_v/F_m). Error bars are one standard deviation from biological triplicates.

Figure 3.1 shows the result from a growth experiment run for 9 days. The results from the experiment are discussed in detail in Sections 3.2.1 and 3.2.2 below. Results on either a strain or light level that are reported as significant are based on Tukey post-hoc tests of the results from the analysis of variance (ANOVA).

3.2.1 Calculated growth rate

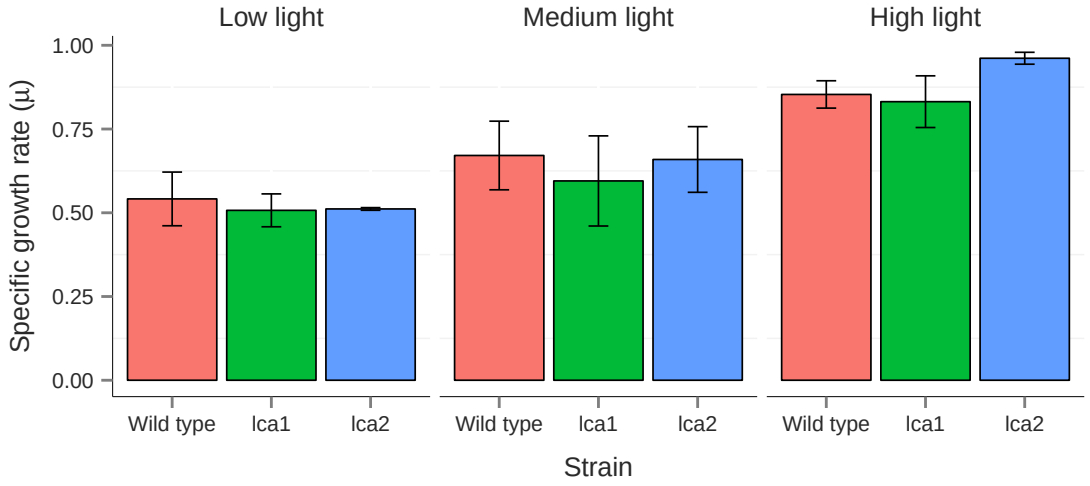


FIGURE 3.2: Specific growth rate (μ) of wild type and lca1 and lca2 measured in the exponential growth phase in cells grown under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates.

The calculated growth rate of the strains was based on the measured cell counts from a 9 day growth experiment (shown in Figure 3.1). The data show the expected pattern of growth after plotting the log transformed counts versus the time (see Appendix A, Figure A.1). Exponential growth is seen on days 2 to 5 under low light, days 1 to 4 under medium light and days 1 to 3 under high light. Clear plateaus indicative of the stationary phase are reached around days 5 to 6 under medium and high light. The pattern is not as clear under low light: the lca2 strain clearly enters the stationary phase under day 7 but both the lca1 and the wild type strain show signs of slow growth even during the last 2 days (days 8-9) of the experiment. Also apparent is a marked slump in cell numbers in the lca1 strain on day 6 under low light. The cause for this drop is unknown but it seems reasonable to assume that the limitations and stress related to the cells just reaching the stationary cells are part of the cause.

TABLE 3.1: The calculated specific growth rate (μ) and doubling time (T_2) based on the time period Δ, starting on day t_0 and ending on day t for the wild type, lca1 and lca2 measured in the exponential growth phase in cells grown under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates.

Parameter	Strain	Light	Δt (t_0, t)	Growth rate (μ)	Doubling time (T_2)
μ & T_2	Wild type	Low	4 (2,5)	0.541±0.08	1.299±0.19
		Medium	3 (1,4)	0.671±0.10	1.048±0.15
		High	2 (1,3)	0.853±0.04	0.814±0.04
	lca1	Low	4 (2,5)	0.507±0.05	1.375±0.14
		Medium	3 (1,4)	0.595±0.13	1.212±0.31
		High	2 (1,3)	0.832±0.08	0.839±0.08
	lca2	Low	4 (2,5)	0.511±0.00	1.355±0.01
		Medium	3 (1,4)	0.659±0.10	1.067±0.15
		High	2 (1,3)	0.961±0.02	0.721±0.01

The increase in growth rate under increasing light is significant (ANOVA, $p < 0.05$) in all strains and shows that the cells are light limited under low and medium light (see Figure 3.2 and Table 3.1). No significant difference (ANOVA, $p > 0.05$) in growth rate was observed on a strain level between the lca strains and the wild type.

3.2.2 Comparison of wild type and lca strains in the exponential growth phase

The following section presents comparisons between the wild type and the lca strains on day 4 of growth when all cultures were in the exponential growth phase (see Section 3.2.1) to allow for more accurate comparisons of the strains under varying light levels under non-limited conditions.

3.2.3 Cell density

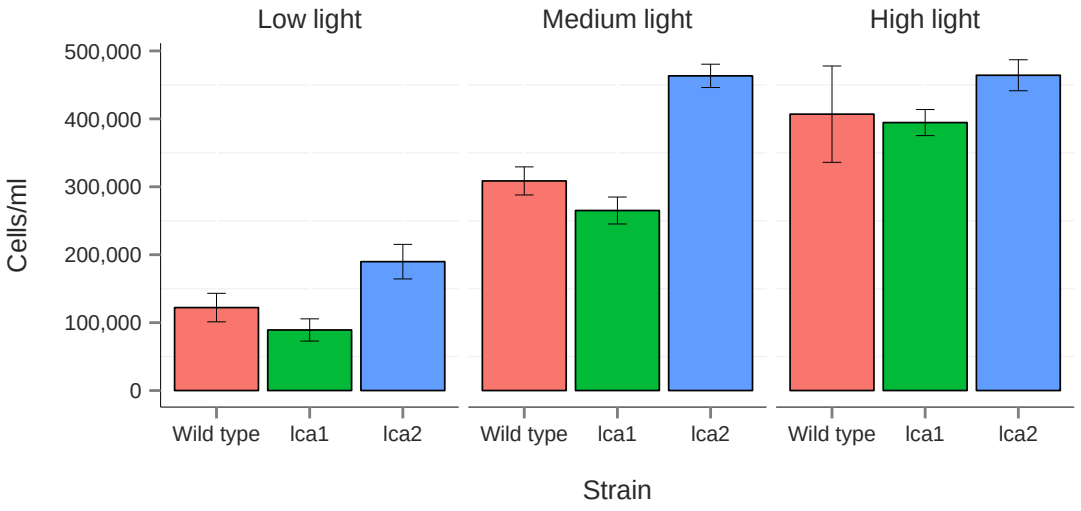


FIGURE 3.3: The mean number of cells per mL in the wild wild type, lca1 and lca2 measured in the exponential growth phase on day 4, in cells grown under low, medium and high light (20, 100 and 1200 µmol photons m^{-2} s^{-1}) during a 9 day experiment. Error bars are one standard deviation from biological triplicates.

The measured cell density on day 4 in the exponential growth phase showed no significant (ANOVA, $p > 0.05$) differences in cell density on a strain level between lca1 and wild type. The difference between lca2 and wild type was significant (ANOVA, $p < 0.05$) under medium light but not under high or low light (see Figure 3.3). As suggested by the calculated growth rate (Table 3.1) all strains show a significant increase in cell density when comparing low and high light results. These measured cell densities and calculated growth rates would suggest that the mutations in the lca strains have caused no significant reduction in the cells viability compared to the wild type from a growth perspective.

3.2.4 Chlorophyll a per cell

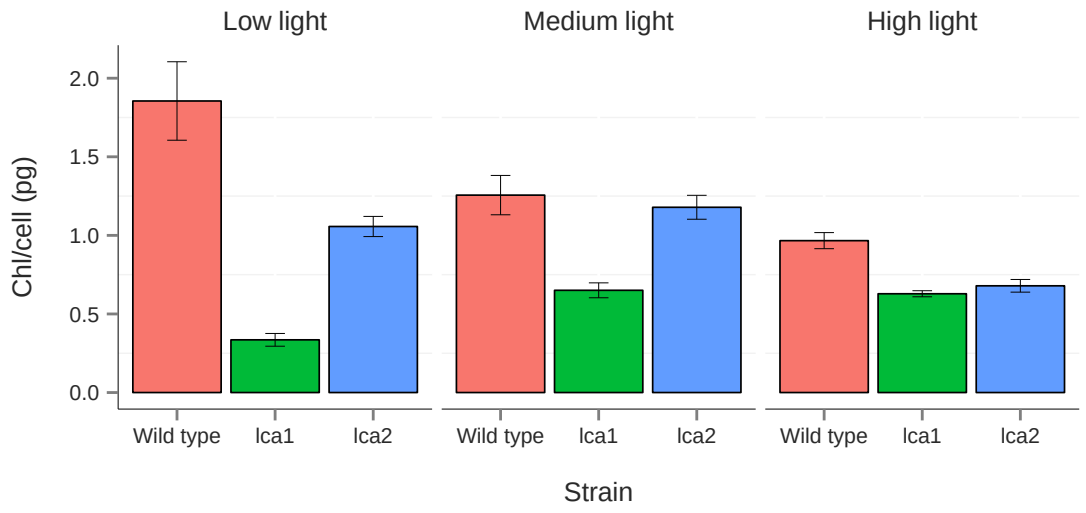


FIGURE 3.4: Chlorophyll a content per cell in the wild type, lca1 and lca2 measured in the exponential growth phase on day 4, in cells grown under low, medium and high light (20, 100 and 1200 µmol photons m^{-2} s^{-1}) during a 9 day experiment. Error bars are one standard deviation from biological triplicates.

The difference in chlorophyll content is significant for both lca strains compared to the wild type (ANOVA, $p < 0.05$) on a strain and a light level. The chlorophyll content per cell gives an insight into the amount of photosynthetic pigment that is maintained in each cell line during exponential growth over a light gradient. Figure 3.4 shows that pigment content increases with decreasing light in both wild type (by 92%) and lca2 (by 56%) - a photo-acclimatory strategy previously observed in *Dunaliella salina* to help cope with light limitations (Neidhardt et al., 1998, Smith et al., 1990). This trend is however not observed in lca1, which displays a different physiology, maintaining a relatively constant chlorophyll content with light.

3.2.5 Functional cross section of photosystem II (σPSII)

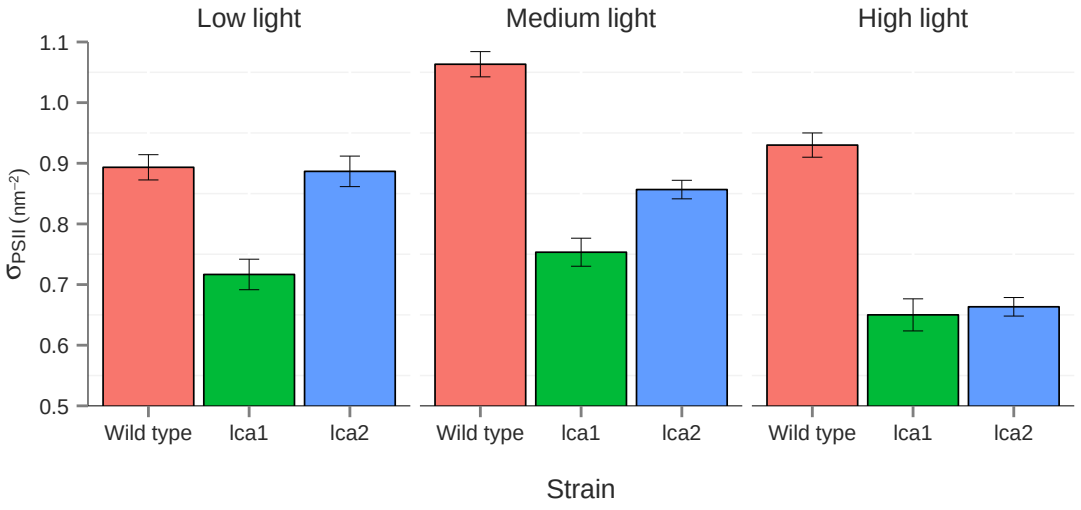


FIGURE 3.5: The mean functional cross section of PSII in the wild type, lca1 and lca2 measured in the exponential growth phase on day 4, in cells grown under low, medium and high light (20, 100 and 1200 µmol photons m^{-2} s^{-1}) during a 9 day experiment. Error bars are one standard deviation from biological triplicates.

The size of the functional cross section of PSII (σPSII) is significantly (ANOVA, p <0.05) smaller in both lca1 and lca2 on a strain and light level compared to wild type. The lca strains are significantly smaller compared to wild type under all light levels with the exception of lca2 under low light. This difference is more pronounced under increasing light, (Figure 3.5), with a significantly smaller functional cross section in lca1 and lca2 but not wild type under high light compared to low light. This confirms that both lca lines reduce their PSII functional cross section in response to increasing light. However, the response is far more pronounced in lca2 (by 34%) than in lca1 (by 16%) suggesting that lca2 has retained the ability to reduce the cross section in response to increasing light. This ability is largely absent in lca1 and only observed under high light conditions.

3.2.6 Photosynthetic efficiency of photosystem II

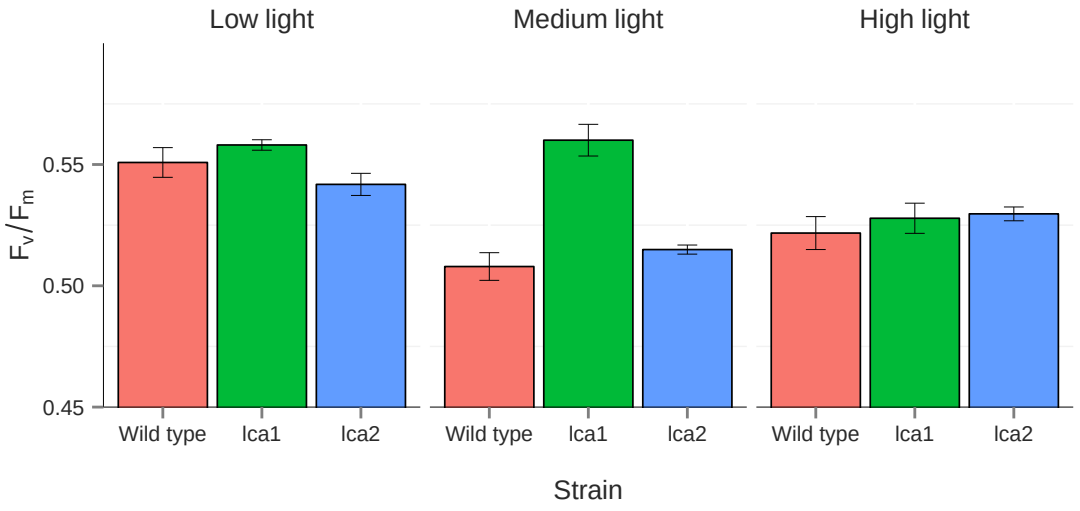


FIGURE 3.6: The mean photosynthetic efficiency of PSII in the wild type, lca1 and lca2 measured in the exponential growth phase on day 4, in cells grown under low, medium and high light (20, 100 and 1200 µmol photons m^{-2} s^{-1}) during a 9 day experiment. Error bars are one standard deviation from biological triplicates.

The photosynthetic efficiency of PSII, as measured by F_v/F_m, remains largely the same in the lca strains compared to the wild type. The only significant (ANOVA, $p < 0.05$) difference on a strain level is seen under medium light with an increased efficiency in lca1 compared to wild type. When comparing the efficiency under low to high light the F_v/F_m is significantly (ANOVA, $p < 0.05$) lower in all strains, which is consistent with an expected reduction in photosynthetic efficiency due to photoinduced damage to photosystem II caused by over absorption of light.

3.3 Further characterization of specific physiology parameters in the exponential growth phase

Based on the understanding gained from the growth experiments more targeted experiments were conceived to help us better understand the differences between the lca1 and lca2 strains compared to the wild type. These experiments was generally run at a single light level in the exponential phase. This generally involved growing the strains under medium light (100 µmol photons m^{-2} s^{-1}) and sampling on day 4 in the later stage of the exponential phase (see Appendix A, Figure A.1) to maximise the number

of cells in the sample, while at the same time minimizing the risk of the limiting conditions (nutritional, physiological) that define the stationary phase from affecting the results (Andersen, 2005).

3.3.1 Rate of oxygen evolution by light intensity

The rate of oxygen evolution under different light intensities gives an insight into how changes in light conditions effects the cells ability to split oxygen through the oxygen evolving complex. This process supplies the electrons that pass through the electron chain during photosynthesis and is therefore rate limiting. Strains with reduced light harvesting complexes have previously been shown to exhibit reduced light inhibition coupled with higher photosynthetic efficiencies under high light conditions (Melis et al., 1998, Mussgnug et al., 2007a). To measure the rate of oxygen evolution, the wild type, lca1 and lca2 strains were grown under medium light (100 µmol photons m^{-2} s^{-1}) for 4 days, after which the cell density and the chlorophyll a content was measured. The cultures were adjusted to one million cells per ml of sample and the rate of oxygen evolution was measured at ten different light intensities from 0 to 2000 µmol photons m^{-2} s^{-1}.

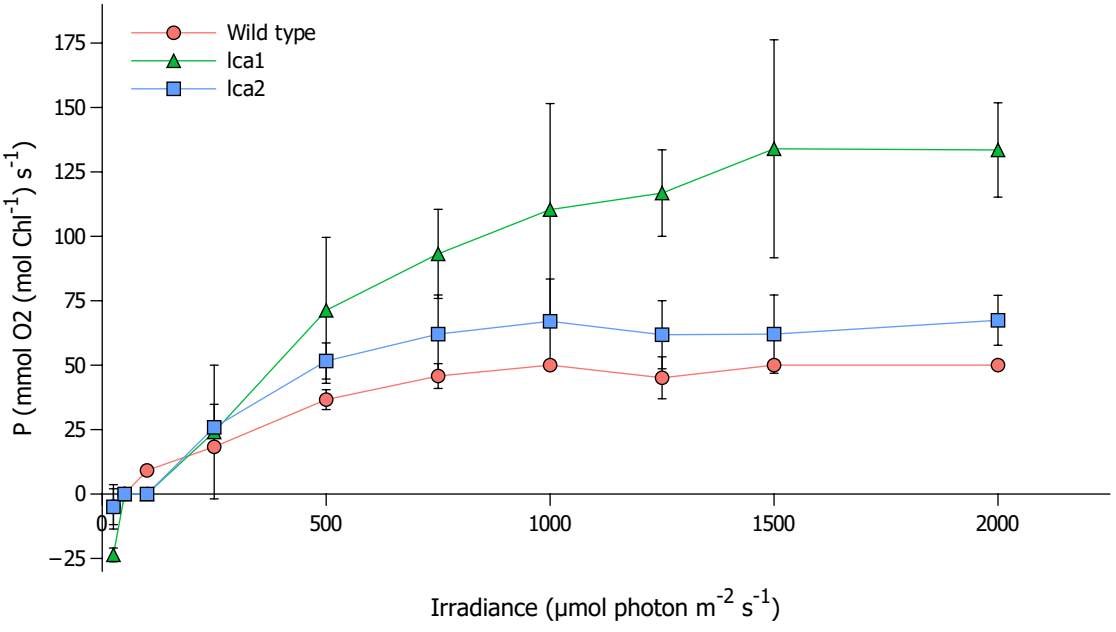


FIGURE 3.7: The rate of oxygen evolution in the wild type, lca1 and lca2 strains grown under medium light (100 µmol photons m^{-2} s^{-1}) as measured under increasing irradiance from 0 to 2000 µmol photons m^{-2} s^{-1} and normalized to chlorophyll a. Error bars are one standard deviation from biological triplicates.

TABLE 3.2: The calculated P_{max}, α and I_k from the chlorophyll a normalised rate of oxygen evolution in the wild type, lca1 and lca2 strains grown under medium light (100 μmol photons m^{-2} s^{-1}). Errors are one standard deviation from biological triplicates.

Parameter	Strain	P_{max} μmol O_2 (μmol Chl a)$^{-1}$ s^{-1}	α	I_k μmol photons m^{-2} s^{-1}
O_2 evolution	Wild type	55.07±3.23	0.115±0.05	523.6 ±205.1
	lca1	159.42±22.10	0.209±0.073	803.33±190.00
	lca2	71.70±16.01	0.146±0.054	517.96±169.60

Figure 3.7 shows the results from the measurements normalized on a per chlorophyll a basis. Both lca strains had a higher rate of oxygen production per chlorophyll on average except under low irradiance levels. The calculated P_{max}, α and I_k values, as estimated for each strain based on the fitted model, (see Section 2.5 in Chapter 2 for more details) are shown in Table 3.2, P_{max} was significantly higher in lca1 compared to wild type (ANOVA, $p < 0.05$) but not lca2. The light intensity at which photosynthesis becomes saturated (I_k) was not significantly higher in any of the lca strains compared to wild type. The significantly higher rate of photosynthesis (P_{max}) suggests a higher photosynthetic conversion efficiency in the lca1 strain compared to the wild type.

3.3.2 PSII electron transport rate by light intensity

Fast repetition rate fluorometry (FRRf) can similarly be used to probe the photosynthetic efficiency by estimating the electron transport rate of PSII under different light intensities. To measure the electron transport rate of PSII, the wild type, lca1 and lca2 strains were grown under medium light (100 μmol photons m^{-2} s^{-1}) and sampled in the exponential phase on day 4. After loading the cultures on the FRRf the PMT voltage was adjusted to keep the maximum fluorescence yield (F_m) signal around 10000 (± 300) before running a rapid light curve (RLC) in 15 steps from 0 to 2046 μmol photons m^{-2} s^{-1}. The electron transport rate was calculated as described in Section 2.4.1 in Chapter 2.

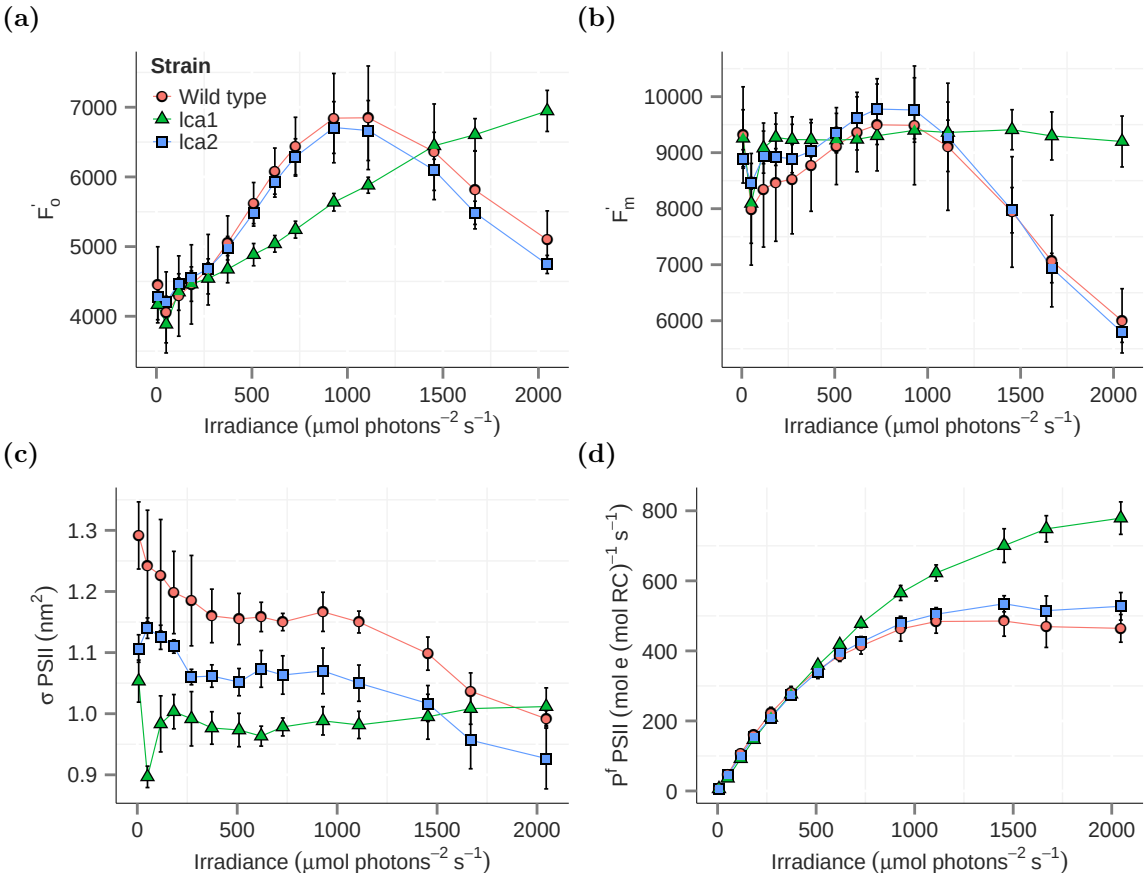


FIGURE 3.8: a) The minimum fluorescence under actinic light (F_o), b) the maximum fluorescence under actinic light (F_m), c) functional cross section under actinic light (σ'PSII) and d) the electron transport rate of PSII P_f as measured by rapid light curves from 0 to 2048 µmol photons m^{-2} s^{-1}. Error bars are one standard deviation from biological triplicates.

TABLE 3.3: The calculated P_{max}, α and I_k parameters based on the electron transport rate of PSII (P_f) in the wild type, lca1 and lca2 strains grown under medium light (100 µmol photons m^{-2} s^{-1}). Errors are one standard deviation from biological triplicates.

Parameter	Strain	P_{max} µmol e (µmol RC)$^{-1}$ s^{-1}	α	I_k µmol photons m^{-2} s^{-1}
P_f (PSII)	Wild type	481.4±47.1	0.884±0.034	543.9±33.1
	lca1	811.0±62.1	0.757±0.010	1071±69.1
	lca2	533.3±27.2	0.817±0.014	652.8±34.4

Figure 3.8 shows the calculated PSII electron transport rate based on the measurements. Both lca strains had similar or higher average electron transport rates compared to the wild type under all levels of irradiance with markedly higher rates in the lca1 strain under higher light. The P_{max}, α and I_k value shown for each strain

was calculated using the same model as previously used to calculate the corresponding parameters for oxygen evolution (see Section 2.5 in Chapter 2 for more details), as shown in Table 3.3, there was a significantly higher P_{max} between the wild type and lca1 (ANOVA, p <0.05) but not lca2. The light intensity at which photosynthesis becomes saturated (I_k) was significantly higher in the lca1 strain but not lca2 compared to wild type. Similar to the results from the oxygen evolution measurements the significantly increased P_{max} in the lca1 strain points to a more efficient conversion of captured light into photochemical energy compared to the wild type.

3.3.3 Functional cross section of photosystem I (σPSI)

To complement the FRRf-based measurements of the functional cross section in PSII, the absolute size of the functional cross section in PSI (σPSI)(ms^{-1}) was estimated using PAM fluorometry. The wild type, lca1 and lca2 strains were grown under medium (100 µmol photons m^{-2} s^{-1}) light for 4 days and concentrated to increase the cell density before measurements using PAM under three different saturating light intensities (5, 10 and 20 mmol photons m^{-2} s^{-1}) and dilutions (1:1, 1:2 and 1:4). The actual cell densities were corrected using the optical absorbance at 635 nm for each sample.

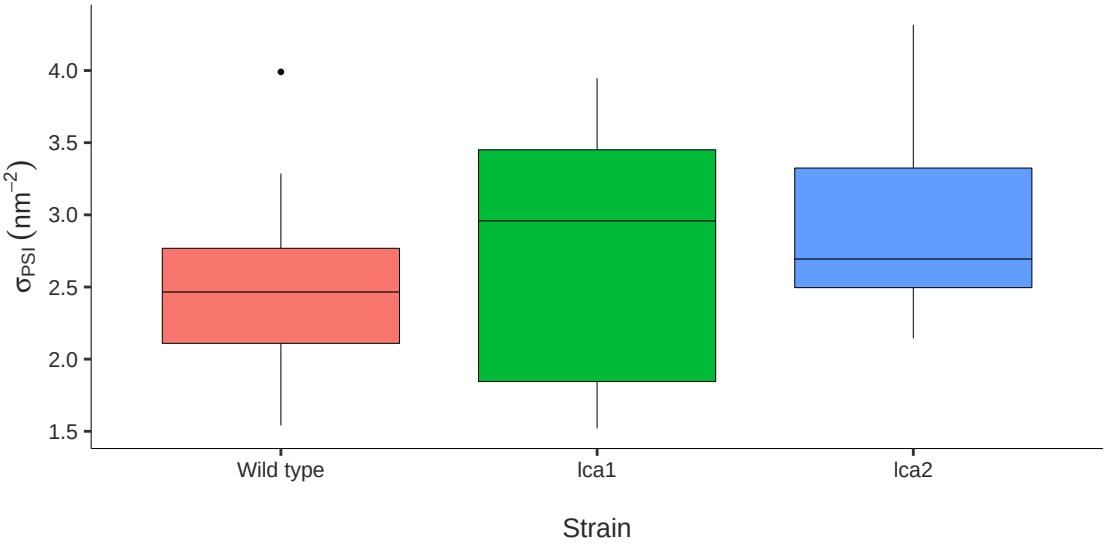


FIGURE 3.9: The calculated functional cross section of PSI (σPSI)(nm^2) as measured under saturating light (5, 10 and 20 mmol photons m^{-2} s^{-1}) in two different sample concentrations (1:1 and 1:2) in the wild type, lca1 and lca2 strain in the exponential growth phase in cells grown under medium light (100 µmol photons m^{-2} s^{-1}).

After correcting for the actual cell density the calculated σPSI show no significant difference (ANOVA, p >0.05) in the size of σPSI in the wild type compared to lca1 or

lca2 (Figure 3.9). Based on these results, the functional cross section of PSI is similar in all three strains when grown under this light level.

3.3.4 Concentrations of photosystem II and I reaction centre and RuBisCo proteins

Western blots were used to estimate the relative proportions of photosystem reaction centres and RuBisCo protein in the lca strains compared to wild type. The relative amount of reaction centres of PSII compared to PSI was estimated using antibodies against the D1 protein of PSII encoded by the *psbA* gene and the iron sulfur binding protein of PSI encoded by the *psaC* gene. Mutant strains of *C. reinhardtii* deficient in the core protein encoded by *psaC* no longer accumulate core subunits of PSI (Takahashi et al., 1991) in a stable manner. Similarly, the DI protein binds the oxygen-evolving centre (OEC) in PSII (Murray, 2012). Both of these core subunits are essential for the stability of each photosystem and the measured amount of protein should be closely related to the actual distribution of photosystems within the chloroplast. To complement these measurements, the amount of RuBisCo was estimated using antibodies against the *rbcL* gene encoding the large subunit of the protein. The wild type, lca1 and lca2 strains were grown under medium (100 μmol photons m^{-2} s^{-1}) light for 4 days, followed by cell counts and whole cell protein extraction. The protein and chlorophyll a content was measured for each sample with gels loaded with equal amounts of total protein.

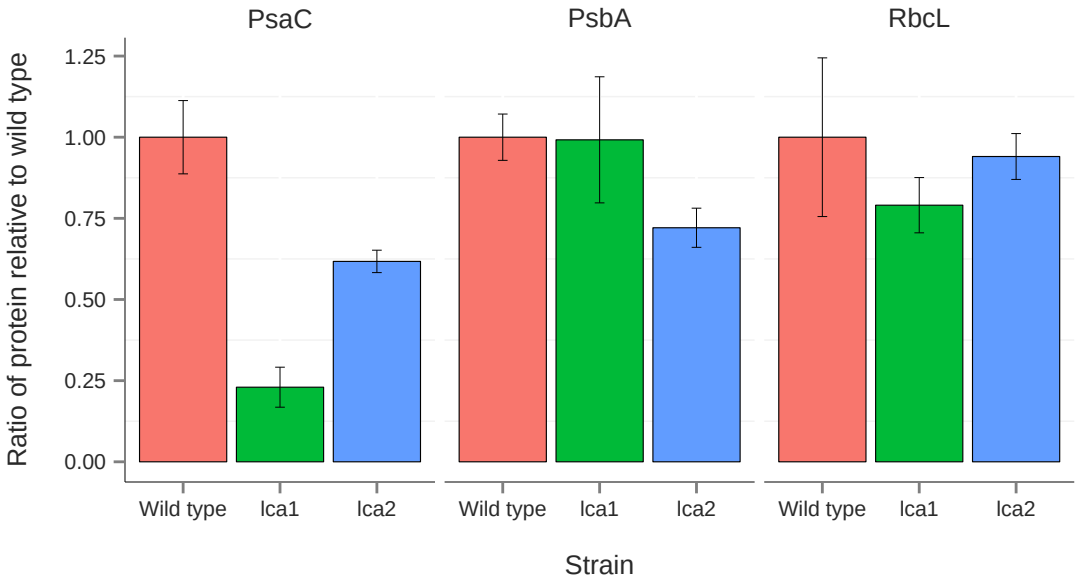


FIGURE 3.10: The relative proportions of PsbA, PsaC and RbcL proteins in the lca1 and lca2 strains compared to the wild type as measured by Western blots from protein extractions of cultures grown under medium light (100 µmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates.

The results in Figure 3.10 show a relatively large reduction (to 23.0% and 61.7%, respectively) in PsaC protein concentration between the lca strains compared to the wild type. The result are significant (ANOVA, $p < 0.05$) with significantly less PsaC in both mutants compared to the wild type. The difference in the concentration of PsbA (99.2% and 72.1%) is less striking and not significant (ANOVA, $p > 0.05$). The amount of RbcL is also somewhat reduced (to 79.1% and 94.1%) in both lca strains but the result is not significant (ANOVA, $p > 0.05$) in either. To conclude, the results from the blots show a significant reduction in the number of photosystem I reaction centres in the lca strains relative to the wild type, but similar levels of photosystem II and RuBisCo proteins.

3.3.5 Pigment composition

High performance liquid chromatography (HPLC) based measurements allow for simultaneous and precise measurements of both the chlorophyll and accessory pigment content in microalgal cells. The wild type, lca1 and lca2 strains were grown under low (20 µmol photons m^{-2} s^{-1}), medium (100 µmol photons m^{-2} s^{-1}) and high light (1200 µmol photons m^{-2} s^{-1}) and sampled in the exponential phase. The cell densities were measured before establishing the pigment content using HPLC.

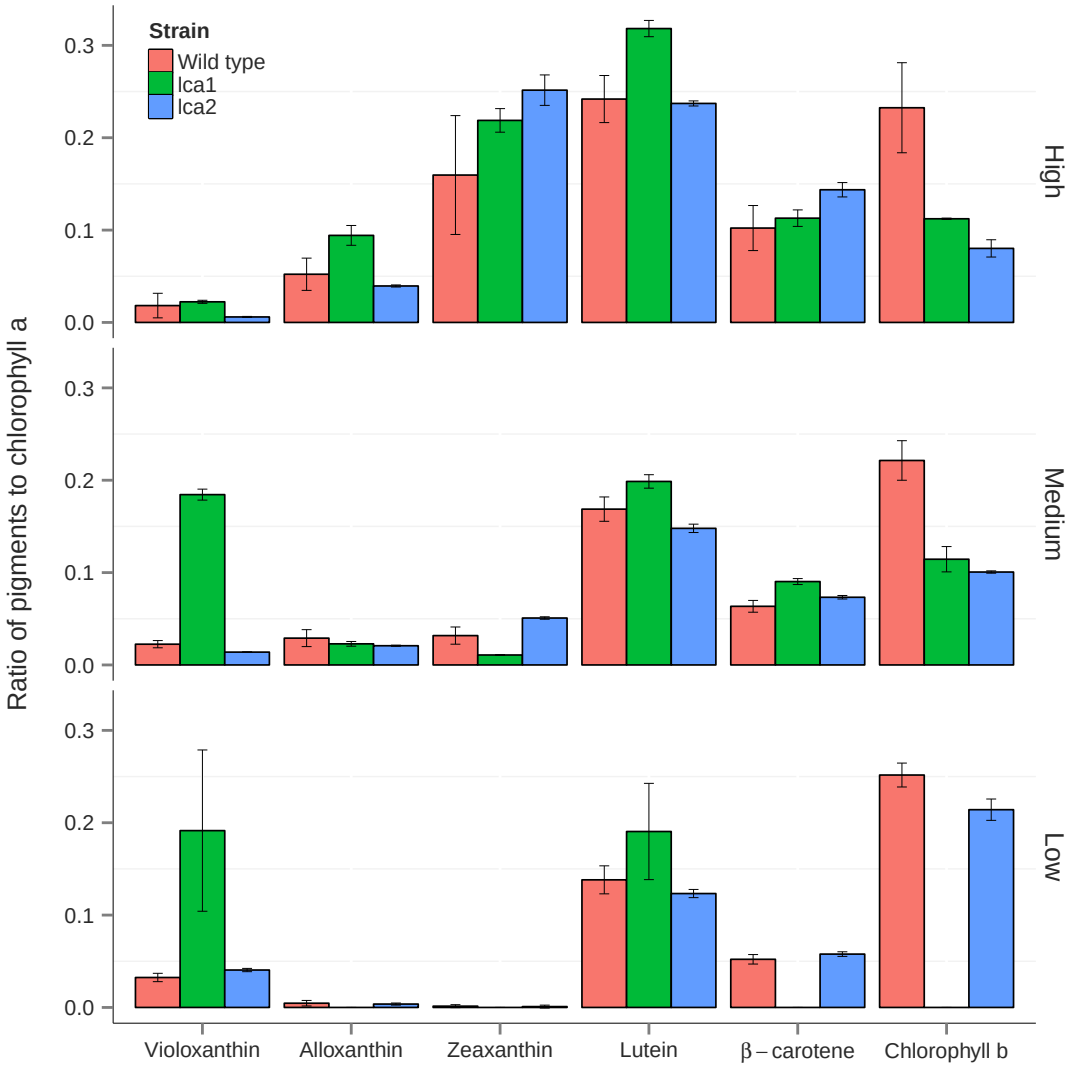


FIGURE 3.11: The pigment composition in the wild type, lca1 and lca2 strains grown under low, medium and high light (20, 100, 1200 µmol photons m^{-2} s^{-1}) normalized to the chlorophyll a content. Error bars are one standard deviation from biological triplicates. Based on data from P. Stephenson.

The original result from the measurement (see Figure A.2 in Appendix A) was normalized on a chlorophyll a basis to show the relationship between chlorophyll a and b as well as to the accessory pigments. As shown in Figure 3.11, the individual accessory pigments and chlorophyll b makes up a fraction of the total pigment content with ratios typically below 0.3 of the chlorophyll a content.

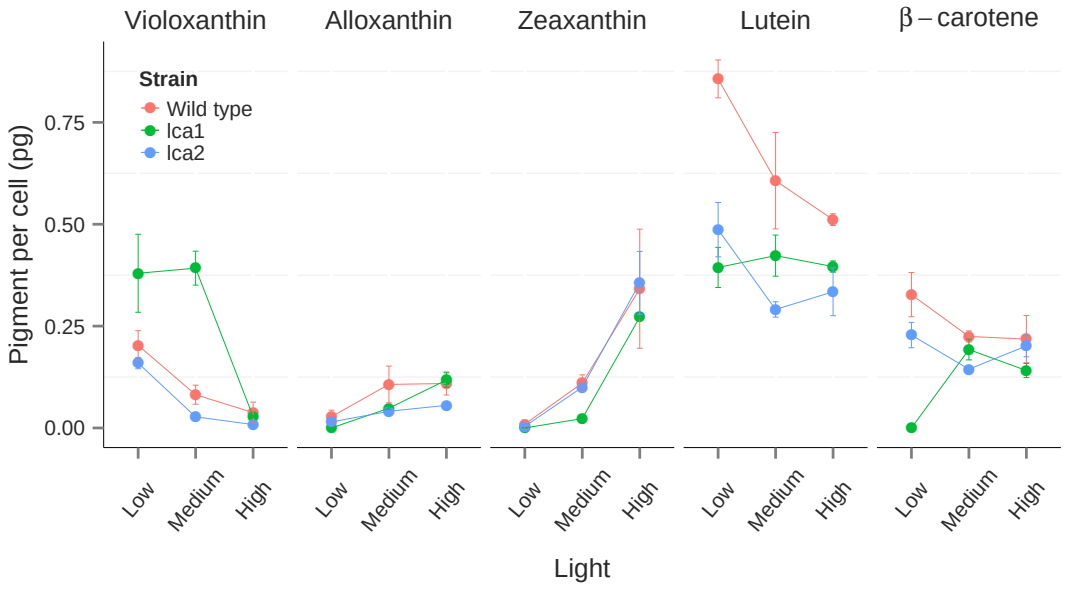


FIGURE 3.12: The amount of accessory pigments (violoxanthin, alloxanthin, zeaxanthin, lutein and β-carotene) in the wild type, lca1 and lca2 strain grown under low, medium and high light (20, 100, 1200 µmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates.

The most apparent difference in the amount of accessory pigment when normalized to the amount of chlorophyll a on a strain level is the significantly (ANOVA, $p < 0.05$) higher violaxanthin content in the lca1 strain under low and medium light compared to both the wild type and the lca2 strain. Violaxanthin is one of the main pigments in the xanthophyll cycle that helps dissipate (quench) excess energy under high light (Jin et al., 2003, Niyogi, 1999). However it has also been suggested that carotenoid pigments have a structural role and that violaxanthin may inhibit chlorophyll aggregation and thereby indirectly hinder the formation of the PSII LHC (specifically LHCIIb) (Phillip et al., 1996). Considering the generally low or even lacking amounts of chlorophyll a and b in the lca1 strain, these finding may be related. Figure 3.12 shows the amount of accessory pigments in the strains without normalization. The amount of violaxanthin under high light compared to medium light is 14 times lower in the lca1 strain compared to 2 and 3 times lower in the wild type and lca2 strains suggesting an underlying difference in the lca1 strains use of accessory pigments. In contrast, the seemingly high lutein content in the wild type compared to lca strains (Figure 3.12) is not evident when normalized to the chlorophyll a content (Figure 3.11).

TABLE 3.4: The chlorophyll a:b ratio in the wild type, lca1 and lca2 strains under low, medium and high light (20, 100, 1200 µmol photons m^{-2} s^{-1}). Errors are one standard deviation from biological triplicates.

Parameter	Strain	Light	Mean
Chlorophyll a:b ratio	Wild type	Low	3.98±0.20
		Medium	4.55±0.43
		High	4.43±0.73
	lca1	Low	N/A
		Medium	8.82±1.10
		High	8.91±0.06
	lca2	Low	4.68±0.25
		Medium	9.94±0.12
		High	12.6±1.55

The measured amounts of chlorophyll a and chlorophyll b from the HPLC measurements of the pigment content were used to calculate the chlorophyll a to b ratios (Chl a:b ratio) shown in Table 3.4. A light driven shift in ratio from low to high light was only observed in the lca2 strain. The change is significant (ANOVA, $p < 0.05$) on a light level, with significant differences in the lca2 strain between light levels. In contrast, the change in ratio by light level is not significant in either the lca1 or the wild type. The change in the chlorophyll a:b ratio on a strain level compared to the wild type was significant in both lca1 and lca2. The difference was significant in both lca1 and lca2 when compared to the wild type under medium and high light but not under low light in lca2. Since no chlorophyll b was found in the lca1 strain under low light, the corresponding chlorophyll a:b ratio could not be calculated for this light level.

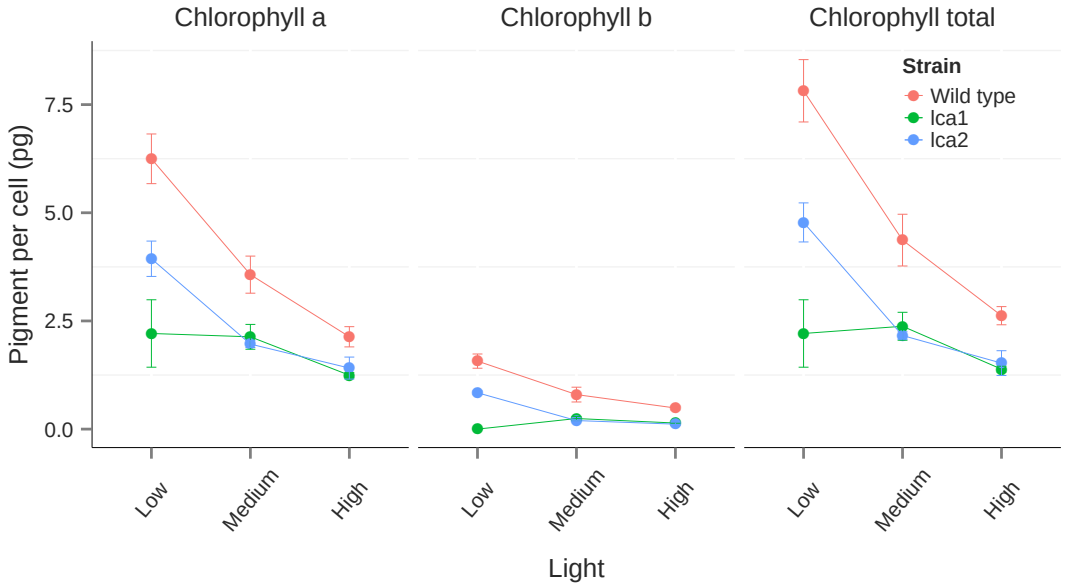


FIGURE 3.13: The amount of chlorophyll a, chlorophyll b and total chlorophyll in the wild type, lca1 and lca2 strains grown under low, medium and high light (20, 100, 1200 µmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates.

Also evident (Figure 3.13) is a significant (ANOVA, $p < 0.05$) light driven response under low to medium light in the amount of chlorophyll a and b in the wild type and lca2 (see Appendix A, Figure A.2). The same response is again lacking in lca1, while significant in the lca2 strain, in support of earlier observations for chlorophyll a (3.4). The wild type and lca2 strain show a similar 3 fold (3.0 and 3.1) increase in the average total chlorophyll compared to 1.6 times in the lca1 strain. The similarity of the response (only offset by the levels of chlorophyll) in the wild type and lca2 strain could suggest that similar mechanisms are responsible for this light driven reduction in chlorophyll content while the same mechanism could possible be either missing or strongly reduced in the lca1 strain.

3.4 Long-term stability of lca strain photophysiology

3.4.1 FRRf-based long term monitoring of stock cultures

The stability of the lca strains over time was monitored regularly using FRRf by measuring the functional cross section (σPSII) and the photosynthetic efficiency of PSII (F_v/F_m) to allow for early discovery of possible reversions in lca photophysiology resulting in a phenotype more closely associated with typical wild type traits.

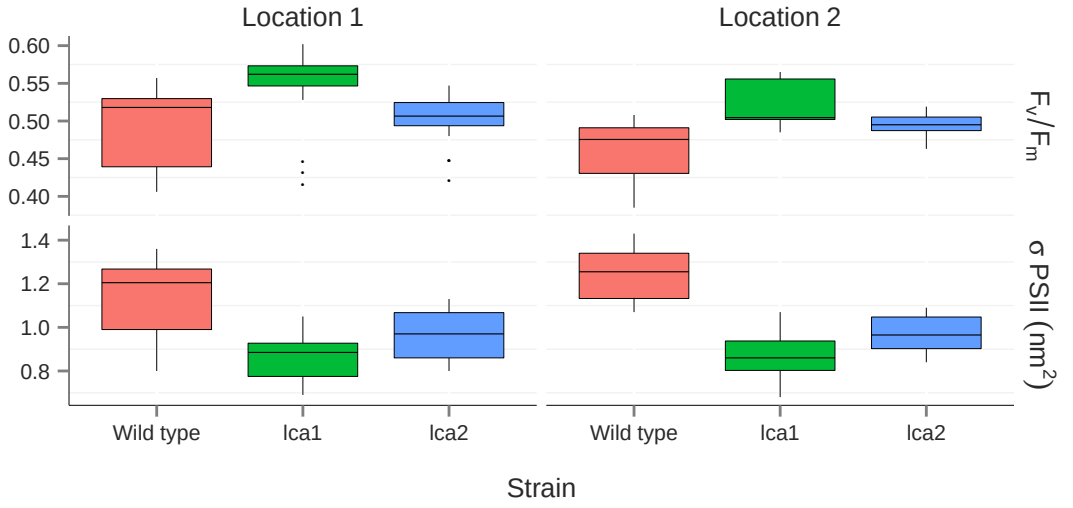


FIGURE 3.14: The difference in photosynthetic efficiency (σPSII) and the functional cross section (F_v/F_m) between the mutant stains and the wild type as measured in stock cultures maintained under medium (100 µmol photons m^{-2} s^{-1}) light in two separate locations from September 2011 to October 2014 and using representative data from monitoring results. Errors are one standard deviation from six measurements in total.

Some variability in σPSII and F_v/F_m was observed in both storage locations but over time cultures grown in both storage locations proved remarkably stable. The difference was significantly different (ANOVA, $p < 0.05$) in location 2 (temperature controlled growth room) with a lower σPSII and higher F_v/F_m in both lca strains. The result for location 1 (temperature controlled incubator) was only significant for lca1. It is worth noting that for all experiments for specific parameters in the growth phase the stock cultures were sourced from location 2. The result would also suggest that the core physiological differences observed in the lca strains compared to the wild type have been maintained throughout this 3 year period, with no observable reversion in phenotype based on these parameters.

3.5 Discussion

The lca1 strain is characterized by a small PSII functional cross section with a low degree of variability under changing light intensities, as seen in Figure 3.5. This is also reflected in the total chlorophyll a content per cell with modest changes under changing light conditions (Figure 3.4). The photosynthetic efficiency measured by F_v/F_m (Figure 3.6) remains equal or higher compared to the wild type. The average electron transport rate of PSII as well as the calculated oxygen evolution per molecule

of chlorophyll a (Figures 3.8 and 3.7) was markedly higher in the lca1 strain compared to the wild type under high light conditions. Compared to the wild type, the lca1 strain also has a strong reduction in the number of PSI reaction centres under medium light (Figure 3.10). The lca2 strain shares a smaller σPSII (Figure 3.5) and amount of chlorophyll a per cell (Figure 3.4) with the lca1 strain, but while both of these parameters remains fairly stable in lca1 they vary in regards to the light intensity in the lca2 strain. The σPSII and the chlorophyll a per cell decreases in response to increasing light intensity, coupled with corresponding decrease in the chlorophyll a:b ratio (Figure 3.4). The measured rates of electron transport and oxygen evolution per chlorophyll a are both higher than in the wild type under higher light conditions but without reaching the values seen in lca1. Similarly to the lca1 strain the lca2 strain also has significantly fewer PSI reaction centres compared to the wild type (Figure 3.10).

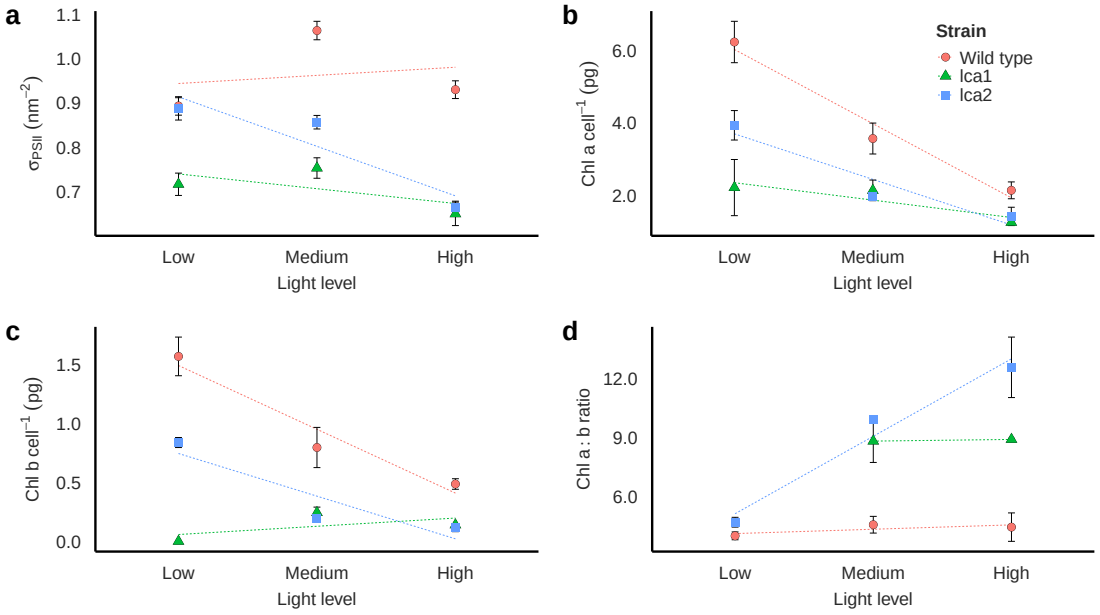


FIGURE 3.15: The changes in a) the size of the functional cross section (σPSII), b) the chlorophyll a content, and c) the chlorophyll b content and d) the chlorophyll a:b ratio in the wild type, lca1 and lca2 strain grown under low, medium and high light (20, 100, 1200 μmol photons m^{-2} s^{-1}). Trended changes (dashed line) were calculated from (least square) linear regressions to the average result for biological triplicates under each light level. Error bars are one standard deviation.

3.5.1 Pigment distribution in the antenna

The chlorophyll a:b ratio measured under low (3.98 ± 0.20) and medium (4.55 ± 0.43) light in the wild type (table 3.4) is slightly higher than the numbers reported by Falkowski and Owens (1980a) under low (2.30) and medium (3.0) light but still comparable to earlier reported results in *D. salina* (Jin and Yokthongwattana, 2003, Smith et al.,

1990, Webb and Melis, 1995). More noteworthy here is the lack of a significant increase in the chlorophyll a:b ratio in both the wild type and the lca1 strain (medium and high light) when grown under higher light intensities (Table 3.4). The lca2 strain on the other hand is shows a typical light driven response with a significantly smaller chlorophyll a:b ratio under low light compared to medium light and high light (Table 3.4). The results also confirm the significant difference between the wild type and lca1 and lca2 on a strain level with an average chlorophyll a:b ratio of 4.32 in the wild type, compared to 8.87 and 9.07 in the lca1 and lca2 strains. This confirms that the chlorophyll antenna in both of the lca strains is significantly smaller under medium and high light compared to the wild type. The lack of measurable chlorophyll b in the lca1 strain under low light would also suggest that the LHC antenna is either completely lacking or very small under this condition.

In contrast to the chlorophyll a:b ratio, the chlorophyll a, per cell in the wild type does show a light driven response, with significantly more chlorophyll a under low light compared to medium and high light (Figure 3.13). A similar light driven and significant change is seen in the lca2 strain when comparing chlorophyll a amounts under low and high light. The trended change by light as seen in Figure 3.15 clearly shows that the wild type and lca2 strains have very similar responses in chlorophyll a and b content under changing light while the same response is seemingly missing in the lca1 strain.

3.5.2 Functional cross section of photosystems

The functional cross section of PSII under light (σ'PSII) has been shown to decrease slowly under increasing light in *D. tertiolecta* (Fujiki et al., 2007). This is similar to the pattern observed in the wild type and the lca2 in the FRRf based Photosynthesis vs. Irradiance (Figure 3.8) measurements. The opposite pattern in the form of a slow increase in σ'PSII was observed in the lca1 strain with the highest σ'PSII measured at 2000 μmol photons m^{-2} s^{-1}. It has been suggested that the decrease in σ'PSII is caused by a combination of non-photochemical quenching (NPQ) involving the xanthophyll cycle and state-transitions (Baker and Bowyer, 1994, Fujiki et al., 2007).

The trended change in the PSII functional antenna (σPSII) (Figure 3.15) suggests that the wild type and the lca1 strain only change the size of the antenna marginally in response to changing light conditions. This is in contrast to the lca2 strain that shows a clear light response with a smaller antenna under increasing light. The measurements of σPSI showed no significant differences in the size of the functional antenna of PSI. The photosynthetic efficiency of PSII (as measured by F_v/F_m) remained

fairly stable throughout the experiments with only a significantly higher F_v/F_m in the lca1 strain compared to the wild type under medium light. The sharp and significant drop in photosynthetic efficiency seen in the wild type and the lca2 strain under medium light compared to low light could be partially related to the increase in σPSII. The overall change in F_v/F_m between light levels is comparable to previously published results (Fujiki et al., 2007).

3.5.3 Non-photochemical quenching

Measurement of the PSII electron transport rate (ETR) using FRRf based rapid light curves, a form of photosynthesis-irradiance curves, presented in section 3.3.2, shows that under the investigated experimental conditions lca1 displayed a greatly restricted ability to undertake NPQ (see Section 1.1.4 for a description) compared to both wild-type and lca2. Both wild-type and lca2 show clear signs of NPQ with a drop in both the minimum F'_o and maximum fluorescence under actinic light F'_m, suggesting activation of a NPQ mechanism and the closure of the number of active PSII at an irradiance around 1000 μmol photons m^{-2} s^{-1}. No similar response in lca1 is observed and F'_o keeps increasing almost linear relative to the irradiance. Subsequently, ETR plateaus out in both wild-type and lca2 around 1000 μmol photons m^{-2} s^{-1} but keeps increasing in lca1, suggesting that the lack of NPQ is not sufficient to result in photoinduced damage at levels high enough to offset the continued increase in electron transport during the time-scale of the experiment, which is apparent on the basis of the continued closure of PSII reaction centres (increase in F'_o).

It has been shown in wild-type *D. tertiolecta* that the inherent (constituent) component of NPQ dominates in the species with a fast response (within seconds) of light exposure (Ihnken et al., 2011). It has also been shown that *D. tertiolecta* displays a unique NPQ with rapid down-regulation under high light (Ihnken et al., 2011). The xanthophyll cycle plays a significant role in *D. tertiolecta* as part of the slower component of NPQ; Casper-Lindley and Björkman (1998) showed that the the violaxanthin pool after 45 min of exposure to high light (1200 μmol photons m^{-2} s^{-1}) is still only partially de-epoxidised to 60 %, and therefore functional, while maximum NPQ was detected after 15 min of exposure. These results show that the xanthophyll cycle in *D. tertiolecta* is normally functional and operating effectively under light levels well above the one observed resulting in saturation of photosynthesis (I_k) based on the calculated ETR in both wild-type (543.9 μmol photons m^{-2} s^{-1}) and lca2 (652.8 μmol photons m^{-2} s^{-1}) (Table 3.3). Interestingly, lca1 also shows an irregular accessory pigment profile with high levels of violaxanthin under low and medium light compared to wild-type

and lca2, but not under high light (see Figure 3.12). This may suggest that the xanthophyll cycle has been affected, either directly, or indirectly by the mutations.

3.5.4 Photosystem stoichiometry

The results from the immunoblots in cells grown under medium light shows that the lca1 strain has a roughly four time reduction in PSI compared to the wild type (see Table 3.10). Assuming an approximate 1:1 PSII:PSI ratio under medium light in the wild type, in line with previously published results (1.23 in *D. salina* at 100 μmol photons m^{-2} s^{-1}), would suggest a PSII:PSI ratio of 4:1 this is higher than the 3:1 ratio reported in *D. salina* under very high light (1700 μmol photons m^{-2} s^{-1}) (Jin and Yokthongwattana, 2003, Smith et al., 1990). The stoichiometric distribution of photosystems (PSII/PSI ratio) is light dependent with the number of PSII more than doubling compared to PSI in *D. salina* when cells are grown under high light compared to medium light (Smith et al., 1990). Under high light (2000 μmol photons m^{-2} s^{-1}) the D1 protein at the centre of PSII has a half-life of only 20 min in *D. salina* and it is likely that the capacity of the repaircycle is the rate-limiting step in PSII under high light (Kim et al., 1993b). However, the reduction in the functional cross section of PSII (σPSII) in the lca should result in less photoinduced damage to the reaction centres, negating the need for excess PSII to PSI reaction centres. The high PSII:PSI ratio could be a response in the lca1 strain to a change caused by the mutation(s) that has resulted in a high light phenotype. The excess PSII to PSI would be the natural response of the cells to excess light conditions. The high PSII:PSI ratio seen in the lca1 strain could possibly be a compensatory response to the small functional cross section of PSII, as previously seen in chlorophyll b deficient (soya bean) mutants (Ghirardi and Melis, 1988), however, the difference seems too large to explain the ratio fully. It has also been suggested that accessory pigments senses the spectral quality of the light and influence the PSII:PSI ratio accordingly. Strains with a substantially altered pigment composition (chlorophyll b or phycobilisome deficient) have been shown to lack this response (Kim et al., 1993a).

3.5.5 Predicted photosynthetic efficiency

Ort and Melis suggested that four different targets or parameters need to be met for truncated light antenna (tla) strains to outperform wild type strains, "To ensure that Tla strains will operate with improved solar energy conversion efficiency, the following analytical tests are useful: (1) the functional PSII and PSI Chl antenna size should be determined to assess the extent of reduction in antenna size, (2) the titer of functional

PSII and PSI reaction centers per cell should be determined to ensure that the density of the photosynthetic electron-transport chains remains unaffected, (3) the quantum yield of photosynthesis must remain high to ensure that the efficiency of photosynthesis under limiting light remains unchanged, and (4) the light-saturated rate of photosynthesis should be inversely proportional to the measured Chl antenna size" (2011). Comparing these targets to the physiological results presented here should help clarify which parameters have likely helped improve the photosynthetic efficiency of the lca strains and which ones might have imparted a negative impact on the performance.

TABLE 3.5: The relative percent change in functional PSII and PSI antenna size, proportion of reaction centres per cell, quantum yield of photosynthesis (P^f (PSII) P_{max}) and the rate of photosynthesis (O_2 evolution P_{max}) compared to the wild type suggested to impact the photosynthetic conversion efficiency.

Question	lca1 (% of WT)	lca2 (% of WT)	Parameter
Reduction in antenna size	70.8/(100)	80.6/(100)	σPSII/σPSI
Reaction centers per cell	99.2/23.0	72.1/61.7	PsaC/PsbA
Quantum yield of photosynthesis	168	111	P^f (PSII) P_{max}
Rate of photosynthesis	289	131	O_2 evolution P_{max}

The strong reduction in lca1 PSI reaction centres (Table 3.5) results in a PSII:PSI ratio normally associated with very high light acclimated cells (as discussed in section 3.5.5). Coupled with the strongly reduced ability of the strain to adjust the functional cross section of PSII (σPSII) and chlorophyll content per cell (Figure 3.15), this phenotype is likely to show optimal growth within a more narrowly defined range of light conditions compared to the wild type. The large increase in the P_{max} as measured both by the electron transport rate and the chlorophyll a normalized rate of oxygen evolution (Table 3.5) allow the strain to capture light energy more efficiently under optimal conditions. Considering that the amount of chlorophyll a in the lca1 strain under all light conditions is below the amounts measured in the wild type under high light, the lca1 strain shows signs of a locked-in high light adjusted phenotype. The changes in the lca2 strain are more restrained, with a small increase in P_{max} coupled with a small and balanced reduction in photosystem reaction centres (Table 3.5) compared to the wild type. The strain also retains a strong light response more similar in nature to the wild type. However, the varying chlorophyll a:b ratio (Figure 3.15) in the lca2 strain compared to a similar ratio under all light levels in the wild type suggest different modes of adjustments. Similar chlorophyll a:b ratios under all light levels in the wild type suggests that the change in chlorophyll a and b is caused not

by a reduction in functional antenna size but rather by the total amount of photosystems per cell. This hypothesis is supported by the measurements of σPSII that remain largely similar under all light levels in the wild type (Figure 3.15). The opposite pattern is observed in the lca2 strain with a strong reduction in σPSII and increase in the chlorophyll a:b ratio under increasing light. This is consistent with changes resulting from a reduction in the size of the peripheral (LHC) antenna. The genetic changes that could account for these observed differences in phenotype and photophysiology between the strains are discussed in Chapter 5 in conjunction with the results from the gene expression analysis.

3.5.6 Continuous growth experiments

Based on the results from the physiological characterisation of the lca strains showing a reduced chlorophyll content per cell and smaller functional PSII antenna (summarized in Figure 3.15), together with the suggested increase in PSII electron flow and oxygen evolution on a per chlorophyll basis from the calculated P_{max} of 168-289 % in lca1 and 111-131 % in lca2 (Table 3.5) it was predicted that the lca strains had the potential to out-compete the wild-type under high light and high cell density conditions. To measure the biomass yield under continuous growth the strains were grown in photobioreactors as described in section 2.2.3 in Chapter 2.

During the trials, the lca1 strain failed to stabilize at growth levels that allowed for continuous dilution of the media and was therefore ultimately removed from any further experiments, while lca2 and wild-type both yielded acceptable and relatively stable growth over time. The reason for this failure is unknown, but could include an increased sensitivity to high light from the lack of observed NPQ, resulting in increased cell death or other complications from the mutations. The Chl a:b ratio in lca2 varied throughout the experiments, during the first experiment it reached levels similar to the wild-type or even lower towards the end of both experiment (Appendix A.3), while during the second experiment which resulted in far higher biomass yields (Appendix A.5a and A.5b) the Chl a:b ratio remained higher than WT (Appendix A.4), as would be expected from a reduced LHC resulting in a lowered chl b binding capacity, during the majority of the experiment. The result in terms of productivity during the first experiment was encouraging with a higher average productivity of 113% of wild-type in one of the lca2 replicates, when measured against the wild-type replicate with the highest productivity. However, the average was still reduced to below wild-type levels, resulting in no significant productivity increase, due to the result from the second lca2 replicate that only reached 56% of the average wild-type maximum (data not shown). This low level of growth likely results from a temporary problem with the individual

cultures, for this reason a second experiment was devised. However, after initial problems with stabilizing the optical density at a predetermined level (2 μg Chl/mL) the light was switched from a sinusoidal light curve to full light, resulting in different culturing conditions. This did result in a doubled productivity but again no significant improvement in the biomass yield. While no significant improvement in productivity could be observed in lca2 under simulated day-light (sinusoidal light curve, 0 - 2000 μmol photons m^{-2} s^{-1}) or using continuous high light (2000 μmol photons m^{-2} s^{-1}), it remains too early to conclude whether or not either of the lca strains could outperform the wild-type under the high light, high cell density conditions that are needed for useful improvements in biomass yield.

3.5.7 Cryopreservation of lca strains

One of the main obstacles that needs to be overcome during the creation of new strains is how to retain or minimize reversion towards a wild-type phenotype. The stability of the mutants depends on the methods used. The original ALBINO3 mutant was shown to be caused by the insertion of a Ds transposon in the affected gene, with a high degree of phenotypic reversion in the presence of the TPase (transposase) gene (Sundberg et al., 1997). Similar, but very low frequency reversions were also observed in *cao/chaos* affecting CpSRP43, which is also caused by insertion of a Ds transposon (Klimyuk et al., 1999). During the development of EMS or gamma radiation mutagenized *Nannochloropsis sp.* strains with altered fatty acid composition only one mutant (JS1) resulting from the gamma ray treatment was shown to be stable out of 20 isolated mutants during a re-screening. However, no further reversion was observed over time in JSI, suggesting a stable phenotype (Schneider et al., 1995). The lca phenotype remained stable during the course of this project, but the possibility of a reversion or contamination always needs to be taken into consideration. Typically, this is addressed by storing the strain in multiple locations as well as by arresting the growth if possible using cryopreservation. Initial trials of cryopreservation using methanol as the cryopreservant (described in Chapter 2, section 2.9.1) resulted in extensive rupturing of cells during the resurrection phase and successful resurrection was limited to wild-type and lca2. This is likely exaggerated in *Dunaliella* by the lack of a cell wall that could help stabilize the outer membrane during freezing. However, later trials using an alternative method involving the encapsulation of the cells in a calcium alginate matrix (Chapter 2, section 2.9.2) resulted in successful cryopreservation of wild-type, lca1 and lca2, albeit at a low survival rate in lca1 based on the long delay (>14 days) in observed growth in resurrected cultures. While resurrection of these cultures from the first batch of calcium alginate encapsulated cells succeeded on multiple occasions,

further attempts of repeating the method using the same conditions failed, which suggests that additional fine-tuning of the protocol is needed. The results show the need for robust methods that can accommodate strains with varying ability to withstand the cryopreservation process.

3.5.8 Conclusion

Table 3.6 show the results from the in-depth photophysiological characterisation of lca1 and lca2 for the three key-parameters, chl cell^{-1}, σPSII and F_v/F_m. The measured chl cell^{-1}, the single most important parameter during the selection of the mutant cells, remain clearly reduced compared to wild-type, regardless of light level, and range from 18-70% of wild-type in lca1 and 57-94% of wild-type in lca2.

TABLE 3.6: The measured chl cell^{-1}, σPSII and F_v/F_m in % of WT, under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}) in lca1 and lca2.

Strain	Light	Chl cell^{-1} (% of WT)	σPSII (% of WT)	F_v/F_m (% of WT)
	Low	18	80	101
lca1	Medium	52	71	110
	High	70	70	101
	Low	57	99	98
lca2	Medium	94	93	101
	High	70	71	102

The small functional antenna of PSII compared to wild-type in both lca1 (70%) and lca2 (71%) under high light is critical from the perspective of increasing the photosynthetic conversion efficiency, the core objective of this project, by helping to reduce excessive light harvesting at the surface of a dense cell culture. The negative effects from over-absorption of light in terms of increased waste of solar energy, either by heat or fluorescent light from NPQ mechanisms and the resulting photoinhibition from photoinduced damage to reaction centres, will be most severe under high light when photosynthesis is fully saturated (Stephenson et al., 2011).

The ability to photoacclimate and increase the size of the PSII antenna under low and sub-saturating light, as seen in lca2 with a σPSII of 99% of wild-type under low light, and should be beneficial in stagnant environments with a strong light gradient from top to bottom within the water column, while under turbulent conditions leading to strong fluctuations in light intensity a constitutively small antenna, as in lca1, is likely

to be beneficial (Moore et al., 2006). The F_v/F_m parameter measuring the photosynthetic efficiency of PSII remained near wild-type in lca2 (98-102%) or slightly higher in lca1 (101-110 %) similar to the results from the original measurements with 106% of wild-type in lca1 and 104% of wild-type in lca2 (Chapter 2, Table 2.11) during the selection process, again suggesting a well preserved phenotype and non-impaired photosynthesis in both lca1 and lca2.

Chapter 4

Construction of transcriptome analysis pipeline

4.1 Introduction

The observed changes in the lca strain phenotype discussed in Chapter 3 are driven by underlying changes in the strain genomes induced by the random mutagenesis. These changes are likely to affect the expression of nuclear encoded genes, with the resulting down- or upregulation of associated protein levels leading to the down-stream effects. The overall cellular levels of gene expression can be assessed by generating transcriptomes, in effect snap-shots of the cells internal levels of messenger RNA at a given time. For this reason, transcriptome libraries were generated for the lca strains and compared to the wild type. Since the main effects of interest are light driven, the transcriptomes were also generated under three different light levels: low, medium and high light, corresponding to 20, 100 and 1200 μmol photons m^{-2} s^{-1}. The transcriptomes were generated using the SuperSAGE protocol, a technology that generates short (26 bp) reads representing full mRNA sequences, in triplicate for a total of 27 libraries (wild type, lca1 and lca2 strains under low, medium and high light). The development of a pipeline for the bioinformatical analysis of these transcriptomes leading to the identification of statistically significant differentially expressed genes is discussed and described in this chapter.

4.1.1 Overview of the SAGE methodology

The SAGE (Serial analysis of gene expression) family of technologies are used to identify genes (gene discovery) and to quantify their expression (expression profiling) and

have been in use in different forms since the development of the technique in 1995 (Harbers and Carninci, 2005, Velculescu et al., 1995). Serial analysis refers to the use of concatenation of shorter sequences to form longer artificial sequences that in turn can be read in a single stretch to allow for more efficient sequencing. At the core of the technology is the use of anchoring and restriction enzymes to capture short sequences (tags) of mRNA that are then used to represent the full cDNA sequence. More recent variants of the technology such as high-throughput SuperSAGE still use the same principle but the use of concatenation of reads to improve the speed of sequencing has been advanced with the use of so-called next generation sequencing (NGS). This vastly improves the sequencing capacity and in effect eliminates the need for concatenation of sequences (Matsumura and Molina, 2012, Matsumura et al., 2003, 2010a, Morozova and Marra, 2008). While many of todays large-scale sequencing projects use variants of the related RNA-Seq technology, SuperSAGE still remains relevant and there are advantages and weaknesses of both technologies (Wang et al., 2009). Since the onset of the technology in 1995, it has been improved to capture longer reads to improve the precision of the technology (Harbers and Carninci, 2005, Matsumura et al., 2003).

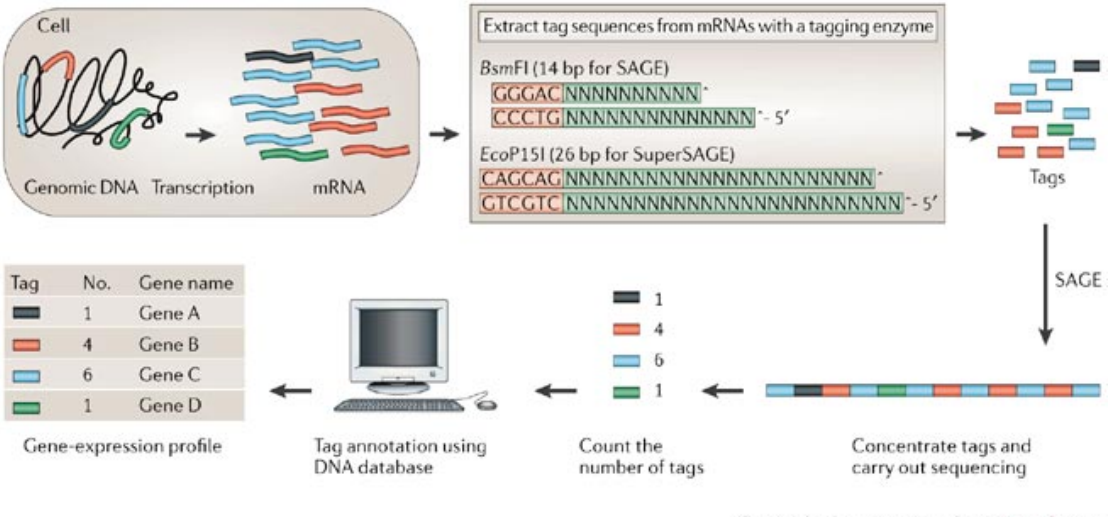


FIGURE 4.1: Simplified representation of the steps involved in analysing SAGE and SuperSAGE transcriptomes. Restriction enzymes extract short representative sequences (tags) from cellular mRNA molecules. The tags are concatenated into longer reads and sequenced. Computer-based analysis is then used to annotate and count the tags to generate a gene expression profile. Figure from Huang et al. (2008).

4.1.2 The SAGE family of gene expression technologies

While many variants of the SAGE family of sequencing technologies exist, a representative technical description of the early versions would involve the following procedures. RNA extracted from cells is used to synthesise cDNA using biotinylated oligo(dT)

primers to select for mRNA. The resulting double-stranded cDNA is digested using a restriction enzyme that cleaves directly 5′ of a 4-bp recognition site (CATG) (Patino et al., 2002). The remaining 3′ portion of the cDNA with the CATG tag at the 5′ end is then recovered by binding to streptavidin-coated beads using affinity purification (Madden et al., 2000). Alternatively, the RNA can be synthesized directly onto beads coated with oligo(dT) primers (Patino et al., 2002). The sample is divided into two separate portions. An adapter sequence with a recognition site for a tagging enzyme (TE), as a primer sequence including a 4 bp cohesive overhang complementing the CATG recognition site are added and ligated to the cDNA. The tagging enzyme (a Type IIS restriction endonuclease) will then digest the cDNA a defined number of bp 3′ of the recognition site and form the SAGE tag. The length of the tag is mainly what sets the early (Sanger sequencing based) SAGE protocols apart and this is directly dependent on the tagging enzyme used. The two samples are mixed and the ends with 2 bp overhangs are repaired using a DNA polymerase and ligated together using a T4 DNA ligase, forming so called ditags. The ditag with an adaptor sequence at each end is PCR amplified using primers recognizing the primer sequences that form part of the adapter sequences. The amplified product is again cleaved using the NlaIII restriction enzyme, resulting in two tags attached back to back with a CATG overhang on each end. These ditags are now ligated together using a T4 DNA ligase forming a polymer of tags. Polymerized tags above a certain size threshold (>500 bp) are selected from a PAGE gel, cloned into a plasmid vector and transformed into *E. coli*. The plasmids insert are finally sequenced.

The original SAGE protocol used enzymes such as FokI or BsmFI that cleave 14 or 15 bp 3′ of the recognition sequence, respectively (Madden et al., 2000). This includes the 4 bp recognition site for the anchoring enzyme, leaving 10-11 bp of the original cDNA. Further developments resulted in the longSAGE protocol (Saha et al., 2002) that uses the MmeI restriction enzyme forming 21 bp tags, as well as SuperSAGE that extends the tag length to 26 bp using EcoP15I (a type III restriction endonuclease) (Matsumura et al., 2003). It is worth noting here that any restriction enzyme can be used. The common use of restriction enzymes with a 4 bp recognitions sequence is based on the calculation that any sequence of this length should occur every $4^4 = 256$ bp on average on a random sequence. This is well below the normal length of genes both in prokaryotes (900-1000 bp, Casjens (1998)) and in eukaryotes. In eukaryotes the gene length varies widely with the species but is commonly larger than the 1700 bp average length seen in yeast (*Saccharomyces cerevisiae*), the 2200 bp average in *Arabidopsis thaliana*, and well below the 4300 bp average in *Chlamydomas reinhardtii* or the 7 to 8 kbp average length in the mouse or human genomes (Hou and Lin, 2009,

Jareborg et al., 2000, Moriyama and Powell, 1998, Peers and Niyogi, 2008, Wortman et al., 2003).

4.1.3 Further improvements using next generation sequencing

With the advent of next generation sequencing that allows for parallel processing of very large number of sequences a new protocol was developed. This high-throughput SuperSAGE protocol (deepSuperSAGE) retains some of the best features of the original SAGE while simplifying the protocol (Matsumura and Molina, 2012, Matsumura et al., 2003, 2010a). The new protocol is almost identical to the original SAGE method up until the formation of ditags (if sequenced on the 454 family of sequencers) or just after the formation of individual tags if using the Illumina or SOLID type of shorter read systems. In the latter case, a second adapter sequence is ligated to the $3'$ end of the tag. In both cases modified adapter sequences harbouring unique synthetic sequences (barcoding) is included in the adapter to allow for the pooling of samples from multiple sources (multiplexing). The ditags or single tags with adapter sequences at both ends are then simply PCR amplifed and purified before sequencing, which removes a number of steps in the protocol while improving the overall speed of the protocol. More importantly this protocol also improves the precision by lowering the number of PCR cycles used in the final step from around 25 cycles down to 5 to 10 cycles, which could otherwise potentially induce a bias due to preferential amplification of amplicons (Matsumura and Molina, 2012, Terauchi and Matsumura, 2008). This evolved SuperSAGE protocol was used to generate the transcriptomes for the analysis of the lca strains.

4.2 Results

4.2.1 SAGE transcriptome library overview

The full dataset consisting of 27 sequenced libraries covering the wild type, lca1 and lca2 strains under low, medium and high light (20, 100 and 1200 μmol photons m^{-2} s^{-1}) in triplicates was processed at the sequencing facility at Iwate Biotechnology Research Centre, Kitakami, Iwate, Japan to remove the adapter sequences and extract the SAGE tag. The size of the tags extracted by the EcoP15I restriction enzyme varies between 26 and 27 bp. For this reason the tag length is normally corrected to 26 bp $3'$ of the recognition site. This is done in silico (using computer-based analysis) and the number of matches for each 26 bp tags is then counted for each library and presented in a single file containing the results for all the libraries (see Appendix B, Table B.1

for an example from the full data set) (Matsumura et al., 2011). The tags found in multiple copies are commonly referred to as unitags in the literature and this term is used in this thesis when appropriate. To clarify for the reader, the unitags are simply the sequences and a single count of a sequence is referred to as a tag.

The size of each of the 27 libraries varied between 801840 to 1984453 tags (counts) with a mean of 1197451 tags per library distributed over 382846 unitags in total. Not all unitags were found in all libraries and the number of unitags with at least one tag count varied between 42824 and 85639 with a mean of 55773 observed unitags per library. Comparing this average number to the total number of observed unitags showed that on average less than one in six unitags was observed in each library (see Table B.2 for more information). This is related to the fact that unitags with very low counts make up the majority of all observed unitags.

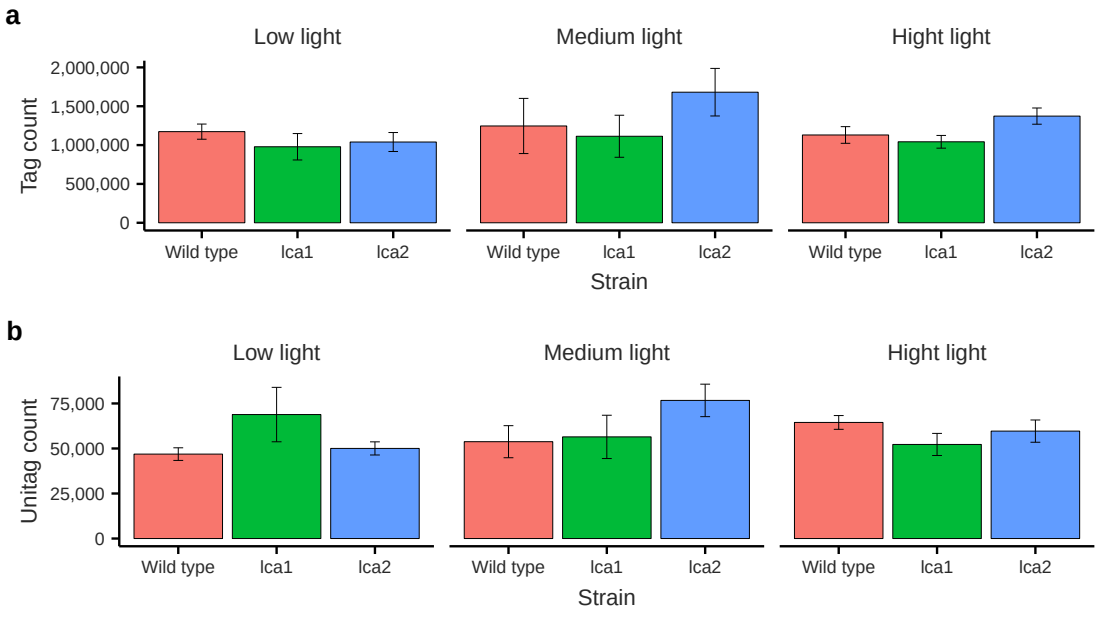


FIGURE 4.2: The mean unprocessed SAGE transcriptome library sizes (a) tag and b) unitag counts) from three replicates each of each of the nine experimental conditions under low, medium and high light for the wild type, lca1 and lca2 strain. The error bars are one standard deviation (n = 3).

The difference in the mean unitag count and sum of tag counts from the three replicates for each condition as shown in Figure 4.2 shows that, while the majority of the libraries have a mean unitag count just above 50000, a few conditions have noticeable higher total unitag counts. This includes the lca1 strain under low light, the lca2 strain under medium light and to some extent the wild type under high light. What is also apparent is that the sum of the tag counts in the lca2 strain under medium light and also under high light is higher than the mean sum of around 1 million tag counts

per library. Also, the lca1 strain has a slightly lower count on average under all conditions. Broken down on a strain and light level (see Figures 4.3 and 4.4) any differences largely disappear, indicating that the differences are present on a single library level rather than on a strain or light level.

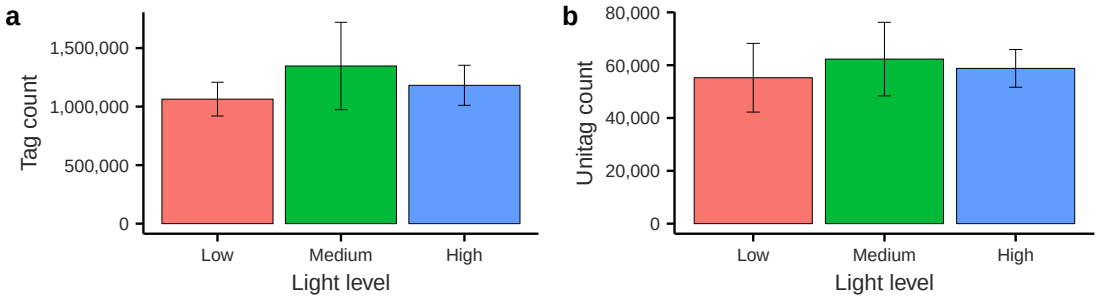


FIGURE 4.3: The mean unprocessed SAGE transcriptome library size (a) tag and b) unitag counts) from the wild type, lca1 and lca2 under low, medium and high light. The error bars are one standard deviation (n = 9).

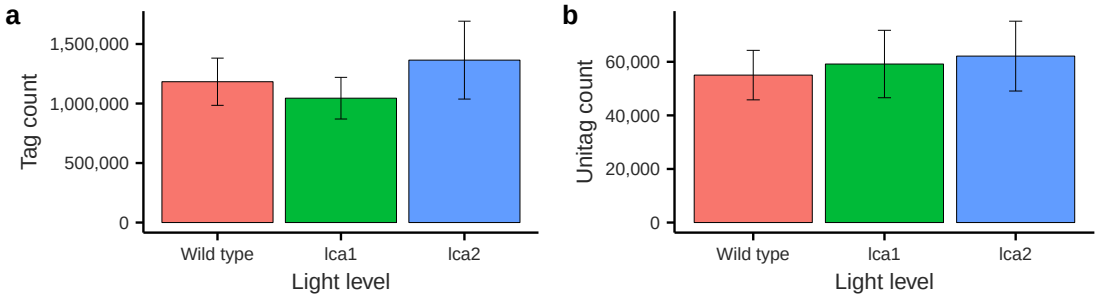


FIGURE 4.4: The mean unprocessed SAGE transcriptome library size (a) tag and b) unitag counts) under low, medium and high light for the wild type, lca1 and lca2 strain. The error bars are one standard deviation (n = 9).

Counting the unitags with only a single tag (sometimes referred to as singletons) in all libraries gives a total of 194646 unitags, which is roughly half of the observed unitags (Yandell and Ence, 2012). From a statistical standpoint these single tags can be safely discarded from any further analysis since no statistically significant results can be found with single hits. However, it makes sense to retain them throughout the mapping and annotation stage to see if they match any previously identified sequences. This can provide more information about whether these sequences are artefacts caused by technical error during the sequencing or actual sequences observed in low numbers.

Unitags with low tag counts make up the majority of the total unitags identified. Figure 4.5 shows the distribution of tag counts in total per unitag as well as the sum of those counts for a number of ranges. Unitags with 1 to 100 tag counts make up 97.4% of the total number of unitags, but the summed tag count for these unitags only accounts for 6.5% of the total number of tag counts. The remaining 2.6% consisting of

10201 unitags accounts for the 93.5% of the total tag counts (see Appendix B, Table B.3 for details). This shows that the data is heavily skewed so that a small number of unitags account for the majority of the total counts, while the majority of the remaining unitags only account for a small fraction of the total tag count.

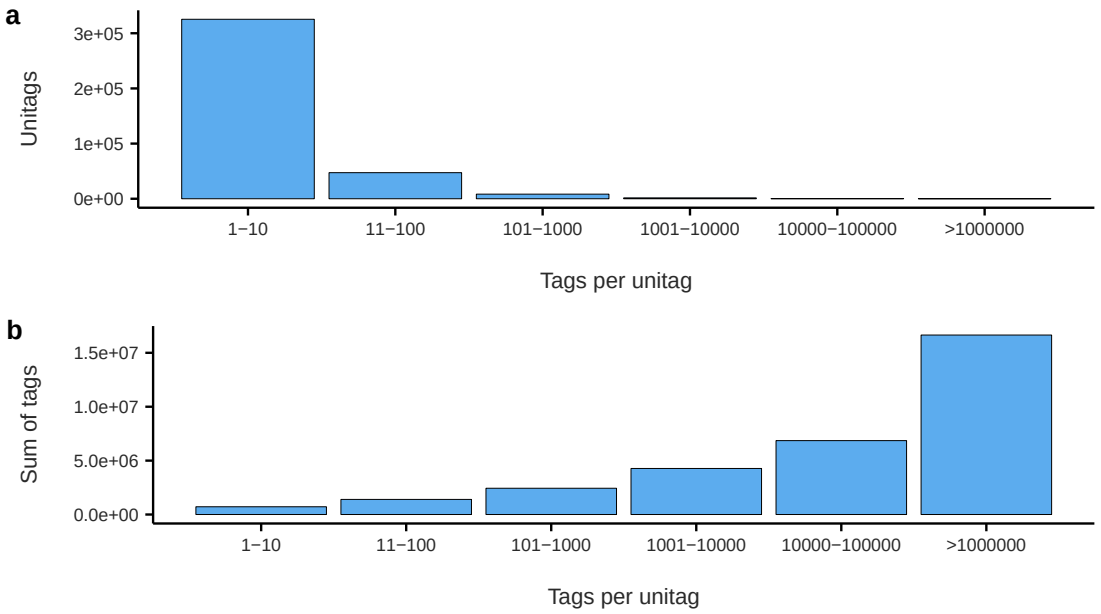


FIGURE 4.5: The distribution of a) unitags and b) the sum of the tags per range of tags per unitag in the combined libraries from the SAGE transcriptome.

Some of the unitags with a hundred or fewer tag counts could potentially be of statistical significance depending on the distribution of the tags throughout the 27 libraries. However, unitags with a tag count in this range with a more evenly distributed tag count across the libraries are unlikely to be of any statistical significance due to the disproportionally larger influence from technical error on small counts (McCarthy et al., 2012). The conclusion is that the majority of the unitags within this range (1 - 100 tags per unitag) that also makes up the majority of the total unitags will not be of any statistical significance. This leaves approximately ten thousand remaining unitags with higher tag counts to account for the majority of the statistically significant differential gene expression that is the primary motivation in analysing this transcriptome.

4.3 Extending the effective sequence length

Different strategies can be used to identify the genes represented by the 26 bp tag sequences produced by the SuperSAGE protocol. The tags can be directly annotated

by searching against a nucleotide database or translated into the corresponding protein sequence and searched against a protein database. The alternative approach is to extend the effective read length by mapping the short tag sequences against longer sequences and then annotating these extended sequences by searching them against a nucleotide or protein database. Searching the tags directly is the most straightforward approach but only possible when a substantial number of annotated genes or a full genome is available from either within the same species or genus. However, no species within the *Dunaliella* genus has yet been sequenced, with the exception of the organelle genomes (mitochondrion and chloroplast) from *Dunaliella salina* the type species of the genus (Smith et al., 2010).

TABLE 4.1: The number of nucleotide and EST sequences available on GenBank for the genus *Dunaliella* by species (*D. salina, D. tertiolecta, D. viridis and D. bioculata*) as well as various related but unidentified sequences belonging to the genus (other) as of March 2015.

Taxa	Nucleotide sequences	EST sequences
D. salina	424	6811
D. tertiolecta	172	0
D. viridis	77	5
D. bioculata	36	0
Other	356	0
Total	1065	6816

As a comparison, a similar search for the genus *Chlamydomonas* yielded 49941 nucleotide sequences and 213376 ESTs.

UniProts UniProtKB protein database lists 541 hits for the genus *Dunaliella* of which 126 are listed as located on the chloroplast and 5 as located on the mitochondrion. A similar search for the genus *Chlamydomonas* lists 15576 entries with 278 located on the chloroplast and 46 located on the mitochondria. Directly comparing these numbers it is apparent that the number of known genes in *Dunaliella* is only a small fraction of the total number of genes found in *Chlamydomonas*, the genome of *C. reinhardtii* has been published (Merchant et al., 2007) with 12264 stable genes found as of release 5.5 by the Joint Genome Institute (JGI) (available from http://phytozome.jgi.doe.gov/pz/portal.html). Out of the 541 listed *Dunaliella* genes nearly all (525) are machine annotated and many are near identical and matched to the same gene which again lowers the actual number of genes available for unitag identification. In addition, only a limited number of genes and ESTs have been sequenced from the genus (see table 4.1) and published to open databases such as NCBI (National Center for

Biotechnology Information). However, the availability of a transcriptome in the form of 409789 sequences allowed us to apply the second approach for unitag identification and map our sequenced tags to longer EST sequences (Rismani-Yazdi et al., 2011). They used the RNA-seq method and sequenced full-length mRNAs with half of the assembled sequences having a read length in excess of 522 bp (Rismani-Yazdi et al., 2011).

The mapping of the tags to the longer mRNA sequences from the transcriptome was done using the python-based software-tool GFESSA (Genome-Free EST SuperSAGE Analysis) that is part of the python based SLiMSuite software package (available from http://slimsuite.blogspot.com.au/2015/05/slimsuite-downloads.html) (Edwards and Palopoli, 2015). The tool was specifically written for the analysis of SAGE datasets for which no genome had been published and instead relies on using homology-based searches against EST libraries. The program uses NCBI's Blast tool to map tags to EST sequences. The settings were set to restrict the mapping tags with a perfect match against an EST. Out of the total 382846 unitags found within the SAGE dataset, 33163 (8.6%) unitags were perfectly mapped to one or more ESTs from the transcriptome dataset. Due to the nature of the EST transcriptome many of the EST sequences are identical, near identical or only differing in size, and in many cases multiple ESTs were mapped to one tag. In total, 73629 of ESTs out of 409789 were mapped to at least one tag.

4.3.1 Annotation of extended reads using BLAST

Annotating the extended reads and then inferring this annotation to the tags allow us to identify the gene sequences represented by the tags. Before this can be done the annotation of the ESTs needs to be carried out. This was performed using BLAST (Basic local alignment search tool) and, specifically, BlastX running on the open computing cluster at the University of Oslo's Bioportal; http://www.bioportal.uio.no). For the BlastX annotation the ESTs were searched against both the NCBI non redundant (NR) protein database as well as the UniProtKB/Swiss-Prot (Swiss-Prot) database (Consortium, 2014, Wheeler et al., 2004). The former database is much more comprehensive and consists of a larger dataset that is fully compiled and maintained using automated processes. The latter on the other hand is manually maintained and curated but is far smaller (Consortium, 2014, Wheeler et al., 2004). The BlastX algorithm translates the nucleotide EST in the six possible reading frames and then searches each translated reading frame against the selected protein database (Gish and States, 1993). The settings used for the search were as follows. The maximum expectation value (E) was set to 1.0E-03 using a BLOSUM62 matrix together with

10 reported scores and alignments per EST. For the search against the NCBI protein database (NCBI-NR) the results were reported in XML format. The resulting XML file containing 156452 alignments was first parsed using a perl script (BlastParse.pl) into a tab delimited text file before being re-parsed using a python script (parse_blast.py) to select the alignment with the best E-value passing a cut-off of a maximum of 1.0E-06. The resulting tab-delimited text file contained 13646 ESTs annotated to the NCBI-NR protein database with an E-value lower than 1.0E-06.

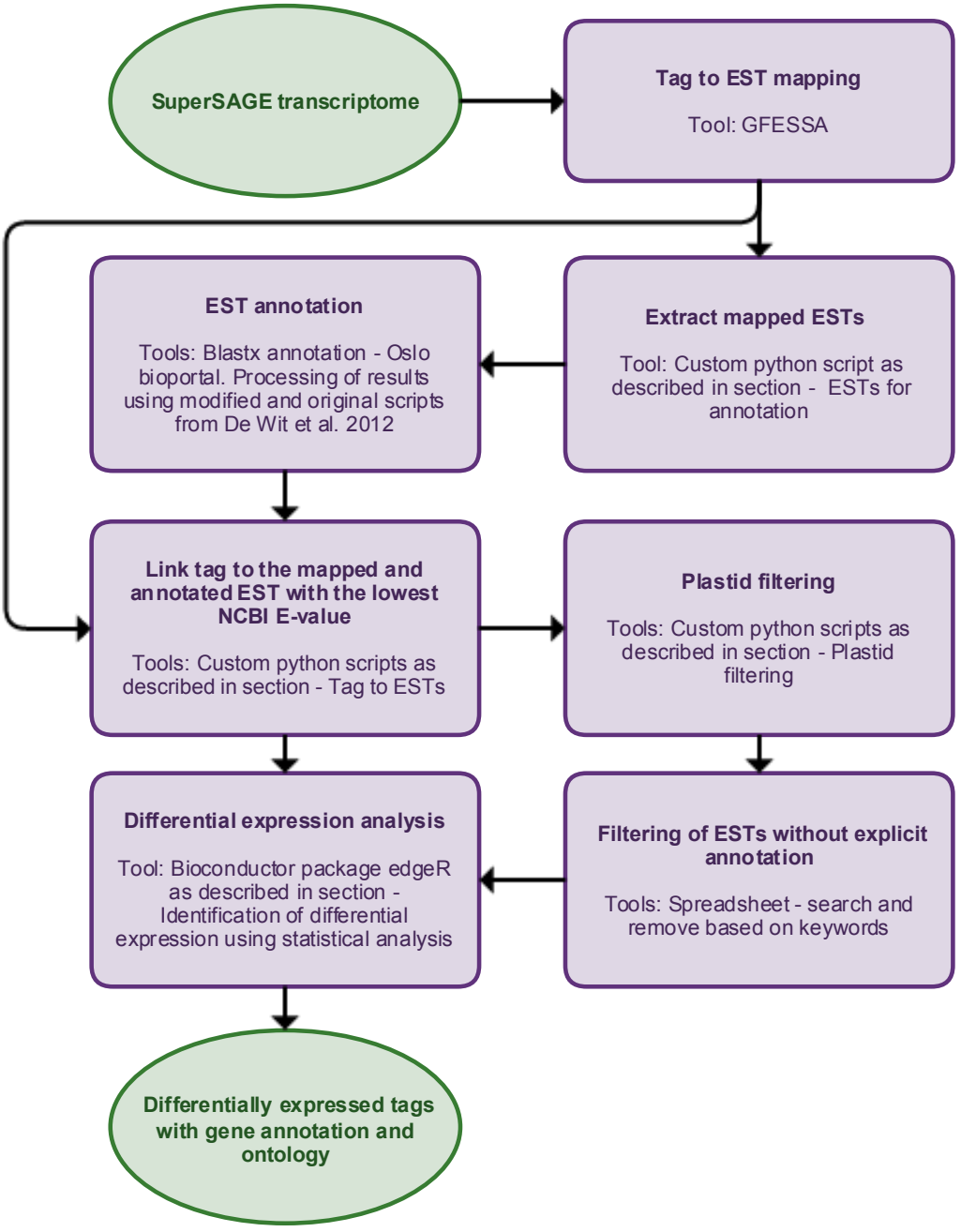


FIGURE 4.6: The SuperSAGE transcriptome was mapped to longer ESTs using GFESSA. All mapped ESTs were then extracted from the full dataset and annotated using BlastX. The mapped and annotated EST for each tag with the lowest NCBI-NR E-value was selected as the representative annotation. Based on this inferred annotation all tags annotated to either chloroplast or mitochondrion encoded genes were removed, followed by removal of all tags without a explicitly stated gene function. All tags were then statistically analysed for differential expression using the bioconductor package edgeR.

For the Swiss-Prot annotation the result were reported directly in tab delimited text format. The results from this annotation together with the NCBI-NR annotation

were processed using a python script (totalannotation.py) that takes the results as input. The resulting file contains the best annotated NCBI-NR and UniProt hit for each sequence (EST in this case) that meets the selected cut-off criteria for the E-value. The script also attaches additional annotation for each sequence that contains a UniProt match. The flat files for each UniProt match were downloaded and added to the sequence annotation. This includes gene ontology (GO) identifiers as well as KEGGKO identifiers if available. All scripts used during this section are part of a guide for analysing RNA-seq data by De Wit et al. (2012) and can be downloaded from http://sfg.stanford.edu/.

The annotated ESTs needed to be linked back to the original count data with mapped EST identifiers. In addition, since the tags in many cases were linked to multiple ESTs some form of selection process was needed to identify the EST deemed most likely to be the best representative of the tag and the gene it originated from. For this reason python scripts were developed to stepwise collect and extract the annotation for the ESTs mapped to each tag. Earlier versions of the scripts (Version 1a) extracted the annotation and then sorted the result for each tag into groups separated by the NCBI identifier and then by NCBI E-value within each group. A later version of the same script (Version 1b) additionally retained only the the EST within each group with the lowest E-value. Based on these results it became apparent that in the vast majority of the cases the annotation for the different ESTs mapped to each tag was very similar or in many cases identical with only the E-value differing. The final version of the script (Version 2) simply selects the EST with the lowest E-value as the representative. The process, as well as example annotations visualizing the process, is summarized in Figures 4.7 and 4.8.

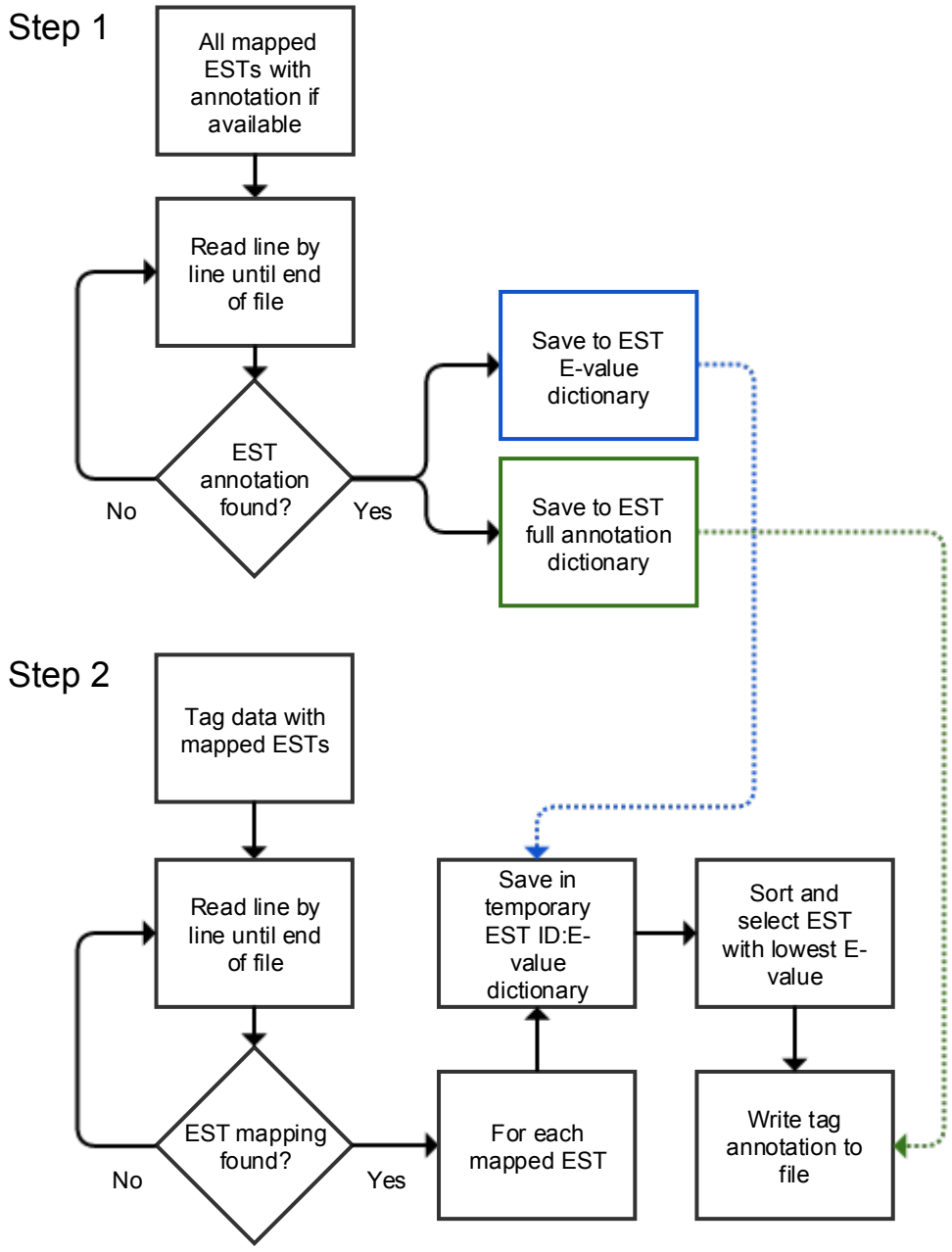


FIGURE 4.7: Simplified representation of the steps performed by the V2 script that links count data with EST mappings with annotated ESTs. In step 1 a file containing ESTs is read line by line and ESTs that have been annotated are moved to two dictionaries, one containing only ESTs (blue outline) and the corresponding E-values and another one containing ESTs and the full annotation (green outline). In step 2 a second file containing the tag counts with mapped ESTs is read line by line and the ESTs mapped to each tag are searched for in EST:E-value dictionary and stored in a third temporary dictionary. This temporary dictionary is then sorted once all ESTs have been read and only the EST with the lowest E-value was retained. This EST is then used to lookup the full annotation in the second dictionary (green outline). Finally the result is written to a file and the process in step 2 is repeated for the next tag.

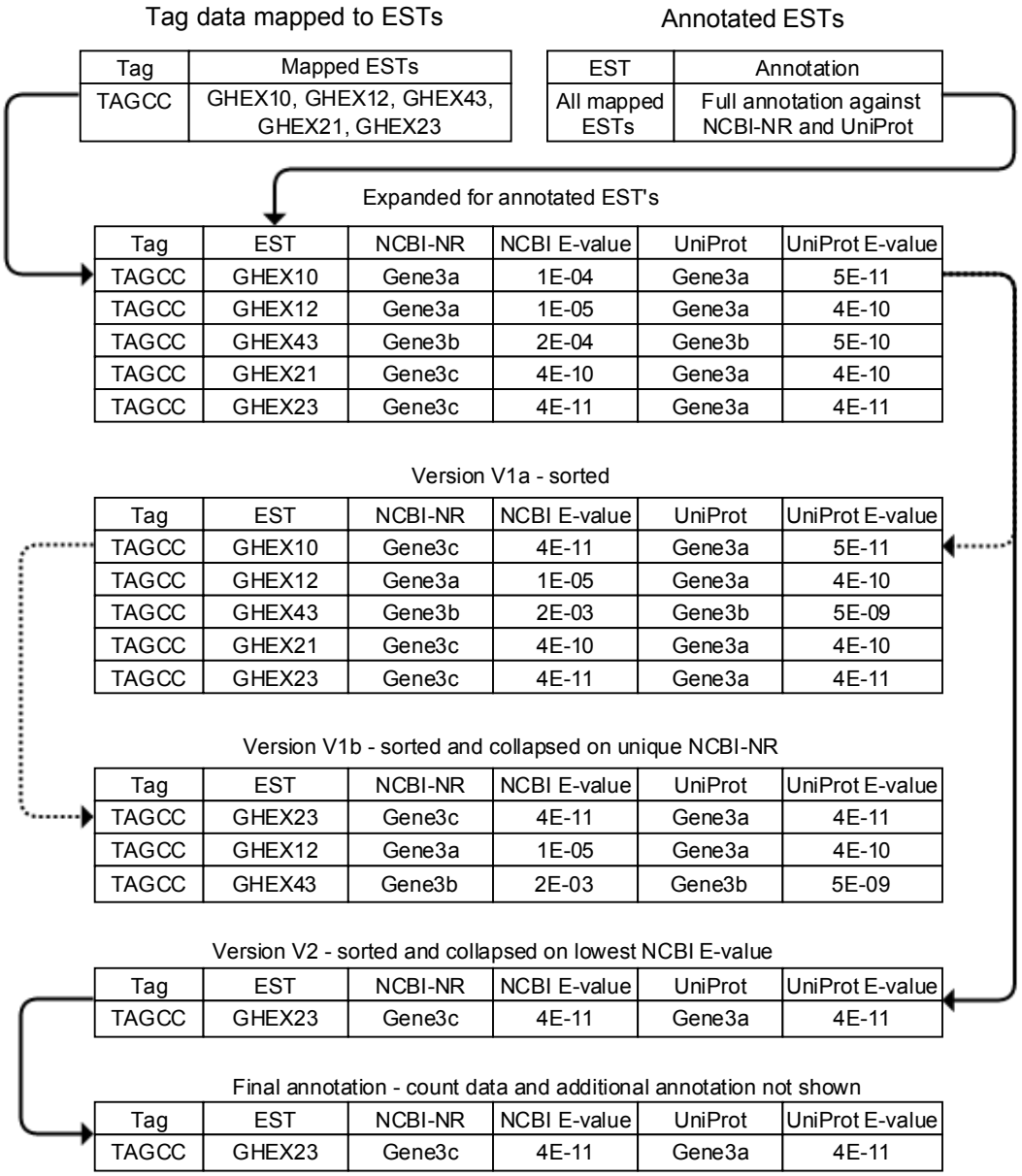


FIGURE 4.8: Flowchart showing the evolution of the output generated by the scripts leading up to the final version used for the linking of tag counts mapped to ESTs to annotated ESTs. The two files containing the tag data mapped to ESTs (top left) and the annotated ESTs (top right) are processed and for each tag all mapped ESTs with an annotation are expanded in all version of the scripts. Version V1a of the script simply sorted the expanded ESTs for each tag by NCBI-NR identifier and E-value while version V1b in addition collapsed the sorted data such that for multiple ESTs with the same NCBI-NR identifier only the EST with the lowest E-value was retained. The final version of the script V2 sorted only based on the NCBI E-value and then selected the EST with the lowest overall E-value. This annotation was then selected as the representative annotation for the linked tag and written to file. The full and final path of the annotation is shown using black lines while dashed lines represents the path explored in version V1a and V1b.

Out of the total 73629 ESTs mapped to a unitag, 13563 were annotated using the

NCBI-NR database while 9104 ESTs was annotated using UniProtKB/Swiss-Prot. Following filtering for conspicuous keywords suggesting a lack of a clear annotation, 7425 ESTs were retained, with the remaining 6138 removed.

4.4 Filtering of organelle encoded genes

The use of oligo (dT) primers during the cDNA synthesis step selects for polyadenylated RNA. Poly(A) tails are added to nuclear encoded mRNA during and after transcription in eukaryotes but is also present in RNA originating from the chloroplast and mitochondria (Rott et al., 2003). While the presence of a poly(A) tail in nuclear encoded mRNA protects the RNA from degradation the opposite has been shown for organelle encoded mRNA. Polyadenylation of RNA in plastids is instead thought to serve as a marker for degradation (Komine et al., 2002, Rott et al., 2003). Arguably it may be the case that detected levels of cDNA annotated as plastid encoded accurately reflects the true levels of RNA. However, this has not been shown so far and it is therefore prudent to filter and remove any tags annotated as plastid encoded to negate this possibility.

Two possible methods to identify these sequences include searching for sequences that are known to be plastid encoded (by location) or by using a homology-based search to identify mRNA lacking transit peptide sequences that are necessary for the import of nuclear encoded preproteins into the chloroplast or mitochondrion. There are strength and weaknesses to both methods. Using a search that is based on the known location of sequences is sound as long as the origin of the sequences can be trusted and is well understood. Arguably this is the case for the majority of the plastid encoded genes that have been studied in detail. Searching by homology on the other hand is considerably more computationally demanding and also dependent on identifying the transit peptide sequences that are known to be involved in import of preproteins across the organelle membranes.

The main approach used in this project was based on identifying organelle-encoded sequences by location. All proteins identified as either of chloroplast or mitochondrial origin in GenBank or SwissProt were searched for in the NCBI protein database. This identified 2162128 sequences in total. Each of these sequences has been assigned a GenInfo Identifier (GI) by NCBI that identifies the sequence. The file containing the annotated ESTs mapped to tags where then searched for matching entries. Sequences containing matches for either the NCBI or UniProt (SwissProt) annotation were filtered from the file using a custom script developed in python for this purpose. The script separates sequences into two files, one with the remaining nuclear-encoded

tags and one with organelle-encoded tags. It also counts hits and reports back the result. See Appendix B, B.1 for the full script. These GI identifiers were then searched against 9358 annotated ESTs that had been mapped to a tag. Based on the results 76 entries were filtered by both the NCBI and the UniProt annotation, 14 entries were filtered based solely on the NCBI entry and a further 194 entries were filtered only by the UniProt entry. In total, 284 tags were identified as organelle encoded and removed from the main dataset.

To further improve the results, a second step of filtering was done using a separate annotation against the *Arabidopsis thaliana* TAIR 10 protein database (TAIR 10). A further 4 sequences, was removed during this process. Again a custom python script was used for the filtering. See Appendix B, B.2 for the full script. While not discussed in detail, this annotation linked 3300 ESTs with at least 100 matched tag counts to TAIR 10. Since the mapping contains descriptors for the chromosomal location of the annotated gene, ESTs mapped to either the mitochondria (ChrM) or chloroplast (ChrC) genomes can be identified and filtered. Out of the 3300 annotated ESTs, 169 were identified of mitochondrial origin with a further 409 identified of chloroplast origin.

The removal of unitags annotated to non-nuclear-encoded genes resulted in 9071 unitags remaining. These were again filtered to retain only unitags with an explicit annotation based on, primarily, the annotation against UniProt or secondarily the annotation against NCBI-NR. This final list containing 6584 unitags with explicit annotations was linked to tags meeting the filtering criteria applied during the edgeR analysis presented below.

4.5 Identification of differentially expressed genes

The swift development of new sequencing technologies during the last decade, also necessitated a similar parallel development of specialised tools to facilitate the resulting data analysis. Some of these tools build on knowledge gained from the analysis of data from microarrays that pose similar difficulties. However, due to the nature of the data, signal intensity-based (i.e. analog) in the case of microarrays that follows a log normal distribution, compared to the digital data attained from NGS sequencing that is normally modelled to a poisson or negative binominal distribution, there are core differences between the analysis required Nowrousian (2010), Ritchie et al. (2015). In addition to how they typically model the data for the analysis of transcriptomes, these tools (see (Soneson and Delorenzi, 2013) for a comparison) differ in how they deal with the normalization of the data. This includes methods for compensating

for differences in library size and composition to allow for more exact detection of significantly differentially expressed genes while controlling the false discover rate (FDR) (Dillies et al., 2013). Two commonly used packages (DESeq and edgeR) for the analysis of transcriptome data generated by NGS sequencing were evaluated for the analysis of this SAGE data. The edgeR package was chosen due to its good support for and the inclusion of well documented case studies using published data for the analysis of multifactorial experiments in the user guide (Anders and Huber, 2010, Robinson and Smyth, 2008). A fairly recent article compares the differences and similarities between these two packages and provides a good overview of the analysis workflow (Anders et al., 2013).

4.5.1 Statistical analysis using the edgeR package

The edgeR (empirical analysis of DGE (differential gene expression) in R) package was used for the statistical analysis of the SAGE transcriptome and is part of the wider bioconductor framework of open source tools developed in the R programming language and includes various tools for the analysis of biological data with an emphasis on high throughput data such as those generated from next generation sequencing (Gentleman et al., 2004). The packages can be downloaded from the main homepage at http://www.bioconductor.org. The edgeR package has been in development since around 2007, with a first release in 2008 (Robinson and Smyth, 2008). While initially developed for the analysis of SAGE data, the methods were applicable to all count-based data with two or more groups of which at least one group had to include replicates. Later releases added support for complex multifactorial experiments (Chen et al., 2014) using generalized linear models (GLM) (McCarthy et al., 2012) and methods for normalizing the data using the TMM method (Robinson and Oshlack, 2010) used for this data set and discussed in more detail in the normalization section below (4.5.3)

For the analysis, the data within edgeR was stored as a list-based DGEList data object. During the initial analysis the count data, additional annotation data and meta data describing the replication and the grouping within the experiment was stored in the object. Calculated normalization factors, estimates of the dispersion within the data set and other results was further added to the object allowing the DGEList to act as a convenient place holder for the associated data.

4.5.2 Filtering of unitags with low tag counts

The data set was filtered for very low count tags. Unitags not meeting a minimum of one tag per million in at least nine libraries (out of 27) were filtered from the data set. This form of filtering, sometimes referred to as a CPM (counts per million) filter, was recently shown to perform most similarly to a newly proposed alternative filtering procedure developed to increase the detection power for medium and highly differentially expressed genes (Rau et al., 2013). The cut-off used was in line with the suggested filtering of tags with less than one CPM per group, in this case the group can be thought of as consisting of either all libraries within one strain (9) or all libraries under one light level (9), which should permit tags present in all libraries in either of these conditions to pass the filtering threshold in most cases. The main goal of the filtering process was to remove any tag that is not found in sufficient number to be of significance during the later testing for tags with differential expression. Using this filtering threshold the total number of unitags within the data set dropped from 382846 to 29120, equal to 7.6% of the original unitags. At the same time the total tag counts within the remaining tags dropped by only 1,205,018 tags out of 32,331,173 meaning that the remaining unitags still represents 96.3% of the total counts. The filtering thereby greatly reduced the size of the data set while retaining nearly all of the useful data. This improved the speed of processing, but more importantly it also improved the estimation of the biological co-variation by removing noise. The total variance between replicates for each gene (or tag in this case) within the dataset is composed of technical and biological variance and denoted as the coefficient of variation (CV), which is the standard deviation divided by the mean (McCarthy et al., 2012). The squared CV equals the squared coefficient for the technical and the biological variation as in equation 4.1 below.

$$Technical\ CV^2 + Biological\ CV^2 = Total\ CV^2 \tag{4.1}$$

The biological CV (BCV) is the actual variance of a gene after removal of all technical noise. BCV remains approximately the same regardless of the abundance of the gene represented by the number of counts, while technical CV decreases with abundance. This means that the variance for unitags (or genes) with low counts are dominated by technical noise, while BCV dominates in high count unitags (McCarthy et al., 2012). In effect, the filtering raises the signal to noise ratio within the dataset by removing technical noise.

Following filtering of the original 6358 unitags with an annotation, only 1313 annotated unitags remained within the 29120 unitags meeting the CPM filter criteria. While

seemingly low these remaining unitags still accounted for 99.5% of the original tag count within the annotated tags.

4.5.3 Normalization of libraries

Transcriptome libraries are typically normalized by adjusting the individual tag counts to the size of the library. This makes sense since the summed tag count of the library is directly proportional to the amount of RNA used during the sequencing (termed library size scaling). However, this normalization or scaling only takes into account the overall size and not the population composition of genes represented by the sequences that form the library. In other words it only corrects for differences between libraries but not within libraries. The default normalization method in edgeR uses the trimmed mean of M-values (TMM) to correct and reduce under and over sampling of genes within each library. The idea behind the method is that the majority of the observed genes are not differentially expressed. Using these genes with a low differential expression, a scaling factor can be calculated for each library that takes into account both the composition and the sequencing depth (library size scale) with the end-goal of minimizing the false discovery rate (FDR) of non differentially-expressed genes (Robinson and Oshlack, 2010). The product of this scaling factor and the library size equals the effective library size. This calculated effective library size is then used during the down-stream analysis within edgeR.

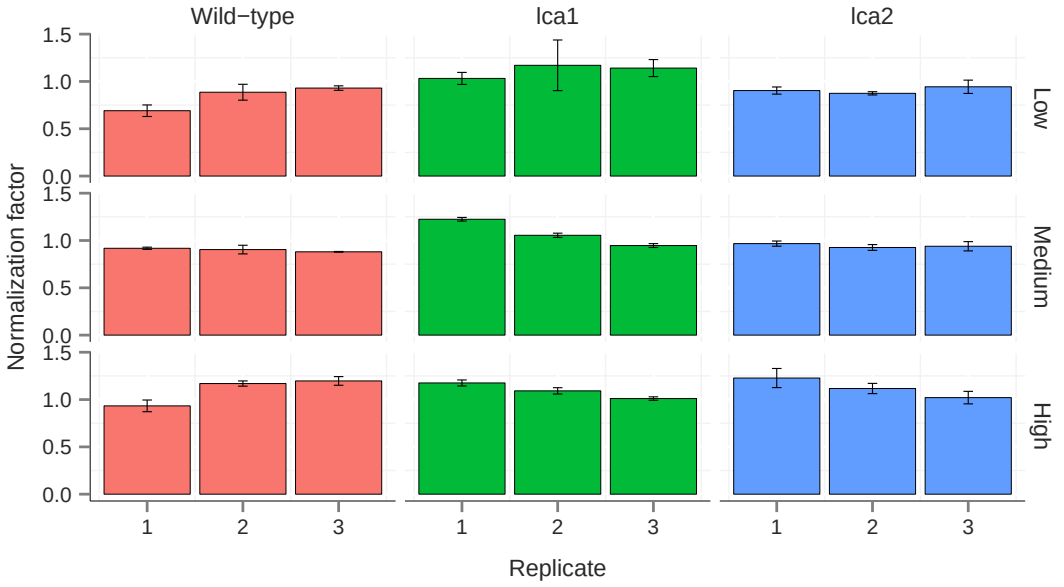


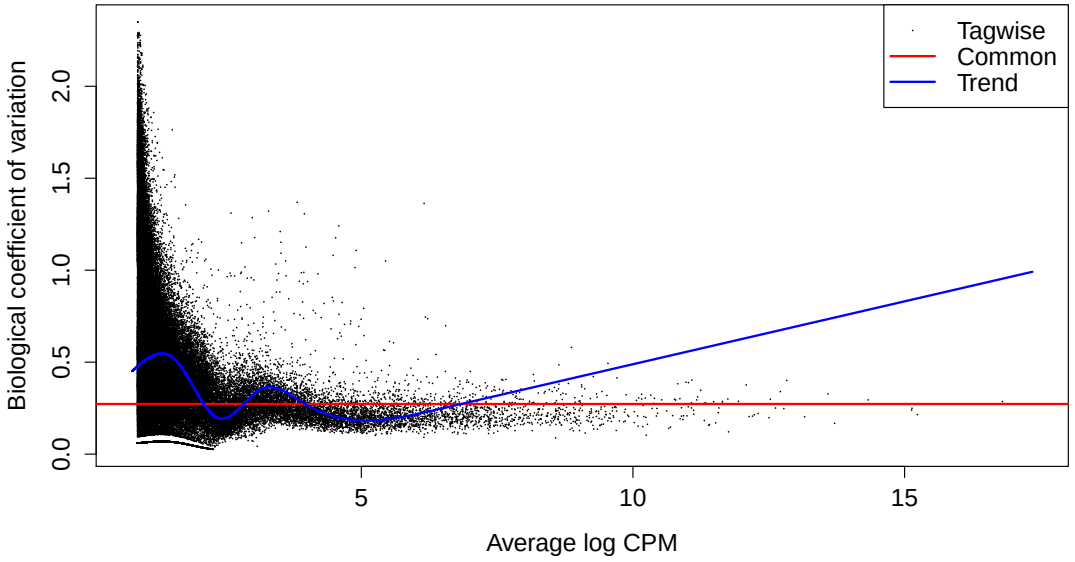
FIGURE 4.9: The mean normalization factor for each of the triplicate SAGE libraries under low, medium and high light for the wild-type, lca1 and lca2 strain. Data shown are based on the mean of the calculated normalization factor for each library using the three different normalization methods (TMM, UQ and RLE) implemented in edgeR. Error bars are one standard deviation.

The edgeR package supports the use of three different normalization methods, in addition to the TMM methods the upper quartile method (UQ) and the relative log expression (RLE) is available. The UQ method uses, as the name implies, the upper quartile of counts to calculate a library scaling factor (Bullard et al., 2010), while the RLE method is the normalization method introduced in the DESeq package and largely similar to the TMM method (Anders and Huber, 2010). A fairly recent study comparing seven of the most commonly used normalization methods, including the three methods above, found that only the TMM and the RLE methods performed satisfactorily under all conditions (Dillies et al., 2013). The mean from the calculated normalization factor using the three methods implemented in edgeR is shown in Figure 4.9 above. The three methods largely generated similar normalization factors with one notable exception using the upper quartile method for the second of the triplicate libraries for lca1 under low light (Figure 4.9), resulting in the increase in standard deviation seen in the figure for this library. Based on these results the TMM method using the default settings was used during the analysis of the SAGE transcriptome.

4.5.4 Estimation of biological co-variation

For multifactorial experiments edgeR uses a Cox-Reid profile-adjusted likelihood (CR) to estimate dispersions (McCarthy et al., 2012). The design matrix describing the experiment, together with the DGEList object that holds the count and annotation data is fitted with a generalized linear model (GLM) that takes into account any existing systematic source of variation. The CR method can then be used to calculate a common dispersion for the whole data set, a trended dispersion by average count size as well as the tagwise dispersion for individual tags. The tagwise dispersion relies on either the common or the trended dispersion, which is used to transform the dispersion for the individual tags in the direction estimated by the common or the tagwise dispersion (Chen et al., 2014). By default edgeR uses the trended dispersion to calculate the tagwise dispersion if it is available. For this analysis the trended dispersion was used to calculate the tagwise dispersion following recommendations in the edgeR user guide. The estimated dispersions using GLM methods for the common, trended and tagwise dispersion for the full data set (before and after filtering of low count tags) are shown in Figure 4.10a and 4.10b.

(A)



(B)

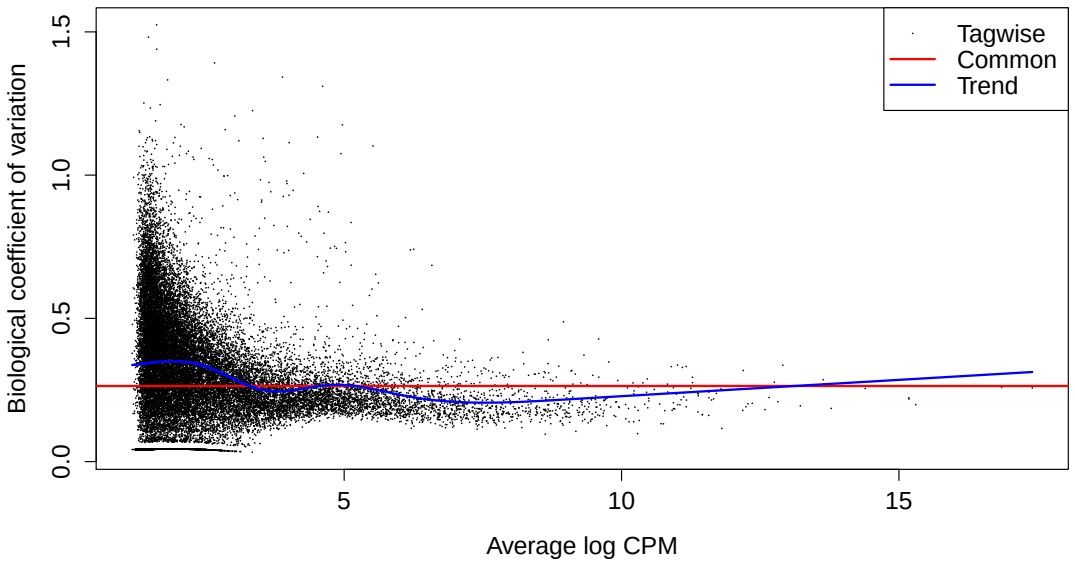


FIGURE 4.10: A. The biological coefficient of variation (shown on the y-axis) as estimated by the overall dispersion for all tags (common) by the average log count per million (CPM) (shown on the x-axis) in the full data set with 382846 tags, trended on a average count size (trended) and for each single tag (tagwise). B. as in (A) but showing the result for the data set filtered to remove low count tags, with 29120 remaining tags.

Based on the results in Figure 4.10a it is apparent that the method for estimating the trended dispersion (estimateGLMTrendedDisp) has over-fitted the dispersion for very high count tags when compared to the common dispersion as well as the tagwise dispersion. After filtering of very low count tags (see the filtering section 4.5.2 for details) a much more reasonable result is achieved (Figure 4.10b). The trended

dispersion now follows the common and the tagwise dispersion. The calculated dispersion based on the filtered dataset with 29120 remaining tags was used for the analysis. What is also apparent is the high degree of variation that is found within very low count tags. This is as expected since these tags will be affected by technical noise to a much higher degree compared to high count tags. The results in Figures 4.10a and 4.10b also shows clearly that for counts with a low number of tags it is unlikely to find any significant differential expression due to the high degree of variation. This further supports the filtering of these tags from the dataset during the statistical analysis.

4.5.5 Visualization of similarities between replicate and inter-group variations using multi-dimensional scaling (MDS)

The relationship between the groups and replicates within the data can be visualized using a multi-dimensional scaling (MDS) plot and shown as a two-dimensional scatter plot. The MDS function (plotMDS) as part of the limma package (Ritchie et al., 2015) averages (using the root mean square (Euclidean distance) since the data are log transformed) the largest $\log_2$ fold change between samples (in absolute terms) for a specific (500 genes by default) number of genes (or tags in this case) with the largest degree of standard deviation (Ritchie et al., 2015).

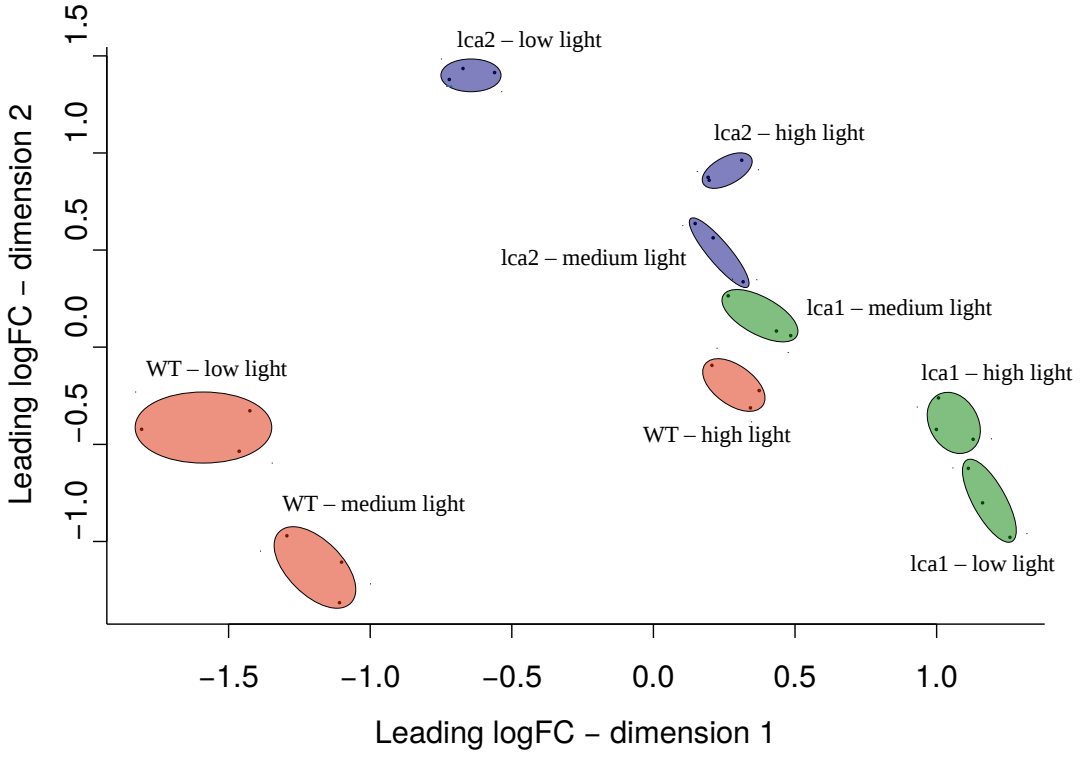


FIGURE 4.11: The difference in leading (maximum) fold change (FC) between the top 500 differentially expressed tags as shown by multi-dimensional scaling (MDS) plots of the filtered data. Each ellipse encircles three replicates from the same condition with the calculated centre for each replicate shown as dots. The plot shows the results for all 27 libraries, with replicates of each wild type (WT), lca1 and lca2 strain under low, medium and high light.

Figure 4.11 shows the resulting MDS plot using the filtered data with 29120 tags. While each library is shown individually in the original plot, they have been grouped manually by encircling the replicate libraries for each group. The libraries falling into each group of replicates are generally closely spaced suggesting low degrees of variance within the groups. On a larger scale each strain seems to group as well, with the wild type, lca1 and lca2 strains occupying the lower left, lower right and top middle part of the figure, again suggesting that clear differences exist among strains.

4.5.6 Statistical testing for differential expression

The testing for differential expression was done using generalized linear model (GLM) likelihood tests following previous protocols and suggestions for multi-factorial experiments (Chen et al., 2014). As previously mentioned, the test automatically takes into account various sources of systematic variation affecting the data set. The testing is performed in a two-step process that first fits a negative binomial GLM model for

each tag using the glmFit function. This is followed by the likelihood ratio statistical test (glmLRT function), which uses the fitted model in conjunction with the design matrix that describes the grouping within the data set. Testing was typically done by contrasting two groups within the design matrix, with the null hypothesis that there is no statistical difference between the selected groups. The results from the test are reported as the calculated fold-change between the tested groups, the result from the likelihood ratio test, the calculated p-value and the corrected false discovery rate. Tags with a FDR below 0.05 were selected as differentially expressed regardless of the level of fold-change. Rather, the significance of the fold-change is considered on an individual tag and gene basis during the reporting and in relation to its reported FDR.

4.6 qPCR verification of relative gene expression

Quantitative real-time PCR (qPCR) was used to verify the results from the SAGE transcriptome. The relative expression (log2 fold-change) for individual genes based on the SAGE transcriptome was compared to the corresponding result for the same genes as measured using qPCR (normalized against the S11 ribosomal gene) as previously described by Pinto et al. (2010). The genes used for the verification was selected from a larger pool of genes related to processes that are either directly or indirectly involved in photosynthesis. The primers for the selected genes showed good primer efficiency across the target range and high specificity with single main peaks in the dissociation curves. The strength of the linear association was calculated using Pearson's correlation coefficient followed by an additional test for the significance of the correlation.

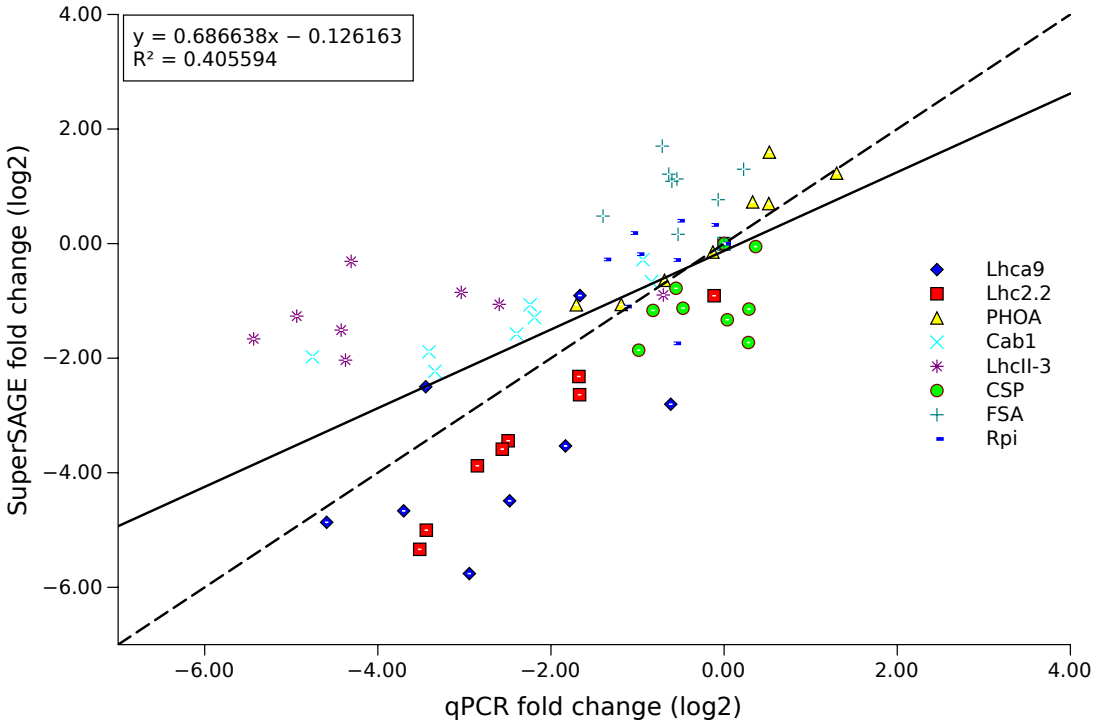


FIGURE 4.12: The correlation for the relative gene expression as measured using SuperSAGE and qPCR. For each gene (*lhca9*, *lhc2.2*, *phoA*, *cab1*, *lhcII-3*, *csp*, *fsa*, *rpi*) the average results (N = 3) under low, medium and high light for the wild-type, lca1 and lca2 is shown relative to the wild-type under low light. The qPCR results are normalized against the *s11* gene. The result from a linear (least square) regression to all results is shown as a dashed line together with the calculated equation and r-squared value. As a reference the solid line show the resulting correlation for a perfect correlation. Based on data from P. Stephenson.

The correlation in the relative fold change measured for 8 selected genes are shown in Figure 4.12. The original qPCR verification using the same RNA used to generate the SAGE libraries, provided significant correlation ($p < 0.05$) for 4 genes out of 8 in total. The combined result as shown in Figure 4.12 was highly significant ($p < 0.001$). The result from this qPCR based verification of the SuperSAGE transcriptome, lends additional support to the results presented here and in Chapter 5.

4.7 Conclusion

The main goal and purpose of generating and analysing the short read SAGE transcriptome was to compare the two lca strains to the wild type under different light conditions and help connect the differences observed on a physiological level with the changes in gene expression that drive the lca phenotype(s). The end result of the analysis, the identification of statistically significant differentially expressed genes will be discussed in Chapter 5.

To conclude, the overall structure of the data in the form of unitags and tags per library was shown to be largely independent of the light conditions or strain (see Figures 4.3 and 4.4). In general, the data contained a large number of unitags with low tag counts. To counter this a CPM filter was applied to remove poorly represented tags. This drove down the technical influence (CV) on the overall variation and allowed for a better estimation of the BCV, the resulting improvement in the common and trended dispersion in turn allows for a better estimation of the tagwise dispersion. To minimize the effects from compositional biases within libraries, a normalization using the trimmed mean of M-values method was applied to the data in addition to the total library size scaling that is automatically applied during the analysis.

Visualisations using MDS plots based on the normalized data showed that replicates were grouped together as expected, suggesting a low degree of variation within the triplicate libraries for the tags showing the highest degree of variation in counts between libraries (see Figure 4.11). The plots also confirmed the clear differences between the strains both on a strain and light level in line with earlier observations based on the photophysiology and general phenotype of the lca strains compared to the wild type.

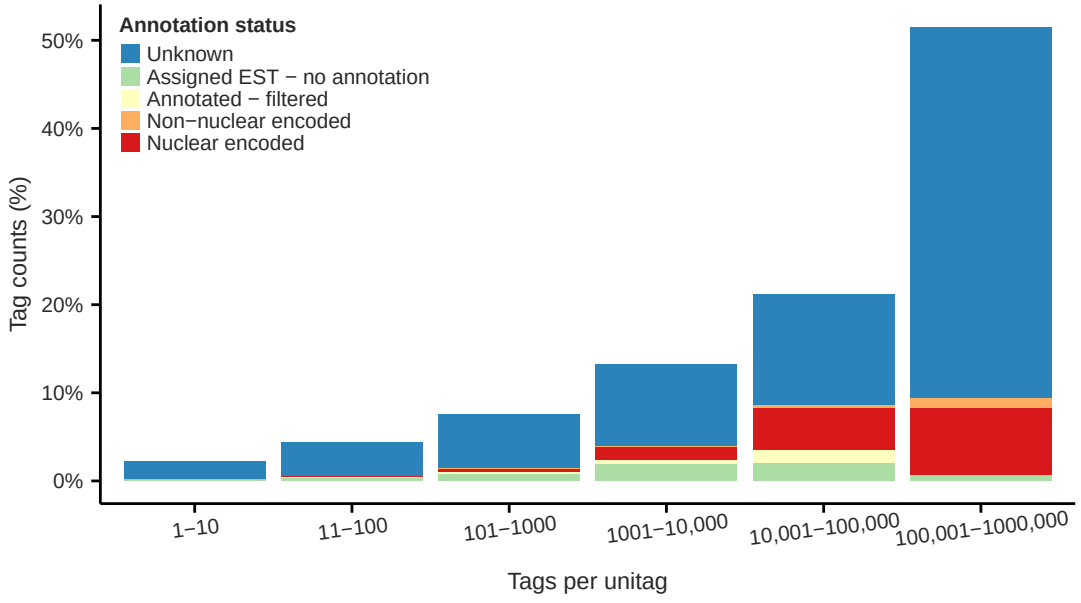


FIGURE 4.13: The annotation status for unitags with inferred annotation to either nuclear- or organelle encoded genes, unitags lacking a clear annotation (annotated - filtered), unitags mapped to non-annotated ESTs and tags not mapped to any EST (unknown) by range of tags per unitag.

The annotation process was complicated by the lack of a published genome of *D. tertiolecta*, resulting in the need for a two step process, with intermediate mapping against

longer sequences from within the same species (see section 4.3.1). This naturally confines the number of possible annotations to the sequences present within this dataset, lowering the total number of sequences that can be annotated. In addition, sequences linked to organelle-encoded genes were removed since mRNA with added poly-A tails originating from the chloroplast or mitochondria are likely to be bound for degradation rather than translation into proteins (see Section 4.4). Figure 4.13 show an overview of the proportion of annotated unitags. Regardless of these limitations, 9071 (shown in green in the figure) annotated unitags linked to nuclear-encoded genes remained after filtering, providing a sizeable database for further higher level analysis and identification of differentially expressed genes (Chapter 5).

Chapter 5

Molecular characterization of lca1 and lca2

5.1 Introduction

The results from the transcriptome analysis of the wild-type and lca strains under low, medium and high light confirm that robust differences exist at the level of the transcriptome among the two low chlorophyll content mutants lca1, lca2 and wild-type. This chapter analyses these difference in detail in an attempt to characterize the molecular basis of the observed phenotypic differences identified in Chapter 3. Further, as transcripts have been obtained across three light level the photoacclimitation strategies of wild-type *Dunaliella* and lca1, lca2 are discussed

Initially, this chapter presents results from high-level comparisons of the pattern of gene expression in the strains under the different conditions followed by extensive gene set enrichment analysis of pathways, processes and functions to pinpoint differences and similarities in the lca strains compared to the wild-type. Specific emphasis is also placed on the regulation of key metabolic processes and pathways including the expression of core photosynthesis and light harvesting genes, the regulation of chlorophyll and carotenoid biosynthesis genes as well as genes involved in the translocation of proteins from nuclear encoded genes into the chloroplast all of which are likely to be relevant in describing the phenotype.

5.1.1 Overview of significant gene expression between conditions

The general pattern of significant differential expression among different conditions on a strain and light level, irrespective of the degree of fold-change, gives a good overview

of where the majority of the transcriptional changes occurs under changing conditions or between strains. The total numbers of significantly (FDR <0.05) up or down-regulated differentially expressed tags, regardless of the level of fold-change are shown in Figures 5.1 and 5.2 below.

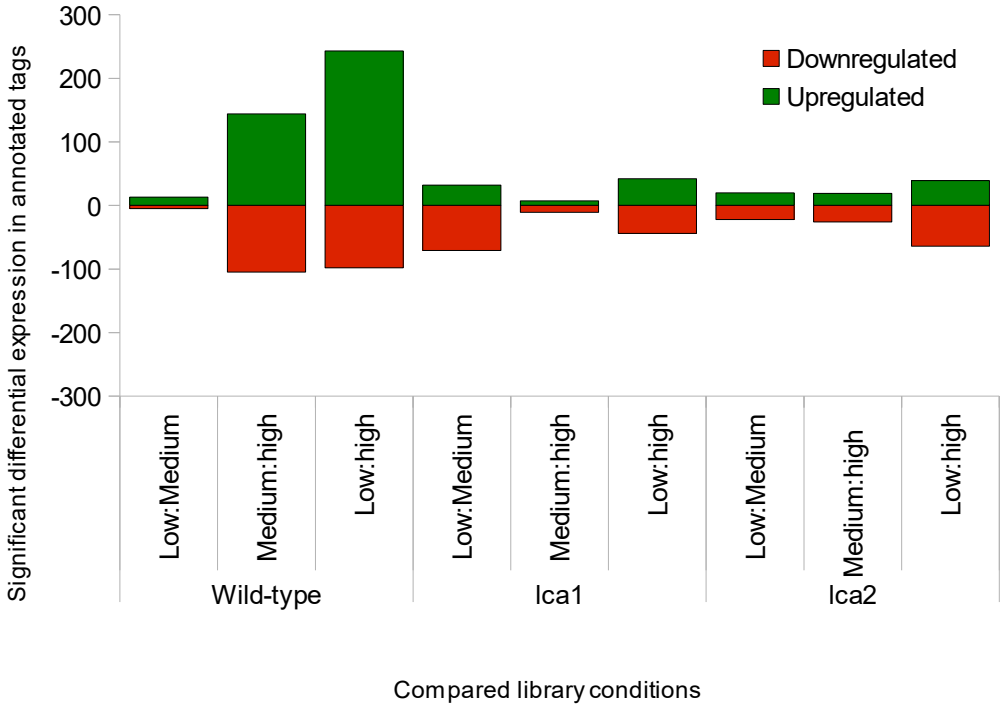


FIGURE 5.1: The total number of significantly (FDR <0.05) up or down-regulated annotated tags in wild-type, lca1 and lca2 between light levels (low vs. medium light, medium vs. high light and low vs. high light).

Figure 5.1 shows that the majority of the significant differential expression of tags in the wild-type occurs at high light (between the light levels medium to high (100 to 1200 μmol photons m^{-2} s^{-1})) and low to high (20 to 1200 μmol photons m^{-2} s^{-1}). The interpretation being that low and medium light exhibit largely similar tag expression in wild-type *D. tertiolecta*. The figure also shows that far more tags are up-regulated than down-regulated at lower light intensities. This general pattern in the level of expression by light agrees with an earlier report suggesting that *Dunaliella* has a light-shade response that is optimized for high light (Falkowski and Owens, 1980b), however, a related article found that the response in the expression of LHC (CAB) proteins in *D. tertiolecta* was largely identical in cells shifted from high light to either low light or darkness (Laroche et al., 1991). Both lca strains show a very different regulatory response between light levels. The most obvious difference is the lower number of significantly differentially expressed (SDE) tags in total, in lca1 and lca2 there are 70% and 75%, respectively, fewer tags compared to the wild-type that show a statistically significant differential expression between low and high light suggesting a much

weaker photoacclimation response in both lca1 and lca2 at the transcriptional level. Also apparent is a change in the direction of regulation with more tags down-regulated than up-regulated in both lca strains. However, while these changes are shared between the lca strains and they both show a similar number of SDE tags between low and high light, they differ in the responses from low to medium and medium to high light. The majority of the regulation in lca1 occurs under low to medium light, compared to medium and high light, while in lca2, the number of regulated tags is similar at these light levels.

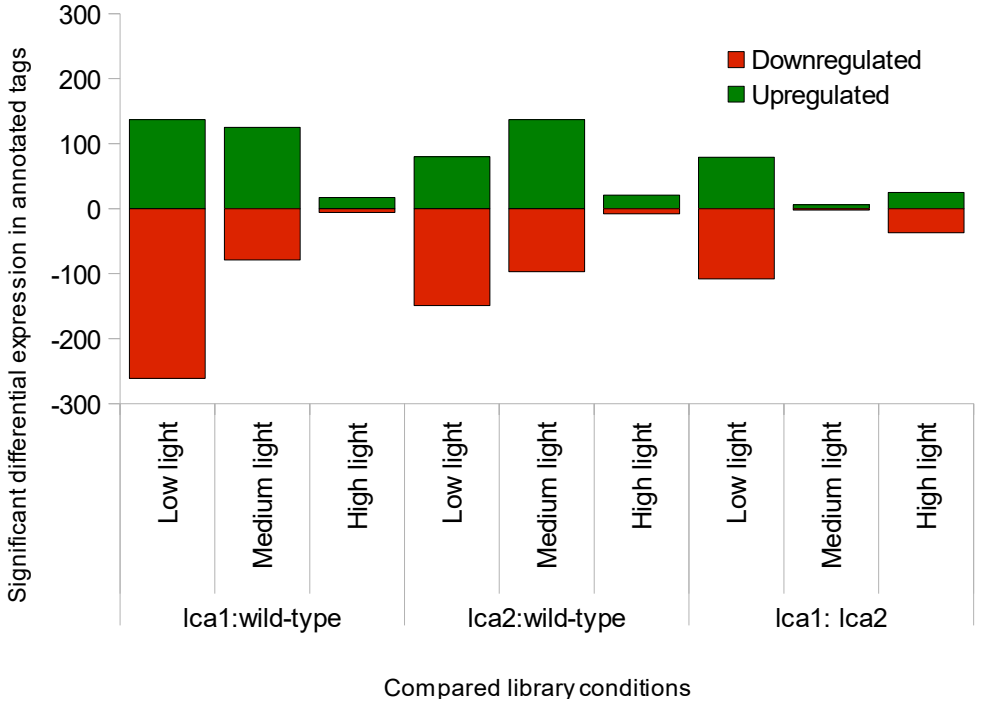


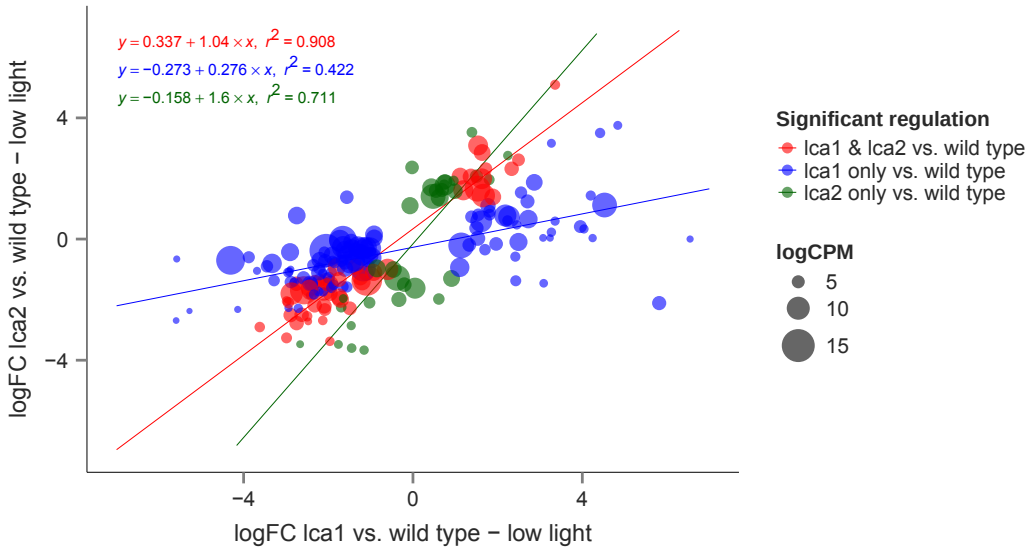
FIGURE 5.2: The total number of significantly (FDR <0.05) up or down-regulated and annotated tags between wild-type and lca1 or lca2 under low, medium or high light.

Figure 5.2 shows the difference between the lca strains and the wild-type under the three measured light levels. The general pattern in regulation between the wild-type and the lca strains is one where the majority of the significant differential expression occurs at low and medium light. In contrast, the lca1 and lca2 strains are clearly different from each other at low and to a lesser extent under high light, while under medium light the mutants show a similar pattern of regulation. The difference between the lca strains under low and high light also shows that the lca1 strain differs from lca2 in that it has more down than up-regulated tags. This basic analysis confirms that while wild-type *Dunaliella* has a strong photoacclimation response at the light levels tested, these have been largely reduced in both lca1 and lca2 at both the phenotypic and transcriptional levels.

5.1.2 Similarity in regulation between mutants and light levels

Based on the results from the previous section (5.1.1) showing a seemingly similar pattern in the number of down and up-regulated tags in lca1 and lca2 under changing light levels (see Figure 5.1) coupled with the low number of SDE tags between lca1 and lca2 (Figure 5.2) under medium light, it is likely that a number of tags show similar expression profiles compared to the wild-type. Tags with a similar fold-change in the lca strains compared to the wild-type under a specific condition should show a strong linear correlation.

(a)



(b)

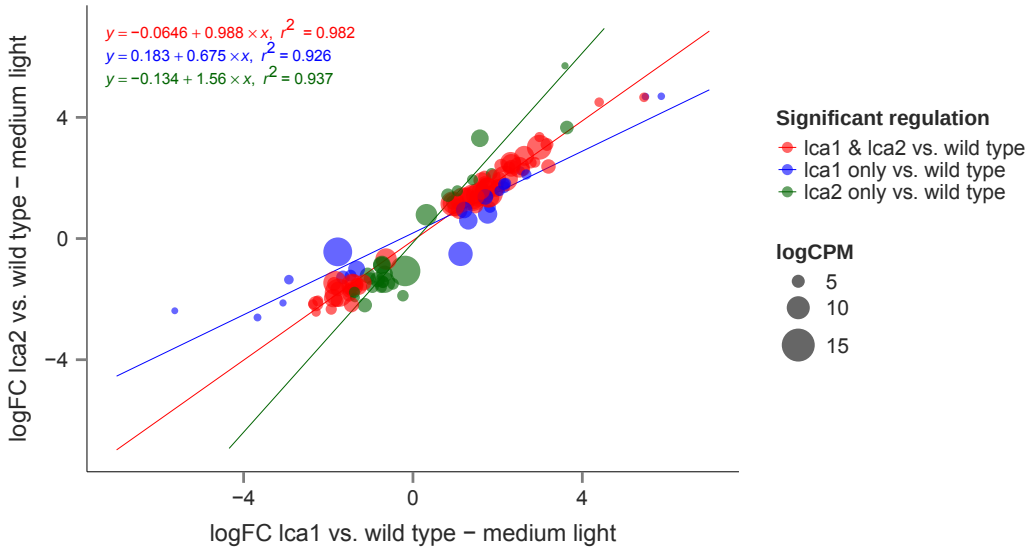


FIGURE 5.3: The relation in log fold-change for SDE tags when comparing lca1 versus wild-type (x-axis) against lca2 versus wild-type (y-axis) under a) low light and b) medium light. SDE tags with an FDR of <0.001 in both comparisons are marked red, while tags meeting the criteria only in lca1 versus wild-type are marked blue, and those only meeting the criteria in lca2 versus wild-type are marked in green. The result from a linear (least square) regression to each distribution is shown as a line together with the calculated equation and r-squared value. In addition the total summed count for each tag is shown as the log count per million (logCPM) as a measure of the overall abundance of the tag within the data set.

Figure 5.3 shows the correlation in tag expression under low and medium light in groups of SDE tags (FDR <0.001) for tags that are either SDE in both lca strains or only in lca1 or lca2 versus the wild-type. High light is not shown due to a low number

of SDE tags under this light level (see Figure 5.4 below for details). The results show a strong correlation with $r^2 > 0.908$ under low light and $r^2 > 0.982$ under medium light in the tags that are SDE in both lca strains from a shared pattern of regulation.

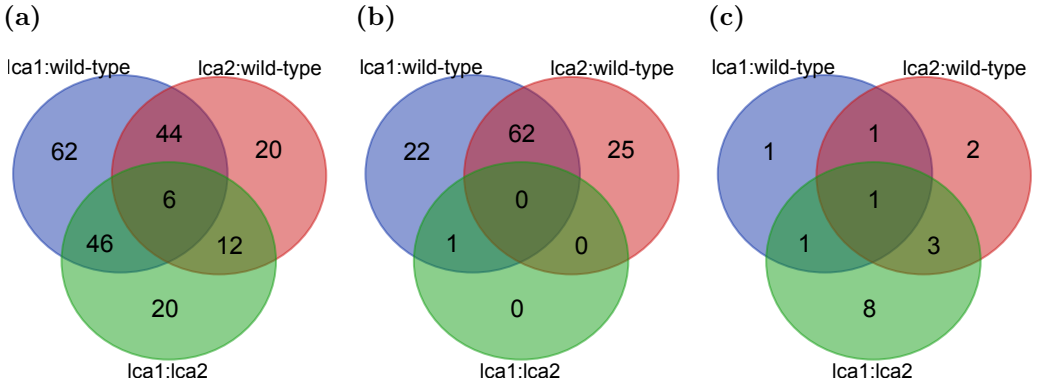


FIGURE 5.4: The number of shared SDE tags (FDR <0.001) in the comparisons lca1 vs. wild-type (blue circle), lca2 vs. wild-type (red circle) and lca1 vs. lca2 (green circle) under a) low, b) medium and c) high light.

Figure 5.3 shows that an increasing proportion of the SDE tags in lca1 and lca2 when compared to wild-type are shared under low to medium light. In addition, Figure 5.4 shows that the difference between lca1 and lca2 decreases from 20 to 0 tags when comparing the results from low to medium light, and slightly increases to 8 tags under high light. To summarize the differences between the lca and wild-type strains decreases under increasing light as measured by the number of SDE tags. In addition, the lca strains are most similar under medium light with no SDE tags between lca1 and lca2 (Figure 5.4).

5.1.3 Significantly differentially expressed tags by cellular component

The overview of the significant differential expression in Figure 5.1 clearly shows that the number of SDE tags involved in photoacclimation are far lower in the lca strains than wild-type. Breaking down this regulation by cellular components and pathways to increase the resolution may help to identify any specific gene or pathway that plays a significant role in the observed pattern of regulation and possibly the phenotype. As part of the annotation process (detailed in Chapter 4), all tags with a significantly matched UniProt ID were also assigned appropriate gene ontology (GO) identifiers linking the tag to cellular components, biological processes and molecular functions. GO identifiers are assigned in a top-down fashion with each level adding further and more detailed information about the specific component, process or function assigned

to the identifier in question. This added layer of information tends to be both useful and highly complex due to the multi-layered nature of ontologies mirroring the complexity of the biological systems that they are describing. To help summarize the information confined within the GO terms, smaller sets of gene ontologies, referred to as GO slims with limited levels of assignment can be used. Figures 5.5 and 5.6 shows some results from the GO terms assigned to significantly regulated tags under selected conditions using a GO slim (slimmed down GO ontology) specific for plants (http://www.geneontology.org/ontology/subsets/goslim_plant.obo).

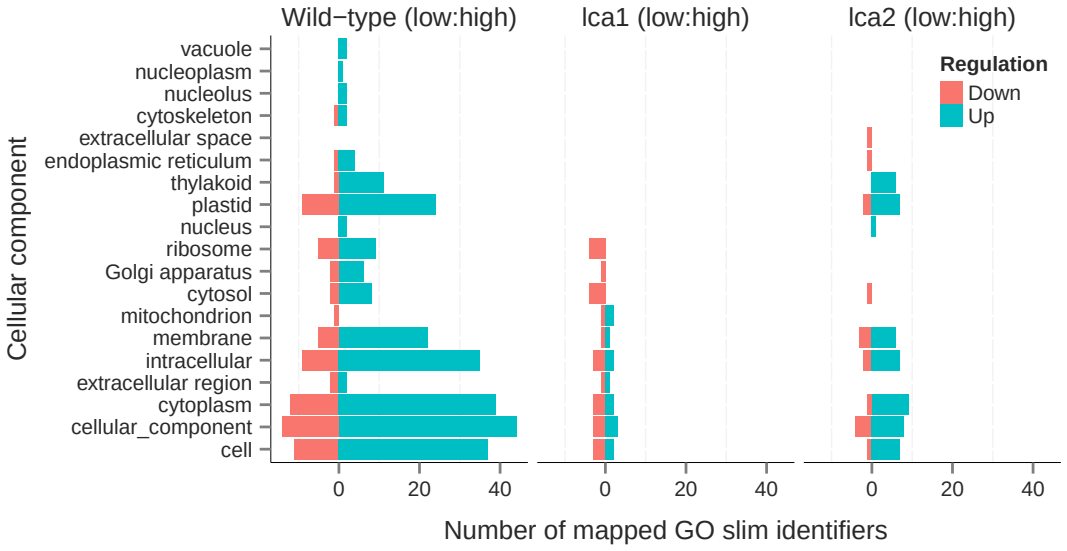


FIGURE 5.5: Number of assigned GO identifiers per GO term for cellular components included in a GO slim set specific for plants with GO identifiers from tags with an FDR <0.001 in each compared condition on a light basis (wild-type, lca1, lca2 under low to high light). GO identifiers linked to tags with a negative fold-change (down regulated) are shown in pink while tags with a positive fold-change (up regulated) under the compared condition are shown in blue. GO terms included in the original GO slim set lacking any assigned GO identifiers have been removed from the set.

Figure 5.5 shows the number of assigned GO terms per cellular component in the wild-type and lca strains from tags that are significantly (FDR <0.001) up or down-regulated under low versus high light. Since terms are assigned in a top-down basis, more general components (cell, cellular-component, intracellular and similar) will be represented by the majority of the tags. Rather, it is the smaller compartments assigned fewer numbers of GO terms that are of a higher interest. As evident in the figure, the plastid component is well represented in both wild-type and lca2, while completely missing from lca1, implying that no significant regulation using a conservative false discover rate of 0.001 is directly linked to the plastid component. The rather modest light-regulated changes that occurs in the lca1 is instead linked to the ribosome, the Golgi apparatus and the cytosol in general. This does not suggest that the

mutations responsible for the lca1 phenotype are not involved in the plastid, instead it suggests that the fold change, reflecting the tag count is static across light levels. In contrast, the changes related to the SDE tags in lca2 by light are clearly localized to the plastid and specifically the thylakoid membrane within the chloroplast. Also of interest is the small number of GO terms linked to the extracellular space that show no significant changes in either the wild-type or lca1. Further, the clear changes seen in ribosome, Golgi and cytosol components, that show strong regulation in the wild-type and modest regulation in lca1 is missing in lca2.

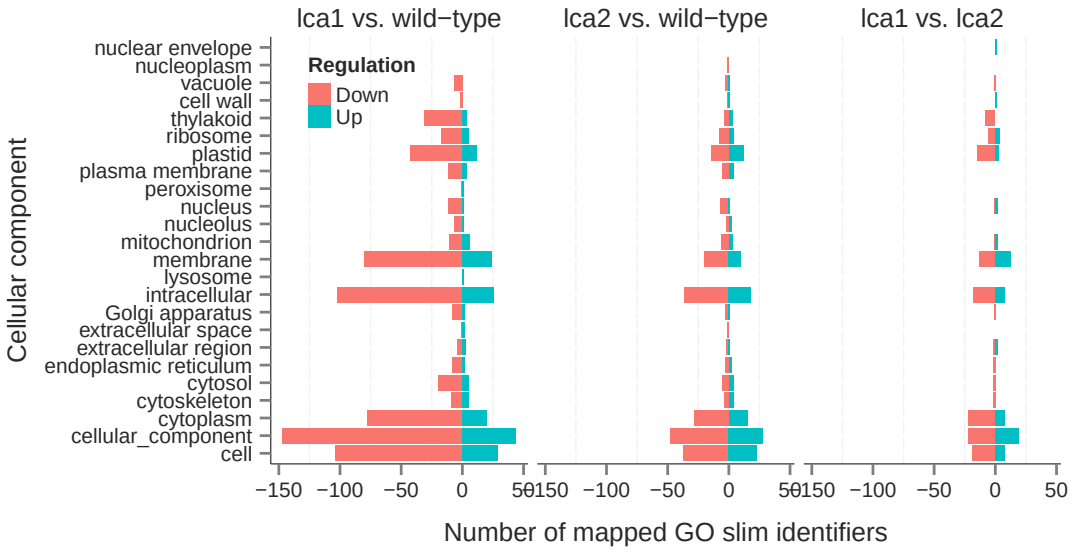


FIGURE 5.6: Number of assigned GO identifiers per GO term for cellular components included in a GO slim set specific for plants with GO identifiers from tags with an FDR <0.001 in each compared condition on a strain basis (lca1 vs. wild-type, lca2 vs. wild-type or lca1 vs. lca2 under low light). GO identifiers linked to tags with a negative fold-change (down regulated) are shown in pink while tags with a positive fold-change (up regulated) under the compared condition are shown in blue. GO terms included in the original plant GO slim set lacking any assigned GO identifiers have been removed from the set.

Figure 5.6 shows the distribution of GO terms derived from tags that are SDE between the wild type and lca strains under low light. The low light condition was selected because this condition shows the greatest difference in regulation. A large fraction of SDE tags between the wild-type and lca1/lca2 are linked to the plastid and thylakoid, suggesting the transcriptome contains relevant information for the molecular characterization of the phenotype. The actual components and groups that are significantly over-represented in relation to the overall pattern of regulation are analysed in the next section.

5.1.4 Enrichment analysis of significantly differentially expressed tags

Changes in gene expression that target specific pathways, processes or components of the cell can be identified using an enrichment analysis. By comparing genes identified under specific conditions versus a background dataset containing all identified genes, pathways assigned either a higher (or sometimes lower) than expected number of genes can be identified and subjected to detailed analysis. The result from an enrichment analysis varies to a certain degree depending on the statistical methods used (Hulsegge et al., 2009), whether or not the significance of the gene expression is taken into consideration, and the significance cut-off when calculating the enrichment score. Results from analysing and identifying possible enrichment in SDE tags deploying different methods largely agreed and the result from the first part of this analysis using the web-based Gene Ontology Enrichment Analysis Software Toolkit (GOEAST, available at http://omicslab.genetics.ac.cn/GOEAST/) (Zheng and Wang, 2008) is presented below.

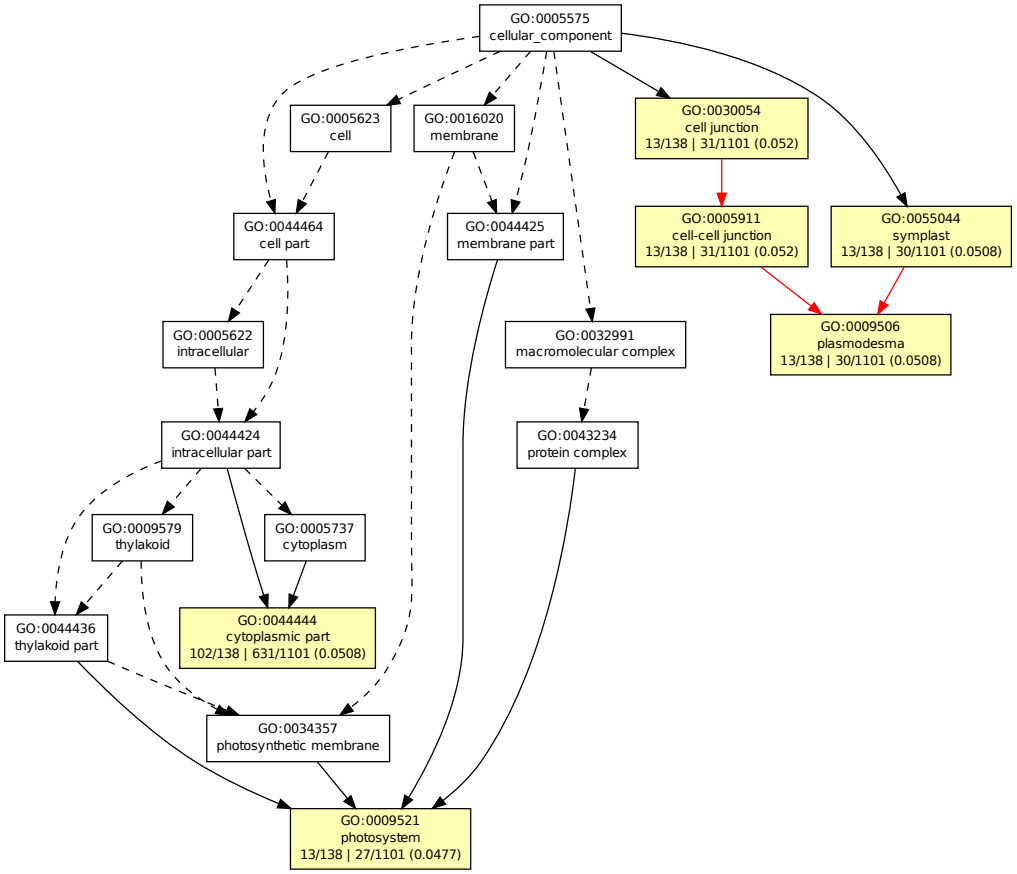


FIGURE 5.7: Enrichment analysis of GO identifiers for cellular components from tags significantly differentially expressed in lca1 versus wild-type under low light. GO terms identified as significantly enriched in GO identifiers linked to these tags are shown in yellow, while non enriched terms are shown in white. The GO terms are hierarchically linked from parent to child term, non significant terms are linked by dashed lines, non-significant to significant terms are linked by solid lines and significant terms are linked by solid red lines. Below enriched terms, the number of mapped GO identifiers out of the total within the dataset are shown, followed by the number of associated GO identifiers out of the total number of identifiers within the full GO database and the calculated p-value for the significance of the enrichment.

The analysis only identified cellular components in lca1 compared to wild-type under low light, as significantly enriched between strains when comparing the lca strains against wild-type or against each other under the same light level. The result from the enrichment test using GOEAST is presented in Figure 5.7 with significantly enriched cellular components marked in yellow. The lowest p-value indicating the significance of the enrichment for each enriched term is represented by the GO term "photosystem", showing that photosynthesis linked tags are most over-represented within the

SDE tags. The second group of interest includes tags mainly linked to ribosomal activity and is related to the annotation of the tags rather than the actual cellular component.

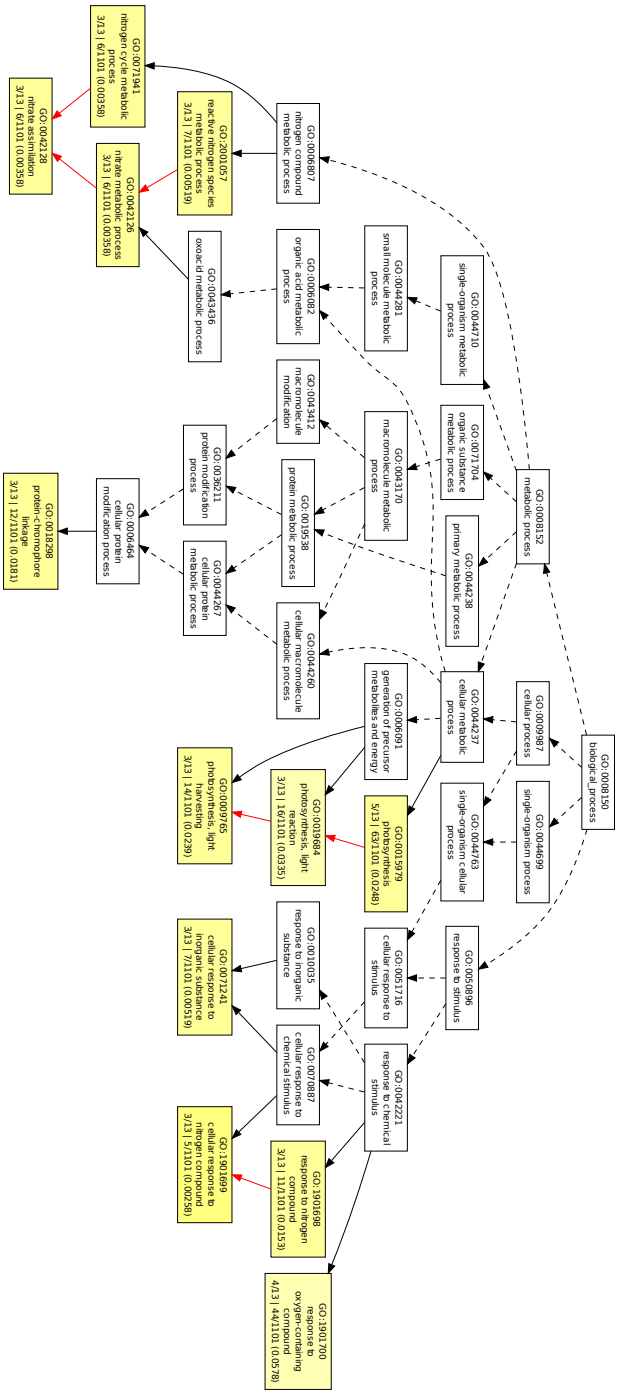


FIGURE 5.8: Enrichment analysis of GO identifiers for biological processes from tags significantly differentially expressed in wild-type under low to medium light. GO terms identified as significantly enriched in GO identifiers linked to these tags are shown in yellow, while non enriched terms are shown in white. The GO terms are hierarchically linked from parent to child term, non significant terms are linked by dashed lines, non-significant to significant terms are linked by solid lines and significant terms are linked by solid red lines. Below enriched terms, the number of mapped GO identifiers out of the total within the dataset are shown, followed by the number of associated GO identifiers out of the total number of identifiers within the full GO database and the calculated p-value for the significance of the enrichment.

The analysis of enriched pathways, processes or functions related to light induced changes within a strain yielded far more significant results. Since the data are extensive and involve a large number of enriched pathways, the results will be discussed in terms of closely related groups of terms showing significant enrichment as well as single terms when appropriate.

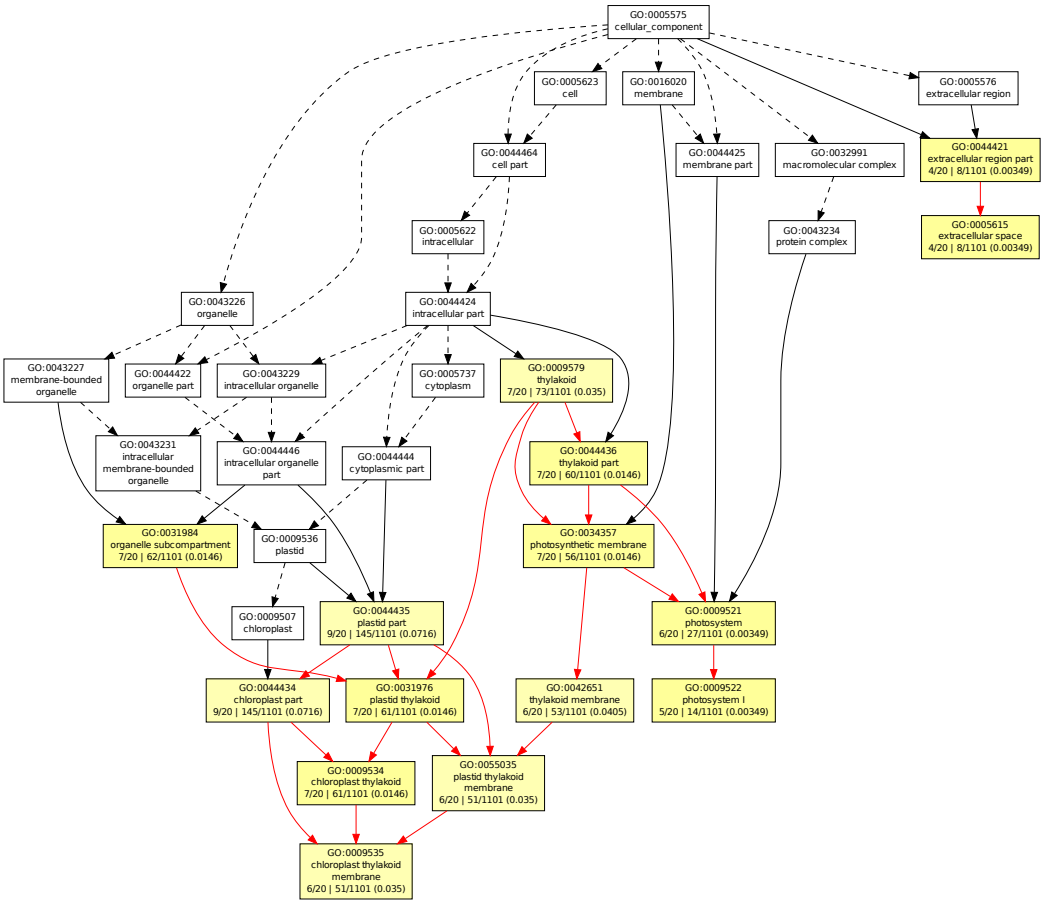


FIGURE 5.9: Enrichment analysis of GO identifiers for cellular components from tags significantly differentially expressed in lca2 under low to medium light. GO terms identified as significantly enriched in GO identifiers linked to these tags are shown in yellow, while non enriched terms are shown in white. The GO terms are hierarchically linked from parent to child term, non significant terms are linked by dashed lines, non-significant to significant terms are linked by solid lines and significant terms are linked by solid red lines. Below enriched terms, the number of mapped GO identifiers out of the total within the dataset are shown, followed by the number of associated GO identifiers out of the total number of identifiers within the full GO database and the calculated p-value for the significance of the enrichment.

The regulation in the wild-type when comparing low to medium light is largely confined to two groups of processes, namely nitrogen metabolic processes, specifically nitrogen transportation, and photosynthesis-related processes including photosystem and LHC related tags (see Figure 5.8). The enrichment in nitrogen processes is likely

to be closely linked to the enrichment (mainly up-regulation) of photosynthesis-related tags under low light. This is expected since many plants increase the amounts and partitioning of nitrogen in chloroplasts in response to low light (Evans, 1989).

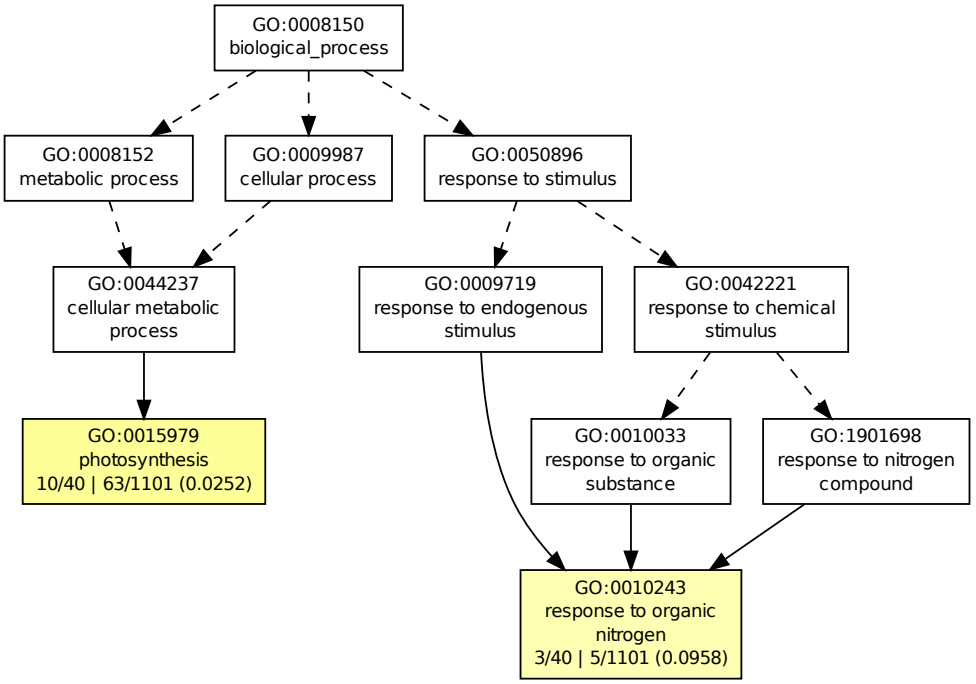


FIGURE 5.10: Enrichment analysis of GO identifiers for biological processes from tags significantly differentially expressed in lca2 under low to medium light. GO terms identified as significantly enriched in GO identifiers linked to these tags are shown in yellow, while non enriched terms are shown in white. The GO terms are hierarchically linked from parent to child term, non significant terms are linked by dashed lines, non-significant to significant terms are linked by solid lines and significant terms are linked by solid red lines. Below enriched terms, the number of mapped GO identifiers out of the total within the dataset are shown, followed by the number of associated GO identifiers out of the total number of identifiers within the full GO database and the calculated p-value for the significance of the enrichment.

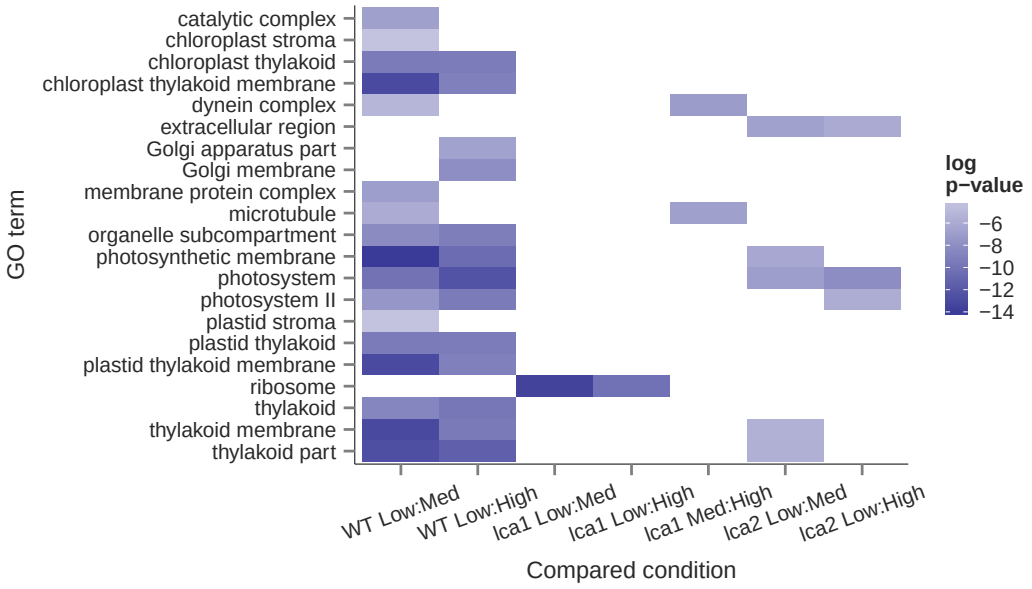
The changes in lca2 under low to medium light are similar to those in the wild-type with a strong emphasis on processes and functions directly involved in photosynthesis when using a stringent FDR cut-off of 0.001 (Figure 5.9), using the more relaxed 0.05 cut-off also picks up the enrichment in nitrogen processes seen in wild-type, as shown in Figure 5.10, which shows the enrichment in biological processes under this condition.

No significant enrichment for any particular process, function or component was discovered in lca1 either under low to medium or low to high light again suggesting that this strain show minimal signs of light induced regulation of cellular components or

processes. Rather transcript levels in this strain are largely independent of the light level suggesting that the cell has either lost the ability to regulate and or sense changes in ambient light.

The results shown so far for enrichment are based on GOEAST enrichment analysis using nuclear encoded tags passing the FDR threshold for significant regulation. As previously mentioned, the result from one enrichment analysis to another to some extents depends not only on the algorithm used to calculate the enrichment scoring and on the type of analysis, but also on the cut-off used when selecting the elements when applicable. Regardless the result from different tools should yield similar results. The results using another enrichment analysis tool, ErmineJ (Lee et al., 2005) (http://erminej.chibi.ubc.ca/) (presented below) that supports a number of different types of enrichment analysis, largely supported the results from GOEAST.

(a)



(b)

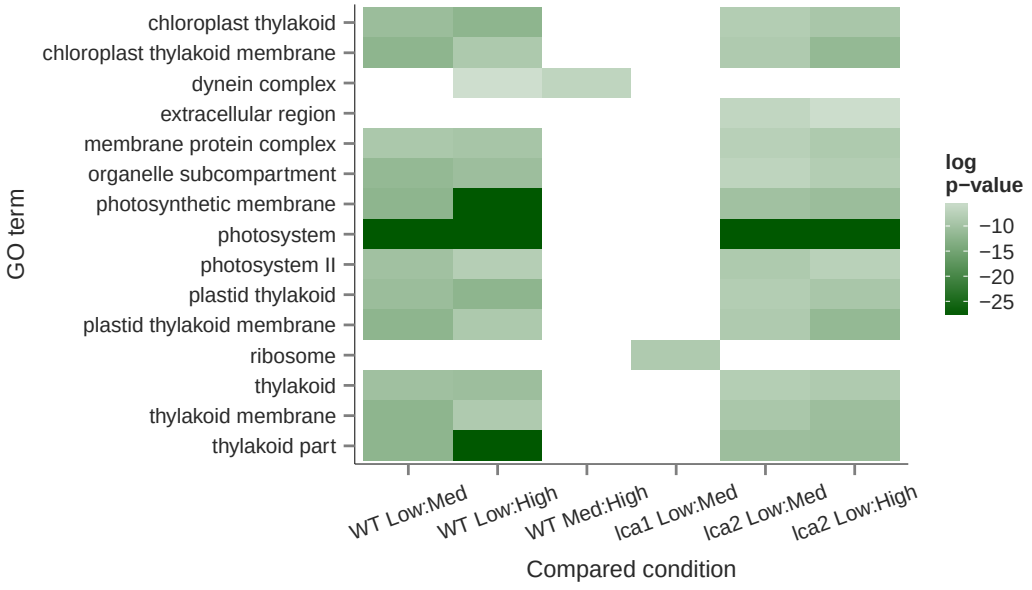


FIGURE 5.11: Enrichment in GO terms for cellular components in the wild-type, lca1 and lca2 strain when comparing the conditions low to medium, low to high and medium to high light using a) receiver operator characteristic (ROC) scoring or b) gene score re-sampling (ROC) to calculate the p-values for the gene-set scoring. Components meeting or exceeding the criteria (multi corrected p-value <0.01) are shown in colour. Conditions for which no genesets meet the criteria have been excluded.

The results from the GOEAST enrichment analysis shown in the previous figures in this section all use sets of genes passing a specific criteria (in this case an FDR of < 0.001) as the input during the analysis which is weighted against a background list (the full set of identified genes or tags). However, tags that do not pass this criteria

are not taken into consideration, neither is the ranking of the tags in the list based on the FDR or the actual FDR itself. Figure 5.11 shows the results from an alternative analysis using a) receiver operator characteristic (ROC) based scoring or b) gene score re-sampling (GSR) to calculate the degree of enrichment (described below). Both types of analysis are non-parametric in relation to the supplied list of genes (i.e. no threshold for the selection of the genes (tags) are used), rather the results are based on either the ranking of the genes (ROC) or the value of the gene-score (FDR in this case) for the GSR analysis (Lee et al., 2005). During the latter analysis a selection of genes from the full supplied list is sampled to estimate the significance of the calculated p-value for genes within the gene-set (Pavlidis et al., 2002). The analysis presented here used 200,000 iterations and full re-sampling according to the recommendations in the tutorial for the software package (http://erminej.chibi.ubc.ca/) and corrects for multi-functional genes that fall into multiple components or categories. The ROC based analysis on the other hand does take the ranking of the gene-score into consideration but places no weight on the actual score (Lee et al., 2005).

The results for the ROC and GSR analyses differ to a certain degree but largely the difference can be related to whether or not a gene-set is considered enriched or not, which is related to how the scoring is calculated in each type of analysis. Figure 5.11 shows the results from changes in gene expression on a light basis. The results largely mimics previous results with wide-ranging enrichment in components related to photosynthesis in the wild-type when comparing conditions low to medium and low to high light. Similar changes are also seen in the lca2 strain under the same conditions, however the changes involve fewer components suggesting that either the number of genes or their rank in the case of ROC or the change in gene-score (GSR) driving the enrichment is lower than in the wild-type. In neither wild-type nor lca2 do the changes from medium to high light result in any significant enrichment other than in dynein complex related components in the wild-type under medium to high light. In lca1 the number of enriched components are far lower regardless of condition, significant enrichment was only seen in ribosomal linked elements under low to medium or low to high light, or in dynein complex and microtubule related components under medium to high light.

To summarize the enrichment in processes, functions and components on a strain basis (i.e. when comparing either lca1 or lca2 to wild-type under one light level) are limited to lca1 when compared to the wild-type. The changes are mainly directly linked to photosynthesis related components, ribosomal activity or transport and regulation

of chloroplast components. Most of this enrichment is driven by tags that are down-regulated in relation to the wild-type under the same condition and visible when comparing changes under low light. While considerable differences exist between the wild-type and lca2 they do not result in significant enrichment of any specific components, functions or processes suggesting that the changes are not closely linked to any specific category but rather cell wide. The light driven changes on the other hand are mainly restricted to the wild-type and lca2, the changes in lca2 by light is very similar to the changes seen in the wild-type but the response is of a smaller magnitude. There are however also a few categories that do show enrichment that may be of specific interest, these include the tags that are involved in the extracellular components seen in Figure 5.10 and Figure 5.11 as well as the lack of enrichment in components related to the Golgi apparatus and cytoplasmic vesicles.

5.2 Nuclear encoded photosynthesis genes

While a number (32 and 31 in *A. thaliana* and *C. reinhardtii*) of the core components of photosystem II (PSII) and I (PSI) as well as the cytochrome b_6f complex and ATP synthase (ATPase) are encoded on the chloroplast genome the majority (35 and 37 in *A. thaliana* and *C. reinhardtii*) and of genes linked to these systems are nuclear encoded (Allen et al., 2011). This includes genes encoding proteins involved in stabilizing and enhancing the function of the oxygen evolving complex of photosystem II, the electron transport system, repair and reduction of photoinduced damage by non-photochemcal quencing (NPQ) of excess light, and also proteins linked to the distal parts of the core such as psaF and psaG that forms parts of the outer core antenna in photosystem I. The photosynthesis genes identified as nuclear encoded (14 in total) in the SAGE transcriptome and part of the core photosynthesis pathway with the exception of genes encoding proteins in the LHC are shown below in Table 5.1.

TABLE 5.1: Nuclear encoded genes that are part of the core photosynthesis pathway with the exception of light harvesting complexes identified in the full SAGE transcriptome. The UniProt identifier linked to the annotated tag is shown together with the identified gene, the resulting protein and the main location of the protein within the pathway.

UniProt	Gene	Protein	Component
Q9SBN6	*psbO*	Oxygen-evolving enhancer protein 1	PSII - OEC
O49344	*psbP*	Oxygen-evolving enhancer protein 2-2	PSII - OEC
Q41643	*psbQ*	Oxygen-evolving enhancer protein 3	PSII - OEC
P49108	*psbR*	Photosystem II 10 kDa polypeptide	PSII
Q02060	*psbS*	Photosystem II 22 kDa protein	PSII
Q9SPI9	*psbW*	Photosystem II reaction center W protein	PSII
Q9LR64	*psb27*	Photosystem II repair protein PSB27-H1	PSII
Q0JG75	*psb28*	Photosystem II reaction center PSB28 protein	PSII
P12356	*psaF*	Photosystem I reaction center subunit III	PSI
P14224	*psaG*	PSI-G: Photosystem I reaction center subunit V	PSI
P18068	*petE*	Plastocyanin	Electron transport
P53991	*petH*	Ferredoxin–NADP reductase	Electron transport
Q93VA3	*petJ*	Cytochrome c6	Electron transport
A8J785	*atpF0B*	ATP synthase subunit b'	ATPase

5.2.1 Photosystem II genes

The oxygen evolving complex (OEC) is formed of three membrane-extrinsic proteins attached on the lumenal side of the thylakoid membrane and encoded by the *psbO*, *psbP* and *psbQ* genes in green algae and higher plants (Pérez-Bueno et al., 2011, Roose et al., 2010). PsbO stabilizes the manganese cluster and allows for high rates of oxygen evolution (Nelson and Yocum, 2006). Two different isoforms of the psbO gene (*psbO1* and *psbO2*) have been found in *A. thaliana* and shown to differ in function, with *psbO1* being the dominant form in wild-type (Murakami et al., 2005). Mutants with impaired PsbO function have been shown to be deficient in oxygen evolution activity and show signs of altered transitions between S-states during cycling between redox states in the OEC (Roose et al., 2010). PsbO is not only involved in the OEC but also plays a structural role, removing the protein reduces the interaction between the two core PSII units while removal of PsbO and or PsbP units shifted LHC units towards the core of PSII (Boekema et al., 2000). In addition, the protein likely stabilizes the complex by providing binding sites for PsbP and PsbQ (Bricker and Frankel, 2003, Tohri et al., 2004), the latter which in turn is bound to PsbP. Plants deficient

in PsbP are highly light sensitive and show a marked decrease in photosynthetic efficiency (20% decrease in F_v/F_m) under 150 μmol photons m^{-2} s^{-1} of light (Ifuku et al., 2005). PsbP have been shown to bind Mn and likely supplies Mn ions to D1 turnover due to light induced damage (Bondarava et al., 2005), which could explain the light sensitivity in PsbP deficient plants. PsbQ which is light inducible and only found in green tissue in *A. thaliana*, on the other hand is not essential for growth in higher plants (Grover et al., 1999, Ifuku et al., 2005). A number of different isoforms of all three genes have been found with suggested different function as reported for PsbO under different environmental conditions (Pérez-Bueno et al., 2011). The smaller (10 kDa) PsbR is sometimes associated with the OEC, elimination of the protein impairs the quantum yield of PSII however the protein is similarly to PsaQ not essential for normal growth and both proteins are likely involved in fine-tuning oxygen evolution and electron transport (Allahverdiyeva et al., 2013). PsbS seems to play a key role in NPQ with rapid quenching (qE) being strongly inhibited or slowed down in *A. thaliana* plants lacking the protein (Kereïche et al., 2010). However, the mechanisms involved in quenching vary between groups. Diatoms, for example, lack PsbS, but retain a strong ability to rapidly quench excess light (Niyogi et al., 2005). PsbW increases in dark grown plants exposed to light and it has been shown that PSII dimeric supercomplexes (containing PSII core dimers and associated LHCII) are lacking from *A. thaliana* plants deficient in the protein (Bishop et al., 2003, Shi et al., 2000). Psb27 is similarly extrinsic and localized on the lumenal side of the membrane. Similarly to PsbW the protein is not vital to normal function and instead involved in the repair of high light induced damage to the PSII core with a possible guiding role during assembly of the complex (Chen et al., 2006). Psb28 share a 16% sequence similarity with PsbW and is situated on the cytoplasmic side of the thylakoid membrane in *Synechosystic sp.* PCC 6803 (Dobáková et al., 2009), the protein is thought to stabilize the CP43-less PSII monomer during assembly of the core complex and mutants with a disrupted *psb28* gene are sensitive to high light and high temperature induced stress (Sakata et al., 2013).

5.2.2 Photosystem I genes

The two identified genes (*psaF* and *psaG*) encoding PSI subunits III (PsaF) and V (PsaG) (Table 5.1) are both intrinsic to the PSI core complex (Dekker and Boekema, 2005) and are located at the periphery of the complex with PsaG, likely bound to PsaB (Zygadlo et al., 2006). Both proteins are likely interacting with the distal LHC antenna (LHCI) (Jensen et al., 2002, Zygadlo et al., 2006). Transformed *A. thaliana* plants with highly reduced PsaG protein content resulted in a 20 to 40% reduction in

PSI core subunits while still displaying a normal phenotype (Jensen et al., 2002). This is explained by a reduction in measured state 1-state 2 transitions of 55%, coupled with a corresponding increase in $NADPH^+$ photoreduction by 48%, increasing the efficiency of PSI and likely compensating for the measured reduction in the number of PSI under optimal light conditions. Apart from a small (5 - 10%) reduction in *lhca1* and *lhca4* no direct effect on the overall light harvesting complex of PSI (LHCI) or chlorophyll b content was observed. However, the results did find indications in support of a reduced stability in the binding of the LHCI to the core complex (Jensen et al., 2002). Both PsaG and PsaF interact with plastocyanin (Pc), PsaF is the likely docking site for Pc and PsaG increases the affinity of Pc for P700 through a currently unknown process (Zygadlo et al., 2005). The lack of PsaG leads to an upstream imbalance in the electron flow from PSII despite the increased efficiency of PSI leading to a far more reduced plastoquinone (Pq) pool. In contrast, the $NADPH^+$ reduction rate in *A. thaliana* mutants with reduced PSA-F content is only 10% of the wild-type level with no fast electron transfer between Pc and P700, the PSI core complex is also destabilized and sensitive to photoinduced damage (Haldrup et al., 2000). PsaG is also interesting from the perspective of mutagenesis since it is imported into the thylakoid membrane not using the Sec/SRP pathways but rather using the so called spontaneous insertion that has also been suggested as the import mechanism for PsbW (Kim et al., 2015). This process depends on the presence of a positively charged protected stromal loop in PsaG. Mutations in this loop have been shown to disrupt the insertion into the thylakoid membrane (Zygadlo et al., 2006).

5.2.3 Electron transport genes

The SAGE analysis also identified three genes encoding proteins involved in the electron transport chain, including plastocyanin, cytochrome c6 and ferrodoxin-NADP reductase. Plastocyanin, generally considered to be the main component responsible for the transfer of reduced electrons from the cytochrome b_6f complex to PSI is complemented by cytochrome c6 in green algae and (Gupta et al., 2002, Sommer et al., 2006, Wood, 1978). These two proteins perform a seemingly interchangeable role with either one being sufficient for efficient electron transport in *A. thaliana* (Gupta et al., 2002). Ferredoxin-NADP reductase (FNR) is thought to play a key role in the allocation of electrons from PSI to linear or cyclic electron flow (Kitayama et al., 1994). In addition, a nuclear encoded component (*atpF0B*) of the ATP synthase in *C. reinhardtii* has been identified (Osnabriick and Sheva, 1998)

5.2.4 Light harvesting complex genes

The accessory light-harvesting complex (LHC) protein, (also known as the chlorophyll a-b binding (CAB) proteins) of PSII and PSI both channels, and modulates, the amount of excitation energy that reaches the core of each photosystem (Grossman et al., 1995, Smith et al., 1990). These apoproteins that make up the majority of the membrane associated proteins, form a scaffold that, in addition to chlorophyll a and b, also binds various carotenoid and xanthophyll pigments in plants and algae (Grossman et al., 1995). While the core complex of both photosystems are strongly conserved in oxygen evolving organisms, the gene families encoding the LHC proteins differ among groups of organisms (Drop et al., 2014). The LHC complexes bind the majority of the total chlorophyll pigments that are responsible for light harvesting, with minimal (functioning) PSII/PSI core units binding around 37 and 95 chl a compared to around 250 to 300 chl a/b when associated with the full LHC complement in the green algae *C. reinhardtii* (Kirst et al., 2011). The actual size of a minimal functional photosystem unit and the complementing LHC differs among species, light conditions and how the minimal unit is defined, however the size and structure of the LHC are central to any effort to increase the photosynthetic efficiency by minimizing the loss of energy due to excessive absorption of light.

The exact number of genes encoding light harvesting complex genes in *D. tertiolecta* or other species of the genus is still unknown. The results of Laroche et al. (1990) suggested that the full (CAB) gene family in *D. tertiolecta* consists of 3-5 members encoding 4 proteins. This analysis identified 4 genes linked to LHCII as inferred from the annotation. The dominant tag based on total counts was linked to the gene *dsLhcII-3*, first isolated from a *D. salina* cDNA library and encoding a 261 amino-acid LHCII CAB protein. The resulting mRNA from this gene was shown to drop markedly in abundance during a shift of cultures from high to low light (Wei et al., 2007). The closest similarity to this gene outside of the *Dunaliella* genus is found in two slightly shorter sequences in *Monoraphidium neglectum* (76%) and *C. reinhardtii* 75%. *dsLhcII-3* cluster together with another LHCII linked gene as seen in Figure 5.12 with the closest match against either a gene encoding a 28.5 kDa protein in *D. tertiolecta* or *dsLhcII-2.1* in *D. salina* as well as a third sequence variant of the latter (*dsLhcII-2.2*) not shown here. This latter group of genes is largely identical, the translated sequence consists of 253 amino-acids which differs in two single amino-acid at position 126 and 241. The two sequences from *D. salina* differ in a single amino-acid in position 126. The sequence from *D. tertiolecta* again differs in position 241 from the previous two sequences, this gene encoding a 28.5 kDa protein was determined to be encoding one out of four CAB apoproteins in *D.tertiolecta* as identified using SDS-PAGE.

The N-terminal amino-acid sequence for these four proteins were sequenced as well with one of them matching the previously mentioned 28 k.5 kDa sequence (Laroche et al., 1990). The remaining three N-terminal sequences does not match any of the sequences listed in Figure 5.12 or other known LHC linked genes from *Dunaliella*. The closest match outside of the genus is found within the Chlorophyceae lineage in *Volvox carteri f. nagariensis* with a sequence similarity of 78% (XP_002952704), and in *C. reinhardtii* with a similarity of 77% (XP_001693987), suggesting a fairly high degree of divergence from other chlorophytes. Loosely associated with this cluster is a gene first isolated in *D. salina* and identified as a minor chlorophyll a/b binding protein, only a small number of tags were matched to this gene and its likely not of significant importance. More interesting is the identification of tags linked to an EST annotated as *lhcb4* since CP29, the proteins encoded by this gene, has not been described in *D. tertiolecta* previously and it has been shown to be involved in state transitions in *C. reinhardtii* (Kargul et al., 2005).

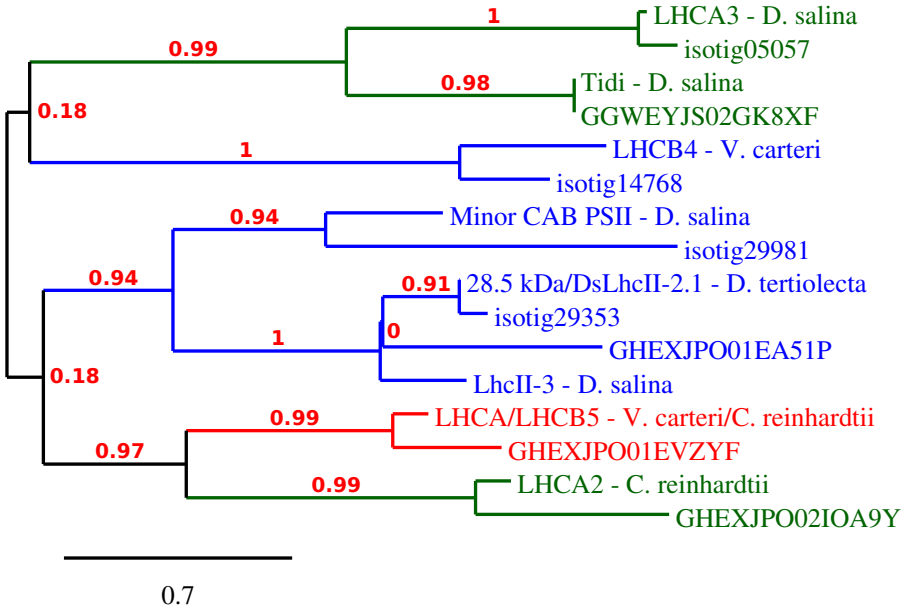


FIGURE 5.12: Un-rooted tree showing the relationship between the identified LHC genes and the matched EST within the SAGE transcriptome. LHCI linked genes are marked in green, and LHCII linked genes are marked in blue, while LHC linked genes with unknown association to either photosystem are shown in red. Branch support values are shown in red numbers.

The most abundant tag linked to a LHCI protein was annotated as *lhca3*, as inferred from a gene identified in *D. salina* (Xu, 2012). The mature protein was calculated to consist of 257 amino-acids with a molecular weight of 27.5 kDa, which matches the weight of one of the three LHCI proteins identified by Sukenik et al. (Sukenik et al., 1988). The LHCI family of proteins in *D. salina* were shown to have a low degree of

similarity with other LHCI proteins and showed a higher degree of divergence within *D. salina* compared to the proteins within the LHCII family in the same species (Xu, 2012). In addition, a small number of tags matched a gene annotated as *lhca2* in *C. reinhardtii* as well as a gene linked to a protein (Tidi) induced under iron deficiency in *D. salina* (Merchant et al., 2007, Varsano et al., 2006), both of these tags were only found in small numbers (<100) but could regardless be of interest under specific conditions. It is worth noting here that a *lhca2* gene has been described in *D. salina* but the EST show a closer match to the homologue in *C. reinhardtii* suggesting that this gene is not the same as described by Xu et al. (Xu, 2012).

TABLE 5.2: Light harvesting complex genes identified in the full SAGE transcriptome. The UniProt identifier linked to the annotated tag is shown together with the identified gene, the resulting protein and the main location of the protein within the pathway.

UniProt	Gene	Protein	Location
A8BDJ0	*lhcII-3*	Major light-harvesting chlorophyll a/b protein 3	LHCII
A1XKU6	*lhcII-2.1*	Major light-harvesting chlorophyll a/b protein 2.1	LHCII
G4WUW1	*minor-lhcII*	Chloroplast minor CAB protein of PSII	LHCII
D8TY82	*lhcb4*	CAB protein of PSII	LHCII
C1K004	*lhca3*	PSI light-harvesting chlorophyll-a/b protein 3	LHCI
A8IKC8	*lhca2*	Light-harvesting protein of photosystem I	LHCI
Q56D18	*tidi*	Chloroplast Tidi	LHCI
Q9SBM9	*lhca*	Light harvesting complex a protein	LHC

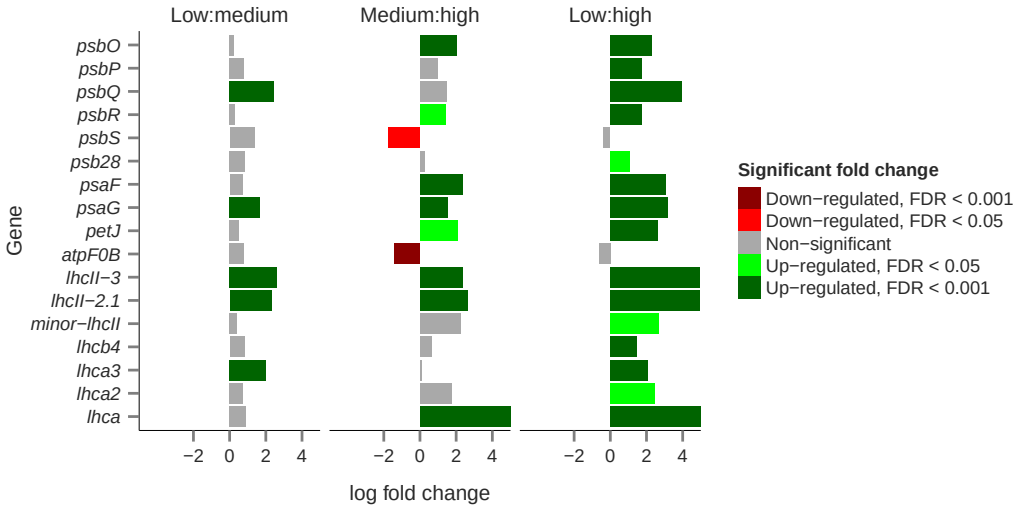
In total, this leaves four sequences annotated as linked to LHCII and three sequences annotated as linked to LHCI as inferred from the annotation, which could encode seven or more proteins depending on the amount of splicing and other post-translational modifications asserted during the passage and translocation of the gene product from these nuclear encoded genes into the chloroplast. However, the result, do match well with the results from Sukenik et al. that found four LHCII and three LHCI apoproteins in *D. tertiolecta* (Sukenik et al., 1988).

5.2.5 Light-driven significant differential expression of core photosynthesis and light-harvesting complex genes in wild-type and lca strains

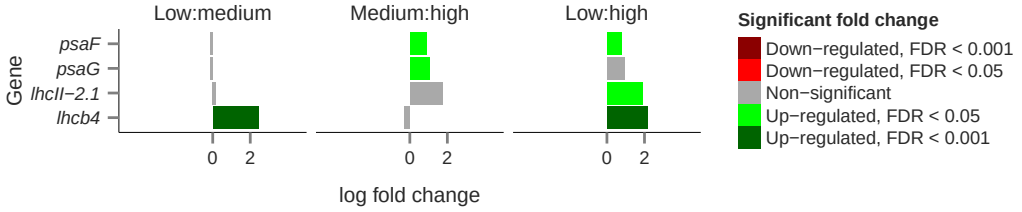
Algal cells can respond to changes in light primarily using two different strategies, they can either change the number of photosystems (n-type) or change the size of each

photosystem (sigma-type) (Falkowski and Owens, 1980b, Six et al., 2008). Wild-type *D. tertiolecta* have previously been shown to respond to a shift from high to low light by increasing the number of PSI reaction centres while largely maintaining the size of each PSI unit, as measured by the Chl/P700 ratio (Falkowski and Owens, 1980b). The change from low to high light (70 to 1900 μmol photons m^{-2} s^{-1}) resulted in a reduction of up to 75% in the number of PSII and PSI reaction centres (Sukenik et al., 1987). These n-type changes are complemented by changes in the Chl a:b ratio, likely caused by the accumulation of apoproteins with different capacities for the binding of Chl b rather than the total abundance of Chl a:b binding apoproteins (Sukenik et al., 1988).

(a)



(b)



(c)

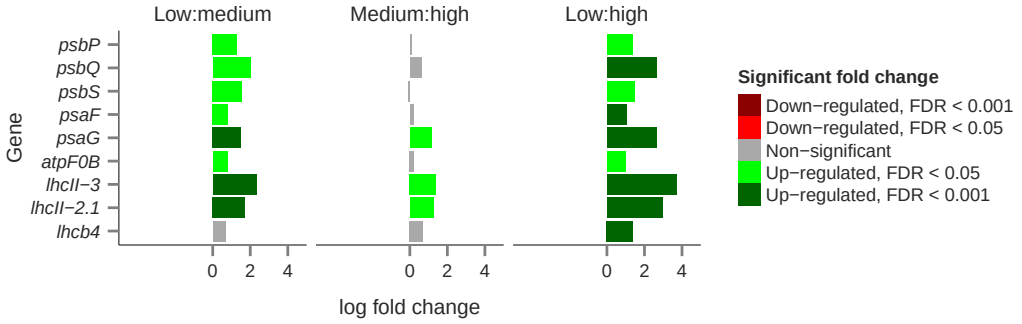


FIGURE 5.13: Significantly differentially expressed core photosynthesis and light harvesting complex genes compared under low to medium, medium to high and low to high light in a) wild-type, b) lca1 and c) lca2. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying an FDR of <0.001 and a lighter colour an FDR of <0.05. The expression is shown as the log fold change.

As evident in Figure 5.13, the wild-type of *D. tertiolecta* shows a strong regulation of gene expression by light, eight of the identified core photosystem and seven of the LHC genes are up-regulated, while none are down-regulated when the light intensity is reduced. A very strong (>16-fold (log 4)) increase is seen in the LHCII/LHCI linked

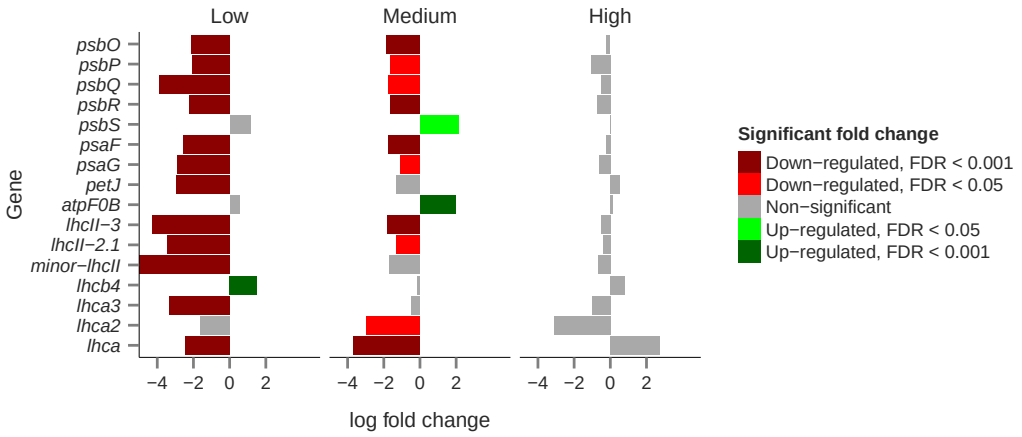
genes involved in PSII light harvesting, which is in contrast to the 2 to 4-fold increases in genes of LHCI, OEC and the transport (*petJ*) and regulation (*psaF* and *psaG*) of electrons to PSI linked genes. The switch from medium to high light as measured by the number of significantly differentially expressed genes induces a stronger response than the switch from low to medium light, suggesting that the main trigger for light regulation is activated at a light intensity at or above medium light. The result may suggest that under medium light the cells up-regulate the expression of genes involved in the OEC (*psbO* and *psbR*) as well as the two major identified LHCII linked genes (*lhcII-3* and *lhcII-2.1*), while the major LHCI linked genes remain on a level comparable to high light with no significant regulation. In addition, the expression of *petJ* encoding Cytochrome c6 and the genes that are likely to be involved in the balancing of electron flow into PSI (*psaF* and *psaG*) are up-regulated, while *atpF0B* encoding a sub-unit of ATPase is down-regulated. This σ-pattern of regulation suggests an increase in the size of the functional antenna of PSII coupled with a reduction in NPQ as suggested by a significant reduction in *psbS*. The major change when comparing low to medium light is again linked to an increase in genes encoding LHCII but also *lhca3* encoding a major LHCI protein and dominant within the identified LHCI linked genes based on tag counts. In addition, *psbQ*, encoding a protein thought to be involved in the optimization of the OEC complex and *psaG* are both significantly up-regulated.

Together these results suggest that wild-type *D.tertiolecta* strongly increases the expression of LHCII linked genes in response to medium light, with a further increase under low light, while the expression of LHCI linked genes are only up-regulated substantially under low light. The strong increase in PSI gene expression and the more modest increase in PSII OEC genes could suggest that the cells are increasing the number of photosystems, with a higher proportion of PSI which likely acts to re-balance the PSII:PSI ratio from typical high light to low/medium light conditions, as previously observed in *D. salina* (Smith et al., 1990). The up-regulation of *psbS* under high light indicating that excess light is absorbed necessitating the use of NPQ does suggest over-absorption of light (Li et al., 2000).

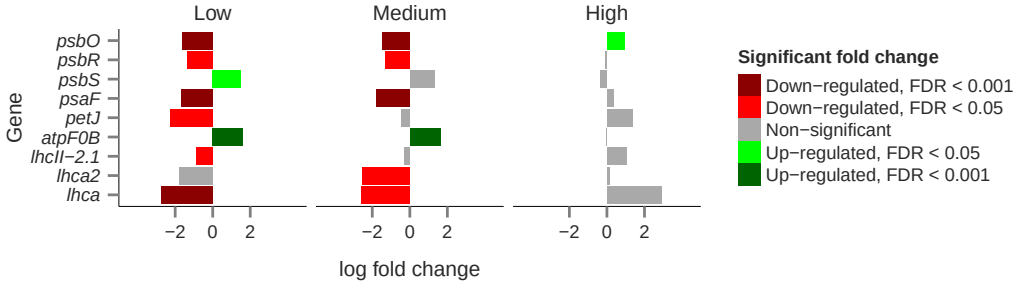
The response in the lca strains is markedly different with fewer genes showing significant regulation in conjunction with lower levels of fold change in general. The lca1 strain increases expression of the PSI genes encoding PsaF and PsaG under medium to high light as well as *lhcb4* under low to medium light. In addition, an increase in lhcII-2.1 is seen under low to high light. The increase in this latter gene suggests a certain increase in LHCII, however, the tags linked to this gene are only found in small numbers compared to lhcII-3 and it is unlikely to result in major changes in the size of LHCII. The increase in the two PSI genes could be linked to a higher PSII:PSI

ratio under high light, reflecting the generally observed pattern in *D. salina* (Smith et al., 1990). More interesting is the strong up-regulation in the *lhcb4* gene under low to medium light since CP29, the protein encoded by this gene, has been linked to state transitions (Tokutsu et al., 2009), suggesting a possibly retained capability of statetransitions in lca1. The light response in lca2 under medium to high light is confined to LHCII linked genes and *psaG* and these genes again increase from low to medium light. Two of the three main OEC genes as well as *psbS* and *psbP* of PSII also show increased expression under low to medium light, however, the lack of an increase in the expression of *psbO* encoding the OEC1 protein that connects the OEC2 and OEC3 subunits encoded by *psbP* and *psbQ* suggests that expression does not result in any increase in PSII units.

(a)



(b)



(c)

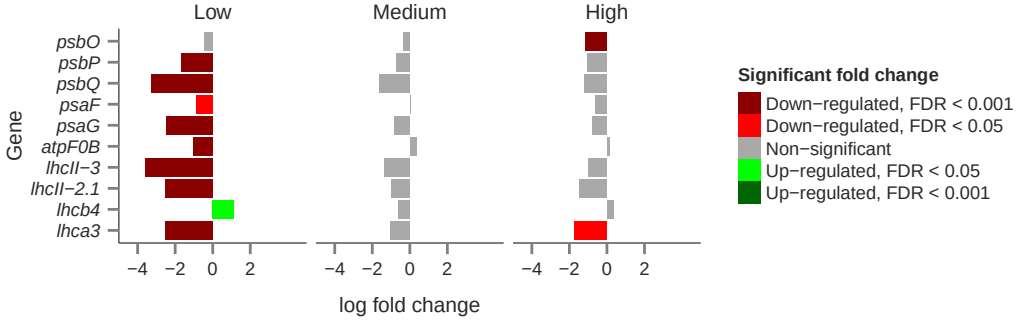


FIGURE 5.14: Significantly differentially expressed core photosynthesis and light harvesting complex genes as compared in a) lca1 vs. wild-type, b) lca2 vs. wild-type and c) lca1 vs. lca2, with the result for each strain compared under low, medium and high light. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying an FDR of <0.001 and a lighter colour an FDR of <0.05. The expression is shown as the log fold change.

When comparing the gene expression in the lca strains directly versus the wild-type it is apparent that the expression levels of photosystem and LHC linked genes in the three strains are largely identical under high light and then increasingly diverge when

compared under lower light levels (Figure 5.14). The gene expression in lca1 is arrested at a level that is largely identical to the expression pattern in the wild-type under high light, resulting in no significant gene expression when comparing the two strains under high light, followed by increasing differences under medium and low light for photosystem and LHC linked genes. In contrast in lca2 where the majority of the changes occurs in the switch from low/medium light to high (see Figure 5.14) the changes are similar to wild-type under both medium and low light. The reduced up-regulation of LHC genes, and particularly LHCII linked genes, in both lca1 and lca2 under low and medium light compared to wild-type is likely to result in a decreased amount of CAB proteins and a reduced chlorophyll binding capacity under low and medium light, mirroring the results observed in the lca phenotype with a reduced Chl content per cell and reduced σPSII (see Chapter 3). When comparing the lca strains directly, lca1 show some differences to lca2 under high light, no differences under medium light and large differences under low light mainly related to the OEC and LHCII elements discussed earlier, again since lca1 is largely arrested the changes are directly related to the changes in lca2 under different light conditions.

5.3 Significant differential expression of chlorophyll and carotenoid metabolism related genes

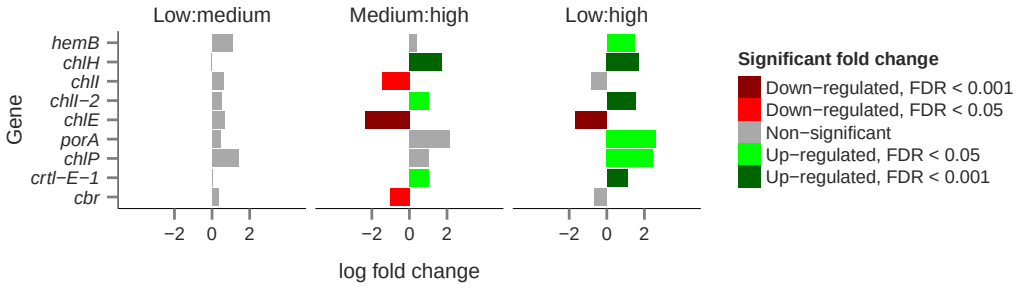
The biosynthesis of chlorophyll involves a chain of more than 15 enzymes that converts glutamic acid precursors via aminolevulinic acid (ALA) into pyrroles that are in turn ligated forming tetrapyrroles and protoporphyrins after further enzymatic action (Gálová et al., 2008, Von Wettstein et al., 1995). Magnesium-chelatases catalyze the insertion of Mg^{2+} into the center of the protoporphyrins, resulting in Mg-protoporphyrins that are ultimately converted to chlorophyll a via chlorophyllide a (Beale, 1999, Von Wettstein et al., 1995). Conversion of chlorophyll a to b is then catalyzed by chlorophyllide a oxidase (CAO) (Porra et al., 1994). Aminolevulinic acid synthesis is a key step in the process and is under both direct gene expression and post-translational control (Czarnecki and Grimm, 2012).

TABLE 5.3: Chlorophyll and carotenoid biosynthesis genes identified in the full SAGE transcriptome. The UniProt identifier linked to the annotated tag is shown together with the identified gene, the resulting protein and the main location of the protein within the pathway.

UniProt	Gene	Protein	Component
Q39566	*gsa*	Glutamate-1-semialdehyde 2,1-aminomutase	ALA
Q42682	*hemB*	Delta-aminolevulinic acid dehydratase	ALA
P42451	*cobA*	Uroporphyrinogen-III C-methyltransferase	Proto IX
Q114U6	*hemE*	Uroporphyrinogen decarboxylase	Proto IX
Q9FNB0	*chlH*	Magnesium-chelatase subunit	Chlorophyll
O22437	*chlD*	Magnesium-chelatase subunit	Chlorophyll
Q94FT3	*chlI*	Magnesium-chelatase subunit	Chlorophyll
A8IKQ6	*chlI-2*	Magnesium chelatase subunit I	Chlorophyll
Q9AR22	*chlE*	Magnesium-protoporphyrin IX monomethyl cyclase 2	Chlorophyll
Q41249	*porA*	Protochlorophyllide reductase	Chlorophyll
Q9ZS34	*chlP*	Geranylgeranyl diphosphate reductase	Chlorophyll
Q52QW3	*crtiso1*	Prolycopene isomerase 1	Carotenoid
O65837	*crtl-E-1*	Lycopene epsilon cyclase	Carotenoid
P27516	*cbr*	Carotene biosynthesis-related protein	Carotenoid

Carotene and chlorophyll biosynthesis are linked through the isoprenoid pathway, which provides the building blocks for lycopene via phytoene and the action of a chain of enzymes with prolycopene isomerase performing the final step (Wise, 2006). Lycopene, which forms a branch point, can be converted to α-carotene or β-carotene through the action of different forms of lycopene cyclases. Further conversion of α-carotene yields lutein while β-carotene can be converted to the xantophyll pigments (Lu and Li, 2008).

(a)



(b)

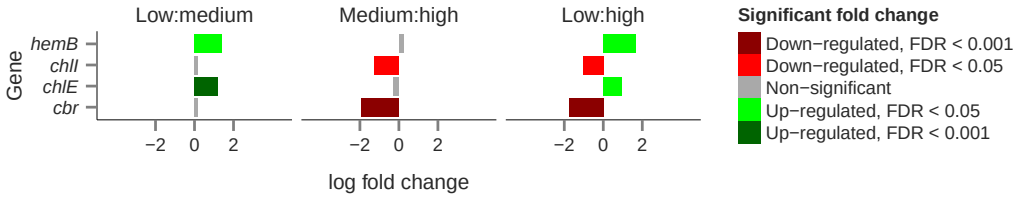


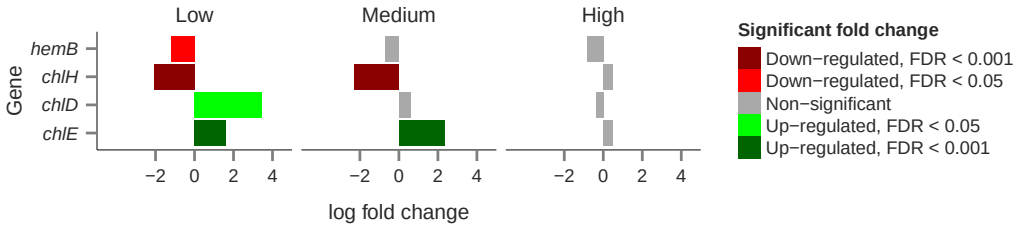
FIGURE 5.15: Significantly differentially expressed chlorophyll and carotenoid biosynthesis related genes compared under low to medium, medium to high and low to high light in a) wild-type, b) lca2. Significantly up-regulated genes are colored green while significantly down-regulated genes are colored red with a dark color signifying an FDR of <0.001 and a lighter color an FDR of <0.05. The expression is shown as the log fold change.

The light regulation in the wild-type strain involves increases in the expression of a gene involved in the synthesis of pyrrole (*hemB*), two subunits of Mg-chelatase as well as *porA* encoding a protein that catalyses the formation of chlorophyllide from protochlorophyllide and *ChlP* that provide pre-cursors for the pathway in the form of phytols. This suggests a general increase in the production of chlorophyll. The significant reduction in *chlE* and lycopene epsilon cyclase, which is involved in the final steps of caroteinoid biosynthesis, was up-regulated while *chlE* encoding a protein that performs a oxygen independent conversion of chlorophyllide to protochlorophyllide was down-regulated. The majority of these changes are driven by the change from medium to high light, shown in Figure 5.15 which mimics the general pattern observed for genes linked to photosynthesis in the wild-type (see Figure 5.13).

The lca1 strain shows no significant regulation of chlorophyll or carotene related genes under different light conditions and instead retains the expression levels observed under high light. The light response in the lca2 strain is confined to a smaller number of genes than in the wild-type and involves a down-regulation of the *cbr* gene encoding a carotene biosynthesis-related protein, this was observed in the wild-type but with a four-fold, compared to two-fold reduction in the expression of the gene. Similar to the wild-type, a reduction in one of the Mg-chelatase subunits encoding genes *chlI* was

observed under medium to high light, coupled with an increase in *hemB* under high light. The *chlE* gene encoding the oxygen-independent version of protochlorophyllide reductase was up-regulated under low to medium light in contrast to the wild-type.

(a)



(b)

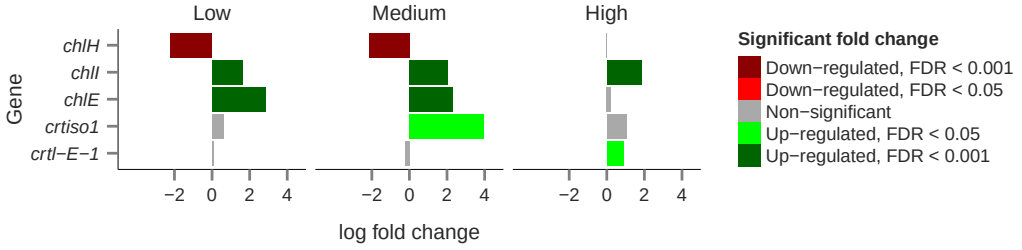


FIGURE 5.16: Significantly differentially expressed chlorophyll and carotenoid biosynthesis related genes as compared in a) lca1 vs. wild-type, b) lca2 vs. wild-type under low, medium and high light. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying an FDR of <0.001 and a lighter colour an FDR of <0.05. The expression is shown as the log fold change.

5.4 Translocation and regulation of chloroplast linked nuclear encoded genes

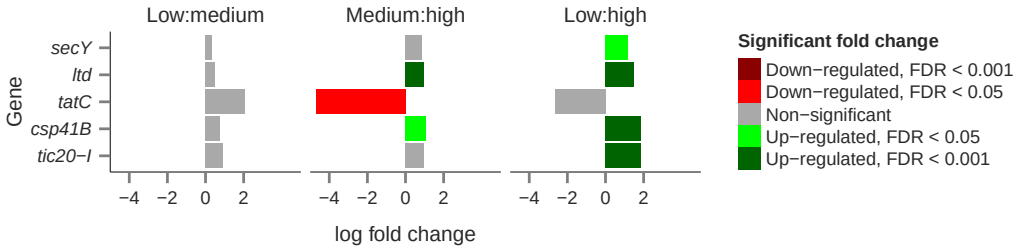
The gene products of nuclear encoded genes targeted for the chloroplast need to be transferred across the inner and outer enveloping membrane of the chloroplast before reaching the stroma. This is made possible by the use of specialized protein transfer complexes (translocons) assisting in the transfer of pre-proteins through the outer (TOC) and inner membrane (TIC) of the chloroplast (Jarvis and Robinson, 2004). Proteins bound for the thylakoid membrane or the lumen are further routed using one of four specific pathways. Proteins destined for the thylakoid are transferred by the SRP (signal recognition particle) including the LHC proteins, or by the spontaneous insertion pathway while proteins bound for the lumen are processed by either the SEC or the TAT (twin arginine translocon) pathways (Robinson et al., 2000).

TABLE 5.4: Nuclear encoded genes related to chloroplast translocation and regulation identified in the full SAGE transcriptome. The UniProt identifier linked to the annotated tag is shown together with the identified gene, the resulting protein and the main location of the protein within the pathway.

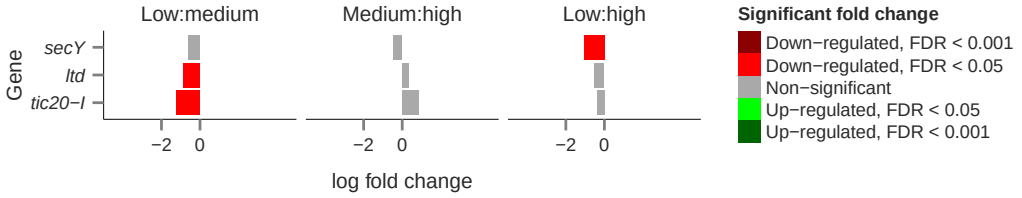
UniProt	Gene	Protein	Component
Q16D42	*secA*	Protein translocase subunit	SEC
P93690	*secY*	Preprotein translocase subunit	SEC
O80842	*cpftsy*	Fused signal recognition particle receptor	SRP
A8JIB9	*srp43*	SRP43/CAO subunit of signal recognition particle	SRP
Q8VY88	*ltd*	LHCP translocation defect	SRP
C4IZX0	*tatC*	Sec-independent protein translocase	TAT
Q944J0	*spcs1*	Signal peptidase complex subunit 1	Spase
Q9LYA9	*csp41a*	Chloroplast stem-loop binding protein of 41 kDa a	RNA bind.
Q9SA52	*csp41b*	Chloroplast stem-loop binding protein of 41 kDa b	RNA bind.
Q9AT00	*tgd3*	Trigalactosyldiacylglycerol 3	Lipid import
Q8GZ79	*tic20-I*	Protein TIC 20-I	TIC

Table 5.4 shows the genes linked to the translocation of pre-proteins into the chloroplast and related processes. Two of the main genes (*secA* and *secY*) of the Sec-dependent pathway involved in the transport of plastocyanin and *psbO* were detected in addition to the gene encoding TatC, which forms part of the TAT complex responsible for the translocation of the OEC associated proteins PsbQ and PsbR into the thylakoid lumen (Yuan and Cline, 1994). Three of the genes in the SRP pathway were also detected, including *srp43* encoding one of the two subunits of the SRP that complexes with LHC proteins entering the chloroplast. The protein encoded by *cpftsy* is involved in the recognition and targeting of this complex to the thylakoid membrane. LTD (from LHCP translocation defect) is thought to mediate the delivery of the emerging pre-protein from the TIC channel (represented by *tic20-I*) to the SRP complex (Ouyang et al., 2011). Also included are the genes encoding two chloroplast stem-loop binding proteins (CSP41a and CSP41b), both of which interact and have been shown to be involved in chloroplast rRNA metabolism and in turn mediate chloroplast transcription and translation, as well as the gene encoding the protein TGD3 involved in the translocation of lipids from the endoplasmic reticulum to the thylakoid (Beligni and Mayfield, 2008, Bollenbach et al., 2009, Lu et al., 2007). The gene *spcs1* encoding a signal peptidase protein is poorly characterized but likely to be involved in the excision of signal peptides during the translocation of proteins into the lumen based on homology.

(a)



(b)



(c)

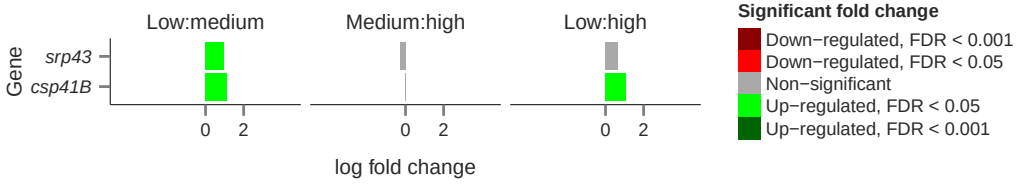
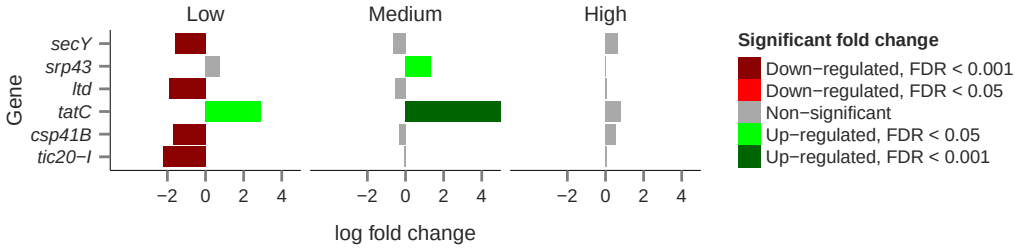


FIGURE 5.17: Significantly differentially expressed translocation and regulatory complex genes compared under low to medium, medium to high and low to high light in a) wild-type, b) lca1 and c) lca2. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying an FDR of <0.001 and a lighter color an FDR of <0.05. The expression is shown as the log fold change.

The results in Figure 5.17 show that less than half of the detected genes are significantly differentially expressed by light. The changes occurring in the wild-type are the strongest under medium to high light with an up-regulation in one of the genes (*ltd*) linked to the SRP pathway but a strong down-regulation of *tatC*. This change in expression level of the SRP linked gene is increased under low to high light, with a significant increase in *tic20-1* linked to the TIC channel and one of the two identified SEC genes. The change may enable an possible increase in the import of the SRP linked LHC proteins as well as Pc and or PsbO, which is in line with the observed results (see Figure 5.13) for the photosynthesis linked genes. However, the strong decrease in *tatC* is not reflected in the expression of the genes (*psbQ* and *psbR*) encoding the proteins translocated by the TAT pathway. In contrast, both of the lca strains show no significant regulation under medium to high light. The regulation is

instead confined to changes occurring under low to medium or high light, with down-regulation of *ltd* and *tic20-I* under low to medium light and of SECY under low to high light. The lca2 strain instead shows a 2-fold up-regulation of *srp43* and *csp41B* linked to the SRP pathway and the import of LHC proteins, compared to a 4-fold up-regulation in the wild-type.

(a)



(b)

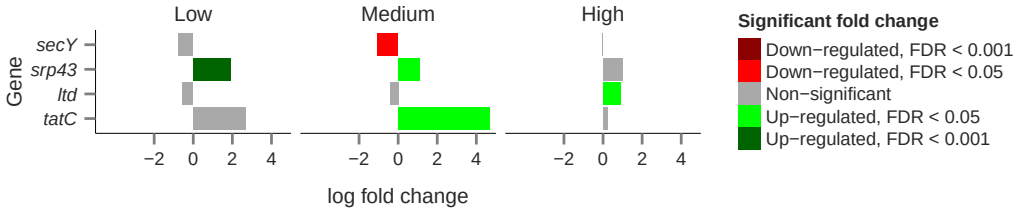


FIGURE 5.18: Significantly differentially expressed chloroplast translocation and regulatory genes compared in a) wild-type vs. lca1, b) wild-type vs. lca2 under low, medium and high light. Significantly up-regulated genes are coloured green while significantly down-regulated genes are coloured red, with a dark colour signifying an FDR of <0.001 and a lighter colour an FDR of <0.05. The expression is shown as the log fold change.

The results when compared on a strain level (as shown in Figure 5.18) further emphasise the large difference in gene regulation between the wild-type and lca1 under low light. As discussed, the wild-type up-regulates the genes encoding the proteins involved in the routing of pre-proteins throughout the chloroplast under low light. This is in contrast to lca1 which show either a down-regulation or no significant regulation of the same genes. The exception is *tatC*, but this is likely a result of the wild-type having a larger functional antenna for PSII, resulting in an increased need for optimization of electron flow through the OEC compared to the lca strains. The increase in *srp43* under low and medium light is likely a result of the delayed light response in lca2 in which the majority of the significant regulation of the LHC and core photosynthesis genes occurs under low light (see Figure 5.13).

5.5 Discussion

The results from the gene expression analysis clearly shows that the photoacclimation response on a genetic level in the lca strains is strongly reduced compared to the pattern seen in the wild-type (see Figure 5.1) as evident by a 70-75% reduction in significantly regulated tags in cells grown under low compared to high light (20 to 1200 μmol photons m^{-2} s^{-1}). The observed differences occur in the response under low and medium light while under high light the response in both lca strains is largely comparable to the wild-type (Figure 5.2). Further analysis also shows that a substantial proportion of the regulated tags are shared between the lca strains, and with a larger proportion being shared under medium light (57%) compared to low light (35%) (Figure 5.4).

This shared regulation of significantly expressed tags is strongly correlated (>0.9) but perhaps more striking is the increase in correlation under high light. Under high light the differences between the strains largely disappear and only a handful of genes show significant regulation (Figure 5.4 c). This pattern suggests that under increasing light, light-induced gene regulation is increasingly similar until a high light genotype largely identical to the wild-type is triggered at an irradiance between medium and high light. Again this agrees with the previously reported observation that *D. tertiolecta* is adapted for high light primarily, similar to terrestrial plants, due to the genus preference for shallow water and resulting in high light conditions, and that its main photoacclimation response occurs in the switch from high light to a lower light condition (low light and darkness largely induce the same initial response) (Falkowski and Owens, 1980b, Laroche et al., 1991). However, more importantly this mechanism still seems to be active in and confining and shaping the light response in the lca strains.

The high level analysis of the regulation of cellular components clearly indicate that the response in the wild-type is centred around the plastid and the thylakoid membrane, this response is much reduced in lca2 and missing from lca1 (Figure 5.5). The significant enrichment in tags linked to photosystem and ribosomal tags when comparing lca1 and wild-type under low light is a direct result of the missing light response in lca1 (Figure 5.7). The changes in lca2 compared to the wild-type are more modest and not strong enough to cause a significant enrichment between the strains, while both the wild-type and the lca2 showed significant light-driven enrichment, again with more modest but similar changes in lca2. No light-driven significant enrichment was observed in lca1.

5.5.1 Photoacclimation response in lca1 and lca2

Based on the the results from the enrichment analysis it is clear that the wild-type and lca2 photoacclimate in response to changing light and these changes are closely linked to the genes involved in photosystems and the thylakoid membrane in particular. Analysis of gene expression of photosynthesis linked genes showed that a switch from high to medium light results in up-regulation of PSII LHC genes followed by a more modest up-regulation of PSI LHC genes under low light. The response in lca2 is delayed and mainly occurs under the switch from medium to low light and is also confined to changes in PSII LHC. The later low light response in PSI observed in the wild-type is missing from lca2 (see Figure 5.13). If this delayed response in the up-regulation of PSII LHC genes and the missing up-regulation of PSI LHC genes is mirrored in the synthesis of LHC apoproteins, the smaller resulting PSII LHC, and the resulting reduction in binding capacity for chlorophyll a and b could explain the difference in σPSII and the lowered total chlorophyll content in the lca2 under low and medium light (see Chapter 3, Figure 3.15). In addition, if lca2 retains PSII LHC components (apoproteins) that are tuned for high light, under medium light, the possibly lowered binding capacity for Chl b in these apoproteins may help explain the change in Chl a:b ratio observed in this strain (Figure 3.15). Using the same assumptions, the lca1 retain the same apoproteins regardless of the light level resulting in a near static σPSII and chlorophyll content per cell.

In contrast, lca1 only show signs of modest photoacclimation that is almost fully confined to changes in the LHC, two genes (*lhcII-2.1* and *lhcb4*) are up-regulated under decreasing light. LhcII-2.1 is only found in small numbers and is likely responsible for the slight increase in σPSII seen under decreasing light (Figure 3.15). The *lhcb4* gene is necessary for state-transitions in *C. reinhardtii* (Kargul et al., 2005, Tokutsu et al., 2009) and it is up-regulated in all three strains under decreasing light, however, in lca1 the strongest up-regulation is seen during the switch from high to medium light while in the wild-type and lca2 this gene is up-regulated under lower light (Figure 5.13). Possibly enabling a transfer of any remaining LHC proteins to PSI in an effort to re-balance the electron flow from PSII to PSI. The high light adapted lca1 with a high PSII to PSI ratio, that based on these results, remain under low and medium light, is likely facing considerable difficulties in balancing the electron flow originating from each pool of photosystems.

While the number of annotated genes in PSI is low, the strong up-regulation of the two genes encoding PSI core units (*psaF* and *psaG*) in the wild-type during the switch from high to low light (Figure 5.13) and complemented by up-regulation of PSII OEC genes under even lower light agrees with a re-balancing of the PSII:PSI ratio which

under high light is skewed towards PSII to absorb the presence of large amounts of deactivated reaction centres from photoinduced damage ((Kim et al., 1993b)) and coupled with a general increase in photosystems. The same response is present in lca2 but again limited to the changes seen in the wild-type under high to medium light, but lacking an up-regulation of OEC genes, which suggests no further increase in photosystems under low light. The only visible response in the lca1 is an up-regulation of the two PSI core genes (*psaF* and *psaG*) which again could either point to a limited re-balancing of the photosystem stoichiometry or a re-balance of the flow of electrons into PSI.

Similarly, the biosynthesis of pigments including chlorophyll and carotenoids largely mimicked the changes by light observed in the three strains. The wild-type gene regulation suggests a strong increase in the biosynthesis of chlorophyll under decreasing light, together with a smaller decrease in the synthesis of carotenoid precursors which again is triggered in the switch from high to medium light (see Figure 5.15). The change in lca2 is shifted towards lower light conditions, with the significant up-regulation of chlorophyll biosynthesis related genes occurring under lower light conditions than in the wild-type (Figure 5.15). Once again, any form of light-driven response is missing and the strain level of gene expression remains stable across light conditions. To summarize, the photoacclimation response in gene regulation generally seems to agree well with the observed phenotype in lca1 and lca2 (Chapter 3).

5.5.2 Regulatory mechanisms of photoacclimation

The changes in the expression of genes involved in the translocation and regulation of chloroplast protein and RNA trafficking may help explain the different pattern of photoacclimation. Both lca strains regulate genes under low and medium light in contrast to the wild-type that shows strong regulation in the switch from high to medium light. The up-regulation in lca2 is confined to genes linked to the SRP responsible for the translocation of pre-LHC proteins. The delayed up-regulation that in turn coincides with the delayed up-regulation of LHC genes may help explain the differences between wild-type and lca2. Two genes previously linked to a phenotype similar to lca1 are significantly down-regulated under low light, including a gene encoding a protein interacting with the SRP. This protein, known as LTD for LHCP translocation defect, does in its mutated form result in a phenotype in *A. thaliana* that shares traits with lca1, this includes a yellow colour from a very low chlorophyll content (5% of WT) and a high Chl a:b ratio (Ouyang et al., 2011). In addition, the gene encoding CSP41B, a chloroplast stem-loop binding protein that associates with CSP41A in stoichiometric proportions and acts as a endonuclease, is strongly down-regulated. Mutants in

A. thaliana deplete in CSP41B display a pale green phenotype from reduced Chl content. Together, these two enzymes (CSP41A and CSP41B) likely interact with and stimulate chloroplast RNA transcription and translation (Beligni and Mayfield, 2008, Bollenbach et al., 2009).

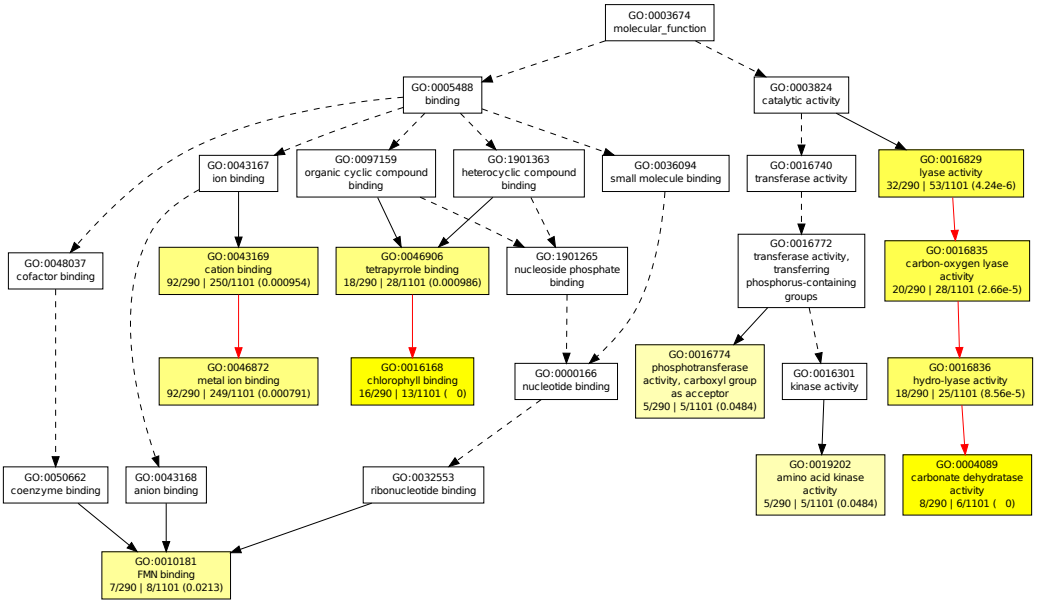


FIGURE 5.19: Enrichment analysis of GO identifiers for molecular functions from tags significantly differentially expressed in wild-type under low to medium light. GO terms identified as significantly enriched in GO identifiers linked to these tags are shown in yellow, while non enriched terms are shown in white. The GO terms are hierarchically linked from parent to child term, non significant terms are linked by dashed lines, non-significant to significant terms are linked by solid lines and significant terms are linked by solid red lines. Below enriched terms, the number of mapped GO identifiers out of the total within the dataset are shown, followed by the number of associated GO identifiers out of the total number of identifiers within the full GO database and the calculated p-value for the significance of the enrichment.

Certain categories enriched in the wild-type but not lca1 or lca2 are interesting from the perspective of regulation, these include enrichment in FMN-binding molecular functions in wild-type under low to medium light (Figure 5.19). Flavin mononucleotide (FMN) is linked to blue-light sensing in plants where it is associated with phototropism and is similar to phototaxis in green algae (Kasahara et al., 2002). Intriguingly it has also been suggested that the rapid induction of chlorophyll-a oxygenase (CAO) and LHCII linked gene transcripts in the shift from high to low light may be controlled by a signaling pathway using a hypohetical blue-light sensor (Masuda et al., 2003). This suggest that the delayed and reduced gene expression of LHCII linked genes in lca2 could possibly be explained by a reduced blue light-sensing capacity. Similarly, the

lack of photoacclimation in lca1 could be explained by a strongly reduced ability to sense light.

While the results are not fully conclusive, there is no direct evidence of a gene knock-out, resulting in the absence or near absence of a gene product. Rather the changes result from differences in gene regulation. The result do not identify any specific mechanisms but point to changes in the lack of regulation of genes involved in the SRP pathway in lca1.

5.5.3 Energy dissipation under high light

Connectivity between PSII centres plays an important role in energy distribution and temporary dissipation under non-saturating light (Ihnken et al., 2011). This connectivity naturally plays a less significant role under high light when most PSII will be working at or near full capacity. This connectivity between PSII units that allows for the transfer of charge from a closed to an open reaction centre may have been affected due to changes in the size and or composition of the peripheral antenna in the lca strains, which would result in a reduced capability to photochemically quench temporary increases in light. The fast component of NPQ, in the form of qE, could alternatively, or in addition, have been affected in lca1. The detected expression levels of the *psbS* gene that have been shown to play a key role in the induction of qE in *A. thaliana* was either equal to or increased in both lca mutants compared to the wild-type under medium light (see Figure 5.14). PsbS mutants typically exhibit defects in the ETR, in contrast to the observations in lca1, suggesting that the decrease in NPQ is unrelated to this gene (Peterson, 2005). The NPQ deficiency in the *A. thaliana* mutant *npq4* has been shown to result from deletion of the *psbS* gene, this mutant displays a normal xanthophyll profile (Li et al., 2000) while *npq1*, another NPQ-deficient mutant, has lost the ability to convert violaxanthin to antheraxanthin and zeaxanthin. This mutant retain 80% of the NPQ and is affected not in qE but in the slow component of NPQ (Niyogi et al., 1997). This suggests that the apparent lack of NPQ in lca1 is likely unrelated to the *psbS* gene. The observed reduction could be related to changes in the xanthophyll cycle as suggested by the irregular pigment profile in lca1 under low and medium light. However, when compared to the 20% decrease in NPQ in *npq1* it seems unlikely that these changes would fully explain the substantial loss of NPQ in lca1. Ihnken et al. (2011) found evidence of a possibly unique and rapid NPQ response in *D. tertiolecta*. The observed changes in lca1 means that the strains could possibly play an important role in the understanding of the high light response in *Dunaliella*. Regardless, this loss of NPQ is likely to have substantial effects on the the general growth of lca1. Reductions in NPQ will increase the availability of energy

(through decreases in heat loss) within the system and the individual cells. This could potentially allow for higher total cell densities as long as the net effects to individual cells, from reduced energy loss from NPQ, is not in turn overshadowed by the resulting increase in photoinduced damage. However, the availability for truncated and low chlorophyll antenna strains to survive and thrive under high light is essential for an increased biomass yield as suggested by de Mooij et al. (2015) and effective NPQ mechanisms may be essential to reduce the amount of photoinduced damage under these conditions.

5.5.4 Comparison against previously described truncated antenna strains

The first of the *tla* (truncated light-harvesting chlorophyll antenna) strains, *tla1*, generated in *C. reinhardtii* using random DNA insertion followed by screening for low chlorophyll fluorescence and or a high Chl a:b ratio was first reported in 2003 (Polle et al., 2003). This was followed by *tla2* in 2011 and *tla3* in (Kirst et al., 2012, 2011). Previously, Amin et al. in 1999, had reported that mutants lacking the SRP (see Chapter 5, section 5.4 for a description of the pathway) SRP43 and SRP45 subunits had reduced levels of several LHCs and reaction centre proteins (Amin et al., 1999). At the same time, and also working with *A. thaliana*, Klimyuk et al. reported that mutants deficient in the SRP43 subunit were chlorotic with reduced levels of chlorophyll and an increased Chl a:b ratio (Klimyuk et al., 1999). In 2002 it was shown that a *C. reinhardtii* mutant deficient in the ALB3 (*alb3*) protein had strongly reduced (>10-fold) LHCs and reduced amounts of PSII (Bellafiore et al., 2002). In addition it was shown that reductions in the size of the LHC resulting from the mutations translated into measurable improvements in the calculated photosynthetic conversion efficiency in some of these mutants, for example, the rate of oxygen evolution in the *tla1* and *tla2* strains normalized on a per chlorophyll basis were 182 and 197% that of the wild-type.

As evident in Table 5.5, many of the previously published mutants target either components of the SRP, which is responsible for the integration of LHC components into the thylakoid membrane after translocation into the chloroplast through the TIC/-TOC complex (Gutensohn et al., 2006) or the LHCs directly via RNA interference. The four (CpSRP43, CpSRP45, CpFTSY and ALB3) proteins currently recognized as part of the chloroplast SRP have previously been targeted by mutations, resulting in various levels of reductions in LHCs and photosystems. Based on the reported viability of the mutants and the observed changes, it is evident that mutations affecting either CpSRP43 and CpSRP45 that complex with LHC preproteins result in more moderate changes to the phenotype of the mutants. The reported CpSRP43 and CpSRP45

TABLE 5.5: Previously published eukaryote mutants with reduced LHCs compared to lca1 and lca2. Based on Kirst and Melis (2014)

Mutant synonym	Mutated protein	Pathway	Species	Genetic modification	Viability	Reference
tla1	tla1	Unknown	*C. reinhardtii*	Random DNA insertion	Viable	(Polle et al., 2003)
ri6	tla1	Unknown	*C. reinhardtii*	RNAi-mediated knock-down	Viable	(Mitra et al., 2012)
stm3LR3	LHCBM (1)	LHC	*C. reinhardtii*	RNAi-mediated knock-down	Viable	(Mussgnug et al., 2007b)
stm6Glc4L01	LHCBM (1-3)	LHC	*C. reinhardtii*	RNAi-mediated knock-down	Viable	(Oey et al., 2013)
stm6Glc4T7	NAB1	LHC	*C. reinhardtii*	Site-directed mutagenesis		(Beckmann et al., 2009)
tla3-cpsrp43	TLA3-CpSRP43	SRP	*C. reinhardtii*	Random DNA insertion	Viable	(Kirst et al., 2012)
cao, *chaos*	CpSRP43	SRP	*A. thaliana*	knockout	Viable	(Amin et al., 1999), (Klimyuk et al., 1999)
ffc/cao	CpSRP54/CpSRP43	SRP	*A. thaliana*	knockout	Viable	(Amin et al., 1999)
ffc	CpSRP54	SRP	*A. thaliana*	knockout	Viable	(Amin et al., 1999), (Kugelmann et al., 2013), (Pilgrim et al., 1998)
tla2-cpftsy	TLA2-CpFTSY	SRP	*C. reinhardtii*	deletion	Viable	(Kirst et al., 2011)
cpftsy	CpFTSY	SRP	*A. thaliana*	knockout	Lethal/ retarded growth	(Asakura et al., 2008), (Kugelmann et al., 2013)
csr1	CpFTSY	SRP	*Zea mays*	knockout	Seedling lethal	(Asakura et al., 2003)
alb3	ALB3	SRP	*A. thaliana*	knockout	Seedling lethal	(Asakura et al., 2008), (Kugelmann et al., 2013), (Sundberg et al., 1997)
ac29	Alb3.1	SRP	*C. reinhardtii*	knockout	Retarded growth	(Bellafiore et al., 2002), (Ossenbühl et al., 2004)
alb3.2	Alb3.2	SRP	*C. reinhardtii*	RNAi	Lethal	Gohre2006
lca1	Unknown	Unknown	*D. tertiolecta*	Chemical mutagenesis	Viable	This study
lca2	Unknown	Unknown	*D. tertiolecta*	Chemical Mutagenesis	Viable	This study

mutants are all viable and the majority of the changes are confined to the LHCs and do not affect the amount of photosystems, the only exception is *ffc*/*cao* in which only *ffc* encoding CpSRP45 in *A. thaliana* results in reduction of photosystems in young plants. This led to the conclusion that cpSRP45 but not cpSRP43 is involved in the biogenesis of chloroplast-encoded proteins (Amin et al., 1999). The CpFTSY subunit is the receptor that recognises the LHCP/SRP complex and, following hydrolysis through the action of GTPase, the LHCP, via mediation of ALB3, is integrated into the thylakoid membrane (Falk et al., 2010, Tzvetkova-Chevolleau et al., 2007). Mutations affecting either CpFTSY or ALB3 are clearly deleterious to a larger extent with lethal or retarded growth in a number of mutants. Also apparent is that these mutants tend to be lethal to a higher degree in higher plants compared to the effects in *C. reinhardtii*, for example the *tla2-cpFTSY* mutant is viable in *C. reinhardtii*, while the homologues, in *A. thaliana* and *Zea mays* are lethal. Similarly mutations to alb3 are lethal in *A. thaliana*, but viable in one out of two variants in *C. reinhardtii*.

Three CpSRP linked genes were detected in the SAGE transcriptome, although none of these genes showed signs of a knock-out they were affected to some extent in the lca mutants. As discussed earlier, in lca1 the regulation is arrested in a wild-type equivalent high light position, while lca2 *cpSRP43* is up-regulated under low light. In addition, both *tla1* and *nab1* were detected but in low enough counts to be removed from filtering from the final results, neither showed any sign of a knock-out in either lca mutant. Down-regulation of the nuclear encoded *tla1* gene results in decreased levels of chlorophyll per cell, smaller LHCs and less chlorophyll b (Polle et al., 2003). The decrease in chlorophyll is disproportionally larger than the decrease in chlorophyll binding capacity of the LHCs, suggesting a possible role in chlorophyll biosynthesis, although the exact mechanism of action remains unknown (Mitra et al., 2012). The RNA binding NAB1 protein suppresses the translation of LHCB gene transcripts, resulting in down-regulation of the PSII LHC (Mussgnug et al., 2005). Both genes were detected in the transcriptome dataset (using a direct tblastn search against matched ESTs) but in low counts with no apparent difference between lca and wild-type gene expression. In summary, none of the previously known mutations affecting the size of LHC's and listed in Table 5.5 showed signs of a knock-down, the SRP pathway is affected to some extent in the lca mutants but the changes are modest and seem to be related to changes in the degree of regulation.

Similarly, comparing the photophysiology of the lca strains with the previously known mutants (Table 5.6), it is apparent that the chlorophyll content per cell in lca1 (ranging from 18 to 70% of wild-type) approaches the lowest measured in other mutants. The lowest content (18%) is measured under low light when growth in lca1 is likely restricted due to energy constraints from the constitutively smaller PSII antenna. The

TABLE 5.6: Photophysiology of previously published eukaryote mutants compared to lca1 and lca2, the Chl total, Chl a:b ratio, growth rate, PSII/PSI antenna size and PSII/PSI is given relative to wild-type. Based on (Kirst and Melis, 2014).

Mutant synonym	Chl total (%)	Chl a:b ratio (%)	Growth rate (%)	PSII/PSI antenna size (%)	PSII/PSI (%)	Reference
tla1	37.5	270		65/50		(Polle et al., 2003)
ri6	30	231		74/50		(Mitra et al., 2012)
stm3LR3	32	208	185			(Mussgnug et al., 2007b)
stm6Glc4L01	50	114	165			(Oey et al., 2013)
stm6Glc4T7	80	106	153	8390/unknown		(Beckmann et al., 2009)
tla3-cpsrp43	16	444		38/52	(Slight reduction)	(Kirst et al., 2012)
cao, *chaos*	57	125		(Reduction of most LHC)	(Normal)	(Amin et al., 1999), (Klimyuk et al., 1999)
ffc/cao	15			(Strong reduction)	(Strong reduction)	(Amin et al., 1999)
ffc	25*			(Reduction of most LHC)	(Reduction)	(Amin et al., 1999), (Kugelmann et al., 2013), (Pilgrim et al., 1998)
tla2-cpftsy	20			63/65	(Slight reduction)	(Kirst et al., 2011)
cpftsy	67			(Strong reduction	(Strong reduction	(Asakura et al., 2008), (Kugelmann et al., 2013)
csr1	12			(Strong reduction	(Strong reduction	(Asakura et al., 2003)
alb3	5			(Strong reduction)	(Strong reduction)	(Asakura et al., 2008), (Kugelmann et al., 2013), (Sundberg et al., 1997)
ac29	30			(Strong reduction)	(Affected to lesser extent)	(Bellafiore et al., 2002), (Ossenbühl et al., 2004)
alb3.2	25-50			(Affected to lesser extent)	(Strong reduction)	(Göhre et al., 2006)
lca1	18-70	194-201	94-98	70-80/100	99/23	This study
lca2	57-94	118-284	94-113	71-99/100	72/62	This study

Chl a:b ratio is comparable to that seen in *tla1/ri6* and the *stm3LR3* mutants with a mutated *tla1* gene or a highly reduced LHCB from RNA interference, suggesting the presence of a small LHC which correlates well with the results from the lca1 LHC gene expression. The measured growth rate of lca1 is similar to wild-type and substantially lower than the increase in growth rate measured in other mutant strains. The measured reduction in the size of PSII antenna is moderate with no observed effect on PSI, while the highly reduced number of PSI compared to PSII is notable none of the previously published mutants show a similar ratio, suggesting that lca1 may result from a mutation that has not been characterized before in relation to truncated antenna mutants.

The changes observed in lca2 are moderate with a small increase in the measured growth rate compared to the wild-type. The most noticeable parameter is the range of the Chl a:b ratio that encompasses nearly the full spectrum reported for previously published mutants, again confirming that lca2 has retained a strong photoacclimation response. Both lca1 and lca2 are clearly comparable to previous published mutants with truncated LHCs in terms of the range of the changes to the LHC. However, lca1 with its lack of photoacclimation response, lack of NPQ and constitutively small LHC shows a combination of changes that together identifies and can be described as a high light locked in phenotype. Similar mutants have recently been described in a patent application (Bailey et al., 2014) that refers to a novel isolated mutant with a locked in high light acclimated (LIHLA) phenotype, with described characteristics that match those of lca1, including a low chlorophyll antenna, a high Pmax and Ek (referred to as Ik in this thesis), reduced NPQ above 500 µmol photons m^{-2} s^{-1} and an equal or 1.5 times higher PSII ETR. The only measured parameter that differs is the size of PSII that can be increased in *LIHLA*, this clearly differs from lca1 but otherwise it seems plausible that these mutants may result from shared or partially overlapping underlying changes in one or more genes.

5.5.5 Conclusion

The observed changes in phenotype, with a reduced chlorophyll content and a smaller light harvesting antenna in both lca strains, are mirrored by changes in the expression of photosynthesis linked genes and associated pigment biosynthesis genes, in particular in LHC genes suggesting the mutations have resulted in direct changes in the regulation of core photosynthesis pathways. The results also show that changes in regulatory genes could help explain aspects of the observed changes in the regulation of photosynthesis. The data are not conclusive enough to draw any clear conclusions about whether the mutations have affected regulatory genes directly or indirectly.

The pattern of gene regulation in lca2 suggest a delayed light response, which in turn results in a delayed increase in PSII LHC under decreasing light. This delayed response also results in the lack of an up-regulation of PSI LHC under low light that is present in the wild-type. The photoacclimation response in lca1 is limited, with minimal increases in PSII LHC and no increase in PSI LHC. This gene regulation matches the observed phenotype closely. Two genes *ltd* and *csp41B* for which mutations have been shown to result in phenotypes with low chlorophyll content (Beligni and Mayfield, 2008, Bollenbach et al., 2009, Ouyang et al., 2011) are strongly down-regulated under low light conditions in lca1. While direct changes in regulatory genes may be involved, another possibility is that the strains have a degraded or missing ability to sense changes in light conditions. In *D. salina* it has been proposed that photoacclimation due to changes in irradiance results from two different signal transduction pathways that regulate the expression of *cao* and *lhcb* (Masuda et al., 2003). Based on the result a similar phenotype may result from the changes to one (lca2) or both pathways (lca1), thereby reducing the strains ability to sense changes in light intensity. Laroche et al. that proposed that "Our results reveal that cab mRNA shows the same initial increase upon removal of high light growth conditions, regardless of whether the cells are shifted to darkness or to LL. These observations suggest that cab genes are suppressed in HL rather than activated by LL." (with LL referring to low light) in experiments using *D. tertiolecta* (Laroche et al., 1991). Based on the lack of regulation in lca1, which resulted in a almost fully locked in high light phenotype and near-static CAB mRNA levels it could be argued that the opposite is true. If correct, the typical (or native baseline) *D. tertiolecta* gene expression results in a high light phenotype, and will remain in this position unless sensing mechanisms confirm a decrease in light. From the perspective of surviving in a high light environment this would be a sensible precaution, since an erroneous increase in the light harvesting capacity might prove fatal from photoinduced damage.

Chapter 6

Synthesis

This thesis describes the characterization of two lca (low chlorophyll antenna) mutants selected using a high-throughput pipeline (described in detail in Chapter 2, section 2.10) that utilizes a combination of mutagenesis and flow cytometry to isolate single mutants with a reduced chlorophyll content (see Figure 6.1 below).

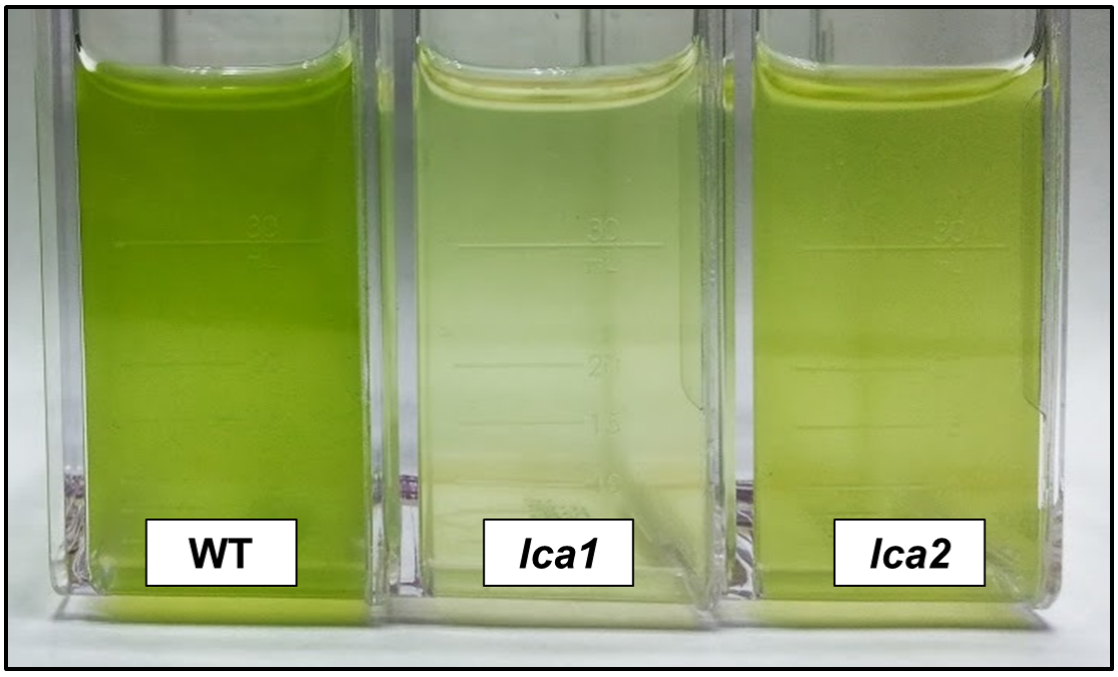


FIGURE 6.1: Cell cultures of *D. tertiolecta*, wild-type, lca1 and lca2, each culture contains similar numbers of cells, emphasising the difference in total chlorophyll per cell.

One of the main questions that needed to be answered in relation to the selection process was whether the pipeline successfully isolated mutants with a stable low chlorophyll content with largely normal to wild-type properties. This relates back to the effects of the mutations. Ideally, the induced changes in phenotype should result from well-defined mutations. This could include either the genes directly responsible for the phenotype or regulatory genes affecting genes of interest. One of the primary concerns was that the observed phenotype may be related to regulatory genes affecting processes far up-stream of the target genes and in doing so may down-regulate not only the pathways of interest but also cell wide metabolic processes. Another concern was that the mutations could have affected separate pathways, some of which could have deleterious effects on the overall conversion efficiency of the cells. As shown by the pathway analysis (Chapter 5, section 5.1.3), the changes were clearly focused on the chloroplasts and the thylakoid membrane in particular. Cell-wide changes, such as those affecting ribosomal activity were confined to lca1 under low light, this observed down-regulation is likely a result of the locked in high-light adjusted phenotype, which will suffer under low light due to energy constrains resulting from a constitutively smaller photosystem antenna that struggles to provide enough energy to the cell. Based on the results presented in this thesis it is clear that the mutations did result in direct changes to the regulation of LHCs and the main regulatory pathway (cpSRP) responsible for the translocation and insertion of LHCP into the thylakoid membrane. The effects on the pathways responsible for the biosynthesis of chlorophyll and accessory pigments mirrored these changes without any major discrepancies.

To summarize the results, the lca1 strain is characterised by low amounts of chlorophyll per cell, a strongly reduced PSII LHC, an apparent lack of NPQ and a minimal photoacclimation response. The lca1 PSII LHC showed a minimal increase in size under decreasing light, while the lca1 PSI LHC showed no corresponding change in size compared to the wild-type, when measured under medium light. This regulation is supported by the gene regulation profile that shows only minimal changes in PSII associated genes and no change in PSI related genes under changing light. In addition, the Chl b content is low or undetectable, suggesting that the LHC is either low or or composed of LHC component with a reduced binding capacity for Chl b. The lca2 strain is also characterised by a low amount of chlorophyll per cell and a reduced PSII LHC. However, the lca2 strain has retained a photoacclimation response that is missing from lca1, and which compared to the wild-type is limited in range. This result is again well supported by the gene regulation profile in the strain, which suggest that the response to decreasing light is delayed in lca2 compared to wild-type, resulting in a gene expression profile in which the majority of the regulation occurs during the

switch from medium to low light rather than high to medium light as seen in the wild-type.

The observed increases in PSII electron transport rate and chlorophyll normalized oxygen evolution in both lca strains did not translate into actual increases in biomass based on the initial results from the growth trials in photobioreactors. These results mirror the recent results of de Mooij et al. (2015) in which four mutant strains with reduced antennas, including *Stm3LR3* and *tla2* did not show any increase in biomass during growth trials. Based on the result, the authors recommended an increased selection for high light resistant cells and collaboration between molecular biologists and engineers, to better match the optimal growth conditions in a photobioreactor with the biological changes occurring in microalgal cells during growth under high light and high cell densities. These findings also highlight the need for further studies using systems biology approaches that trace the partitioning of energy and metabolic carbon compounds from the source to sinks, such as starch metabolism, fatty acid synthesis, or photoinduced damage to give a few examples, throughout the organism. Studies in this direction have already yielded important results (Peers, 2014) and combining these findings with knowledge gained from projects like this should help to further focus the efforts.

Another question that needs to be answered is whether this form of pipeline is the most suitable for the generation of new mutant strains with specific properties. Based on the findings in this project the pipeline successfully generated and isolated strains with the specific properties selected for. The two lca strains characterized in detail in this thesis both resulted in measurable gains (PSII ETR, oxygen evolution rate) from the reduction in the chlorophyll antenna, although it could not be shown that these increases translated into the theorized improvements in biomass. Further refinements of the selection process that directly links these basic selection parameters to the end-goal may be needed. This could include selection for increased fatty acid or starch content in addition to high light resistance as suggested by de Mooij et al. (2015), which could be included as part of the selection process with minimal changes to the current protocol. One important observation in relation to this hypothetical future development is the speed at which the process identified the constraints in gene expression by light in *D. tertiolecta* (see Chapter 5, section 5.3). One of the main justifications behind this high-throughput pipeline was the identified need to expand the number of marine microalgal species available for biotechnological purposes. This high-throughput pipeline could possibly be used as a means of identifying the inherent constraints and the level of plasticity for parameters of interest in novel microalgal species. By generating shallow and multiplexed transcriptomes from mutants selected

under different conditions, it may be possible to outline the constraints of each parameter for a comparatively low cost.

The analysis did not identify the underlying mutations that drove the changes in the phenotypes, as expected due to the lack of a published genome and the technique (SuperSAGE) used for the sequencing of the transcriptome. The use of SuperSAGE transcriptomes are favourable in terms of cost, allowing for replication of libraries (Matsumura et al., 2010b), but the short reads (26 bp), means that the chances of successfully discovering mutations are reduced. While unlikely due to the nuclear encoding of LHC proteins and the genes involved in the regulatory pathways, mutations may still have affected the small number of organelle encoded genes. The RNA library preparations used, selected for mRNA with attached poly-(A) tails and will preferentially select for nuclear encoded genes. The organelle encoded mRNA are still selected for to a certain extent but results do not represent a true organelle gene expression (see Chapter 4, section 4.4 for a discussion). Complementary sequencing of organelle transcriptomes may be necessary to fully elucidate the effects of the mutations. Regardless of these possible short-comings, the results from the analysis of the transcriptome data showed a good degree of correlation with the results from the characterisation of the photophysiology.

6.1 Conclusion

The mutagenesis and selection process used to generate the lca mutants successfully selected mutants with the targeted properties, indicating that the method could have a range of future applications for the generation of mutants with specific and well defined properties. However, the results also shows that although all selection criteria were met, the lca mutants did not result in increases in biomass. This suggests that the fairly stringent selection criteria used (low chlorophyll content, retained cell size and viability) did result in the local hypothesized improvements (increased PSII electron transport and oxygen evolution on a chlorophyll basis), but not the overall cell-wide improvements in biomass yield. This necessitates the need for an improved understanding of the downstream energy and fixed carbon partitioning throughout the cell. It also shows the need for improved selection for the end goal, which could include improved fatty acid or biomass yield. Regardless, the selection process successfully and efficiently selected novel and phenotypically stable mutants. The characterization of these mutants showed that one of the mutants, lca1 has a locked in high light phenotype with a constitutively smaller antenna, with a minimal photoacclimation response and a reduced NPQ response. In contrast in lca2, the antenna is

reduced in comparison to wild-type but with a retained photoacclimation response and an equal or better growth rate.

Appendix A

Photophysiological characterization

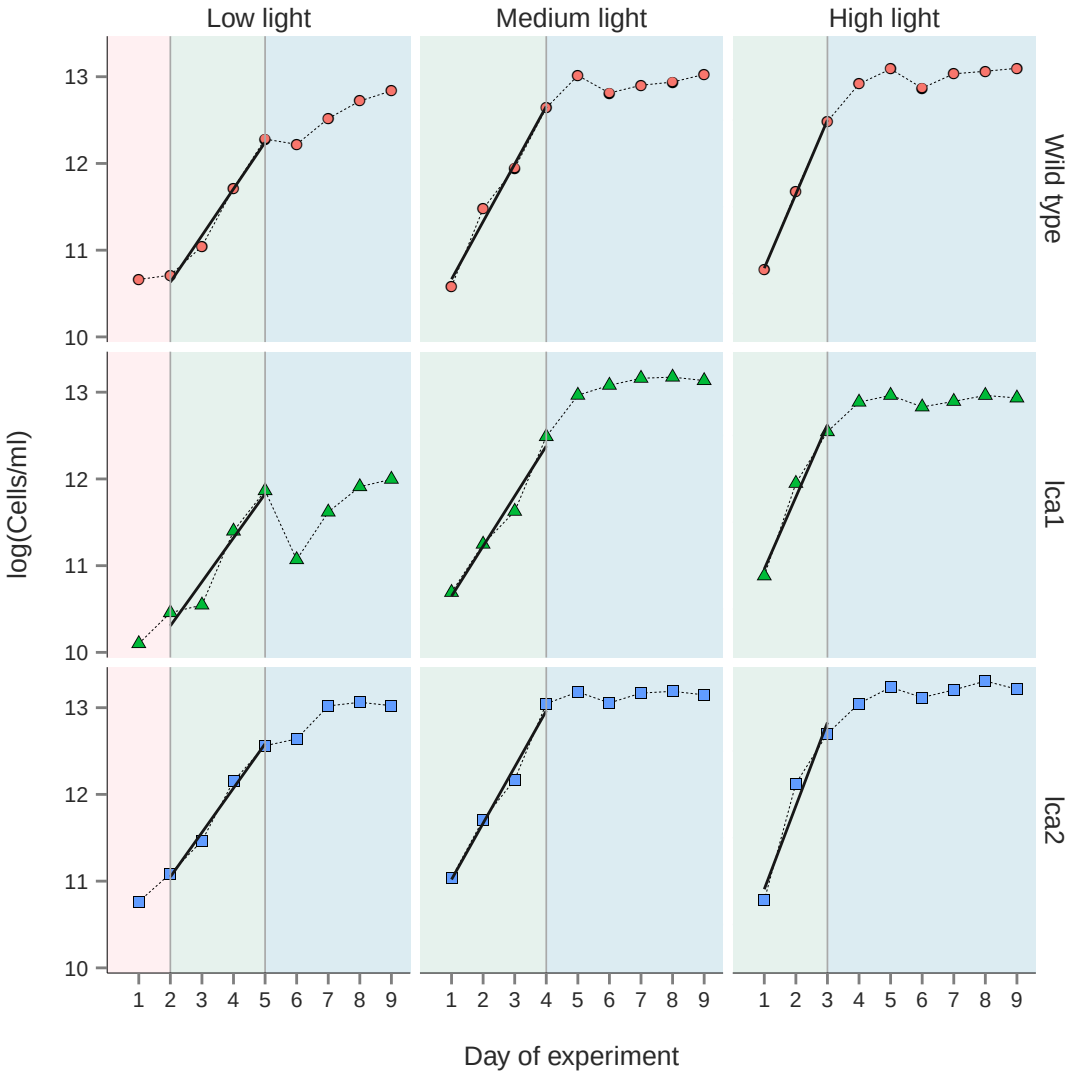


FIGURE A.1: The growth rate of the lca1, lca2 and wild type strains under low, medium and high light. The cell counts have been log transformed to better visualize the lag phase (light pink), exponential phase (light green) and stationary phase (light blue) of the cells. The fit of a linear regression of the counts in the exponential phase which gives us the growth rate is also shown using a solid line.

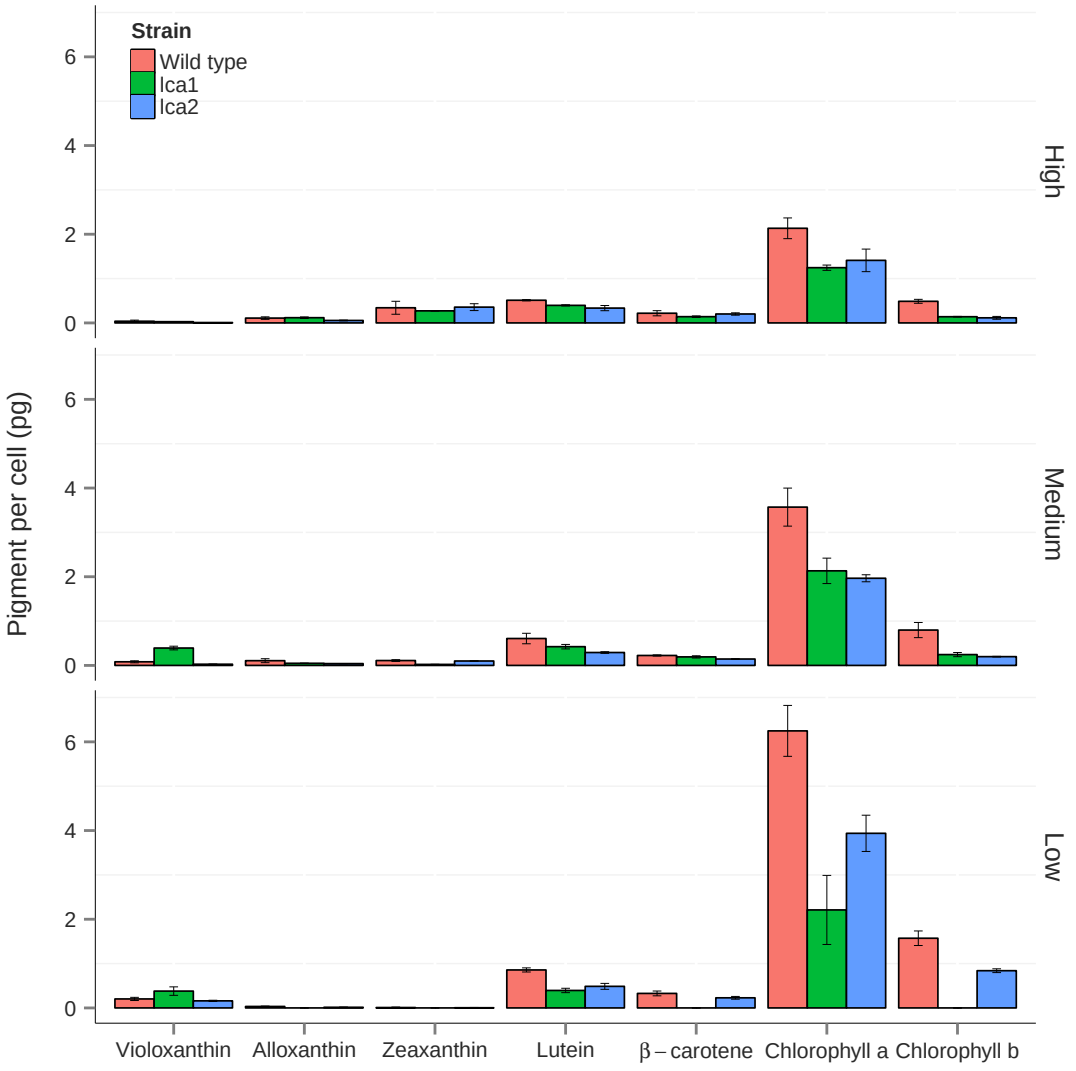


FIGURE A.2: The pigment composition in the wild type, lca1 and lca2 strains grown under low, medium and high light (20, 100, 1200 μmol photons m^{-2} s^{-1}). Error bars are one standard deviation from biological triplicates. Based on data from P. Stephenson.

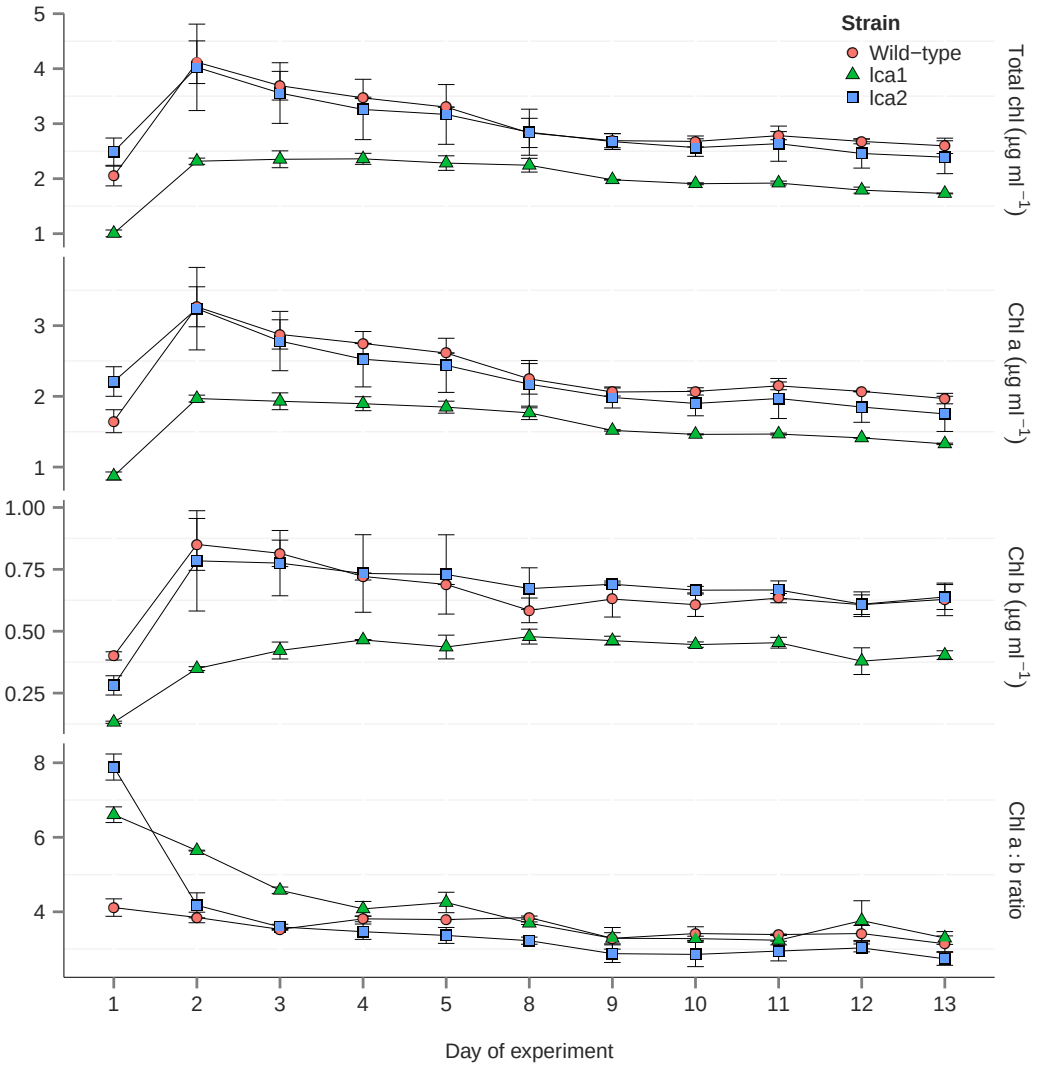


FIGURE A.3: The measured total chlorophyll, chlorophyll a and b and the ratio of chlorophyll a and b in the wild-type, lca1 and lca2 under 13 days of growth in a photobioreactor under simulated day-light (sinusoidal light curve from 0 - 2000 µmol photons m^{-2} s^{-1})

.

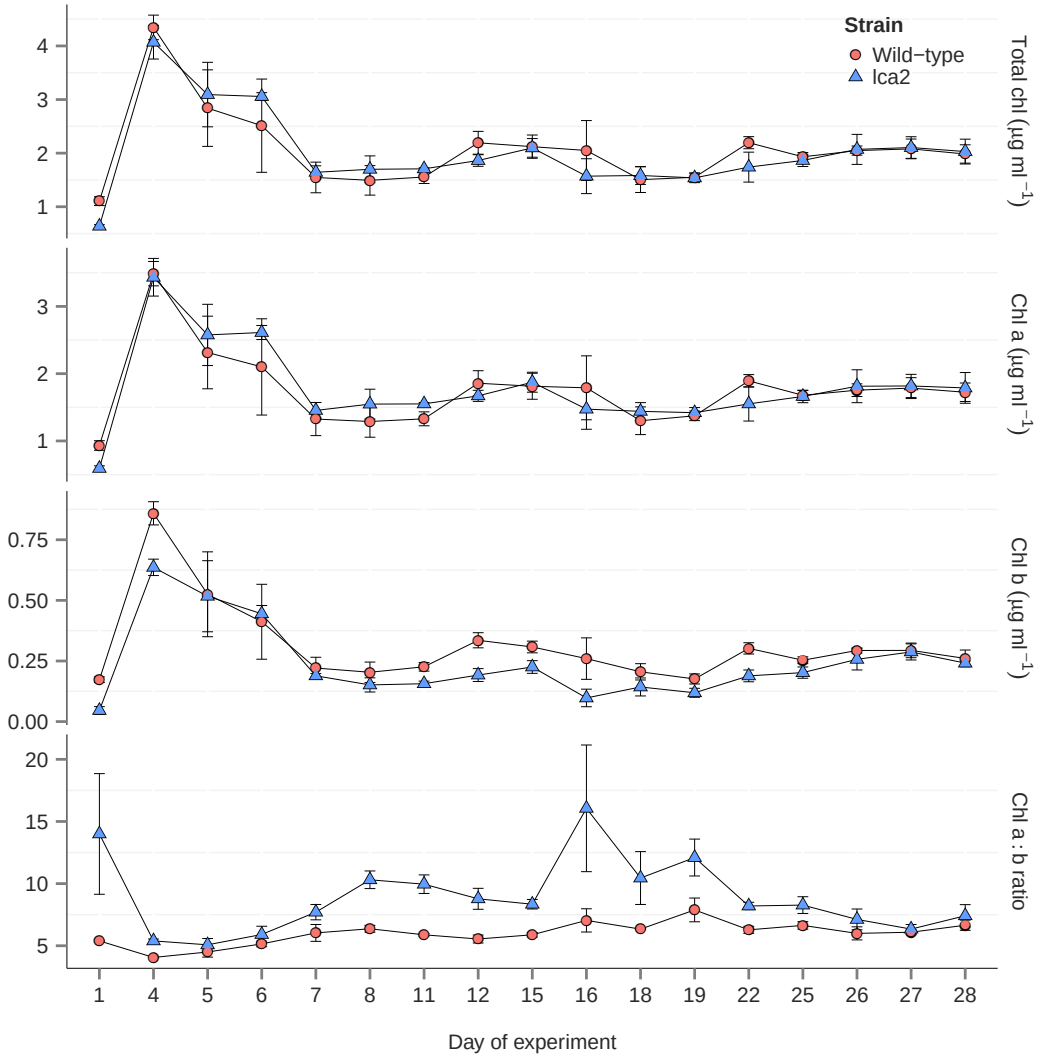
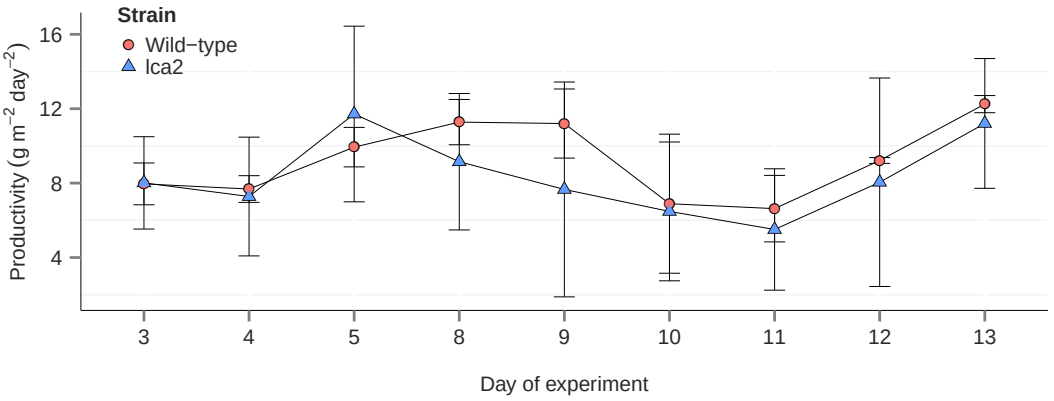


FIGURE A.4: The measured total chlorophyll, chlorophyll a and b and the ratio of chlorophyll a and b in the wild-type and lca2 under 28 days of growth in a photo-bioreactor under full light (2000 µmol photons m^{-2} s^{-1})

.

(A)



(B)

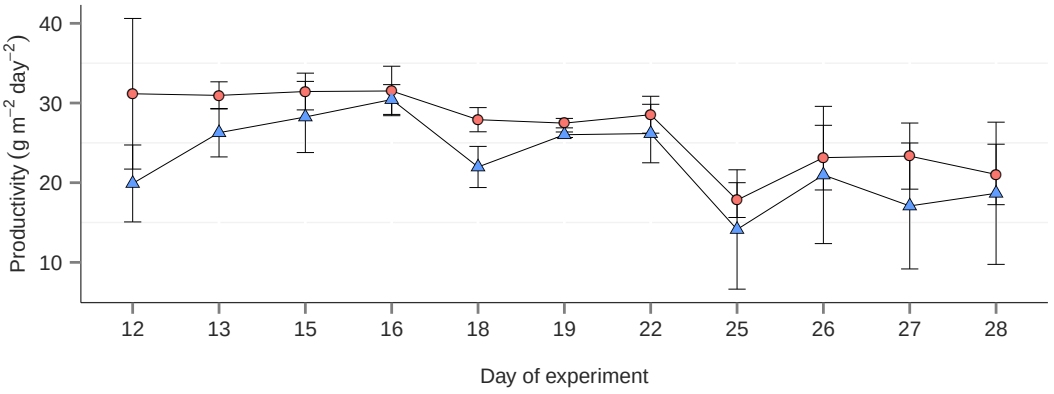


FIGURE A.5: A) The measured productivity (g m^{-2} $^{-1}$) in wild-type and lca2 during 13 days of growth in a photobioreactor under simulated day-light (sinusoidal light curve from 0 - 2000 µmol photons m^{-2} s^{-1}). B. Same as (A) but showing the measured productivity during 16 days of of growth under full light (2000 µmol photons m^{-2} s^{-1}).

Appendix B

Construction of transcriptome analysis pipeline

Table B.1: Unitags (Tag) and tag counts from the full SAGE library. The example shows the tag count for 24 tags identified in two replicates in the wild type under low (WTL1 and WTL2) and medium light (WTM1 and WTM2).

Tag	WTL1	WTL2	WTM1	WTM2
CATGGCGCGCCGCAGGCGGCGGCGGA	0	0	1	0
CATGGCGCGGCTGTCTCCTAAGCAGG	1	3	4	12
CATGGCGCTGCTGGTCGGATTTGGGG	0	0	2	0
CATGGCGGCCTGGCCCCAGTGCGCAA	4	5	3	12
CATGGCGGCTGCCAGGGCAGCACTGG	8	3	6	8
CATGGCGGTGCGCATCGTCAAGCTCG	0	2	2	3
CATGGCGGTGCGCATCGTCAAGTACG	7	9	11	5
CATGGCGGTGCTGAGCAACGTGCTTC	9	4	3	5
CATGGCGTGGCCACATCTTGGTCCTC	6	2	6	5
CATGGCGTTGGCGCCGGAAATCCCTG	0	0	0	0
CATGGCTACCTCTTGTAGTCAGCATT	0	0	0	3
CATGGCTACGCGCCTCACCAGGTGCG	2	0	4	2
CATGGCTACGTTTTTCCCTCAACTGC	0	0	0	0
CATGGCTAGCCTGGCGGCTGTGGATT	3	3	3	2
CATGGCTCAGTCTCGCTGCTGTGCGC	2	3	2	4
CATGGCTCCCAGCGCGATGCGCTGCT	5	2	0	7
CATGGCTCGACGCTCTGCTCCCACGT	10	4	6	10
CATGGCTCGTCACGCCGGGAAGCAGC	7	12	4	8
CATGGCTCTCCCTGGGGGCTTCCCTC	3	10	4	6
CATGGCTGATAACGACTTTCCCGGCT	0	1	4	3
CATGGCTGCTGCAGCGCAGGCACAGC	3	6	4	3
CATGGCTGCTGGGAGCGTTGGGCTTT	6	4	3	7

Table B.2: The total observed number of unitags and the sum of the tag count for each of the 27 libraries from the wild type, lca1 and lca2 strains under low, medium and high light in triplicate.

Strain	Light	Replicate	Total observed unitags	Sum of counts
Wild type	Low	1	42824	1238997
		2	48790	1219386
		3	49009	1060898
	Medium	1	47719	852538
		2	49593	1342685
		3	63983	1542570
	High	1	62250	1189393
		2	68864	1194741
		3	62245	1006256
lca1	Low	1	69271	1140982
		2	83712	992346
		3	53503	801840
	Medium	1	50917	807296
		2	70217	1319708
		3	48172	1213800
	High	1	58365	1127761
		2	52260	1034790
		3	46084	964461
lca2	Low	1	54079	1122762
		2	49013	1096600
		3	46998	899420
	Medium	1	67615	1372320
		2	76814	1685890
		3	85639	1984453
	High	1	66552	1466137
		2	57720	1260096
		3	54656	1393047

Table B.3: The distribution of tags per unitag in the combined libraries from the SAGE transcriptome in six different categories ranging from a single tag to a million tags per unitag and more. The total number of unitags within each category and the percent of unitags per category out of the total, together with the summed tag count tag counts as percent of total per category.

Tags per unitag	Unitags	Unitags (%) of total	Tags	Tags (%) of total
1 - 10	325343	85.0	718836	2.2
11 - 100	47302	12.4	1400681	4.3
101 - 1000	8370	2.2	2435924	7.5
1001 - 10000	1558	0.4	4271349	13.2
10001 - 100000	250	0.1	6854346	21.2
$\geq$ 100001	23	0.0	16650037	51.5

```
#!/usr/bin/python

# import modules
import fileinput
import csv

# File input
Source_name = "NCBI_filt_lowest_EST_by_evalue.csv"
Source_file = fileinput.input([Source_name])

Filter_name = "All_sequence_gi_Chloro_Mito_NCBI_UniProt.txt"
Filter_file = fileinput.input([Filter_name])

# File output
Result_filtered = "Filtered_output_combined.txt"
Result_filtered = open(Result_filtered, "a")

Result_remaining = "Remaining_output_combined.txt"
Result_remaining = open(Result_remaining, "a")

Entries_list = []
Total_count = 0
Remaining = 0
NCBI_filtered = 0
UniProt_filtered = 0
Multiple_filtered = 0
Exceptions = 0

# Iterate through the filter file and append entries to a list
for line in Filter_file:
  Entries = line.strip()
  Entries_list.append(Entries)

# Iterate through each line in the source file and tab-delimit.
for line in Source_file:
  Source_hit = line.split("\t")
  Source_NCBI = Source_hit[36].split("|")
  Source_NCBI = Source_NCBI[1:2]
  Source_NCBI = ', '.join(Source_NCBI)
  Source_UniProt = Source_hit[38]
  Total_count += 1

  if Source_hit[0] == "Tag":
    Result_filtered.write(line)

  elif Source_NCBI not in Entries_list and Source_UniProt not in Entries_list:
    Result_remaining.write(line)
    Remaining += 1

  elif Source_NCBI in Entries_list and Source_UniProt in Entries_list:
    Result_filtered.write(line)
    Multiple_filtered += 1

  elif Source_NCBI in Entries_list:
```

```
55      Result_filtered.write(line)
56      NCBI_filtered += 1
57
58    elif Source_UniProt in Entries_list:
59      Result_filtered.write(line)
60      UniProt_filtered += 1
61
62    else:
63      Exceptions += 1
64
65  print "Results summary:", "\n"
66  print "Entries in source file:", Total_count - 1
67  print "Entries in filter file:", len(Entries_list)
68  print "Entries filtered by both NCBI and UniProt ID:", Multiple_filtered
69  print "Remaining entries filtered only by NCBI ID:", NCBI_filtered
70  print "Remaining entries filtered only by UniProt ID:", UniProt_filtered
71  print "Remaining entries after filtering:", Remaining
72  print "Total number of processed entries in source file:", NCBI_filtered, "+",
      UniProt_filtered, "+", Multiple_filtered, "+", Remaining, "=", NCBI_filtered
      + UniProt_filtered + Multiple_filtered + Remaining
73  print "Number of exceptions found:", Exceptions, "\n"
74
75  # Close open files
76  Source_file.close()
77  Filter_file.close()
78  Result_filtered.close()
79  Result_remaining.close()
80
81  print "End of script!"
```

Listing B.1: Filtering of plastid linked tags - stage 1

```
#!/usr/bin/python

# import modules
import fileinput
import csv

# File input
Source_name = "Remaining_output_combined.txt"
Source_file = fileinput.input([Source_name])

Filter_name = "Arabidopsis_plastid_linked_ESTs.txt"
Filter_file = fileinput.input([Filter_name])

# File output
Result_filtered = "Filtered_output_AT.txt"
Result_filtered = open(Result_filtered, "a")

Result_remaining = "Remaining_output_AT.txt"
Result_remaining = open(Result_remaining, "a")

Entries_list = []

# Iterate through the filter file and append entries to a list
for line in Filter_file:
  Entries = line.strip()
  Entries_list.append(Entries)

print len(Entries_list)

# Iterate through each line in the source file and tab-delimit.
for line in Source_file:
  Source_hit = line.split("\t")
  Source_ID = Source_hit[34]
  Source_ID = Source_ID.strip('"')

  if Source_ID not in Entries_list:
    Result_remaining.write(line)

  elif Source_ID in Entries_list:
    Result_filtered.write(line)
    print "Filter list hit:", Source_ID

  else:
    print "Exception found, please check for errors"

# Close open files
Source_file.close()
Filter_file.close()
Result_filtered.close()
Result_remaining.close()

print "End of script!"
```

Listing B.2: Filtering of plastid linked tags - stage 2

```
#!/usr/bin/python

# import modules
import fileinput
import csv
from Bio import SeqIO

# Open the source file to be read (one line at a time)
Source_name = "Mapped_AT_high_frequency_tags.csv"
Source_file = fileinput.input([Source_name])

# Open the fasta formatted AT protein file
Search_name = "TAIR10_pep_20101214.fasta"

# Create and open output file
Result_name = "EST_Protein_Prediction.tdt"
Result_file = open(Result_name, "a")

# Create dictionary containing all protein sequences in fasta format, protein AT
    number is the key
handle = open(Search_name, "rU")
AT_protein_dict = SeqIO.to_dict(SeqIO.parse(handle, "fasta"))

# Iterate through each line in the source file and tab-delimit.
for line in Source_file:
    Source_Line = line.split()

# Clean up the protein identifier to make it searchable
    Source_protein = Source_Line[3]
    Source_protein = Source_protein[11:]

# Write complete header (look for first line identifier)
    if "TopHit" == Source_Line[3]:
        Result_file.write(line.rstrip())
        Result_file.write('\tAT identifier  Symbols Description Location and
    Length\n')

# Search for protein identifier in AT protein dictionary
    if Source_protein in AT_protein_dict.keys():

# Split the description part of the dictionary using the | character
        parts = AT_protein_dict[Source_protein].description.split("|")

# Write the result to a new file
        Result_file.write(line[:-2])
        Result_file.write('\t')
        Result_file.write(AT_protein_dict[Source_protein].id)
        Result_file.write('\t')
        Result_file.write(parts[1])
        Result_file.write('\t')
        Result_file.write(parts[2])
        Result_file.write('\t')
        Result_file.write(parts[3])
        Result_file.write('\n')
```

```
53
54 # Close open files
55 Source_file.close()
56 handle.close()
57 Result_file.close()
58
59 print "End of script!"
```

Listing B.3: Annotation of A. thaliana linked tags using TAIR 10

```
1  #!/usr/bin/python
2
3  # import modules
4  import fileinput
5  import sys
6
7  # Open the annotation file to be read (one line at a time)
8  Annotation_file = fileinput.input([sys.argv[1]])
9  # Open the mapping file to be read (one line at a time)
10 Mapping_file = fileinput.input([sys.argv[2]])
11 # Name, create and open the result file for writing in append mode
12 Result_file = open(sys.argv[3], "a")
13
14 # Variables, lists and dictionaries
15 EST_line_dict = {}
16 EST_dict = {}
17 EST_list = []
18 Temp_list = []
19 Temp_line = []
20 Final_list = []
21 Final_dict = {}
22 EST_temp_dict = {}
23 count_line = 0
24
25 # ---------- Part 1 - Process annotation file ----------#
26
27 # Iterate through each line in the annotation file, tab-delimit and select the
     key and value
28 for line in Annotation_file:
29
30   Annotation_Line = line.split('\t')
31
32   if Annotation_Line[3] != 'topnrEvalue' and Annotation_Line[3] != 'No_sig_nr_hit
       ' and Annotation_Line[8] != 'eValue' and Annotation_Line[8] != 'No_evalue':
33
34     NCBI_ID_entries = Annotation_Line[2].split('|')
35     EST = Annotation_Line[0]
36     NCBI_ID = NCBI_ID_entries[1]
37     NCBI_eValue = Annotation_Line[3]
38     Swiss_ID = Annotation_Line[6]
39     Swiss_eValue = Annotation_Line[8]
40     Temp_list = [EST, NCBI_ID, NCBI_eValue, Swiss_ID, Swiss_eValue]
41     EST_dict[EST] = Temp_list
42     EST_list.append(Temp_list)
43
```

```
    # Separate dict that holds the full line for printing
    EST_line_dict[EST] = line

  elif Annotation_Line[0] == 'ContigName':
    Annotation_header = line

  Temp_list = list()

# ---------- Part 2a - Process mapping file ----------#

# Iterate through each line in the mapping file, tab-delimit and clean up the EST entries for each tag
for line in Mapping_file:
  line = line.rstrip('\r\n')
  Mapping_Line = line.split('\t')
  EST_entries = Mapping_Line[32].split(';')

  # Print the header to the result file
  if Mapping_Line[0] == 'Tag':
    Mapping_header = line
    Result_file.write(Mapping_header)
    Result_file.write('\t')
    Result_file.write(Annotation_header)

  # Iterate through the EST entries and search for hits in the EST dict, if found move them to a new temp dict
  else:
    for item in EST_entries:
      if item in EST_dict:
        EST_temp_dict[item] = EST_dict[item]

    # Check if the EST_temp_dict contains any entries
    if len(EST_temp_dict) >= 1:

      #Write a newline between each tag-group and count the total number of tags with annotated EST's
      Result_file.write("\n")
      count_line += 1
      for item in EST_entries:
        if item in EST_dict:
          # Create a temporary list of lists containing the entries in the EST_temp_dict
          Temp_line = EST_dict[item]
          Temp_list.append(Temp_line)

    # Use this commented section for a print out of the results without collapsing to single entries
    #Final_list = Temp_list

    # Sort the Temp_list by NCBI's gi then e-value
    Temp_list = sorted(Temp_list, key=lambda item: (item[1], float(item[2])))

    # Collapse list to single entries with lowest e-value for each NCBI gi ID (ex '323434961')
    Identifier = 0
```

```
    Prev_identifier = 0
    for item in Temp_list:
      Identifier = item[1]
      if Identifier != Prev_identifier:
        Final_list.append(item)

      Prev_identifier = Identifier

    # Re-sort the Temp_list by e-value
    Final_list = sorted(Final_list, key=lambda item: float(item[2]))

    # ---------- Part 3 - Print out to result file----------#

    # Iterate over the keys in the final dict
    for item in Final_list:
      Result_file.write(line)
      Result_file.write('\t')
      Result_file.write(EST_line_dict[item[0]])

    # Clear lists for next line
    EST_temp_dict.clear()
    Final_list = list()
    Temp_line = list()
    Temp_list = list()

# Close open files
Annotation_file.close()
Mapping_file.close()
Result_file.close()

print "Total number of annotated tags", count_line
print "Script finished!"
```

Listing B.4: Assigning EST annotation to tag

```
# Load the edgeR bioconductor package
library(edgeR)

#Read in original SAGE data with tags and counts for each library
datafile <- read.delim("../Data/MT_tag_comp.tdt.txt",row.names="Tag", header=TRUE
    )

#Read in the annotation data
annotfile <- read.csv("../Data/Annotated_tags_filtered.csv", row.names="Tag",
    header=TRUE)


#Create a vector with nine factorial levels that groups the libraries by
    replicates
Design <- factor(c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,6,6,6,7,7,7,8,8,8,9,9,9))

#Create default data storage object in edgeR, linking together the tag data, the
    annotation and the design
Data_Full <- DGEList(counts=datafile, genes=annotfile[rownames(datafile),], group
    =Design)
```

```
16
17 #Create a design matrix that show edgeR how the experiments are related
18 Matrix <- model.matrix(~0+Design, data=Data_Full$samples)
19
20 #Filter low counts tags (retain any tag with one count or more per million in 9
       or more of the 27 libraries) and create a new DGEList object that holds only
       the tags that meets this requirements.
21
22 keep <- rowSums(cpm(Data_Full) >= 1) >= 9
23 Data_filt <- Data_Full[keep,]
24
25 #Reset the library sizes (inherited from the full data object) with the new total
        library counts in the filtered object.
26 Data_filt$samples$lib.size <- colSums(Data_filt$counts)
27
28 #Show the number of tags in original and the filtered object
29 dim(Data_Full)
30 dim(Data_filt)
31
32 #Calculate the normalization factors for each library depending on the library
       size and show the library size (lib.size) and the calculated normalisation
       factor (norm.factors)
33
34 Data_filt <- calcNormFactors(Data_filt)
35 Data_filt$samples
36
37 #Calculate the common, trended and pairwise dispersion of the tags
38 Data_filt <- estimateGLMCommonDisp(Data_filt, Matrix)
39 Data_filt <- estimateGLMTrendedDisp(Data_filt, Matrix)
40 Data_filt <- estimateGLMTagwiseDisp(Data_filt, Matrix)
41
42 #Fit the GLM model to the DGEList object using the design matrix
43 fit_Data_filt <- glmFit(Data_filt, Matrix)
44
45 #Find differentially expressed genes (in this example by contrasting the wild
       type versus the lca2 strain under low light)
46 glm_L2vsWT_L <- glmLRT(fit_Data_filt, contrast=c(-1, 0, 0, 0, 0, 0, 1, 0, 0))
47
48 #Save the result sorted by lowest false discovery rate (FDR) to a comma separated
        file.
49 write.csv(topTags(glm_L2vsWT_L, n=nrow(Data_filt))$table, file="glm_L2vsWT_L.csv"
       )
50
51 #Visualize the calculated biological dispersion and show the resulting plot
52 plotBCV(Data_filt)
53
54 #Visualize the relationship between the libraries using multidimensional scaling
       and show the resulting plot
55 plotMDS(Data_filt)
```

Listing B.5: edgeR analysis example

Bibliography

Allahverdiyeva, Y., Suorsa, M., Rossi, F., Pavesi, A., Kater, M. M., Antonacci, A., Tadini, L., Pribil, M., Schneider, A., Wanner, G., Leister, D., Aro, E. M., Barbato, R., and Pesaresi, P. (2013). Arabidopsis plants lacking PsbQ and PsbR subunits of the oxygen-evolving complex show altered PSII super-complex organization and short-term adaptive mechanisms. *Plant Journal*, 75(4):671–684.

Allen, J. F., de Paula, W. B. M., Puthiyaveetil, S., and Nield, J. (2011). A structural phylogenetic map for chloroplast photosynthesis. *Trends in Plant Science*, 16(12):645–655.

Alonso, D. L., Segura del Castillo, C. I., Grima, E. M., and Cohen, Z. (1996). FIRST INSIGHTS INTO IMPROVEMENT OF EICOSAPENTAENOIC ACID CONTENT IN PHAEODACTYLUM TRICORNUTUM (BACILLARIOPHYCEAE) BY INDUCED MUTAGENESIS. *Journal of Phycology*, 32(2):339–345.

Alve, E. and Nagy, J. (1986). Estuarine foraminiferal distribution in Sandebukta, a branch of the Oslo Fjord. *The Journal of Foraminiferal Research*.

Amin, P., Sy, D. a., Pilgrim, M. L., Parry, D. H., Nussaume, L., and Hoffman, N. E. (1999). Arabidopsis mutants lacking the 43- and 54-kilodalton subunits of the chloroplast signal recognition particle have distinct phenotypes. *Plant physiology*, 121(1):61–70.

Anders, S. and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome biology*, 11(10):R106.

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., and Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nature protocols*, 8(9):1765–86.

Andersen, R. A. (2005). *Algal culturing techniques*. Academic Press.

Andersson, M. X. and Dörmann, P. (2008). The Chloroplast: Interactions with the Environment. In *The Chloroplast: Interactions with the Environment*, chapter Chloroplas, pages 125–158.

Asakura, Y., Hirohashi, T., Kikuchi, S., Belcher, S., Osborne, E., Yano, S., Terashima, I., Barkan, A., and Nakai, M. (2003). Defects in the Biogenesis of Thylakoid Membranes. *Online*, 16(January):1–14.

Asakura, Y., Kikuchi, S., and Nakai, M. (2008). Non-identical contributions of two membrane-bound cpSRP components, cpFtsY and Alb3, to thylakoid biogenesis. *Plant Journal*, 56(6):1007–1017.

Azuma, T., Harrison, G. I., and Demain, a. L. (1992). Isolation of a gramicidin S hyperproducing strain of Bacillus brevis by use of a fluorescence activated cell sorting system. *Applied microbiology and biotechnology*, 38:173–178.

Bailey, S., MCCARREN, J., LIEBERMAN, S. L., MEUSER, J. E., ROMANO, A. E., Yee, D., Soriaga, L., Brown, R. C., WEISSMAN, J. C., and PRINCE, R. C. (2014). Algal mutants having a locked-in high light acclimated phenotype.

Baker, N. R. and Bowyer, J. R. (1994). *Photoinhibition of Photosynthesis: From Molecular Mechanisms to the Field*. Environmental plant biology series. BIOS Scientific Publishers.

Barber, J., Nield, J., Morris, E., Zheleva, D., and Hankamer, B. (1997). The structure, function and dynamics of photosystem two. *Physiologia Plantarum*, 100(4):817–827.

Barlow, R., Cummings, D., and Gibb, S. (1997). Improved resolution of mono- and divinyl chlorophylls a and b and zeaxanthin and lutein in phytoplankton extracts using reverse phase C-8 HPLC. *Marine Ecology Progress Series*, 161:303–307.

Barredo, J. L. (2012). *Microbial Carotenoids From Fungi: Methods and Protocols*. Methods in Molecular Biology. Humana Press.

Beale, S. I. (1999). Enzymes of chlorophyll biosynthesis. *Photosynthesis Research*, 60(1):43–73.

Beckmann, J., Lehr, F., Finazzi, G., Hankamer, B., Posten, C., Wobbe, L., and Kruse, O. (2009). Improvement of light to biomass conversion by de-regulation of light-harvesting protein translation in Chlamydomonas reinhardtii. *Journal of biotechnology*, 142(1):70–7.

Beer, S., Björk, M., and Beardall, J. (2014). *Photosynthesis in the Marine Environment*. EBL ebooks online. Wiley.

Beligni, M. V. and Mayfield, S. P. (2008). Arabidopsis thaliana mutants reveal a role for CSP41a and CSP41b, two ribosome-associated endonucleases, in chloroplast ribosomal RNA metabolism. *Plant Molecular Biology*, 67(4):389–401.

Bellafiore, S., Ferris, P., Naver, H., Göhre, V., and Rochaix, J.-D. (2002). Loss of Albino3 leads to the specific depletion of the light-harvesting system. *The Plant cell*, 14(9):2303–2314.

Ben-Amotz, a. and Avron, M. (1983). On the Factors Which Determine Massive beta-Carotene Accumulation in the Halotolerant Alga Dunaliella bardawil. *Plant physiology*, 72(3):593–597.

Ben-Shem, A., Frolow, F., and Nelson, N. (2003). Crystal structure of plant photosystem I. *Nature*, 426(6967):630–635.

Birmingham, B. C., Coleman, J. R., and Colman, B. (1982). Measurement of photorespiration in algae. *Plant physiology*, 69(1):259–262.

Bishop, C., Purton, S., and Nugent, J. (2003). Molecular analysis of the Chlamydomonas nuclear gene encoding PsbW and demonstration that PsbW is a subunit of photosystem II, but not photosystem I. *Plant Molecular Biology*, 52(2):285–289.

Björn, L. and Govindjee (2008). The Evolution of Photosynthesis and Its Environmental Impact. In Björn, L., editor, *The Evolution of Photosynthesis and Its Environmental Impact*, chapter Photobiolo, pages 255–287. Springer New York.

Blankenship, R. E. (2002). *Molecular Mechanisms of Photosynthesis*. Wiley.

Boekema, E. J., Van Breemen, J. F. L., Van Roon, H., and Dekker, J. P. (2000). Conformational changes in photosystem II supercomplexes upon removal of extrinsic subunits. *Biochemistry*, 39(42):12907–12915.

Bollenbach, T. J., Sharwood, R. E., Gutierrez, R., Lerbs-Mache, S., and Stern, D. B. (2009). The RNA-binding proteins CSP41a and CSP41b may regulate transcription and translation of chloroplast-encoded RNAs in Arabidopsis. *Plant Molecular Biology*, 69(5):541–552.

Bolton, J. R. (1977). Photochemical conversion and storage of solar energy. *Journal of Solid State Chemistry*, 22(1):3–8.

Bondarava, N., Beyer, P., and Krieger-Liszkay, A. (2005). Function of the 23 kDa extrinsic protein of Photosystem II as a manganese binding protein and its role in photoactivation. *Biochimica et biophysica acta*, 1708(1):63–70.

Borowitzka, M. and Siva, C. (2007). The taxonomy of the genus Dunaliella (Chlorophyta, Dunaliellales) with emphasis on the marine and halophilic species. *Journal of Applied Phycology*, 19:567–590.

Borowitzka, M. a. (1999). Commercial production of microalgae: ponds, tanks, tubes and fermenters. *Journal of Biotechnology*, 70(1-3):313–321.

Bricker, T. M. and Frankel, L. K. (2003). Carboxylate groups on the manganese-stabilizing protein are required for efficient binding of the 24 kDa extrinsic protein to photosystem II. *Biochemistry*, 42(7):2056–2061.

Bullard, J. H., Purdom, E., Hansen, K. D., and Dudoit, S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics*, 11:94.

Butler, W. L. (1978). Energy Distribution in the Photochemical Apparatus of Photosynthesis. *Annual Review of Plant Physiology*, 29(1):345–378.

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. *Annual review of genetics*, 32:339–377.

Casper-Lindley, C. and Björkman, O. (1998). Fluorescence quenching in four unicellular algae with different light-harvesting and xanthophyll-cycle pigments. *Photosynthesis Research*, 56(3):277–289.

Cerutti, H., Ma, X., Msanne, J., and Repas, T. (2011). RNA-Mediated Silencing in Algae: Biological Roles and Tools for Analysis of Gene Function. *Eukaryotic cell*, 10(9):1164–72.

Chaturvedi, R. and Fujita, Y. (2006). Isolation of enhanced eicosapentaenoic acid producing mutants of Nannochloropsis oculata ST-6 using ethyl methane sulfonate induced mutagenesis techniques and their characterization at mRNA transcript level. *Phycological Research*, 54(December 2004):208–219.

Chaturvedi, R., Uppalapati, S., Alamsjah, M. A., and Fujita, Y. (2004). Isolation of quizalofop-resistant mutants of Nannochloropsis oculata (Eustigmatophyceae) with high eicosapentaenoic acid following N-methyl-N-nitrosourea-induced random mutagenesis. *Journal of Applied Phycology*, 16:135–144.

Chen, H., Zhang, D., Guo, J., Wu, H., Jin, M., Lu, Q., Lu, C., and Zhang, L. (2006). A Psb27 homologue in Arabidopsis thaliana is required for efficient repair of photodamaged photosystem II. *Plant Molecular Biology*, 61(4-5):567–575.

Chen, M. and Blankenship, R. E. (2011). Expanding the solar spectrum used by photosynthesis. *Trends in Plant Science*, 16:427–431.

Chen, Y., Lun, A. T. L., and Smyth, G. K. (2014). Differential Expression Analysis of Complex RNA-seq Experiments Using edgeR. In *Statistical Analysis of Next Generation Sequence Data*, number January, pages 1–25.

Chitnis, P. R. (2001). I: Function and Physiology. *Annual Review of Plant Physiology and Plant Molecular Biology*, 52(1):593–626.

Clarens, A. F., Resurreccion, E. P., White, M. a., and Colosi, L. M. (2010). Environmental life cycle comparison of algae to other bioenergy feedstocks. *Environmental Science and Technology*, 44(5):1813–1819.

Consortium, T. U. (2014). UniProt: a hub for protein information. *Nucleic Acids Research*, 43(D1):D204–D212.

Croce, R. and van Amerongen, H. (2014). Natural strategies for photosynthetic light harvesting. *Nature chemical biology*, 10(7):492–501.

Czarnecki, O. and Grimm, B. (2012). Post-translational control of tetrapyrrole biosynthesis in plants, algae, and cyanobacteria. *Journal of Experimental Botany*, 63(4):1675–1687.

de Mooij, T., Janssen, M., Cerezo-Chinarro, O., Mussgnug, J. H., Kruse, O., Ballottari, M., Bassi, R., Bujaldon, S., Wollman, F.-A., and Wijffels, R. H. (2015). Antenna size reduction as a strategy to increase biomass productivity: a great potential not yet realized. *Journal of Applied Phycology*, 27(3):1063–1077.

De Wit, P., Pespeni, M. H., Ladner, J. T., Barshis, D. J., Seneca, F., Jaris, H., Therkildsen, N. O., Morikawa, M., and Palumbi, S. R. (2012). The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular ecology resources*, 12(6):1058–1067.

Dekker, J. P. and Boekema, E. J. (2005). Supramolecular organization of thylakoid membrane proteins in green plants. *Biochimica et biophysica acta*, 1706(1-2):12–39.

Demirbas, M. F. (2011). Biofuels from algae for sustainable development. *Applied Energy*, 88(10):3473–3480.

Dheda, K., Huggett, J. F., Bustin, S. a., Johnson, M. a., Rook, G., and Zumla, A. (2004). Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *BioTechniques*, 37(1):112–119.

Diatchuk, V., Lotan, O., Koshkin, V., Wikstroem, P., and Pick, E. (1997). Inhibition of NADPH oxidase activation by 4-(2-aminoethyl)-benzenesulfonyl fluoride and related compounds. *The Journal of biological chemistry*, 272:13292–13301.

Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, N. S., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., and Jaffrézic, F.

(2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14(6):671–683.

Dobáková, M., Sobotka, R., Tichý, M., and Komenda, J. (2009). Psb28 protein is involved in the biogenesis of the photosystem II inner antenna CP47 (PsbB) in the cyanobacterium Synechocystis sp. PCC 6803. *Plant physiology*, 149(2):1076–1086.

Douglas, R. H. and Ballou, C. E. (1980). Isolation of Kluyveromyces lactis mannoprotein mutants by fluorescence-activated cell sorting. *Journal of Biological Chemistry*, 255(12):5979–5985.

Drop, B., Webber-Birungi, M., Yadav, S. K. N., Filipowicz-Szymanska, A., Fusetti, F., Boekema, E. J., and Croce, R. (2014). Light-harvesting complex II (LHCII) and its supramolecular organization in Chlamydomonas reinhardtii. *Biochimica et biophysica acta*, 1837(1):63–72.

Durão, P., Aigner, H., Nagy, P., Mueller-Cajar, O., Hartl, F. U., and Hayer-Hartl, M. (2015). Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nature chemical biology*, 11(2):148–55.

Edwards, R. and Palopoli, N. (2015). Computational Prediction of Short Linear Motifs from Protein Sequences. In Zhou, P. and Huang, J., editors, *Computational Peptidology SE - 6*, volume 1268 of *Methods in Molecular Biology*, pages 89–141. Springer New York.

Eisentraut, A. (2010). SUSTAINABLE PRODUCTION OF SECOND-GENERATION BIOFUELS. Potential and perspectives in major economies and developing countries.

Evans, J. R. (1989). Photosynthesis and nitrogen relationship in leaves of C3 plants. *Oecologia*, 78:9–19.

Falk, S., Ravaud, S., Koch, J., and Sinning, I. (2010). The C terminus of the Alb3 membrane insertase recruits cpSRP43 to the thylakoid membrane. *Journal of Biological Chemistry*, 285(8):5954–5962.

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Högberg, P., Linder, S., Mackenzie, F. T., Moore, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W. (2000). The global carbon cycle: a test of our knowledge of earth as a system. *Science (New York, N.Y.)*, 290(5490):291–296.

Falkowski, P. G. (1980). *Primary Productivity in the Sea*. Environmental science research. Plenum Press.

Falkowski, P. G. (1998). Biogeochemical Controls and Feedbacks on Ocean Primary Production. *Science*, 281(5374):200–206.

Falkowski, P. G., Katz, M. E., Knoll, A. H., Quigg, A., Raven, J. a., Schofield, O., and Taylor, F. J. R. (2004). The evolution of modern eukaryotic phytoplankton. *Science (New York, N.Y.)*, 305(5682):354–60.

Falkowski, P. G. and LaRoche, J. (1991). ACCLIMATION TO SPECTRAL IRRADIANCE IN ALGAE. *Journal of Phycology*, 27(1):8–14.

Falkowski, P. G. and Owens, T. G. (1980a). Light- Shade Adaptation. *Plant Physiology*, 66:592–595.

Falkowski, P. G. and Owens, T. G. (1980b). Light-Shade Adaptation : TWO STRATEGIES IN MARINE PHYTOPLANKTON. *Plant physiology*, 66(4):592–595.

Falkowski, P. G. and Raven, J. A. (1997). *Aquatic Photosynthesis*. Life science. Ecology. Blackwell Science.

Fenchel, T. (2008). The microbial loop - 25 years later. *Journal of Experimental Marine Biology and Ecology*, 366(1-2):99–103.

Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J., and Iwata, S. (2004). Architecture of the photosynthetic oxygen-evolving center. *Science (New York, N.Y.)*, 303(March):1831–1838.

Finazzi, G., Moreau, H., and Bowler, C. (2010). Genomic insights into photosynthesis in eukaryotic phytoplankton. *Trends in Plant Science*, 15:565–572.

Fromme, P., Jordan, P., and Krauß, N. (2001). Structure of photosystem I. *Biochimica et Biophysica Acta - Bioenergetics*, 1507(1-3):5–31.

Fujiki, T., Suzue, T., Kimoto, H., and Saino, T. (2007). Photosynthetic electron transport in Dunaliella tertiolecta (Chlorophyceae) measured by fast repetition rate fluorometry: Relation to carbon assimilation. *Journal of Plankton Research*, 29:199–208.

Gálová, E., Šalgovičová, I., Demko, V., Mikulová, K., Ševčovičová, A., Slováková, ., Kyselá, V., and Hudák, J. (2008). A short overview of chlorophyll biosynthesis in algae. *Biologia*, 63(6):947–951.

Gentleman, R. C., Gentleman, R. C., Carey, V. J., Carey, V. J., Bates, D. M., Bates, D. M., Bolstad, B., Bolstad, B., Dettling, M., Dettling, M., Dudoit, S., Dudoit, S., Ellis, B., Ellis, B., Gautier, L., Gautier, L., Ge, Y., Ge, Y., Gentry, J., Gentry, J., Hornik, K., Hornik, K., Hothorn, T., Hothorn, T., Huber, W., Huber, W., Iacus, S., Iacus, S., Irizarry, R., Irizarry, R., Leisch, F., Leisch, F., Li, C., Li, C., Maechler, M., Maechler, M., Rossini, A. J., Rossini, A. J., Sawitzki, G., Sawitzki, G., Smith, C., Smith, C., Smyth, G., Smyth, G., Tierney, L., Tierney, L., Yang, J. Y. H., Yang, J. Y. H., Zhang, J., and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80.

Ghirardi, M. L. and Melis, A. (1988). Chlorophyll b deficiency in soybean mutants. I. Effects on photosystem stoichiometry and chlorophyll antenna size. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 932:130–137.

Gill, S. S. and Tuteja, N. (2010). Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant physiology and biochemistry : PPB / Société française de physiologie végétale*, 48(12):909–30.

Gilmore, a. M. (1997). Mechanistic aspects of xanthophyll cycle-dependent photoprotection in higher plant chloroplasts and leaves. *Physiologia Plantarum*, 99(1):197–209.

Gish, W. and States, D. J. (1993). Identification of protein coding regions by database similarity search. *Nat Genet*, 3(3):266–272.

Göhre, V., Ossenbühl, F., Crèvecoeur, M., Eichacker, L. A., and Rochaix, J.-D. (2006). One of two alb3 proteins is essential for the assembly of the photosystems and for cell survival in Chlamydomonas. *The Plant cell*, 18(6):1454–1466.

Goldblatt, C., Lenton, T. M., and Watson, A. J. (2006). Bistability of atmospheric oxygen and the Great Oxidation. *Nature*, 443(7112):683–686.

Gorbunov, M. Y., Kolber, Z. S., Lesser, M. P., and Falkowski, P. G. (2001). Photosynthesis and photoprotection in symbiotic corals. *Limnology and Oceanography*, 46(1):75–85.

Goyal, A. (2007a). Osmoregulation in Dunaliella, Part I: Effects of osmotic stress on photosynthesis, dark respiration and glycerol metabolism in Dunaliella tertiolecta and its salt-sensitive mutant (HL 25/8). *Plant Physiology and Biochemistry*, 45(9):696704.

Goyal, A. (2007b). Osmoregulation in Dunaliella, Part II: Photosynthesis and starch contribute carbon for glycerol synthesis during a salt stress in Dunaliella tertiolecta. *Plant Physiology and Biochemistry*, 45(9):705710.

Green, B. and Parson, W. W. (2003). *Light-Harvesting Antennas in Photosynthesis*. Advances in Photosynthesis and Respiration. Springer.

Green, B. R. (1988). The chlorophyll-protein complexes of higher plant photosynthetic membranes or Just what green band is that? *Photosynthesis Research*, 15:3–32.

Greene, E. A., Codomo, C. A., Taylor, N. E., Henikoff, J. G., Till, B. J., Reynolds, S. H., Enns, L. C., Burtner, C., Johnson, J. E., Odden, A. R., Comai, L., and Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics*, 164(2):731–740.

Grossman, A. R., Bhaya, D., Apt, K. E., and Kehoe, D. M. (1995). Light-harvesting complexes in oxygenic photosynthesis: diversity, control, and evolution. *Annual review of genetics*, 29:231–88.

Grover, M., Gaur, T., Kochhar, a., Maheshwari, S. C., and Tyagi, a. K. (1999). Nucleotide sequence of psbQ gene for 16-kDa protein of oxygen-evolving complex from Arabidopsis thaliana and regulation of its expression. *DNA research : an international journal for rapid publication of reports on genes and genomes*, 6(3):173–177.

Guillard, R. R. (1975). Culture of Phytoplankton for Feeding Marine Invertebrates. In Smith, W. and Chanley, M., editors, *Culture of Marine Invertebrate Animals SE - 3*, pages 29–60. Springer US.

Guillard, R. R. L. and Ryther, J. H. (1962). STUDIES OF MARINE PLANKTONIC DIATOMS: I. CYCLOTELLA NANA HUSTEDT, AND DETONULA CONFERVACEA (CLEVE) GRAN. *Canadian Journal of Microbiology*, 8(2):229–239.

Gupta, R., He, Z., and Luan, S. (2002). Functional relationship of cytochrome c(6) and plastocyanin in Arabidopsis. *Nature*, 417(6888):567–571.

Gutensohn, M., Fan, E., Frielingsdorf, S., Hanner, P., Hou, B., Hust, B., and Klösgen, R. B. (2006). Toc, Tic, Tat et al.: Structure and function of protein transport machineries in chloroplasts. *Journal of Plant Physiology*, 163(3):333–347.

Haldrup, A., Simpson, D. J., and Scheller, H. V. (2000). Down-regulation of the PSI-F subunit of photosystem I (PSI) in Arabidopsis thaliana. The PSI-F subunit is essential for photoautotrophic growth and contributes to antenna function. *Journal of Biological Chemistry*, 275(40):31211–31218.

Hall, D. and Rao, K. (1994). *Photosynthesis*. Studies in Biology. Cambridge University Press.

Hallman, a. (2007). Algal transgenics and biotechnology. *Transgenic Plant Journal*, 1(1):81–98.

Harbers, M. and Carninci, P. (2005). Tag-based approaches for transcriptome research and genome annotation. *Nature methods*, 2(7):495–502.

Hasan, S. S., Yamashita, E., Baniulis, D., and Cramer, W. a. (2013). Quinone-dependent proton transfer pathways in the photosynthetic cytochrome b6f complex. *Proceedings of the National Academy of Sciences of the United States of America*, 110(11):4297–302.

Hejazi, M. a., Holwerda, E., and Wijffels, R. H. (2004). Milking microalga Dunaliella salina for beta-carotene production in two-phase bioreactors. *Biotechnology and bioengineering*, 85(5):475–81.

Hill, R. (1937). Oxygen Evolved by Isolated Chloroplasts. *Nature*, 139(3525):881–882.

Hill, R. and Bendall, F. (1960). Function of the Two Cytochrome Components in Chloroplasts: A Working Hypothesis. *Nature*, 186(4719):136–137.

Horton, P., Ruban, a. V., and Walters, R. G. (1996). Regulation of Light Harvesting in Green Plants. *Annual Review of Plant Physiology and Plant Molecular Biology*, 47(1):655–684.

Hou, Y. and Lin, S. (2009). Distinct gene number-genome size relationships for eukaryotes and non-eukaryotes: Gene content estimation for dinoflagellate genomes. *PLoS ONE*, 4(9).

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., and Darzins, A. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant journal : for cell and molecular biology*, 54(4):621–39.

Huang, H., Cai, L., and Wong, H. W. (2008). Clustering Analysis of SAGE Transcription Profiles Using a Poisson Approach. In *METHODS IN MOLECULAR BIOLOGY*.

Huesemann, M. H., Hausmann, T. S., Bartha, R., Aksoy, M., Weissman, J. C., and Benemann, J. R. (2009). Biomass productivities in wild type and pigment mutant of Cyclotella sp. (Diatom). *Applied biochemistry and biotechnology*, 157(3):507–26.

Hulsegge, I., Kommadath, A., and Smits, M. a. (2009). Globaltest and GOEAST: two different approaches for Gene Ontology analysis. *BMC proceedings*, 3 Suppl 4:S10.

Ifuku, K., Yamamoto, Y., Ono, T.-A., Ishihara, S., and Sato, F. (2005). PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. *Plant physiology*, 139(3):1175–1184.

Ihnken, S., Kromkamp, J. C., and Beardall, J. (2011). Photoacclimation in Dunaliella tertiolecta reveals a unique NPQ pattern upon exposure to irradiance. *Photosynthesis research*.

Jareborg, N., Birney, E., and Durbin, R. (2000). Mouse and Human Gene Pairs Comparative Analysis of Noncoding Regions of 77 Orthologous Mouse and Human Gene Pairs. *Genome Research*, pages 815–824.

Jarvis, P. and Robinson, C. (2004). Mechanisms of protein import and routing in chloroplasts. *Current Biology*, 14(24):1064–1077.

Jensen, P. E., Rosgaard, L., Knoetzel, J., and Scheller, H. V. (2002). Photosystem I activity is increased in the absence of the PSI-G subunit. *Journal of Biological Chemistry*, 277(4):2798–2803.

Jin, E., Feth, B., and Melis, A. (2003). A mutant of the green alga Dunaliella salina constitutively accumulates zeaxanthin under all growth conditions. *Biotechnology and bioengineering*, 81(1):115–24.

Jin, E. and Yokthongwattana, K. (2003). Role of the reversible xanthophyll cycle in the photosystem II damage and repair cycle in Dunaliella salina. *Plant physiology*, 132(May):352–364.

Kargul, J., Turkina, M. V., Nield, J., Benson, S., Vener, A. V., and Barber, J. (2005). Light-harvesting complex II protein CP29 binds to photosystem I of Chlamydomonas reinhardtii under State 2 conditions. *The FEBS journal*, 272(18):4797–4806.

Kasahara, M., Swartz, T. E., Olney, M. a., Onodera, A., Mochizuki, N., Fukuzawa, H., Asamizu, E., Tabata, S., Kanegae, H., Takano, M., Christie, J. M., Nagatani, A., and Briggs, W. R. (2002). Photochemical properties of the flavin mononucleotide-binding domains of the phototropins from Arabidopsis, rice, and Chlamydomonas reinhardtii. *Plant physiology*, 129(2):762–773.

Kehoe, D. M. and Gutu, A. (2006). Responding to color: the regulation of complementary chromatic adaptation. *Annual review of plant biology*, 57:127–150.

Kereïche, S., Kiss, A. Z., Kouřil, R., Boekema, E. J., and Horton, P. (2010). The PsbS protein controls the macro-organisation of photosystem II complexes in the grana membranes of higher plant chloroplasts. *FEBS Letters*, 584(4):759–764.

Kim, J. H., Glick, R. E., and Melis, a. (1993a). Dynamics of Photosystem Stoichiometry Adjustment by Light Quality in Chloroplasts. *Plant physiology*, 102(1):181–190.

Kim, J. H., Nemson, J. a., and Melis, a. (1993b). Photosystem II Reaction Center Damage and Repair in Dunaliella salina (Green Alga) (Analysis under Physiological and Irradiance-Stress Conditions). *Plant physiology*, 103(1993):181–189.

Kim, S. J., Robinson, C., and Mant, A. (2015). Sec/SRP-independent insertion of two thylakoid membrane proteins bearing cleavable signal peptides. *FEBS Letters*, 424(1):105–108.

Kim, Y., Schumaker, K. S., and Zhu, J.-K. (2006). EMS mutagenesis of Arabidopsis. *Methods in molecular biology (Clifton, N.J.)*, 323(6):101–3.

Kirst, H., Garcia-Cerdan, J. G., Zurbriggen, A., Ruehle, T., and Melis, A. (2012). Truncated photosystem chlorophyll antenna size in the green microalga Chlamydomonas reinhardtii upon deletion of the TLA3-CpSRP43 gene. *Plant physiology*, 160(4):2251–60.

Kirst, H., García-Cerdán, J. G., Zurbriggen, A. S., and Melis, A. (2011). Assembly of the Light-Harvesting Chlorophyll Antenna in the Green Alga Chlamydomonas reinhardtii Requires Expression of the TLA2-CpFTSY Gene. *Plant physiology*, 158(2):930–45.

Kirst, H. and Melis, A. (2014). The chloroplast signal recognition particle (CpSRP) pathway as a tool to minimize chlorophyll antenna size and maximize photosynthetic productivity. *Biotechnology advances*, 32(1):66–72.

Kitayama, M., Kitayama, K., and Togasaki, R. K. (1994). A cDNA clone encoding a ferredoxin-NADP+ reductase from Chlamydomonas reinhardtii. *Plant physiology*, 106(4):1715–1716.

Klimyuk, V. I., Persello-Cartieaux, F., Havaux, M., Contard-David, P., Schuenemann, D., Meiherhoff, K., Gouet, P., Jones, J. D., Hoffman, N. E., and Nussaume, L. (1999). A chromodomain protein encoded by the arabidopsis CAO gene is a plant-specific component of the chloroplast signal recognition particle pathway that is involved in LHCP targeting. *The Plant cell*, 11(1):87–99.

Kolber, Z. and Falkowski, P. G. (1993). Use of active fluorescence to estimate phytoplankton photosynthesis in situ. 38(8):1646–1665.

Komine, Y., Kikis, E., Schuster, G., and Stern, D. (2002). Evidence for in vivo modulation of chloroplast RNA stability by 3'-UTR homopolymeric tails in Chlamydomonas reinhardtii. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6):4085–90.

Koressaar, T. and Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, 23(10):1289–1291.

Kruse, O., Rupprecht, J., Mussgnug, J. H., Dismukes, G. C., and Hankamer, B. (2005). Photosynthesis: a blueprint for solar energy capture and biohydrogen production technologies. *Photochemical & photobiological sciences : Official journal of the European Photochemistry Association and the European Society for Photobiology*, 4(12):957–970.

Kugelmann, M., Fausser, A., Ossenbühl, F., and Brennicke, A. (2013). Phenotypes of Alb3p and carotenoid synthesis mutants show similarities regarding light sensitivity, thylakoid structure and protein stability. *Photosynthetica*, 51(1):45–54.

Kulp, T. R., Hoeft, S. E., Asao, M., Madigan, M. T., Hollibaugh, J. T., Fisher, J. C., Stolz, J. F., Culbertson, C. W., Miller, L. G., and Oremland, R. S. (2008). Arsenic(III) fuels anoxygenic photosynthesis in hot spring biofilms from Mono Lake, California. *Science (New York, N.Y.)*, 321(5891):967–970.

Larkum, A. W. D., Ross, I. L., Kruse, O., and Hankamer, B. (2011). Selection, breeding and engineering of microalgae for bioenergy and biofuel production. *Trends in biotechnology*, pages 1–8.

Laroche, J., Bennett, J., and Falkowski, P. G. (1990). Characterization of a cDNA encoding for the 28.5-kDa LHCII apoprotein from the unicellular marine chlorophyte, Dunalieila tertiolecta. *Gene*, 95:165–171.

Laroche, J., Mortain-Bertrand, a., and Falkowski, P. G. (1991). Light Intensity-Induced Changes in cab mRNA and Light Harvesting Complex II Apoprotein Levels in the Unicellular Chlorophyte Dunaliella tertiolecta. *Plant physiology*, 97:147–153.

Lee, H. K., Braynen, W., Keshav, K., and Pavlidis, P. (2005). ErmineJ: tool for functional analysis of gene expression data sets. *BMC bioinformatics*, 6:269.

Li, X. P., Björkman, O., Shih, C., Grossman, a. R., Rosenquist, M., Jansson, S., and Niyogi, K. K. (2000). A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature*, 403(6768):391–395.

Li, Z., Wakao, S., Fischer, B. B., and Niyogi, K. K. (2009). Sensing and responding to excess light. *Annual review of plant biology*, 60:239–260.

Livak, K. J. and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif*, 25(4):402–408.

Lu, B., Xu, C., Awai, K., Jones, a. D., and Benning, C. (2007). A small ATPase protein of Arabidopsis, TGD3, involved in chloroplast lipid import. *Journal of Biological Chemistry*, 282(49):35945–35953.

Lu, S. and Li, L. (2008). Carotenoid metabolism: Biosynthesis, regulation, and beyond. *Journal of Integrative Plant Biology*, 50(7):778–785.

Madden, S., Wang, C., and Landes, G. (2000). Serial analysis of gene expression: from gene discovery to target identification. *Drug Discovery Today*, 5(9):415–425.

Marsac, N. T. (1977). Occurence and nature of chromatic adaption in cyanobacteria. *J Bacteriol*, 130(130 (1)):82–91.

Masuda, T., Tanaka, A., and Melis, A. (2003). Chlorophyll antenna size adjustments by irradiance in Dunaliella salina involve coordinate regulation of chlorophyll a oxygenase (CAO) and Lhcb gene expression. *Plant molecular biology*, 51(5):757–71.

Matsumura, H. and Molina, C. (2012). DeepSuperSAGE: High-Throughput Transcriptome Sequencing with Now- and Next-Generation Sequencing Technologies. In *Tag-based Next Generation Sequencing*.

Matsumura, H., Reich, S., Ito, A., Saitoh, H., Kamoun, S., Winter, P., Kahl, G., Reuter, M., Kruger, D. H., and Terauchi, R. (2003). Gene expression analysis of plant host-pathogen interactions by SuperSAGE. *Proceedings of the National Academy of Sciences of the United States of America*, 100:15718–15723.

Matsumura, H., Yoshida, K., Luo, S., Kimura, E., Fujibe, T., Albertyn, Z., Barrero, R. a., Krüger, D. H., Kahl, G., Schroth, G. P., and Terauchi, R. (2010a). High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PloS one*, 5(8):e12010.

Matsumura, H., Yoshida, K., Luo, S., Kimura, E., Fujibe, T., Albertyn, Z., Barrero, R. a., Krüger, D. H., Kahl, G., Schroth, G. P., and Terauchi, R. (2010b). High-throughput superSAGE for digital gene expression analysis of multiple samples using next generation sequencing. *PLoS ONE*, 5(8).

Matsumura, H., Yoshida, K., Luo, S., Krüger, D. H., Kahl, G., Schroth, G. P., and Terauchi, R. (2011). High-Throughput SuperSAGE. *Methods in molecular biology*, 687(7):135–146.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research*, 40(10):4288–97.

Melis, A. (2009). Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science*, 177(4):272–280.

Melis, A., Neidhardt, J., and Benemann, J. (1998). Dunaliella salina (Chlorophyta) with small chlorophyll antenna sizes exhibit higher photosynthetic productivities and photon use efficiencies than normally pigmented cells. *Journal of Applied Phycology*, pages 515–525.

Mendoza, H., de la Jara, A., Freijanes, K., Carmona, L., Ramos, A. A., de Sousa Duarte, V., and Serafim Varela, J. C. (2008). Characterization of Dunaliella salina strains by flow cytometry: a new approach to select carotenoid hyperproducing strains. *Electronic Journal of Biotechnology*, 11(4).

Merchant, S., Prochnik, S., and Vallon, O. (2007). The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science*, 318(5848):245–250.

Minowa, T., Yokoyama, S.-y., Kishimoto, M., and Okakura, T. (1995). Oil production from algal cells of Dunaliella tertiolecta by direct thermochemical liquefaction. *Fuel*, 74(12):17351738.

Mitra, M., Kirst, H., Dewez, D., and Melis, A. (2012). Modulation of the light-harvesting chlorophyll antenna size in Chlamydomonas reinhardtii by TLA1 gene over-expression and RNA interference. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 367(1608):3430–43.

Moore, A. (2008). Biofuels are dead: long live biofuels(?) - part two. *New biotechnology*, 25(2-3):96–100.

Moore, C. M., Suggett, D. J., Hickman, A. E., Kim, Y. N., Tweddle, J. F., Sharples, J., Geider, R. J., and Holligan, P. M. (2006). Phytoplankton photoacclimation and photoadaptation in response to environmental gradients in a shelf sea. *Limnology and Oceanography*, 51(2):936–949.

Moriyama, E. N. and Powell, J. R. (1998). Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. *Nucleic Acids Research*, 26(13):3188–3193.

Morozova, O. and Marra, M. A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics*, 92(5):255–64.

Mostafa, S. (2012). Microalgal biotechnology: prospects and applications. In *Plant Science*, chapter Microalgal.

Mueller-Cajar, O. and Whitney, S. M. (2008). Directing the evolution of Rubisco and Rubisco activase: first impressions of a new tool for photosynthesis research. *Photosynthesis research*, 98(1-3):667–75.

Müller, P., Li, X. P., and Niyogi, K. K. (2001). Non-photochemical quenching. A response to excess light energy. *Plant physiology*, 125(4):1558–1566.

Murakami, R., Ifuku, K., Takabayashi, A., Shikanai, T., Endo, T., and Sato, F. (2005). Functional dissection of two Arabidopsis PsbO proteins PsbO1 and PsbO2. *FEBS Journal*, 272(9):2165–2175.

Murchie, E. H., Pinto, M., and Horton, P. (2009). Agriculture and the new challenges for photosynthesis research. *New Phytologist*, 181(3):532–552.

Murray, J. W. (2012). Sequence variation at the oxygen-evolving centre of photosystem II: a new class of 'rogue' cyanobacterial D1 proteins. *Photosynthesis research*, 110(3):177–84.

Mussgnug, J. H., Thomas-Hall, S., Rupprecht, J., Foo, A., Klassen, V., McDowall, A., Schenk, P. M., Kruse, O., and Hankamer, B. (2007a). Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. *Plant biotechnology journal*, 5(6):802–14.

Mussgnug, J. H., Thomas-Hall, S., Rupprecht, J., Foo, A., Klassen, V., McDowall, A., Schenk, P. M., Kruse, O., and Hankamer, B. (2007b). Engineering photosynthetic light capture: impacts on improved solar energy to biomass conversion. *Plant biotechnology journal*, 5(6):802–14.

Mussgnug, J. H., Wobbe, L., Elles, I., Claus, C., Hamilton, M., Fink, a., Kahmann, U., Kapazoglou, a., Mullineaux, C. W., Hippler, M., Nickelsen, J., Nixon, P. J., and Kruse, O. (2005). NAB1 Is an RNA Binding Protein Involved in the Light-Regulated Differentail Expression of the Light-Harveting Antenna of Chlamydomonas reinhardtii. *The Plant Cell*, 17(December):3409–3421.

Neidhardt, J., Benemann, J., Zhang, L., and Melis, A. (1998). Photosystem-II repair and chloroplast recovery from irradiance stress: relationship between chronic photoinhibition, light-harvesting chlorophyll antenna size and. *Photosynthesis research*, pages 175–184.

Nelson, N. and Yocum, C. F. (2006). Structure and function of photosystems I and II. *Annual review of plant biology*, 57:521–565.

Nisbet, E. G. and Nisbet, R. E. R. (2008). Methane, oxygen, photosynthesis, rubisco and the regulation of the air through time. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1504):2745–2754.

Niyogi, K., Bjorkman, O., and Grossman, A. (1997). Chlamydomonas Xanthophyll Cycle Mutants Identified by Video Imaging of Chlorophyll Fluorescence Quenching. *The Plant cell*, 9(8):1369–1380.

Niyogi, K. K. (1999). PHOTOPROTECTION REVISITED: Genetic and Molecular Approaches. *Annual review of plant physiology and plant molecular biology*, 50:333–359.

Niyogi, K. K., Li, X. P., Rosenberg, V., and Jung, H. S. (2005). Is PsbS the site of non-photochemical quenching in photosynthesis? *Journal of Experimental Botany*, 56(411):375–382.

Nowrousian, M. (2010). Next-generation sequencing techniques for eukaryotic microorganisms: sequencing-based solutions to biological problems. *Eukaryotic cell*, 9(9):1300–10.

Oey, M., Ross, I. L., Stephens, E., Steinbeck, J., Wolf, J., Radzun, K. A., Kügler, J., Ringsmuth, A. K., Kruse, O., and Hankamer, B. (2013). RNAi Knock-Down of LHCBM1, 2 and 3 Increases Photosynthetic H2 Production Efficiency of the Green Alga Chlamydomonas reinhardtii. *PLoS ONE*, 8(4).

Oren, A. (2005). A hundred years of Dunaliella research: 19052005. *Saline systems*, 1(1):2.

Ort, D. R. and Melis, A. (2011). Optimizing antenna size to maximize photosynthetic efficiency. *Plant physiology*, 155(1):79–85.

Osnabriick, U. and Sheva, B. (1998). Chlamydomonas reinhardtii. *Trends in Genetics*, 14(11):S.18–S.19.

Ossenbühl, F., Göhre, V., Meurer, J., Krieger-Liszkay, A., Rochaix, J.-D., and Eichacker, L. a. (2004). Efficient assembly of photosystem II in Chlamydomonas reinhardtii requires Alb3.1p, a homolog of Arabidopsis ALBINO3. *The Plant cell*, 16(7):1790–1800.

Ouyang, M., Li, X., Ma, J., Chi, W., Xiao, J., Zou, M., Chen, F., Lu, C., and Zhang, L. (2011). LTD is a protein required for sorting light-harvesting chlorophyll-binding proteins to the chloroplast SRP pathway. *Nature communications*, 2:277.

Parker, M. S., Mock, T., and Armbrust, E. V. (2008). Genomic insights into marine microalgae. *Annual review of genetics*, 42:619–645.

Patino, W. D., Mian, O. Y., Hwang, P. M., and Tomaselli, G. F. (2002). Serial analysis of gene expression: Technical considerations and applications to cardiovascular biology. *Circulation Research*, 91:565–569.

Pavlidis, P., Lewis, D. P., and Noble, W. S. (2002). Exploring gene expression data with class scores. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 474–485.

Pearson, G., Lago-Leston, A., Valente, M., and Serrão, E. (2006). Simple and rapid RNA extraction from freeze-dried tissue of brown algae and seagrasses. *European Journal of Phycology*, 41(1):97–104.

Peers, G. (2014). Increasing algal photosynthetic productivity by integrating ecophysiology with systems biology. *Trends in biotechnology*, (Box 1):1–5.

Peers, G. and Niyogi, K. K. (2008). Pond scum genomics: the genomes of Chlamydomonas and Ostreococcus. *The Plant cell*, 20(March):502–507.

Pérez-Bueno, M. L., Barón, M., and García-Luque, I. (2011). PsbO, PsbP, and PsbQ of photosystem II are encoded by gene families in Nicotiana benthamiana. Structure and functionality of their isoforms. *Photosynthetica*, 49(4):573–580.

Pérez-Pérez, M. E., Lemaire, S. D., and Crespo, J. L. (2012). Reactive oxygen species and autophagy in plants and algae. *Plant physiology*, 160(1):156–64.

Pessarakli, M. (2001). *Handbook of photosynthesis*, volume 40.

Peterson, R. (2005). PsbS genotype in relation to coordinated function of PS II and PS I in Arabidopsis leaves. *Photosynthesis Research*, 85(2):205–219.

Phillip, D., Ruban, A. V., Horton, P., Asato, A., and Young, A. J. (1996). Quenching of chlorophyll fluorescence in the major light-harvesting complex of photosystem II: a systematic study of the effect of carotenoid structure. *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1492–7.

Pilgrim, M. L., Van Wijk, K. J., Parry, D. H., Sy, D. a. C., and Hoffman, N. E. (1998). Expression of a dominant negative form of cpSRP54 inhibits chloroplast biogenesis in Arabidopsis. *Plant Journal*, 13(2):177–186.

Pinto, P. I., Matsumura, H., Thorne, M. A., Power, D. M., Terauchi, R., Reinhardt, R., and Canário, A. V. (2010). Gill transcriptome response to changes in environmental calcium in the green spotted puffer fish. *BMC genomics*, 11:476.

Pisal, D. S. and Lele, S. S. (2005). Carotenoid production from microalga, Dunaliella salina. *Indian Journal of Biotechnology*, 4(4):476–483.

Planavsky, N. J., Asael, D., Hofmann, A., Reinhard, C. T., Lalonde, S. V., Knudsen, A., Wang, X., Ossa Ossa, F., Pecoits, E., Smith, A. J. B., Beukes, N. J., Bekker, A.,

Johnson, T. M., Konhauser, K. O., Lyons, T. W., and Rouxel, O. J. (2014). Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nature Geoscience*, 7(4):283–286.

Polle, J. E. W., Kanakagiri, S., Jin, E., Masuda, T., and Melis, A. (2002). Truncated chlorophyll antenna size of the photosystems - A practical method to improve microalgal productivity and hydrogen production in mass culture. *International Journal of Hydrogen Energy*, 27(11-12):1257–1264.

Polle, J. E. W., Kanakagiri, S.-D., and Melis, A. (2003). tla1, a DNA insertional transformant of the green alga Chlamydomonas reinhardtii with a truncated light-harvesting chlorophyll antenna size. *Planta*, 217(1):49–59.

Porra, R. J., Schäfer, W., Cmiel, E., Katheder, I., and Scheer, H. (1994). The derivation of the formyl-group oxygen of chlorophyll b in higher plants from molecular oxygen. Achievement of high enrichment of the 7-formyl-group oxygen from 18O2 in greening maize leaves. *European journal of biochemistry / FEBS*, 219(1-2):671–679.

Pulz, O. and Gross, W. (2004). Valuable products from biotechnology of microalgae. *Applied Microbiology and Biotechnology*, 65:635–648.

Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryotic cell*, 9(4):486–501.

Ramos, A. a., Polle, J., Tran, D., Cushman, J. C., Jin, E.-S., and Varela, J. C. (2011). The unicellular green alga Dunaliella salina Teod. as a model for abiotic stress tolerance: genetic advances and future perspectives. *Algae*, 26(1):3–20.

Rasmussen, B., Fletcher, I. R., Brocks, J. J., and Kilburn, M. R. (2008). Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature*, 455(7216):1101–1104.

Rau, A., Gallopin, M., Celeux, G., and Jaffrézic, F. (2013). Data-based filtering for replicated high-throughput transcriptome sequencing experiments. *Bioinformatics*, 29(17):2146–2152.

Rismani-Yazdi, H., Haznedaroglu, B. Z., Bibby, K., and Peccia, J. (2011). Transcriptome sequencing and annotation of the microalgae Dunaliella tertiolecta: Pathway description and gene discovery for production of next-generation biofuels. *BMC Genomics*, 12(1):148.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. pages 1–13.

Robinson, C., Robinson, C., Woolhead, C., Woolhead, C., Edwards, W., and Edwards, W. (2000). Transport of proteins into and across the thylakoid membrane. *Journal of experimental botany*, 51 Spec No(February):369–74.

Robinson, M. D. and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3):R25.

Robinson, M. D. and Smyth, G. K. (2008). Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics*, 9(2):321–332.

Roose, J. L., Yocum, C. F., and Popelkova, H. (2010). Function of PsbO, the photosystem II manganese-stabilizing protein: Probing the role of aspartic acid 157. *Biochemistry*, 49(29):6042–6051.

Rothschild, L. J. (2008). The evolution of photosynthesis...again? *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1504):2787–2801.

Rott, R., Zipor, G., Portnoy, V., Liveanu, V., and Schuster, G. (2003). RNA polyadenylation and degradation in cyanobacteria are similar to the chloroplast but different from Escherichia coli. *The Journal of biological chemistry*, 278(18):15771–7.

Ryan-Keogh, T. J., Macey, A. I., Cockshutt, A. M., Moore, C. M., and Bibby, T. S. (2012). The cyanobacterial chlorophyll-binding-protein IsiA acts to increase the in vivo effective absorption cross-section of PSI under iron limitation. *Journal of Phycology*, 48(1):145–154.

Saha, S., Sparks, A. B., Rago, C., Akmaev, V., Wang, C. J., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2002). Using the transcriptome to annotate the genome. *Nature biotechnology*, 20:508–512.

Sakata, S., Mizusawa, N., Kubota-Kawai, H., Sakurai, I., and Wada, H. (2013). Psb28 is involved in recovery of photosystem II at high temperature in Synechocystis sp. PCC 6803. *Biochimica et Biophysica Acta - Bioenergetics*, 1827(1):50–59.

Schenk, P. M., Thomas-Hall, S. R., Stephens, E., Marx, U. C., Mussgnug, J. H., Posten, C., Kruse, O., and Hankamer, B. (2008). Second Generation Biofuels: High-Efficiency Microalgae for Biodiesel Production. *BioEnergy Research*, 1(1):20–43.

Schneider, J., Livne, A., Sukenik, A., and Roessler, P. (1995). A MUTANT OF NANNOCHLOROPSIS DEFICIENT IN EICOSAPENTAENOIC ACID PRODUCTION. *Phytochemistry*, 40(3):807–814.

Senge, M., Ryan, A., Letchford, K., MacGowan, S., and Mielke, T. (2014). Chlorophylls, Symmetry, Chirality, and Photosynthesis. *Symmetry*, 6(3):781–843.

Sheehan, J., Dunahay, T., Benemann, J., and Roesller, P. (1998). A look back at the US department of energy's aquatic species program: biodiesel from algae. Technical report, National Renewable Energy Laboratory.

Shi, L. X., Lorković, Z. J., Oelmüller, R., and Schröder, W. P. (2000). The low molecular mass PsbW protein is involved in the stabilization of the dimeric photosystem II complex in Arabidopsis thaliana. *Journal of Biological Chemistry*, 275(48):37945–37950.

Six, C., Finkel, Z. V., Rodriguez, F., Marie, D., Partensky, F., and Campbell, D. a. (2008). Contrasting photoacclimation costs in ecotypes of the marine eukaryotic picoplankter Ostreococcus. *Limnology and Oceanography*, 53(1):255–265.

Smith, B. M., Morrissey, P. J., Guenther, J. E., Nemson, J. a., Harrison, M. a., Allen, J. F., and Melis, a. (1990). Response of the Photosynthetic Apparatus in Dunaliella salina (Green Algae) to Irradiance Stress. *Plant physiology*, 93:1433–1440.

Smith, D. R., Lee, R. W., Cushman, J. C., Magnuson, J. K., Tran, D., and Polle, J. E. W. (2010). The Dunaliella salina organelle genomes: large sequences, inflated with intronic and intergenic DNA. *BMC plant biology*, 10:83.

Sommer, F., Drepper, F., Haehnel, W., and Hippler, M. (2006). Identification of precise electrostatic recognition sites between cytochrome c6 and the photosystem I subunit PsaF using mass spectrometry. *Journal of Biological Chemistry*, 281(46):35097–35103.

Soneson, C. and Delorenzi, M. (2013). A comparison of methods for differential expression analysis of RNA-seq data. *BMC bioinformatics*, 14(1):91.

Spolaore, P., Joannis-Cassan, C., Duran, E., and Isambert, A. (2006). Commercial applications of microalgae. *Journal of Bioscience and Bioengineering*, 101(2):87–96.

Stephenson, P. G., Moore, C. M., Terry, M. J., Zubkov, M. V., and Bibby, T. S. (2011). Improving photosynthesis for algal biofuels: toward a green revolution. *Trends in biotechnology*, 29(12):615–623.

Sukenik, A., Bennett, J., and Falkowski, P. (1987). Light-saturated photosynthesis Limitation by electron transport or carbon fixation? *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 891(3):205–215.

Sukenik, A., Bennett, J., and Falkowski, P. (1988). Changes in the abundance of individual apoproteins of light-harvesting chlorophyll a/b-protein complexes of photosystem I and II with growth irradiance in the marine chlorophyte Dunaliella tertiolecta. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, 932:206215.

Sundberg, E., Slagter, J. G., Fridborg, I., Cleary, S. P., Robinson, C., and Coupland, G. (1997). ALBINO3, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria. *The Plant cell*, 9(5):717–730.

Takahashi, Y., Goldschmidt-Clermont, M., Soen, S. Y., Franzén, L. G., and Rochaix, J. D. (1991). Directed chloroplast transformation in Chlamydomonas reinhardtii: insertional inactivation of the psaC gene encoding the iron sulfur protein destabilizes photosystem I. *The EMBO journal*, 10(8):2033–2040.

Tanniou, A., Turpin, V., and Lebeau, T. (2012). Comparison of cryopreservation methods for the long term storage of the marine diatom Haslea ostrearia (simonsen). *Cryobiology*, 65(1):45–50.

Teodoresco, E. (1905). Organisation et developpement du Dunaliella, nouveau genre de Volvocacee-Polyblepharidee. *Beih. z. bot. Centralbl.*, 18:215232.

Terauchi, R. and Matsumura, H. (2008). SuperSAGE: The Most Advanced Transcriptome Technology for Functional Genomics. *The Handbook of Plant Functional Genomics: Concepts and Protocols*.

Thai, T., Doan, Y., and Obbard, J. P. (2012). Enhanced intracellular lipid in Nannochloropsis sp. via random mutagenesis and flow cytometric cell sorting. *Algal Research*, 1(1):17–21.

Till, B. J., Reynolds, S. H., Greene, E. a., Codomo, C. a., Enns, L. C., Johnson, J. E., Burtner, C., Odden, A. R., Young, K., Taylor, N. E., Henikoff, J. G., Comai, L., and Henikoff, S. (2003). Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome research*, 13(3):524–30.

Tohri, A., Dohmae, N., Suzuki, T., Ohta, H., Inoue, Y., and Enami, I. (2004). Identification of domains on the extrinsic 23 kDa protein possibly involved in electrostatic interaction with the extrinsic 33 kDa protein in spinach photosystem II. *European Journal of Biochemistry*, 271(5):962–971.

Tokutsu, R., Iwai, M., and Minagawa, J. (2009). CP29, a monomeric light-harvesting complex II protein, is essential for state transitions in Chlamydomonas reinhardtii. *Journal of Biological Chemistry*, 284(12):7777–7782.

Tzvetkova-Chevolleau, T., Hutin, C., Noël, L. D., Goforth, R., Carde, J.-P., Caffarri, S., Sinning, I., Groves, M., Teulon, J.-M., Hoffman, N. E., Henry, R., Havaux, M., and Nussaume, L. (2007). Canonical signal recognition particle components can be bypassed for posttranslational protein targeting in chloroplasts. *The Plant cell*, 19(5):1635–1648.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, 40(15):1–12.

van der Giezen, M. (2011). Mitochondria and the Rise of Eukaryotes. *BioScience*, 61(8):594–601.

Varsano, T., Wolf, S. G., and Pick, U. (2006). A chlorophyll a/b-binding protein homolog that is induced by iron deficiency is associated with enlarged photosystem I units in the eucaryotic alga Dunaliella salina. *Journal of Biological Chemistry*, 281(15):10305–10315.

Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995). Serial Analysis of Gene Expression. *Science*, 270(5235):484–487.

Vinayak, V., Manoylov, K., Gateau, H., Blanckaert, V., Hérault, J., Pencréac'h, G., Marchand, J., Gordon, R., and Schoefs, B. (2015). Diatom Milking: A Review and New Approaches. *Marine Drugs*, 13(5):2629–2665.

Von Wettstein, D., Gough, S., and Kannangara, C. G. (1995). Chlorophyll Biosynthesis. *The Plant cell*, 7(7):1039–1057.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63.

Webb, M. R. and Melis, a. (1995). Chloroplast Response in Dunaliella salina to Irradiance Stress (Effect on Thylakoid Membrane Protein Assembly and Function). *Plant physiology*, 107(1 995):885–893.

Wei, L., Cao, Y., Bai, L., Liang, X., Deng, T., Li, J., and Qiao, D. (2007). Cloning and expression of a gene coding for the major light-harvesting chlorophyll a/b protein of photosystem II in the green alga Dunaliella salina. *Journal of Applied Phycology*, 19(1):89–94.

Welschmeyer, N. a. (1994). Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. *Limnology and Oceanography*, 39(8):1985–1992.

Wheeler, D. L., Church, D. M., Edgar, R., Federhen, S., Helmberg, W., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Suzek, T. O., Tatusova, T. a., and Wagner, L. (2004). Database resources of the National Center for Biotechnology Information: update. *Nucleic acids research*, 32(Database issue):D35–D40.

White, A. L. and Jahnke, L. S. (2002). Contrasting effects of UV-A and UV-B on photosynthesis and photoprotection of beta-carotene in two Dunaliella spp. *Plant & cell physiology*, 43(8):877–884.

Wingler, a., Lea, P. J., Quick, W. P., and Leegood, R. C. (2000). Photorespiration: metabolic pathways and their role in stress protection. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 355(1402):1517–1529.

Wise, R. R. (2006). *The Structure and Function of Plastids*, volume 23 of *Advances in Photosynthesis and Respiration*. Springer Netherlands, Dordrecht.

Wood, P. M. (1978). Interchangeable copper and iron proteins in algal photosynthesis. Studies on plastocyanin and cytochrome c-552 in Chlamydomonas. *European journal of biochemistry / FEBS*, 87(1):9–19.

Wortman, J. R., Haas, B. J., Hannick, L. I., Smith, R. K., Maiti, R., Ronning, C. M., Chan, A. P., Yu, C., Ayele, M., Whitelaw, C. a., White, O. R., and Town, C. D. (2003). Annotation of the Arabidopsis genome. *Plant physiology*, 132:461–468.

Xu, H. (2012). Light intensity dependent expression of Lhca gene family encoding LHCI in Dunaliella salina. *African Journal of Microbiology Research*, 6(32):6165–6171.

Yandell, M. and Ence, D. (2012). A beginner's guide to eukaryotic genome annotation. *Nature reviews. Genetics*, 13(5):329–42.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. L. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*, 13(1):134.

Yuan, J. and Cline, K. (1994). Plastocyanin and the 33-kDa subunit of the oxygen-evolving complex are transported into thylakoids with similar requirements as predicted from pathway specificity. *Journal of Biological Chemistry*, 269(28):18463–18467.

Zheng, Q. and Wang, X. J. (2008). GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. *Nucleic acids research*, 36(Web Server issue):358–363.

Zhu, X. G., Long, S. P., and Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology*, 19(2):153–159.

Zipfel, W. and Owens, T. G. (1991). Calculation of absolute photosystem I absorption cross-sections from P700 photo-oxidation kinetics. *Photosynthesis research*, 29(1):23–35.

Zygadlo, A., Jensen, P. E., Leister, D., and Scheller, H. V. (2005). Photosystem I lacking the PSI-G subunit has a higher affinity for plastocyanin and is sensitive to photodamage. *Biochimica et Biophysica Acta - Bioenergetics*, 1708(2):154–163.

Zygadlo, A., Robinson, C., Scheller, H. V., Mant, A., and Jensen, P. E. (2006). The properties of the positively charged loop region in PSI-G are essential for its "spontaneous" insertion into thylakoids and rapid assembly into the photosystem I complex. *Journal of Biological Chemistry*, 281(15):10548–10554.