Transcription factor engineering for the enhancement in the production of carbon storage molecules in *Chlamydomonas reinhardtii*

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Science and Engineering

2018

Javiera A. Ziehe Moreira

School of Earth and Environmental Sciences

TABLE OF CONTENTS

ABSTRACT

The depletion of fossil fuels and the problems of greenhouse gas pollution has pushed the development of alternative sustainable biofuels including from microalgae. Microalgae can produce high abundance of lipids and carbohydrates that can be utilised as a biofuel feedstock. However, improvements are required throughout the production process for microalgae biofuel to become commercially viable. One of the major bottlenecks is low metabolite productivity. Microalgae cultivation conditions such as nutrient starvation that can boost the accumulation and storage of lipids and carbohydrates in microalgae also compromise growth. To overcome this problem, genetic engineering may offer solutions to improve the production of these macromolecules without compromising growth. For genetic engineering to be successful, a better knowledge of the molecular mechanism involved in the biosynthesis of carbon storage molecules is required. Transcription factors (TF) are of interest as potential targets for genetic manipulation since their engineering could provide stronger up-regulation of metabolic pathways. In an effort to contribute to the identification of key TFs involved in carbon storage metabolism, this study focused on the characterization of two TFs from *Chlamydomonas reinhardtii*. PSR1 is a TF that has previously been identified as a regulator of lipid and starch metabolism, however, the exact mechanisms of regulation are unknown. By using the electrophoretic mobility assay (EMSA) I tested the ability of PSR1 to bind to the promotor region of three genes involved in the carbon storage metabolism. PSR1 was able to bind specifically to the regulatory elements in the promotors of these three genes. These results showed the DNA binding activity of PSR1 for the first time and indicates a mechanism though which PSR1 is regulating starch and lipid metabolism in *C. reinhardtii*. A previously uncharacterised TF, MYB2 was predicted from genomic data to be upregulated under different stress conditions. MYB2 showed close phylogenetic relationship and structural similarity with MYB-type TFs from *Arabidopsis thaliana* that are regulators of various environmental stress responses. To determine if MYB2 is involved in the regulation of starch and lipid metabolism, MYB2 overexpression and knockdown lines were generated and characterised. Surprisingly, both knockdown and overexpression lines showed similar phenotypes with regard to carbon storage content. Both sets of these MYB2 modified lines showed an increase in the lipid accumulation under stress condition, particularly under nitrogen (N) limitation where there was a more than 2 fold increase in lipid content when compared to control lines. These results suggest that MYB2 is involved in the regulation of carbon storage metabolism. However, its role appears to be more complex than expected and potential mechanisms of genetic robustness and compensation in its function are proposed and discussed. These results expose the complex regulatory networks involved in the regulation of the carbon storage metabolism in microalgae. In conclusion, results from this investigation have provided important insights into the role of two TFs involved in the lipid and starch biosynthesis, which could potentially be targets for genetic engineering for the enhancement of lipid and starch production in microalgae in the future.

DECLARATION

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

COPYRIGHT STATEMENT

I. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

II. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made only in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance Presentation of Theses Policy You are required to submit your thesis electronically Page 11 of 25 with licensing agreements which the University has from time to time. This page must form part of any such copies made.

III. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

IV. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=2442 0), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.library.manchester.ac.uk/about/regulations/) and in The University's policy on Presentation of Theses.

ACKWLEDGEMENTS

Undertaking this PhD has been a truly life changing experience. It is a journey that started with moving from the other side of the world and ended with a countless number of enriching experiences. It would have been impossible to finish my PhD without the support and guidance that I received from so many people.

Firstly, I would like to express my gratitude to my supervisor, Jon Pittman for the countless hours that he dedicated to my thesis. His comments, suggestions and constructive criticism throughout this entire project helped to shape my thesis and my career as a scientist. But above all, I would like to thank him for always making me feel lucky for having him as a supervisor. It is a common theme amongst PhD students to complain about supervisors. However, Jon never gave me a reason to complain, he was incredibly understanding and patient and his support kept me going during difficult times. Special thanks for keeping me sane during the crazy days before my submission and for the late nights of corrections.

Secondly, I would like to thank my wonderful family. Being so far away from home made the PhD journey difficult at times. However if I had not been so far away, I would have never seen the extents of their endless love and support. Even though I know they never quite understood what my PhD was about, they were my biggest cheerleaders. Thanks to my mum for her daily phone calls that made me feel that there was no distance that a mum could not cross. To my dad for always guiding me and giving me wise advices. To my big brother Luchito, for honouring the title of "big brother" by being the most caring and supportive person and for giving me one of the biggest joys of my life, my beautiful niece Magda. To my little sister Pauzi, for being my confidant and my friend. Thanks to them for visiting me from so far away and giving me the possibility to travel back home every year. It was the fuel I needed to carry on.

I deeply thank my boyfriend Eddie for all his support throughout my PhD. Special thanks for the patience and support during the days in which I was writing my thesis. For picking me up at crazy times at Uni, for cooking the most delicious meals and for standing my bad moods during stressful times. I could not have asked for a better partner through this journey. His love and support made Manchester feel like home. Thanks also to his family, especially Lynne O'Sullivan for taking the time to proofread the earlier drafts of my thesis.

I would also like extend my gratitude to all the members of the Pittman lab for all their help and support. Special thanks to my friend and lab mate Oscar who made my lab days a thrill.

Finally I would like to thank my sponsors. Firstly CONYCIT, the Chilean government institution that founded my PhD. Secondly to my parents, Pauzi and Luchito who have also supported me financially throughout this process.

Chapter 1

General Introduction

1.1 Introduction

Rapid population growth and advances in technology have led the world to depend on the use of energy. This dependence has led to an energy world crisis due to the exhaustive use of the main source of today's energy, fossil fuels. During the past century the most important sources of energy worldwide has been the fossil fuels: petroleum, followed by coal, then natural gas (Ritchie and Roser, 2019). The problem is that this source of energy is non-renewable and while the demand is increasing every day, the available resources are rapidly decreasing, and they will be over eventually. Some studies have predicted that only 50 years of oil reserve are left (Owen *et al.*, 2010). Moreover, the extraction of energy from fossil fuels requires burning them, which releases gases that contribute to the greenhouse effect, with CO_2 being the biggest contributor (US EPA, 2019). Yearly, billions of tonnes of CO_2 are released into the environment from the burning of fossil fuels, turning this into an unhealthy and unsustainable scenario. In front of this daunting situation, the solution has become obvious; an alternative source of energy must become priority. Renewable sources of energy contribute today to the 19.3% of the total world energy consumption, with the rest coming from fossil fuels (REN21., 2018). Although currently low and far from ideal, this proportion of energy derived from renewable sources promises to increase in the upcoming years. Currently, there are many sources of renewable energy such as wind, solar, hydro and bioenergy (Owusu and Asumadu-Sarkodie, 2016). In contrast to fossil fuels, renewable energy sources exist over wide geographical areas and release little or no particles and gases that cause air pollution or negatively impact human health, contributing to the mitigation of climate change (Mohtasham, 2015).

This project is mainly focus in bioenergy as a renewable source, particularly, the optimization of biofuel production. Biofuels are defined as a type of energy produced from biological material or biomass. This biomass can be obtained from a variety of feedstocks including sugar, starch, animal fats, and vegetable oil (Ho *et al.*, 2014). Biofuels are simply a combination of hydrocarbon chains containing no sulphur, or aromatic substances associated with fossil fuels (York, 2008). They are considered a renewable energy source because they are obtained from inexhaustible natural sources (biomass) that can be "renewed" in short periods

of times. Depending on the source of feedstock for production, biofuels can be classified as first, second and third generation biofuels (Lee & Lavoie, 2013).

First generation biofuels, also called "conventional" biofuels correspond to fuels obtained directly from plant biomass. The most widely produced biofuels in this generation are bioethanol and biodiesel (Lee and Lavoie, 2013). Bioethanol is obtained from starch or sugar elements of edible plants such as wheat, corn and sugar cane, while biodiesel is produced from oilseed crops such as from palm oil, rapeseed and soybean (Timilsina and Shrestha, 2011; Rutz and Jannsen, 2007). The technologies to obtain first generation biofuels from the above feedstocks are quite simple, involving either fermenting the sugars into ethanol (Lin and Tanaka, 2006) or breaking up the fatty oils through transesterification into diesel (Balat, Balat and Öz, 2008). The large availability of the feedstock and efficient technologies of extraction, have made the production of first generation biofuels successful. In fact to date, they are the only biofuels produced at commercial industrial scale (York, 2008). The leading producers of biofuels from first generation are United States and Brazil, accounting for 84% of global ethanol production and 26% of biodiesel production between them in 2017 (IEA, 2018). Both ethanol and biodiesel are blended with fossil fuels and used mainly as transportation fuels. However, the use of these types of biofuels has raised some concerns based on the feedstocks used, including the impact it may have on biodiversity and land use and the competition with food crops (Mohr and Raman, 2013).

Second generation biofuels, also called "advanced" biofuels correspond to fuels obtained from non-food crops, agricultural and forest residues (Naik *et al.*, 2010). The main feedstock used for the production of this kind of fuel is lignocellulosic biomass such as crop residues or purpose-grown grasses or woody crops (Lee & Lavoie, 2013). The principal second generation biofuel is the ethanol produced from cellulosic biomass (Robak and Balcerek, 2018), however, biobutanol (Kumar and Arvind, 2017), dimethyl ether (DME) (Wang *et al.*, 2011) and biohydrogen (Cheng *et al.*, 2011) are also emerging as viable biofuels obtained from this same feedstock. These biofuels are usually obtained according to two different approaches, generally referred to as "thermo" or "bio" pathways (Sims *et al.*, 2010). Currently, second generation biofuels are not cost effective

and there are several technical issues that need to be addressed before they can be commercially available (Eisberg, 2006). However, the large availability and abundancy of lignocellulosic plant biomass and the fact that it is a non-food material, makes it a promising source for the production of biofuels.

Third generation biofuels, also called "oilgae" correspond to fuels, which use algal biomass (macroalgae or microalgae) as the main source of feedstock, but can also be derived from other microorganisms such as oil-rich bacteria, cyanobacteria or yeasts (Maity *et al.*, 2014). These photosynthetic algal organisms use the energy from sunlight and CO_2 to produce oxygen and biomass that can be converted into biofuels. The diversity of fuel that can be obtained from microalgae, the non-competition with food crops and the reduced land use, amongst other advantages, have made microalgae an interesting feedstock for biofuel production. Nevertheless, the high costs of cultivation, maintenance and bioprocessing of algae cultures are a barrier that needs to be overcome for algae biofuel to be commercially available (Han *et al.*, 2015).

1.2 Algal biofuels

Algae are mainly aquatic eukaryotic organisms that survive by synthesising their own sugars using sunlight as an energy source through the process of photosynthesis (Odjadjare, Mutanda and Olaniran, 2017). They are the most important photosynthesizing organisms on Earth. They capture more of the sun's energy and produce more oxygen than all plants combined (FAO, 2009). The term "algae" is generally applied to eukaryotic single (microalgae) or multicellular organisms (macroalgae or "seaweed") whose cells contain a distinct membrane-bound nucleus, although some definitions of algae also include prokaryotic cyanobacteria (Day *et al.*,1999). Most algae cells have rigid cell walls composed largely of cellulose and/or other polymers and are organized into separate membrane organelles, including a nucleus, mitochondria and chloroplast. In most algae the primary photosynthetic pigment is chlorophyll but also many contain secondary pigments that are responsible for their colour; carotenoids, which are brown or yellow, and the phycobilins, which are red or blue (Brennan and Owende, 2010). The global number of algal species is estimated to be over a

million species (Guiry, 2012), including red, green and brown algae. Algae biomass has the potential to be used in a variety of applications including biofuels, cosmetics, pharmaceuticals, nutrition and food additives, aquaculture, and pollution prevention (Brennan & Owende, 2010). Of particular interest of this project is the lipids and carbohydrates present in the biomass. The stored oil content of algae (principally as triacylglycerol (TAG)) can be harvested and converted into biodiesel (Chisti, 2007), whilst the algae's stored carbohydrate content (principally starch) can be fermented to produce bioalcohols (Özçimen and Inan, 2015). Microalgae biofuel (third generation biofuels) offers several advantages that make them more appealing when compared to other feedstocks, including:

- Microalgae are easy to cultivate. They are adaptable microorganisms that can grow in a wide range of light and temperature sources as well as grow in areas unsuitable for agricultural purposes. Microalgae are able to grow either in freshwater, saline water or wastewater. Moreover, they are able to use contaminants like NH_4^+, NO_3^-, PO_4^{3-} as nutrient sources, and in turn can be useful as an alternative water treatment method (Mata, Martins, & Caetano, 2010).
- Microalgae can be used as a feedstock to generate several non-toxic, renewable biofuels including biodiesel (Kumar *et al*., 2014); bioalcohols (in particular bioethanol and biobutanol) (Shah and Sen, 2011); photobiologically produced biohydrogen; hydrocarbons as biomethane, bioethane and biopropane; and syngas (Li *et al.*, 2008).
- Microalgae have higher productivity when compared to conventional forestry or agricultural crops. The main reason for this higher productivity is that they can grow extremely rapidly, with some species doubling their biomass within 24 hours. Moreover, in some species oil content can exceed 80% by weight of dry biomass (Miller *et al.*, 2010).
- Cultivation of microalgae requires much less land than other biodiesel feedstocks of agricultural origin. This makes the competition for arable soil with other crops, in particular for human consumption; greatly reduced (Mata, Martins and Caetano, 2010).
- Microalgae are not typically a food crop. Therefore microalgae cultivation does not compete with food consumption (Dragone *et al.*, 2010).

Currently, a variety of species of microalgae have been considered and investigated for biofuel production, including *Nannochloropsis* spp. (Liu, Song and Qiu, 2017), *Dunaliella* spp. (Mehdi, Sohi and Eghdami, 2014), *Chlorella* spp. (Guccione *et al.*, 2014), diatoms (e.g. *Phaeodactylum* spp.), (d'Ippolito *et al.*, 2015) and *Chlamydomonas reinhardtii* (Scranton *et al.*, 2015). However, microalgae biofuels are still not commercially available. Despite all the advantages discussed above, several barriers need to be overcome in order for microalgae biofuel to become an economically viable alternative. The main challenges are presented below:

- Yields: The success for biofuel production using microalgae depends on content of carbohydrates and lipids in microalgae biomass. Several studies have reported that the composition of biomass depends on environmental conditions (Ji *et al.*, 2011; Markou and Georgakakis, 2011; Ho, Chen and Chang, 2012). Microalgae have demonstrated the ability to increase carbohydrate and lipid content when exposed to different types of environmental stress (e.g., high light stress, temperature stress, nutrient starvation stress, and salt stress) (Zhu, Li and Hiltunen, 2016a). Unfortunately, these conditions have a negative impact on microalgae growth, affecting the yields of biomass (Tan and Lee, 2016). Therefore, improving lipid and carbohydrate content without affecting the growth is the key for achieving high yields of biofuel production. A strategy to overcome this issue is to use a two-stage cultivation. In the first stage microalgae are cultivated under ideal conditions that promote its growth, and a second stage is dedicated to increase the lipid and carbohydrate content by exposing them to a physiological stress (Rodolfi, Chini Zittelli, Bassi, Padovani, Biondi, Bonini and Mario R Tredici, 2009; Dragone *et al.*, 2011). However, the available technologies to transfer microalgae from one phase to another are not cost-effective. Further experimentation is required to find ways to achieve high lipid and carbohydrate productivity without compromising growth. This would ultimately have a significant impact on the overall production cost of microalgae biofuels.
- Cultivation: Cultivation of microalgae at large scale can be done either in open or closed systems. Open systems are usually ponds or tanks that are exposed to the environment. The simplicity and low costs of the set up and maintenance of open ponds have made them a cost effective alternative and has attracted the

attention of several companies (Costa and de Morais, 2014). However, open ponds are susceptible to contaminants that can restrict the commercial production of microalgae. Moreover, the inefficient stirring mechanisms and low mass transfer rates leads to lower productivity (Ling *et al.*, 2009). On the other hands, closed cultivation systems are mainly conducted in photobioreactors. Currently, there are many types of photobioreactors (eg. tubular, vertical, flat-plate, annular, fermenter-type). All of them provide microalgae with a controlled environment that reduces the risk of contamination, CO_2 losses and require smaller areas, resulting in higher productivities compared to open pond cultivation (Acién *et al.*, 2017). However, the use of photobioreactors increases the production costs; the illumination, the CO_2 feed and the circulator systems are among the additional cost included in the use of these systems (Norsker *et al.*, 2011). Another problem associated with algae cultivation is the high costs of cultivation medium nutrient feed. Microalgae require nitrogen, phosphorus and other micronutrients for optimal growth (Markou, Angelidaki and Georgakakis, 2012). The fertilizer costs together with freshwater contribute up to 50% of the total biomass production impacting directly in the economic viability of algae biofuels (Singh and Das, 2014). The use of wastewater for the growth of microalgae has become an attractive alternative for producing algae biomass from low-cost waste based nutrient media (Posadas *et al.*, 2017). Recycling of the nutrients from wastewater can help to reduce the water and fertilizer costs of microalgae cultivation while reducing environmental footprints and improving water quality.

- Production process: The harvesting of the algal biomass, extraction and purification of lipids and carbohydrates for the conversion to biofuels also present challenges. First, as microalgae are small microorganisms suspended in water, the harvesting process is a challenging task. Algae biomass constitutes only about 0.1% of the total culture volume in open ponds and can increase when using photobioreactors to 1% (Marrone *et al.*, 2018). Thus, the collection of very small fractions of algae from large volumes of water can require high energy inputs increasing the production costs. Currently, there are several methods used to harvest microalgae; (1) separation techniques such as spontaneous settling of the cells, sedimentation, flocculation, and floatation and (2) concentration techniques such as centrifugation and filtration (Al hattab, Ghaly and Hammoud, 2015).

However, while effective, some of these techniques are not suitable for large scale production and have high costs, are energy intensive, and can require significant maintenance. In reality a combination of harvesting techniques is used, however the efficiency of them tend to be strain dependent and further research is required in this area (Milledge and Heaven, 2013). After the harvesting, another key part of the process that has presented challenges is the lipid extraction. The disruption of microalgae cell walls to obtain the intracellular lipids and the separation of them from the rest of the biomass can present significant barriers that can impact the overall production costs. Drying the microalgae biomass have proven effective for the extraction of lipids, however it is extremely energy intensive. Hence, technologies using wet biomass lipid extraction are preferable (Lee *et al.*, 2017). Solvent extraction does not require the algal biomass to be fully dried and this method can successfully extract oil from intact cells of several species. However, there are environmental concerns about the use of toxic solvents (Cooney, Young and Nagle, 2009). Other approaches that have been developed include use of supercritical CO_2, enzymatic extraction, and ultrasonic cell disruption, among others (Kapoore *et al.*, 2018). Again, these methods can be very costly and require skilled labour. Finally, after harvesting and isolation of the oils and carbohydrates, each is converted to biofuel forms like biodiesel, biogasoline, ethanol, butanol, and biohydrogen. Due to the similarities with petroleum, the challenges of algae oil conversion to biofuels are similar than those associated by petroleum companies. However, the high protein content of algae biomass and the limitation of conversion of the current technologies, biofuel from microalgae required further catalytic removal of atoms such as oxygen, nitrogen, and sulphur (Yang *et al.*, 2016).

In summary, microalgae have the potential to become an economically viable feedstock for biofuel production, although improvements in the overall production process are required for this technology to become a sustainable and cost effective biofuel alternative.

1.3 Metabolic pathways leading to biofuel production in microalgae

Microalgae biomass is rich in lipids, carbohydrates and proteins. Studies have reported that algae can contain around 20-80% of lipids, 10-40% of carbohydrates

and 10–50% of proteins on a dry matter basis (% dry weight). These percentages vary depending on the microalgae species, cultivation method and conditions (El-Dalatony *et al.*, 2017). As above mentioned, carbohydrates from algae biomass can be converted into bioethanol and lipids into biodiesel through processes of fermentation and transesterification, respectively. Understanding the basic characteristics of carbohydrate and lipid metabolism in microalgae is necessary for developing more effective approaches to increase the biofuel productivity.

Lipid biosynthesis pathway

Lipid biosynthesis in microalgae uses the energy provided by photosynthesis and CO_2 to generate glycerate-3-phosphate (G3P), which is the precursor of several storage molecules. G3P is converted to pyruvate and later to acetyl-CoA by the enzymatic complex pyruvate dehydrogenase (PDC). Acetyl-CoA is the initiator molecule from which fatty acid (FA) synthesis starts. Acetyl-CoA is carboxylated to form malonyl-CoA, a reaction that is catalysed by the acetyl-CoA carboxylase enzyme (ACCase) located in the plastid or in the cytosol of the microalgae. Within the plastid, malonyl-CoA is transfered to the acyl carrier protein (ACP) that is part of the fatty acid synthase complex (FAS). Then, the 3-ketoacyl-CBP synthase (KAS) catalyses the condensation of malonyl-ACP with an acetyl group to form ketobutyryl-ACP, which is then converted to butyryl-ACP through a series of reactions of reduction-dehydration-reduction. The butyryl-ACP is processed in the same way as malonyl-ACP, and after several repetitions of the cycle, 16C palmitoyl-ACP is generated, which is then condensed with 2 carbon atoms from acetyl-CoA to produce the 18 carbon FA stearoyl-ACP. This acyl-ACP generated in the plastid could be used for the synthesis of membrane structural lipids of the organelle or transferred to monoacylglycerol-3-phosphate to generate diacylglycerol (DAG). These transformations prevent this molecule from leaving the plastid and serve as a substrate for TAG. Acyl groups that do not stay in the plastid fulfil structural functions and are converted into free saturated or monounsaturated FA (Baba and Shiraiw, 2013). Finally, free FA can be saturated by the fatty acid desaturase enzymes, to generate polyunsaturated FA which enter the endoplasmic reticulum (ER) as acyl-CoA where they can be modified into

different types of lipids, such as phosphatidylcholine, phosphatidylethanolamine or glycerolipids, such as TAG (Liu and Benning, 2013).

When acyl-CoA enters the ER they can follow the Kennedy pathway to synthesize TAG or be processed by different acyltransferases. The Kennedy pathway corresponds to a series of successive reactions beginning with the addition of an acyl group from an acyl-CoA to a glycerol 3-phosphate in the position 1, generating a lysophosphatidic acid (LPA) (Cagliari *et al.*, 2011). After this reactions, an acyl group can be added to the position 2 of the LPA to generate phosphatidic acid (PA), which is then phosphorylated to generate DAG. The last step, considered the limiting step for the synthesis of TAG is the addition of a third acyl group to LPA by the acyl-CoA:diacylglycerol acyltransferase (DGAT) enzyme, specific for this pathway. TAGs will be finally stored in vesicles and are considered the main energy storage molecules in microalgae (Sharma, Schuhmann and Schenk, 2012a).

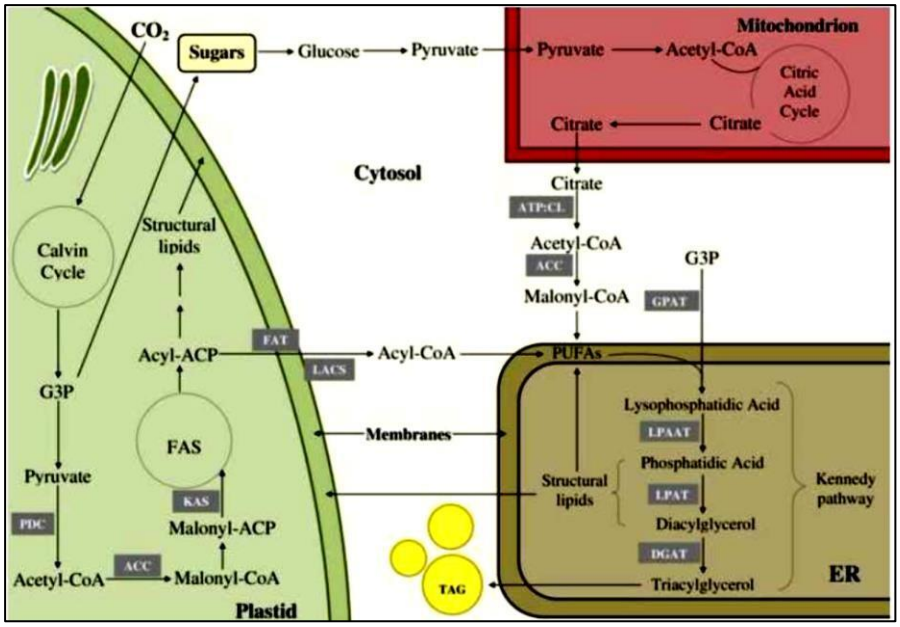

Figure 1.1: A simplified scheme of lipid synthesis in microalgae (Taken from Bellou *et al.*, 2014).

1.3.2 Carbohydrate biosynthesis pathway

Carbohydrates are common energy and carbon storage products in algae, mostly in the form of starch-type polysaccharides. It is stored in the plastids of microalgae and is considered one of the principal sources of chemical energy (Raven and Beardall, 2003). Starch is a homopolysaccharide of glucose units joined by glycosidic bonds forming linear (amylose) and ramified structures (amylopectin) (Berg *et al.*, 2002). Starch biosynthesis in microalgae occurs in the chloroplast by the action of several enzymes with various isoforms in each step of the pathway. The synthesis is initiated by fructose-6-phosphate, a product of photosynthetic carbon fixation via de Calvin cycle. Phosphoglucose isomerase 1 (PGI) catalyses the conversion of fructose-6-phosphate into glucose-6-phosphate and then into glucose-1-phosphate by the action of the phosphoglucoisomerase enzyme (PGM) (Fig. 1.2). The rate limiting step of the starch biosynthesis is considered to be the conversion to ADP-glucose, by the reaction of glucose 1-phosphate with ATP, catalysed by the ADP-glucose pyrophosphorylase (AGPase). Following the production of ADP-glucose, the starch synthesis involves reactions catalysed by at least three different enzymes; starch synthases (SS), starch branching (BE) enzymes and starch debranching enzymes (DBE) (Fig. 1.2). However, the exact steps involved in the synthesis of the different types of starch may vary within different species (Delatte *et al.*, 2005).

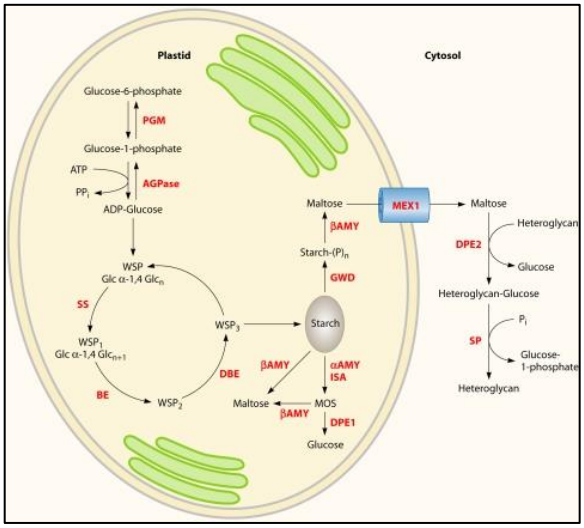

Figure 1.2: Starch biosynthesis pathway in microalgae. Metabolites formed in each steps are shown in black and enzymes in red. PGM, Phosphoglucoisomeras; AGPase, ADP-glucose pyrophosphorylase; SS, Starch synthases; BE, branching enzymes; DBE, debranching enzymes; AMY, Amylases; DPE, disproportionating enzymes; GWD, Glucan-water dikinases (Taken from Radakovits *et al.*, 2010).

1.4 Environmental factors leading to lipid and starch accumulation in microalgae

Microalgae metabolic pathways have been shown to be highly impacted by environmental factors such as light intensity, salinity, temperature and nutrient availability (Cheng and He, 2014). In most of the cases, the adjustment in the metabolic pathways in response to stress leads to the accumulation of carbohydrates and lipids (Khan, Shin and Kim, 2018). This has proven to be a very useful strategy to study the pathways involved in the accumulation of these storage molecules. Nutrient availability is considered to be an easy and affordable technique to modify the carbon storage metabolism in microalgae (Dragone *et al.*, 2011). Controlling the nutrients in the culture media is relatively easy. Therefore, several studies have used this approach to have a better understanding of the mechanisms leading to the accumulation of lipids and carbohydrates. Nutritional requirements can vary within microalgae species; nitrogen, and phosphorus are the major macronutrients required by all microalgae.

1.4.1 Phosphorus stress

Phosphorus (P) is the 11^{th} most abundant element on earth and it is usually found in the form of phosphate. Due to its essential role in biological processes, P is one of the most dispersed elements in nature. It is a key component of cell tissues of plants and animals and is necessary for the structure, growth and propagation of all living organisms. In microalgae it accounts for 1-3% of their dry weight and it is assimilated as inorganic orthophosphate (PO_4) (Grossman and Aksoy, 2015). However, even though P is an abundant element on earth, it has low availability in most environments, and it is not unusual for algae to become limited from this essential nutrient. This scenario had led microorganism such as microalgae to adapt quickly and develop mechanism to cope with the lack of available P. In P rich conditions, microalgae store excess phosphorus within the cells in the form of polyphosphate granules, so they can prolong their growth in periods of P limitation. In P limited conditions, as in other nutrient stress conditions, the cells adjust their cellular metabolism and reduce cell division and growth to survive. In addition to this general response shared with other stress responses, microalgae cells have developed specific responses to P deficiency, which are mainly focussed on the

scavenging and recycling of P (Kamalanathan and Pierangelini, 2015; Mühlroth *et al.*, 2017).

Scavenging mechanisms maximize the capture of P through the synthesis and secretion of phosphatases, and the up-regulation high affinity PO_4 transporters. For example, the secreted alkaline phosphatase encoded by the *PHOX* gene in *C. reinhardtii*, is substantially up-regulated under P-depleted conditions (Bajhaiya *et al.*, 2016b). Also in *C. reinhardtii* under P depletion PO_4 uptake is driven through up-regulated high affinity P transporters including members of the PTB gene family that exhibit homology to animal like Na^+/Pi transporters. For example, *C. reinhardtii* has reported a 20-fold increase in the expression of the PTB2 P high affinity channel under P depleted conditions (Moseley, Chang and Grossman, 2006). When P becomes limited, phospholipids are replaced by non-phosphorus membrane lipids, such as sulfoquinovosyldiacyl-glycerol (SQDG), by either breaking down the existing phospholipid and/or by de novo synthesis of non-phosphorus lipids. A substantial response to P starvation is through the increased in carbon storage metabolism, initially an increase in starch biosynthesis and storage and then subsequently and increase in TAG biosynthesis and storage. For example, in *C. reinhardtii* P starvation was reported to yield over a 10-fold increase in starch accumulation and an a approximate 5-fold increase in TAG accumulation, after 7 days growth during P limited conditions (Bajhaiya *et al.*, 2016b).

1.4.2 Nitrogen stress

Nitrogen (N) is another essential element for various bioprocesses in microalgae, including for protein and chlorophyll synthesis. The decrease in the synthesis of key proteins from photosystem I and II (PSI and PSII) highly affects the photosynthetic rates in cells under N starvation. In addition, there is a decrease in the synthesis of pigments such as chlorophyll and carotenoids (Schmollinger *et al.*, 2014; Kamalanathan and Pierangelini, 2015). Under N starvation, microalgae direct the carbon flux towards the synthesis of storage molecules such as lipids and carbohydrates, leading to an intracellular accumulation of these molecules (Sharma, Schuhmann and Schenk, 2012). The increase in lipid content is the most significant effect on the metabolism of N starved cells, although studies have also reported that it causes carbohydrates accumulation (Odjadjare, Mutanda and

Olaniran, 2017; Rehman and Anal, 2019). The level of increase in lipid metabolism due to N depletion depends on the microalgae species. For example, there was a slight increase observed in *Phaeodactylum tricornutum* in response to N starvation in which TAG content increased from 69 to 75% (Alonso *et al.*, 2000), while highly significant increases in TAG have been observed in species such as *Chlorella vulgaris*, which increases TAG content up to in response to N starvation by 40% (Rodolfi *et al*., 2009).

1.4.3 Salinity stress

Salinity is another factor affecting carbon metabolism in microalgae. Algae cells increase the synthesis of storage molecules such as carbohydrates and lipids to protect them under osmotic stress. Low molecular weight carbohydrate such as trehalose and sucrose have reported to increase in response to salt stress (Shamala *et al.*, 2006; Markou, Angelidaki and Georgakakis, 2012). Similarly, lipids have been reported to increase in several microalgae species in response to salt stress, in particular saturated fatty acids, which are of great interest to the biodiesel industry (Chen, Jiang and Chen, 2008; Pandit, Fulekar and Karuna, 2017). Species like *Dunaliella* have been used as a model due to the fact that they can grow with concentrations of NaCl near to saturation point. Increasing the NaCl concentration from 0.5 M to 1 M produced an increase in lipid content from 60% to 67% in *Dunaliella tertiolecta* (Takagi, Karseno and Yoshida, 2006). In freshwater species such as *C. reinhardtii*, salt stress has also been observed to increase TAG yield (Siaut *et al.*, 2011). However, although salt stress increases the accumulation of carbon storage molecules, it hinders their growth leading to poor biofuel production yields. A recent transcriptomic analysis of cells under salt stress may present potential solutions to this issue (Wang *et al.*, 2018). The identification of key genes involved in salt stress response opens the door to the development of genetic-engineering based solutions that could improve biofuel yields.

1.5 Strategies for lipids and starch production improvement in microalgae

1.5.1 Metabolic engineering

One approach to increase TAG biosynthesis in microalgae has been the overexpression of enzymes directly involved in FA and TAG biosynthesis such as ACCase and DGAT. Previous studies in bacteria, higher plants and fungi have proven the feasibility of this approach (Courchesne *et al.*, 2009). However, even though the overexpression of these target enzymes has been successful, in many cases, no considerable increase in lipid production was detected. For example, ACCase was isolated (Roessler, 1990) and successfully overexpressed in the diatoms *C. cryptica* and *N. saprophila*, but as in the other organisms, no significant increase of lipid accumulation was observed (Dunahay, Jaruis and Roessler, 1995). This result seems to be true for most species because no increase in lipid accumulation was reported even though significant increase of the ACCase and/or intermediate products such as fatty acids was commonly observed. So far, results from overexpression of ACCase suggest that probably the committing step catalysed by this enzyme is not the rate-limiting step in the fatty acid synthesis pathway. On the other hand, the overexpression of DGAT in different species increased TAG accumulation as well as lipid droplet formation (Jako et al., 2014). This result could be explained by the fact that DAG, the substrate of DGAT, could be allocated for either phospholipid biosynthesis or TAG formation. Probably the overexpression of DGAT would direct more DAG to TAG formation rather than phospholipid formation (Courchesne *et al.*, 2009). In fact, studies in plants overexpressing DGAT have revealed that there is an increase in TAG synthesis as well as from fatty acid (Thelen and Ohlrogge, 2002). However, despite all the studies pointing at DGAT as an important rate-limiting step in lipid biosynthesis, there have been mixed successes of overexpression of this enzyme in microalgae.

1.5.2 Transcription Factor engineering

Another approach that has emerged recently is Transcription Factor Engineering (TFE), in which a metabolic pathway is studied in the context of the whole organism (Fig. 1.3). This approach aims to use the down or upregulation of

transcription factors (TFs) responsible for regulating specific metabolic pathways involved in the accumulation of a target such as lipids for their overproduction (Bajhaiya, Ziehe and Pittman, 2017). TFs are proteins that are able to interact either directly or indirectly with the DNA and control the rate of transcription of a particular group of genes. The control exerted by TFs could result in enhancement, or repression in the production of particular metabolites (Grotewold, 2008). The advantage of this new approach is that TFE unlike the other approaches does not focus only on one target gene but on a set of genes involving one or several metabolic pathways. By having this broader control of a particular pathway, TFE can potentially overcome the limitations of the other approaches.

Previous experiments manipulating transcription factors to enhance the production of a particular metabolite have been carried out, focusing on products such as therapeutic proteins in mammalian cells (Reik *et al.*, 2007), flavonoids in *Arabidopsis thaliana* (Sharma, Schuhmann, & Schenk, 2012), alkaloids for pharmaceutical products in plant (Endt, Kijne, & Memelink, 2002), among others. All of these studies, in which a particular TF was overexpressed, resulted in enhancement of the production of the desire metabolites.

It is important to note that for engineering a particular TF, it is crucial to know in which particular pathway is involved, in order to obtain the desired outcomes. Thus, for the purpose of increasing lipid production it is necessary to know which are the TFs involved in lipid metabolism. To date, there is not much information about TFs regulating lipid metabolism of microalgae. About 147 putative TFs and 87 putative transcription regulators (TRs) have been identified in *C. reinhardtii* (Thiriet-Rupert *et al.*, 2016). However, the biological function of only a small number of them has been determined (Riaño-Pachón *et al.*, 2008). The identification of TFs involved in lipid metabolism in other organisms can give us insights into the potential of TFE for the enhancement of lipid production. For example, sterol regulatory element binding protein (SREBP) has been established as the master regulators of lipid homeostasis in mammals (Eberlé *et al.*, 2004). In plants, studies reported that SebHLH protein, a member of the bHLH family of TFs, might play a key role in the transcriptional regulation of genes related to lipid biosynthesis during seed development (Kamisaka *et al.*, 2007). Another important finding was the soybean Dof-type type (DNA binding with one finger) TF gene. It

was reported that this gene was correlated with the lipid content in soybean seeds. Particularly GmDof4 and GmDof11, two members of the family were found to increase the content of total fatty acids and lipids of Arabidopsis seeds by upregulating genes that are associated with the biosynthesis of fatty acids (Wang et al., 2007). Moreover, Dof-type TF family sequences were also identified in a variety of representative organisms from different taxonomic groups: the unicellular green alga *C. reinhardtii*, the moss *Physcomitrella patens*, the club moss *Selaginella moellendorffii*, the gymnosperm *Pinus taeda*, the dicotyledoneous *A. thaliana* and the monocotyledoneous angiosperms *Oryza sativa* and *Hordeum vulgare* (Moreno-Risueno *et al.*, 2007). Another study worth mentioning is the overexpression of WRINKLED1 (WRI1) in *A. thaliana*. This transcription factor, previously related with the regulation of seed oil biosynthesis, was overexpressed together with the downregulation of APS1, a gene known to encode ADP-glucose pyrophosphorylase that participates in starch biosynthesis. Results of this study showed a production of 5.8 times more oil in vegetative tissues compared with plants with either WRI1 overexpression or APS1 alone (Sanjaya *et al.*, 2011). This study confirmed the feasibility of TFE, however, there is still a long way to go in microalgae. There is an urgent need of identifying key TFs involved in the regulation of lipid metabolism in order to successfully develop TFE. Fortunately, there are an important number of available techniques to identify and manipulate TFs. For characterization and determination of TFs DNA specificities, the technologies available includes electrophoretic mobility shift assay (EMSA) and the DNase I protection assay (Hattori *et al.*, 2007). Recently, other more advanced techniques have been developed that creates global maps of specific protein-DNA interactions in cells such as microarray (ChIP-chip) or by sequencing (ChIP-seq) (Wederell *et al.*, 2008; Martin, Vining and Dombrowski, 2018; Scott *et al.*, 2018).

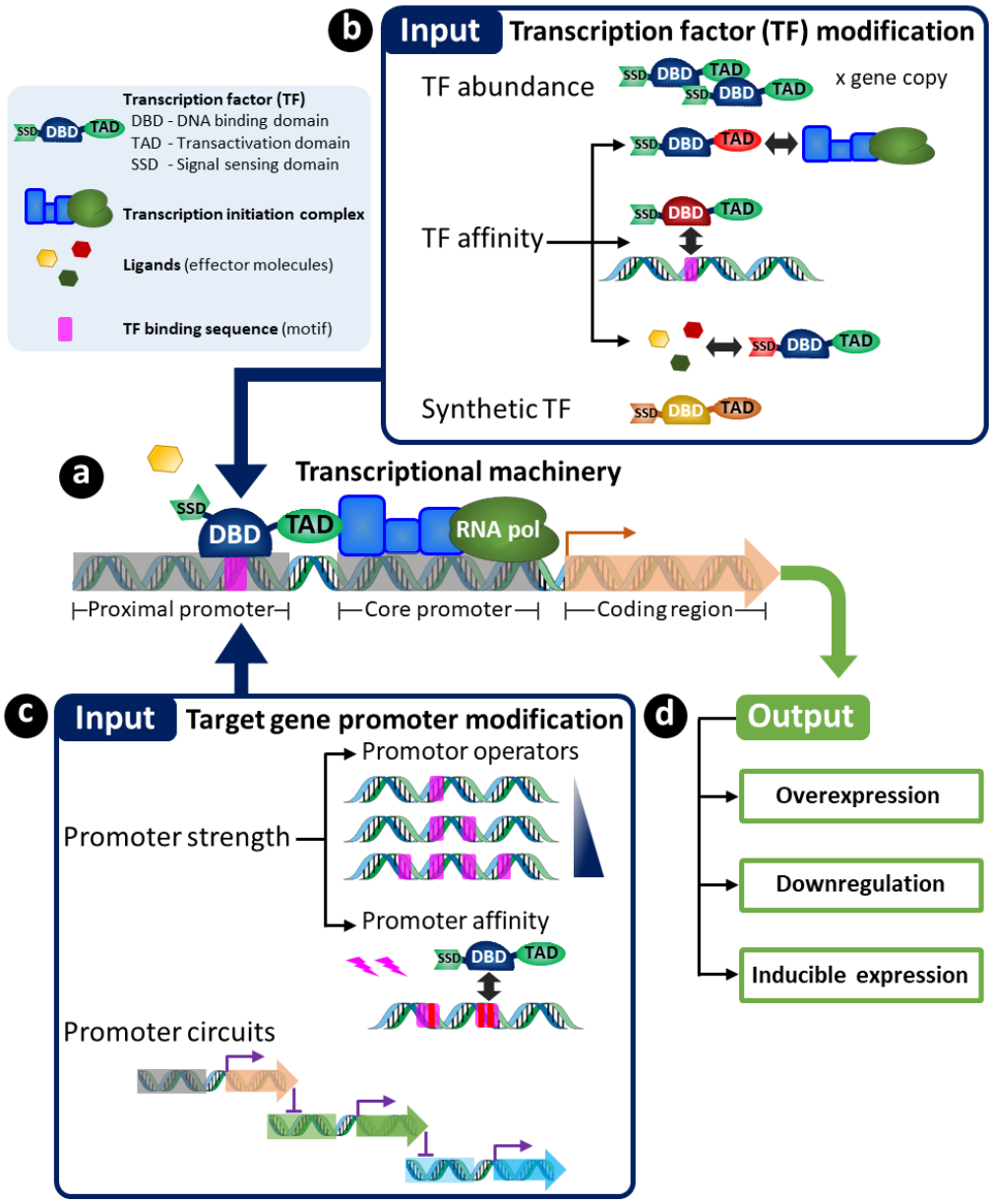

Figure 1.3: Overview of Transcriptional factor engineering (TFE). In eukaryotes, transcription factors (TFs) control the gene expression by binding at the DNA level to a specific sequences (motif) within the gene promoters and by promoting (as an activator), or blocking (as a repressor) the recruitment of proteins complexes in the core machinery necessary to start transcription (including RNA polymerase) (A). TFE comprises the manipulation of TFs function through direct modification of the TF structure altering its affinity to interact either with the DNA (through modification of the DBD), with the protein complexes of the core machinery (through modification of its TAD) or by changing its affinity with ligand molecules (through modification of its SSD). Alternatively, synthetic biology has allowed the design of TFs with desire affinities and specificities with gene promotors (B). Modifications on the sequence of gene promoters will also affect TFs function by altering its binding affinity with promotor specific motifs, or by increasing the TFs recruitment with the insertion of motifs which finally will affect the promoter strength (C). Altogether, these approaches of TFE aims on generating tunable outcomes in which genes can be either overexpress, downregulated or induced (D).

1.6 Aims and Objectives

The overall aim of this project is to understand aspects of the regulatory mechanisms of carbon storage metabolism in microalgae. The understanding of these regulatory mechanisms could open new alternatives for improving the production of carbon storage molecules useful for the biofuel industry (Courchesne *et al.*, 2009). The main focus of this investigation was directed towards the study of transcription factors (TFs) that could potentially be involved in the up-regulation of lipid and starch metabolism in *C. reinhardtii*. Knowing the regulators of these processes provides new targets for genetic engineering. It is though that by manipulating TFs that are important in the regulation of these pathways, the carbon flux could be directed towards the production of carbon storage molecules (Bajhaiya, Ziehe and Pittman, 2017). In contrast to targeting individual enzymes, the manipulation of TFs can give a more global control to metabolic pathways, bringing stronger outcomes.

To accomplish this overall aim, the first objective of this investigation (addressed in Chapter 3) was to study the transcriptional targets of the PSR1 TF. This TF has previously been linked with the regulation of the metabolism of carbon storage molecules, both starch and TAG, however, the mechanisms underlying this regulation are unknown (Bajhaiya *et al.*, 2016b). Using electrophoretic mobility assay (EMSA) I tested the ability of PSR1 to bind to the promoter regions of selected genes involved in starch and TAG metabolism. The results of this study showed that PSR1 is able to bind specifically to the promoter regions of these genes. This is the first evidence of the TF binding activity of PSR1 and the potential mechanism through which it regulates its downstream targets involved in carbon storage metabolism.

The second objective (addressed in Chapter 4) was focus on the study of the MYB TF family. Member of this TF family in higher plants and microalgae have been found upregulated in response to different environmental stresses, suggesting a potential pivotal role in the control of stress response in *C. reinhardtii* (Schmollinger *et al.*, 2014a; Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b). As a first step to determine if stress response could be linked with the regulation of stress-induced carbon storage metabolism, bioinformatic gene identification and a phylogenetic analysis was performed. By analysing the phylogenetic relationships between MYB proteins in *C.*

reinhardtii and the well characterized MYB TF family in *A. thaliana,* I was able to infer functionality. Interestingly, MYB TFs from *C. reinhardtii* showed high homology with members of the *A. thaliana* TF family clade that have been involved in different responses to stress. This is in accordance with transcriptomic analysis showing upregulation of these TFs under these conditions. These analyses gave us more confident of the involvement of MYB TF in responses to stress, making them good targets to study their potential links with the metabolism of carbon storage molecules.

The third objective (addressed in Chapter 4 and 5) was to functionally characterize a member of the MYB TF family. MYB2, a member of the MYB TF in *C. reinhardtii*, was found to be highly upregulated under nitrogen (N) and phosphorus (P) starvation, and was chosen for its potential involvement in the carbon storage metabolism. MYB2 overexpression and knockdown lines were generated and characterized for their lipid and starch content. Surprisingly, results showed that both lines presented similar patterns of carbon storage accumulation under stress. Compared to controls, both lines increase more than 2 fold their lipid content in response to low N and both showed a moderate increase in the lipid content under low P and salt stress. These results evidenced for the first time the involvement of a MYB TF in the regulation of carbon storage metabolism in microalgae. However, its role appears to be more complex, and potential redundant roles are hypothesised for this TF.

Chapter 2

Materials and methods

2.1 Materials

2.1.1 Chemicals and reagents

Ampicillin	Sigma Aldrich
Bradford reagent	Bio-Rad
Bovine serum albumin (BSA)	Sigma Aldrich
CSPD, ready-to-use	Sigma Aldrich
Digoxigenin wash and block buffer set	Roche
Dimethyl sulfoxide (DMSO)	Sigma Aldrich
DNA ladder	Bioline
dNTP mix	Thermo Fisher Scientific
Dithiothreitol (DTT)	Sigma Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Sigma Aldrich
Ethanol	Thermo Fisher Scientific
Gel loading dye, purple (6X)	New England Biolabs
HEPES	Thermo Fisher Scientific
Hexane	Sigma Aldrich
Imidazole	Sigma Aldrich
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Sigma Aldrich
Kanamycin	Thermo Fisher Scientific
Lysogeny broth (LB broth)	Thermo Fisher Scientific
NaCl	Thermo Fisher Scientific
NEB buffer 4	New England Biolabs
Nile Red	Sigma Aldrich
Protease inhibitor cocktail	Sigma Aldrich
Tri reagent	Sigma Aldrich
Tris-base	Thermo Fisher Scientific
Safeview	NBS Biologicals

2.1.2 Enzymes

Restriction enzymes	New England Biolabs
My Taq Red mix	Bioline
Superscript III	Invitrogen

Kapa HIFI	Kapa Biosystems
DNase I	Invitrogen
T4 DNA ligase	Roche
Alkaline phosphatase	Roche

2.1.3 Antibodies

Anti-digoxigenin-AP, Fab fragments	Roche
6x-HIS tag monoclonal	Thermo Fisher Scientific
Anti-mouse HRP	Thermo Fisher Scientific

2.1.4 Solutions

Chlamydomonas culture media

Hutner's trace elements: 50 g EDTA disodium salt in 250 mL, 22 g $ZnSO_4$. $7H_2O$ in 100 mL, 11.4 g H_3BO_3 in 200 mL, 5.06 g $MnCl_2$. $4H_2O$ in 50 mL, 1.61 g $CoCl_2$. $6H_2O$ in 50 mL, 1.57 g $CuSO_4$. $5H_2O$ in 50 mL, 1.10 g $(NH_4)6Mo_7O_{24}$. $4H_2O$ in 50 mL, 4.99 g $FeSO_4$. $7H_2O$ in 50 mL and making up to 1 L volume (Hutner, 1950).

TAP salt solution: 15 g l^{-1} NH_4Cl, 4 g L^{-1} $MgSO_4$. $7H_2O$, and 2 g L^{-1} $CaCl_2$. $2H_2O$.

Potassium phosphate solution: 288 g L^{-1} K_2HPO_4 and 144 g L^{-1} KH_2PO_4. pH was adjusted to 7.0

Standard TAP: One litre of standard TAP medium was prepared by adding 25 mL of TAP salt solution, 0.375 mL of potassium phosphate solution, 1 mL of Hutner's trace elements mixture and 1 mL of glacial acetic acid (to final concentration of 18 mM), buffered to pH 7.2 with 20 mM Tris and made to 1L with de-ionized water. Solid medium was prepared by the addition of 1.5 % (w/v) agar. Modified TAP medium was prepared with reduced total nitrogen (N) or total phosphorus (P) by modifying the TAP salt solution with a reduced concentration of NH_4Cl to give 0.7 mM ammonium in low N TAP medium, or modifying the potassium phosphate solution to give 0.01 mM phosphate in low P TAP medium. In the low P TAP medium, KCl was added to maintain the potassium concentration.

Buffers for competent bacteria preparation

TFB1: 30 mM CH_3CO_2K, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 100 mM $RbCl_2$ and 15% glycerol. pH was adjusted to 5.8.

TFB2: 10 mM MOPS, 75 mM $CaCl_2$, 10 mM RbCl and 15% glycerol. pH was adjusted to 6.5.

Buffers for EMSA

HEDG buffer: 25 mM HEPES, 2 mM EDTA and 10% glycerol.

Added fresh: 1 mM DTT

HEDGK buffer: 25 mM HEPES, 2 mM EDTA, 10% glycerol and 50 mM KCl

Added fresh: 1 mM DTT

Tris/Borate/EDTA (TBE) buffer: For 10x, 108 g of Tris and 55 g of Boric acid were dissolved in 900 ml of ddH_2O. Then 40 ml of 0.5 M Na_2EDTA (pH 8.0) were added. Volume was adjusted t 1 L.

Buffers for protein expression

Lysis buffer: 25 mM Tris-HCl and 100 Mm NaCl. pH was adjusted to 8.

Added fresh: Protease inhibitor cocktail

Buffers for protein purification

Binding buffer: 20 mM sodium phosphate, 50 mM imidazole and 0.5 M NaCl. pH was adjusted to 7.4.

Elution Buffer: 20 mM sodium phosphate, 100-500 mM imidazole and 0.5 M NaCl. pH was adjusted to 7.4.

Buffers for genomic DNA extraction

Cetyl trimethyl ammonium bromide (CTAB) buffer: 2 % (w/v) CTAB, 100 mM Tris-HCl (pH 8), 1.4 M NaCl, 20 mM EDTA (pH 8) and 2 % (v/v) beta-mercaptoethanol.

2.1.5 Microalgae strains

Chlamydomonas reinhardtii strains CC-125 (CCAP11/32C) (mt+, *nit1*, *nit2*), referred to throughout as wild type, and *cw15* (CCAP11/32CW15+) (mt+, *cw15*, *nit1*, *nit2*, *ARG7*) were obtained from the UK Culture Collection of Algae and Protozoa (CCAP), Oban, Scotland, UK, while strains CC-4350 (mt+, *cw15*, *nit1*, *nit2*, *arg7-8*) and CC-4351 (mt+, *cw15*, *arg7-8*) were obtained from the Chlamydomonas Resource Center, at the University of Minnesota, USA.

The *MYB2* knockdown lines (MYB-KD) used in this study were previously generated in the Pittman lab (Tan, Bajhaiya and Pittman, unpublished) by constructing a *MYB2-pChlamiRNA3* plasmid containing a gene-specific target amiRNA sequence against *MYB2*, using the method as described by Molnar et al. (2009), then by transforming the cell wall deficient *cw15* strain with *pChlamiRNA3* plasmid (control strain) or *MYB2-pChlamiRNA3* plasmid using the glass bead method (Kindle, 1990) and selected on TAP-agar medium containing 10 μg mL^{-1} paromomycin.

2.1.6 Primers and Oligonucleotides

Table 2.1 Oligonucleotides for EMSA

Name	Oligonucleotide (5'-3')
PTB2 forward	TAATGATGTAGGATATACCTTAACAGCATGCA
PTB2 reverse	TGCTGTTAAGGTATATCCTACATCATTA
PTB2 mutant forward	TAATGATGTAGGCAATACCTTAACAGCATGCA
PTB2 mutant reverse	TGCTGTTAAGGTATTGCCTACATCATTA
SSS1 forward	ATGGCAAAGCATATTCACTGCCACGGGTGCA
SSS1 reverse	CCCGTGGCAGTGAATATGCTTTGCCAT
SSS1 mutant forward	ATGGCAAAGCTAACTCACTGCCACGGGTGCA
SSS1 mutant reverse	CCCGTGGCAGTGAGTTAGCTTTGCCAT
SSS5 forward	ACGACCACTCGGCATATACATTCTAGGTGCA
SSS5 reverse	CCTAGAATGTATATGCCGAGTGGTCGT
SP2 forward	CCACGCCCGGGAATATGCCCTTCATGCA
SP2 reverse	TGAAGGGCATATTCCCGGGCGTGG
GPD4 forward	TCTGCAGGGTATATGCCGTACCAGATATTGCA
GPD4 reverse	ATATCTGGTACGGCATATACCCTGCAGA
GPD4 mutant forward	TCTGCAGGGTGCATGCCGTACCAGATATTGCA
GPD4 mutant reverse	ATATCTGGTACGGCATGCACCCTGCAGA

Table 2.2 Primers for expression analysis by RT-PCR

Name	Primer (5'-3')
MYB2 forward	TCGTTGTGCGTAGCTTTTTG
MYB2 reverse	ACATACGAAACTCGGCATCC
MYB1 forward	GCTGATCTGGAACGGGCTGGG
MYB1 reverse	CTGGCTGTTTACTGCGGCACTGC
rBCL forward	ATGTCACCACAAACAGAGACTAAAGC
rBCL reverse	AGTCCACCGCGTAGACATTCAT
Actin forward	CAGTAGGAGGCATAGGGTTTGG
Actin reverse	TCAACGAATTGGGGTGTGTG

Table 2.3 Primers for MYB2 Overexpression construct cloning

Name	Primer (5'-3')
$MYB2_c$ forward	AAAGCTAGCCCGTAATTCTTGACCCTGAC
$MYB2_c$ reverse	AAAGAATTCCAAGGGCACTGGATGTCC

2.1.7 Commercially available kits

Total Starch Assay Kit	Megazyme International
Midi prep	Qiagen
PCR purification kit	Qiagen
DIG Oligonucleotide Tailing Kit, 2nd Generation	Sigma Aldrich
TNT® Coupled Wheat Germ Extract System	Roche

2.2 Methods

2.2.1 Microalgae cultivation conditions

Strains were cultivated using Tris-acetate-phosphate (TAP) medium (Harris, 1989). All the strains were grown at 22 °C under cool white fluorescent lamps (150 µmol m^{-2} s^{-1}) following a 16h : 8h dark : light regime on an orbital shaker (120 rpm) or on a solid agar plate. Before inoculation of cell cultures, all the strains and cell lines were grown in standard TAP medium up to late exponential phase. The starting cell density was kept same for all treatments (~65 x 103 cells mL^{-1}) and cells were cultured in triplicate. For stress induction, 30 ml of exponential phase cultures grown in standard TAP medium, were centrifuged at 600 g for 5 min at room temperature, washed with low P or low N media to remove externally bound P or N, and then inoculated into either low P or low N TAP media to give an initial cell density (determined by optical density (OD) at 680 nm) of 0.05. For salt stress experiments, a concentrated solution of NaCl (5 M) was added at appropriate volumes as necessary to obtain concentrations of 100 mM, 150 mM and 200 mM of NaCl into early exponential phase (day 3) cultures

2.2.2 Physiological parameters determination

Cell growth, total chlorophyll and carotenoid concentration of strains

Cell growth was monitored by OD measurements at 680 nm (OD680) using a Thermo Spectronic Aquamate UV-spectrophotometer throughout the cultivation periods (as indicated in the results sections). Total chlorophyll (chlorophyll *a* and *b*) and carotenoid measurement was performed using ethanol extraction. A 5 mL volume of cells at different stages of growth were pelleted by centrifugation (1500 *g* for 10 min), resuspended in 95% (v/v) ethanol and vortexed to extract the pigments. The concentration of total chlorophyll and carotenoids was determined using a UV–visible spectrophotometer by measuring the OD of the extract at 470 nm, 664 nm, and 649 nm against an ethanol blank, and was calculated by the formula described previously (Lichtenthaler 1987):

Chl *a* = 13.36 x OD_{664} - 5.19 x OD_{649}

Chl *b* = 27.43 x OD_{649} - 8.12 x OD_{664}

Carot = (1000 x OD_{470} - 2.13 c Chl *a* - 97.64 x Chl) / 209

Protein measurement

A stock solution of BSA of 100 mg mL^{-1} was prepared and diluted to concentrations ranging from 5 mg mL^{-1} to 30 mg mL^{-1}. A 2 µL sample from each concentration was diluted with 798 µL of distilled water and 200 µL of Bradford reagent was added and incubated for 20 min. Measurements were taken at OD 595 nm using a UV–visible spectrophotometer (Jenway) then a protein standard curve was plotted and the trend line equation (OD_{595} value = 0.0542 X concentration) was obtained. At different stages of growth a 5 mL volume of cells were pelleted by centrifugation (1500 *g* for 10 min) and total protein was determined by resuspending the harvested cells in extraction buffer containing 30 mM Tris HCl, pH 7.5 and 1 µL of protease inhibitor cocktail followed by two rounds of freeze-thaw in liquid nitrogen. The extract was centrifuged at 12000 *g* for 15 min at room temperature. A volume of supernatant (diluted or undiluted as necessary) was added to 200 µL of Bradford reagent for protein quantification following the manufacturer's instructions (Bio- Rad). Total protein was determined using the BSA standard curve.

Fluorescent staining of neutral lipid content

For the measurement of neutral lipid content, a modified procedure of the original method was used (Chen et al. 2009). Algal samples were obtained at different stages of growth (as indicated in the results sections) and were normalised to an OD_{680} value of 0.5 then 740 µL of the normalised algal culture was stained with 10 µL Nile red solution at a concentration of 50 µg mL^{-1} in 25% (v/v) DMSO and mixed by rapid inversion and incubated in darkness for 10 min at room temperature. Samples were analysed using a Fluoromax-4 Spectrofluorometer (Horiba, Glasgow, UK), using a laser excitation line at 530 nm, and emission was collected at 575 nm.

Total lipid measurement

Quantification of the total lipid content was determined by solvent extraction in a ST-243 Soxtec™ unit (FOSS). Hexane (ACS spectrophotometric grade, ≥ 98.5 %) was used as extracting solvent since it has shown to perform well as an extracting agent of neutral lipids (TAGs) induced under nitrogen-deprived conditions (McNichol *et al.*, 2012). Prior to lipid extraction, dried cell pellets were manually pulverised using mortar and pestle alternated with liquid nitrogen supply. Cells were weighed and placed in cellulose extraction thimbles (26 x 60 mm, 603, Whatman®), and the lipids were then extracted in a three-stage program (extraction 2 h, rinsing 40 min, and solvent recovery 20 min). Extracted lipids were allowed to cool down to room temperature in a desiccator and the concentration was then calculated gravimetrically.

Starch measurement

Starch was quantified from cell samples using the Total Starch Assay Kit. At different stages of growth a 5 mL volume of cells were pelleted by centrifugation (1500 *g* for 10 min) then chlorophyll was removed by washing in 95% (v/v) ethanol and incubating at 85°C for 5 min followed by centrifugation at 13000 *g* for 10 min. Cell pellets were resuspended in 200 µL 80% (v/v) ethanol, 500 µL DMSO and incubated at 90 °C for 1 h in a Thermo shaker (PHMT, Grant Bio, UK) to solubilise the starch. The extract was digested with amyloglucosidase. A 3 mL volume of GOPAD reagent (Total Starch Assay Kit) was added to digested sample and to D-glucose standards at concentrations ranging from 30 µg mL^{-1} to 200 µg mL^{-1}.

Samples were incubated for 20 min at 50 °C and absorbance was measured against the reagent blank at 508 nm using a UV–visible spectrophotometer. The glucose standard curve was plotted and the trend line equation (Absorbance = 0.0349 X concentration) was obtained. Starch concentration from the cell extracts was determined using the D-glucose standard curve, and values were multiplied by 162/180 (adjustment for free D-glucose to anhydro D-glucose) to calculate total starch concentration.

Fourier transform-infrared (FT-IR) spectroscopy

Cell cultures harvested at different stages of growth and treatment were harvested and centrifuged at 1500 *g* for 20 min then the supernatant was removed and the wet weight was recorded and normalised to 60 mg mL^{-1} by re-suspension in Milli-Q (Millipore) water. For each replicate sample, 2 x 20 μL cell suspension was added onto a 96-well silicon microplate, with each sample being oven dried at 40 °C for 30 min before the second volume was added. An Equinox 55 FT-IR spectrometer (Bruker Corporation), equipped with a deuterated triglycerine sulfate detector with a HTS-XT high-throughput microplate extension was used to collect the FT-IR spectra from the plate. Spectra were collected twice over the wavenumber range 4000 – 600 cm^{-1}, following previously reported settings (Muhamadali et al., 2015). Spectra were pre-processed by baseline correction for CO_2 peaks using extended multiplicative signal correction algorithm (EMSC2) (Martens and Stark, 1991). All pre-processed data were subjected to principal components analysis (PCA) using MATLAB (The MathWorks Inc, USA). Band assignments were determined as described previously (Driver et al., 2015). The lipid band height (1745 cm^{-1} wave number; υC=O ester functional groups), amide I band height (1655 cm^{-1} wave number; υC=O of amides) and carbohydrate band heights (1160, 1086, 1050, and 1036 cm^{-1} wave numbers; υC-O-C and υC-O of carbohydrates) were measured individually and ratios of these band heights (lipid:amide I and carbohydrate:amide I) were calculated.

2.2.3 EMSA assay

PSR1 protein expression

The full length PSR1 cDNA with an N-terminal 6xHIS-3HA-tag was commercially synthetized (Genescript) and used to express PSR1 protein using two expression systems:

1. - Bacteria protein expression and purification: The full-length PSR1 cDNA (Genescript) was cloned into the pETHIS vector by blunt end ligation. The resulting pETHIS-PSR1 vector was transformed into *E. coli* BL21 (DE3) cells and colonies were checked by restriction analysis to identify the orientation of the cDNA inserts.

For PSR1 expression, an overnight starter culture was grown at 37°C in 20 ml LB medium supplemented with 50 μg mL^{-1} kanamycin. PSR1 expression was initiated by inoculating 1 L of LB (supplemented with 50 μg mL^{-1} kanamycin) with 20 ml of the starter culture. The bacteria were grown at 37°C in a shaking incubator to an OD_{600} of 0.5-0.6, and then the IPTG was added to a concentration of 0.5 mM to induce PSR1 expression. After 4 hours of IPTG induction, cells were harvested by centrifugation at 5,000 *rpm* for 20 minutes.

For protein extraction, bacteria pellets were resuspended in 15 mL of lysis buffer (see materials) and incubated on ice for 1 hour. The resulting suspension was disrupted by sonication on ice for 5 min with pulses of 45 sec on and 15 sec off at 25% of the maximal power and centrifuged for 30 min at 15000 rpm at 4˚C.

For protein purification, a 5 ml HisTrap™ HP column (GE Healthcare) was used following the manufacturer's instructions. Briefly, columns were equilibrated with the low-imidazole binding buffer (see materials) and the sample (clear supernatant from previous step) was applied to the column with a syringe at a flow rate of 5 mL min^{-1}. Columns were washed with binding buffer until the absorbance reached a steady baseline (at least 10–15 column volumes). Finally, the PSR1 His-tagged protein was eluted with elution buffer containing a linear gradient from 100–500 mM of imidazole (see Materials). Protein content of fractions was determined using the method of Bradford. PSR1 purity was determined by western blot using an anti HIS antibody.

2. – In vitro protein expression using wheat germ extract: The full-length PSR1 cDNA (Genescript) was cloned into the pTNT vector (Promega) at the EcoRI

and XbaI restriction sites. The resulting pTNT-PSR1 vector was used for the in vitro protein expression using the TNT® Coupled Wheat Germ Extract System following the manufacturer's instructions (Promega). Briefly, 1 µg of pTNT-PSR1 vector was mixed with the TNT® Wheat Germ Extract, T7 RNA polymerase and Amino Acid Mixture in one reaction mix and incubated at 30°C for 2 hours. Protein content of the extract was determined using the method of Bradford and PSR1 expression was determined by western blot using an anti HIS antibody.

Probe annealing and labelling

The sequences of the oligonucleotides used for generating EMSA probes are listed in Table 2.1 of the Materials section. Oligonucleotides were annealed to its complimentary strand to form the double strand oligonucleotide probes use for the EMSA. 2 µL of each complementary oligonucleotide (1 µg μL^{-1}) were combined with 5 µL of 10X NEB buffer 4 and ddH_2O to a final volume of 50 µL. The reaction was incubated in a PCR block at 98°C for 5 minutes and then removed to allow to cool slowly to room temperature.

Probes were labelled using the DIG Oligonucleotide Tailing Kit, 2nd Generation following manufacturer's instructions. In brief, for the generation of a long digoxigening labelled tail, 100 ng of the double strand oligonucleotide were combined with a mixture of DIG-dUTP and dATP, reaction buffer and the terminal transferase enzyme. For the generation of a short digoxigenin labelled tail, 100 ng of the double strand oligonucleotide were combined with DIG-dUTP, reaction buffer and the terminal transferase enzyme. Both reactions were incubated at 37°C for 15 minutes and then placed on ice. The enzyme activity was stopped by adding 0.2 M EDTA.

For the determination of the labelling efficiency, dot blot was performed. A series of dilutions of the DIG labeled probes (1000 pg μL^{-1}, 100 pg μL^{-1}, 10 pg μL^{-1} and 1 pg μL^{-1}) were spotted directly on a positively charged nylon membrane (Roche) and the DNA was immobilized on the *membrane by* UV *crosslinking.* For reference, dilutions of a DIG-labeled control DNA were included on the same membrane. After crosslinking, membranes were rinse briefly in washing buffer (see Materials section) and incubated in blocking buffer for 30 minutes. After blocking,

membranes were incubated for 30 minutes with DIG antibody diluted in blocking buffer (1:1000), washed extensively in washing buffer and equilibrated for 3 minutes in detection buffer. The digoxigenin signal was detected using the CSPD, ready-to-use and exposed to autoradiography film for 2-3 minutes.

EMSA reaction, running and exposure

The PSR1 protein and the DNA probes were incubated in a binding reaction at 30°C as follow:

	Control	With competitor	With no competitor	Reaction
HEDG (3µl) + HEDGK (1µl) (with DTT)	4 µl	4 µl	4 µl	10-30 minutes
PSR1 protein (80 nM)	0 µl	2.5 µl	2.5 µl	
Non-specific competitor Salmon Sperm (1µg/µl)	1 µl	1 µl	1 µl	2-5 minutes
Wild Type or Mutant DIG-labelled probes (1 ng/ µl)	1 µl	1 µl	1 µl	10-20 minutes
Cold competitor (unlabelled probe) (100 ng/ µl)	0 µl	< =1.5 µl	0 µl	
dH_2O	4 µl	0 µl	1.5 µl	
Total	10 µl	10 µl	10 µl	10 µl

The DNA protein reaction mixtures were loaded onto a native polyacrylamide gel and were subjected to electrophoresis (2.5 hours at 80 volts in small gels and 5 hours at 200 volts in large gels) in 0.5x TBE running buffer. A separate lane containing gel loading dye, purple (6X) was ran in parallel with the samples to judge how far the free probe had ran. *Following gel* electrophoresis the free and protein-bounded DNA *were transferred* from *the gel to* a positively charged nylon membrane (Roche) at *100 V* for 1 *hours* at 4ºC in 0.5 x TBE buffer. After electrotransfer the blot was crosslinked using UV light and incubated in blocking buffer for 30 minutes. The membrane was then incubated with the with DIG antibody diluted in blocking buffer (1:1000), washed extensively in washing buffer and equilibrated for 3 minutes in detection buffer. The digoxigenin signal was detected using the CSPD, ready-to-use and exposed to autoradiography film for 2-3 minutes.

2.2.4 Nucleic acids extraction

Genomic DNA extraction

A 5 mL volume of cells was centrifuged at 6000 *g* for 2 min and resuspended in 500 µL of Cetyl trimethyl ammonium bromide (CTAB) buffer. Cells were vortexed and incubated at 65 °C for 1 h then 500 µL of phenol/chloroform/isoamylalcohol (25:24:1) (Fisher) was added and vortexed again. The precipitate was centrifuged at 10000 *g* for 5 min then the upper aqueous phase was transferred into a fresh Eppendorf tube and 0.5 mL of isopropanol was added. The precipitated DNA was pelleted by centrifugation at 13000 *g* for 20 min. Supernatant was removed and the DNA pellet was washed with 200 µL 70% ethanol, and resuspended in 50 µL of distilled water. Quantification of DNA was performed using a Nano Drop 3300 (Thermo Scientific, USA).

RNA extraction

A 25 mL volume of cells was centrifuged at 6000 *g* for 2 min and RNA was isolated from the cells using TRI Reagent following manufacturer's protocol. Essentially, 1 ml of TRI Reagent was added to the cell pellet and mixed using a pipette then centrifuged at 12000 *g* for 10 min at 4 °C, followed by incubation at 25 °C for 5 min. A 0.2 ml volume of chloroform was added and vortexed then reincubated for 5 min at 25 °C then centrifuged at 12000 *g* for 10 min at 4 °C. The upper aqueous phase was transferred into a fresh Eppendorf tube and 0.5 mL of isopropanol was added. The precipitated RNA was pelleted by centrifugation at 13000 *g* for 20 min. Supernatant was removed and the RNA pellet was washed with 200 µL 70% ethanol, and resuspended in 100 µL of RNase-free distilled water. Quantification of RNA was performed using a Nano Drop 3300 (Thermo Scientific, USA).

2.2.5 Expression analysis by RT-PCR

Total RNA was DNase treated and 1 µg was reverse transcribed to cDNA using a mix of oligo dT and specific reverse primer and the Superscript III reverse transcriptase according to the manufacturer's instructions. For the detection of expression levels of *MYB1* and *MYB2*, 1 µL of the total cDNA synthesised from the different samples was used as the template for the PCR amplification using MyTaq Red Mix. The primers used in the PCR reaction were designed using Primer3

(http://primer3.wi.mit.edu/) and synthesised by Eurofins Genomics. *rBCL* primer sequences were used for the amplification of a housekeeping gene and to confirm successful cDNA synthesis, and were taken from de Vere et al. (2012). The PCR cycling conditions used were as follows: 95 °C for 3 minutes, 35 cycles of 95 °C for 15 seconds, 50 °C for *rBCL* for 1 minute or 55 °C for *MYB2* for 30 seconds and 72 °C for 30 seconds. After the 35 cycles, a final step of 72°C for 5 minutes was added. RT-PCR products were analysed by electrophoresis on a 1.5 % agarose gel and imaged with a UV transilluminator following staining with SafeView dye.

2.2.6 MYB2 Overexpression construct cloning

The full length *MYB2* cDNA was PCR amplified using Kapa HIFI using the MYB2F (5′-AAA GCT AGC CCC GTA ATT CTT GAC CCT GC-3'; NheI restriction enzyme site underlined) and MYB2R (5′-AAA GAA TTC CAA GGG CAC TGG ATG TCC-3′; EcoRI restriction enzyme site underlined) primers, added at 0.5 μM concentration each. The PCR reaction was initiated at 95 °C for 1 min, and was run for 35 cycles (95 °C for 15 s, 54 °C for 15 s, 72 °C for 10 s). The reaction products were resolved on a 1 % (w/v) agarose gel, and observed under UV illumination following staining with SafeView dye. Positive PCR products were sequenced in full (GATC-Biotech) using MYB2F and MYB2R primers. The *MYB2* cDNA was a 1751 bp fragment including all 3 exons and partial 5' and 3' untranslated regions. The cDNA was digested with NheI and EcoRI prior to plasmid ligation, using 1x restriction enzyme buffer and 0.5 μL of each restriction enzyme and incubated at 37 °C for 2 h before DNA purification using the Qiagen PCR purification kit following the manufacturer's instructions.

The *MYB2* cDNA was cloned into the pCB740 plasmid which contains the *ARG7* gene marker, and the *HSP70B* gene downstream of the *HSP70A/RBCS2* fusion promoter (Schroda et al., 2000), and which has been successfully used previously for over-expression in *C. reinhardtii* (Bajhaiya et al., 2016b). pCB740 plasmid was prepared by isolation from *E. coli* using a Qiagen Midi-Prep kit following the manufacturer's instructions. The *HSP70B* gene was removed by digestion with NheI and EcoRI as described above. After restriction digestion, the linearized plasmid DNA was dephosphorylated to prevent self-ligation. The following was added in a tube: 8.8 μL of plasmid DNA (20-100 ng), 2 μL of alkaline phosphatase, 1.2 μL of alkaline dephosphorylation buffer (10x) then incubated at 37 °C for 1 h. The alkaline

phosphatase was inactivated by heating reaction mix at 65°C for 5 min. PCR product was purified before ligation using Qiagen PCR purification kit. DNA ligation was performed with a 1:3 ratio between plasmid DNA and *MYB2* insert DNA. The following were combined in a tube: 2 μL of ligation buffer (1x), 1 μL of plasmid DNA (20-50 ng), 3 μL of insert DNA (70-100 ng), and 1 μL of T4 DNA ligase. The reaction mixture was incubated at 16 °C for 4 h. The ligation reaction of pCB740-MYB2 plasmid was transformed into competent *E. coli*.

After incubation overnight at 37 °C in LB agar containing 50 mg mL^{-1} ampicillin, bacteria colonies were screened by PCR using MYB2 primers as described above. Positive colonies were grown at 37 °C in liquid LB medium containing 50 mg mL^{-1} ampicillin. After plasmid replication, plasmid isolation was performed using a Qiagen Midi-Prep kit following the manufacturer's instructions, and DNA purity and concentration was confirmed using a NanoDrop 3300 Spectrophotometer (Thermo Scientific, USA). Sequencing of purified plasmid confirmed the presence of the *MYB2* gene within the pCB740-MYB2 plasmid using the MYB2 primers. The junction site between the *HSP70A-RBCS2* tandem promoter and the *MYB2* gene at the 5′ end (NheI site) was 108 bp before the ATG start codon, and 224 bp after the TGA stop codon at the 3' end.

2.2.7 Bacteria and algae transformation

Chlamydomonas transformation

To generate *MYB2* over-expression (MYB2-OE) lines and vector control lines, CC-4350 or CC-4351 cells were prepared for particle bombardment by spreading of 50 mL of cells grown to exponential phase onto TAP agar lacking arginine. Particle bombardment with the pCB740-MYB2 plasmid (Fig. 5.1), or empty-vector control (pCB740) was performed using a PDS-1000 He Particle delivery system (Bio Rad, France) as described previously (Boynton and Gillham, 1993), except using 0.6-μm gold microcarriers, a 6 cm target distance between the muzzle of the gene gun and the cells, and either 900-p.s.i. or 1100-p.s.i. rupture disks (Bio-Rad). After transformation, cells were incubated for 14 days for selection of arginine prototrophic colonies. PCR from genomic DNA and RT-PCR from cDNA was performed using MYB2 primers as described above to confirm positive transformants and transgene expression level.

Preparation of chemically competent *E. coli*

Two 2.5 L flasks with 500 mL of sterilized LB media were inoculated with 5 mL of cells from an overnight culture of *E. coli* DH5α. Flask were shaken at 37°C for 2-3 hours until they reach an OD_{600} of 0.4-0.6. Cultures were centrifuged at 4000 g for 10 min at 4°C. Supernatant was discarded and the pellet was resuspended in 40% of the original culture volume (200 ml) using ice-cold TFB1 buffer (see Materials), and incubated in ice for 5 minutes. Suspended cells were centrifuged again at 4000 g for 10 min at 4°C. Pellet was resuspended in 4 % of the original culture volume (20 ml) using ice-cold TFB2 buffer (see Materials) and incubated in ice for 25 minutes.

***E. coli* transformation**

Plasmids were transformed into either in lab-made chemically competent *E. coli* DH5α or commercially available chemically competent XL1-Blue cells (Agilent Technologies, USA) according to the manufacturer's instructions. 100 µL of competent cells and 1.7 µL of beta-mercaptoethanol were added to pre chill 15 mL Falcon tube, gently mixed and incubated on ice for 10 minutes. 10-20 ng of ligated vector was added to the reaction mixture tube and incubated on ice for 30 min. The mixture was heat shocked at 42°C for 45 sec then immediately returned to ice for 2 min. 900 µl of pre-heated (42°C) sterile LB medium was added and tube was incubated at 37°C for 1 h with shaking at 200 rpm. 150 µL of the transformation mixture was platted on LB agar plates containing appropriate antibiotics, IPTG (1mM) and X-gal (5%) (for the XL1-Blue cells). Plates were incubated overnight at 37°C and blue-white screening was performed.

2.2.8 Bioinformatics

MYB identification

To identify the maximum number of MYB domain-containing proteins, the Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de/) was used to search the characteristic MYB domains in the proteome of *C. reinhardtii*. SMART uses profile-hidden Markov models built from multiple sequence alignments to identify protein domains in protein sequences. The SMART database identified fragments of putative proteins in

the *C. reinhardtii* proteome containing the MYB domain. To obtain the whole-protein sequences, a BLAST was performed using the Phytozome database (http://phytozome.jgi.doe.gov) with all of the candidate MYB fragments obtain with SMART against the *C. reinhardtii* proteome. Second, the Pfam database was used to confirm if each of the candidate MYB sequences was a member of the MYB family. To avoid overestimation of the number of MYBs, all of the candidate MYB genes were searched in the PlantTFDB (PlnTFDB, version 3.0) (http://plntfdb.bio.uni-potsdam.de/v3.0/index.php?sp_id=CRE4) and confirmed manually for the presence of the MYB domain. The compilation of the databases results allowed us to obtain a putative list of TFs.

To examine the structural similarities among the putative MYB proteins in *C. renhardtii*, the conserved motifs of MYB proteins were investigated. Their complete amino acid sequences were subjected to The Multiple Expectation Maximization for Motif Elicitation analysis (MEME) (Bailey *et al*. 2009) using the following parameters: optimum motif width was set from 6 to 50; and the maximum number of motifs was set to identify 10 motifs.

Phylogenetic analysis

For phylogenetic analysis of MYB gene sequences, amino acid sequences for *Arabidopsis thaliana* and *C. reinhardtii* genes were obtained from Phytozome and multiple sequence alignments were performed using Clustal Omega using the Phylip output. Phylogenetic analysis was performed using the maximum likelihood method under the WAG + F model of amino acid substitution and using the fast bootstrap approach (Stamatakis et al., 2008) to determine tree confidence, using RAxML version 7.1 (Stamatakis, 2006). For bootstrapping, 1,000 iterations were used. The tree was viewed using the FigTree program.

2.2.9 Statistical analysis

Differences between cell lines and stress treatments were assessed using one-way or two-way ANOVA as appropriate in GraphPad Prism v 7. When significant

differences at a confidence level of 95% were detected, the Tukey-Kramer post-hoc test was applied to determine significant groups.

Chapter 3

Identification of PSR1 transcriptional targets involved in lipid and starch metabolism

3.1 Introduction

With attempts to elucidate the molecular mechanisms that control responses under phosphorus (P) starvation, Shimogawara et al., (1992), studied *C. reinhardtii* mutants that exhibit abnormal responses to P starvation. The mutants were isolated by their ability to survive an environment with high concentrations of ^{32}P and their inability to grow in a media with glucose-1-phosphate under P deprived conditions. These characteristics were attributed to the low scavenging abilities of the mutant; the ability to survive high concentration of ^{32}P was attributed to the lower P uptake of the mutant; the inability to grow in glucose-1-phosphate media was attributed to the low extracellular phosphate activity compared with the wild type (WT) strain under P starvation conditions, and thus its inability to process esterified P. This mutant was identified as *psr1* (phosphorus starvation response) and the protein encoding for the gene was detected by screening clones, transformed with cosmid libraries, that rescue the mutant phenotype and allow them to grow in glucose-1-phosphate media. Further studies on *PSR1* expression reinforce its role as an essential regulator of the P starvation response. WT lines of *C. reinhardtii* exhibit increase transcript levels of *PSR1* after 4-8 hours of P starvation, with a peak of accumulation at 8 hours (Wykoff *et al.*, 1999). The same study, using immunocytochemical techniques, was able to locate PSR1 mainly in the nucleus under both nutrient- replete and P-starvation conditions. Moreover, studies on *psr1* complementation lines that rescue the WT phenotype, reported an increase in PSR1 levels by >10-fold after 24 hours of P starvation, which correlates with *PSR1* transcript levels (Bajhaiya *et al.*, 2016b). Together, these results confirmed the importance of the PSR1 transcription factor (TF) in the response to P deprivation and was the first regulator of P metabolism in a photosynthetic eukaryote to be identified.

As in microalgae, plants preferably assimilate P as inorganic phosphate, which is commonly limited in the environment. Consequently, plants have also developed adaptive responses to cope with the limited P availability. Biochemical and metabolic adaptations are similar to the ones observed in microalgae including increased availability of inorganic phosphate, increase in P uptake through high affinity P transporters, increased expression of nucleases and phosphatases and phospholipid remodelling (Misson *et al.*, 2005; Nilsson, *et al*., 2010). Also, plants have reported developmental responses such as root growth and architecture that

promotes exploitation of soil P resources and include increases in root / shoot ratio, root hair proliferation and length, and lateral root number (Lambers and Plaxton, 2015). Several genes have been identified as responsible for the acclimation of plants to P starvation (e.g. genes encoding high affinity P transporters and phosphatases), however, only a few regulators have been reported. Phosphorus starvation response 1 (PHR1), a transcription factor belonging to the G2-like family, was the first gene involved in the control of P responses in vascular plants to be characterized at the molecular level. This regulator was first identified in *Arabidopsis thaliana* (*A. thaliana*) through the study of mutants with impaired P starvation responses including the induction of P starvation-responsive genes and the increase in anthocyanin accumulation (Rubio *et al*., 2001). PHR1 is located in the nucleus and its expression levels exhibit an increase in response to P starvation, however, unlike PSR1 this increase is only moderate. It binds as a dimer to the imperfect palindromic sequence GNATATNC present in the promoters of P starvation response genes (Rubio *et al*., 2001).

While little is known about how PSR1 is regulated and how its activity is controlled in response to P starvation, there is considerably more information of its orthologue in plant, PHR1. Comparisons of PSR1 with PHR1 revealed that they present a similar DNA binding domain sequence and coiled-coil domain sequences that function in protein-protein interaction and DNA binding (Rubio *et al*., 2001) (Fig. 3.1). PSR1 is a protein of 752 aa with a conserved putative DNA binding domain of G2-like proteins. Unlike PHR1, PSR1 contains three glutamine rich regions in the C-terminal half, which have been identified as critical for the positive regulatory activity of several different eukaryotic transcription factors (Wykoff *et al.*, 1999). The presence of a DNA binding domain and an activator domain suggest that PSR1 may work directly activating the transcriptional activity of P-starvation-responsive genes. To date, two different regulation systems had been identified through which PHR1 is regulating its downstream targets: one is an indirect mechanism in which PHR1 activates a miRNA (miR399), which in turn mediates the inactivation of PHO2, a E2 conjugase in charge of repressing a set of P starvation-induced genes (including P transporters) under P replete conditions (Bari *et al.*, 2006); and the other one is a direct mechanism in which PHR1 binds directly to the cis regulatory element, P1BS (GNATATNC) in the promoter regions of P starvation-induced genes (Bustos *et al.*, 2010). However, even though the PHR1 regulation

system has been found to be highly conserved in other higher plants (Rubio *et al*., 2001), it is still unclear how conserved this regulation mechanism is in microalgae. Differences between their protein sequences (glutamine rich domain in PSR1), in the abundance of transcripts upon P starvation (PSR1 transcript is highly upregulated, PHR1 is upregulated only moderately), and the absence of conserved regulation intermediates (PHO2) in *C. reinhardtii* (Bari *et al.*, 2006), suggest that higher plants may have evolved a more complex regulation system in response to stress.

Although no sequence specific DNA binding site has been determined for PSR1, putative PSR1 targets have been proposed based on transcriptional analysis of *C. reinhardtii* under P-replete conditions compared with P starvation (Moseley *et al*., 2006; Bajhaiya *et al.*, 2016b). Also due to the homology between PSR1 and PHR1 and the conserve coiled-coil and DNA binding domain, GNATATNC, the PHR1 binding site (P1BS), has been proposed as a putative PSR1 binding site. Among the specific responses to P starvation in *C. reinhardtii*, P scavenging through the expression of phosphatases (Riegman *et al.*, 2000; Kruskopf and Plessis, 2004) and high affinity uptake transporters has been well reported (Shimogawara *et al.*, 1999; Moseley *et al*., 2006). PHOX, the major alkaline phosphatase expressed under P starvation conditions has reported an increase over 5000-fold in its transcript levels under P-starvation conditions, an increase that is not detected in the *psr1* mutant under the same conditions (Moseley *et al*., 2006; Bajhaiya *et al.*, 2016b). Moreover, the PHOX promoter contains the putative P1BS in an intronic region of its sequence. From the ten PTB-type high affinity Na^+/P transporters encoded in the *C. reinhardtii* genome, at least seven were found to be more than >2 fold up-regulated by low-P in WT at day-3 and 5 of P starvation, particularly PTB2 with an increase of over 20-fold (Moseley *et al*., 2006). Similarly, other genes well known for their role in *C. reinhardtii* acclimation to P starvation, were induced in a PSR1-dependent manner, including light-harvesting complex proteins (LHC) and proteins involved in oxidative protection such as alternative oxidase (AOX1) and putative plastid terminal oxidase (PTOX1) *(*Moseley *et al.*, 2006).

Another set of genes found to be highly upregulated in P starvation conditions are genes involved in the carbon metabolism, specifically lipid and starch (Bajhaiya *et al.*, 2016b). As previously mentioned, it is well reported the increase of storage molecules such as starch and lipids under nutrient stress conditions and how valuable these molecules are for the biofuel production (Hu et al., 2008). Currently,

RNA-seq data together with bioinformatics have given valuable information about the correlation between PSR1 and starch and lipid gene expression. Bajhaiya et al. (2016b) reported that *psr1* mutants exhibit a great reduction in the starvation-induced starch and lipid accumulation capacity that is restored when PSR1 recovers its function (*psr1*:PSR1 complementation lines). The expression of genes from lipid metabolism like GPD1, GPD4, DGTT2, DGTT3 and PDH1 and from the starch metabolism like SSS1, SSS5, ISA3, SP2, AMA3 were found to be upregulated under P-limited treatment only in WT with highly reduced expression in the psr1 mutant. Interestingly, these genes also present the P1BS site, suggesting a PSR1 induced regulation. However, there is no information about the mechanism of how PSR1 could be regulating these pathways. This study, therefore aimed to determine if PSR1 mediates the direct regulation of genes involved in starch and lipid biosynthesis under P starvation by binding to their promoter region as a TF. Based on previous data from the Bajhaiya et al. (2016b) study, one gene involved in the P starvation response (PTB2), one gene involved in lipid metabolism (GPD4) and 3 genes involved in starch metabolism (SSS1, SSS5 and SP2) were chosen for analysis.

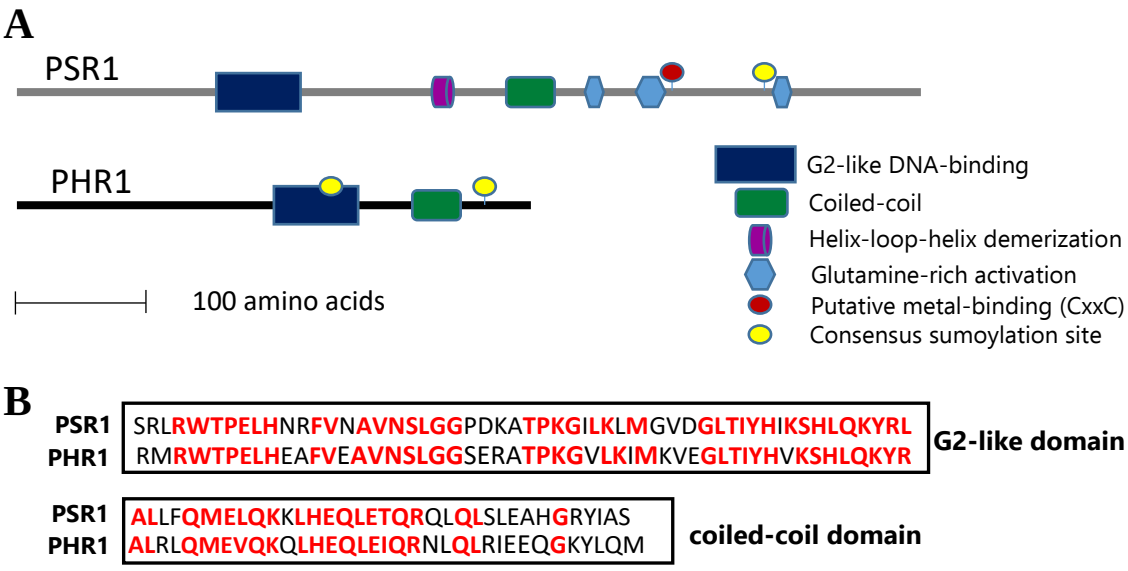

Figure 3.1: PSR1 sequence and comparison with PHR1 TF. (A) Schematic representation of the protein domains in *C. reinhardtii* PSR1 and *A. thaliana* PHR1. (B) Alignments of the sequences of the G2-like DNA binding domain and coiled-coil domain of PSR1 and PHR1. Identical residues are highlighted in red.

3.2 Results

3.2.1 EMSA assay standardization

The interaction of PSR1 with its target genes is the key to determine how this TF is regulating lipid and starch biosynthesis under P-starvation. To study this, we can use the Electrophoretic mobility shift assay (EMSA), also known as the gel retardation assay, which is one of the most popular and sensitive techniques to determine DNA-protein interaction (Alves and Cunha, 2012). This assay relies on the ability of a protein (either purified or in a cellular/nuclear extract) to bind in a specific manner to a labelled DNA probe and retard its migration during the electrophoresis of a native polyacrylamide gel. The retardation of the migration of the DNA generates a shift that can be detected through the labelling of the probe (Alves and Cunha, 2012). Although this is a very well-established technique, successful gel shifts may require the optimization of several parameters that can affect the sensitivity and reliability of the assay. Critical steps for EMSA assay that required standardization with PSR1 TF and its putative targets are presented below.

3.2.1.1 Protein expression standardization

The first and one of the most crucial steps for the EMSA is the generation of the PSR1 TF protein. Considering that PSR1 is a eukaryotic protein, the first choice to express it was by using an eukaryotic expression system, in this case the cell-free system based on the wheat germ embryo. The system coupled the transcription and translation in one step using wheat germ embryos containing all of the components for translation in a concentrated dried state and ready for protein synthesis as soon as the germination process starts (Harbers, 2014). Synthetically generated PSR1 cDNA with a 6HIS-3HA tag at the N-terminus (Fig. 3.2 A) was cloned into the pTNT vector (Fig. 3.2 B) (according to the manufacturer's instructions) and mixed with the wheat germ extract. Protein expression was checked in the soluble and insoluble fractions and analysed by western blotting using an anti-His antibody. However, expression of PSR1 was not successful using this system (data not shown). There was no detectable protein at the predictive molecular mass in the soluble or insoluble fraction. It was thus decided to use the traditional *Escherichia coli* BL21 (DE3) system for protein expression for its cost-effectiveness and reliability of use (Rosano and Ceccarelli, 2014). The synthetically generated PSR1 cDNA fragment this time

was cloned into the pET-HIS vector (Fig. 3.2 C), which was then transformed into *E. coli* BL21 (DE3). The obtained protein was purified using a His Trap affinity column. Western blot analyses using the anti-His antibody showed the presence of a band at ~100 kDa, consistent with previous studies (Wykoff *et al.*, 1999) indicating that expression and purification of PSR1 was successful (Fig. 3.2 D).

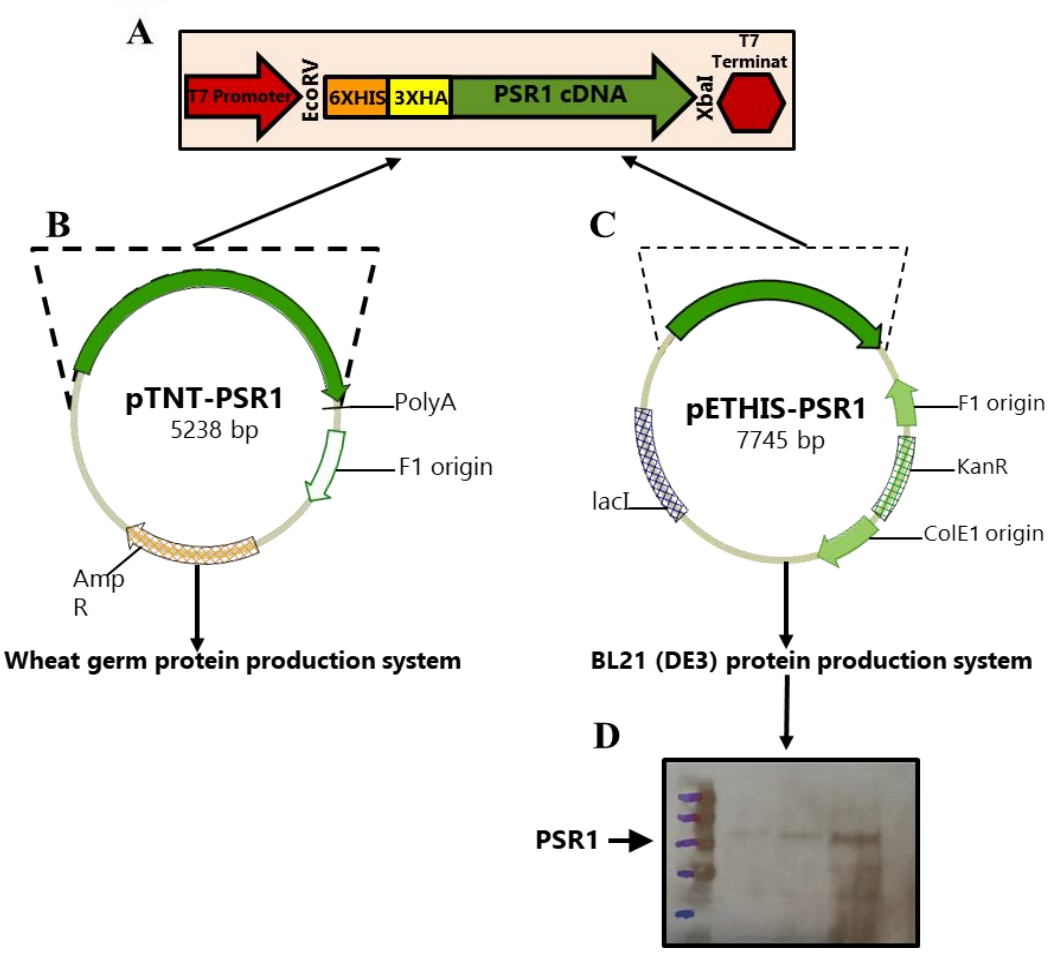

Figure 3.2: Constructs used for the PSR1 protein expression. **(A)** Synthetically generated PSR1 cDNA containing a 6HIS-3HA N-terminal tagged. **(B)** The recombinant cDNA was cloned into the pTNT vector and produced the pTNT-PSR1, used for the PSR1 expression with the TNT® Coupled Wheat Germ Extract System, **(C)** or into the pET-HIS vector to produce the pET-HIS-PSR1 used for the PSR1 expression with BL21 (DE3) *E.coli* system. **(D)** Serial dilution of the purified PSR1 protein (0.8 ng, 8 ng and 80 ng) analysed by western blot using anti HIS-tag antibody. Arrow indicates the position of migration of the PSR1 protein (~100 kDa).

3.2.1.2 Probe labelling standardization

Another crucial aspect of the EMSA assay is the labelling of the probe. The ideal label should have a high sensitivity of detection and should not interfere with the formation of the DNA-protein complex. The traditional method used for labelling probes in EMSA assay is ^{32}P 5′-end labelling (Hellman and Fried, 2007), however because of its high costs and danger of use, other non-radioactive methods have been developed (Harris, 1991). One method that has proven high specificity of detection is the use of digoxigenin (DIG). In this method nucleotide triphosphate analogues containing the digoxigenin molecule are incorporated either randomly throughout the DNA sequence or at the 5' or 3' ends (Höltke *et al.*, 1995). The DIG-labelled probe is detected using an antibody against digoxigenin (anti-DIG) to which alkaline phosphatase has been conjugated. Considering that the DNA-protein interaction should not be disrupted by the label, in this assay the DIG-label was added at the 3'end of the DNA probe. Initially, following the manufacturer's recommendation and with the aim of having a strong signal for the detection of the probe, a long tail of DIG-nucleotides was added to the 3' end of the DNA probes (Fig.3.3 A). Two PSR1 target probes (PTB2 and PTB2 mutant probe) and two controls (one previously labelled and the other one labelled in parallel with the PSR1 target probes), all with similar lengths, were tested for the standardization of the labelling method. The success of the labelling was firstly confirmed by a dot blot in which dilutions of the labelled probe (4 serial dilutions) were spotted along a positively charged nylon membrane (Fig.3.3 B). The comparison of the spot intensity of the controls with the PSR1 target probes matched very closely, indicating the success of the DNA labelling to a similar yield of the controls. The next step was to run the labelled probes into the EMSA gel. Electrophoresis was run using a 6% polyacrylamide native gel (following literature recommendations for typical EMSA gels) for 1 hour, and the probe was detected by western blot using an anti-DIG antibody. However, the results were not as expected, and none of the tested probes ran to the bottom of the gel, leaving a smeared pattern (Fig.3.3 C). The results did not improve by decreasing the probe concentration (Fig.3.3 C), by using a lower percentage gel (4%) or by running the gel for longer times (data not shown). Accordingly, it was then decided to try another approach in the probe labelling by adding a shorter DIG-labelled tail. This was achieved by only adding DIG-dUTP nucleotides that added a tail of up to 3 nucleotides to the DNA probe (Fig. 3.4 A). As previously, the success

of the probe labelling was confirmed by dot blot (data not shown), and later run for 1 hour in a 6% native polyacrylamide gel. This time the result was different (no smeared pattern) but negative again, because the short DIG-tail labelled probe was not detectable (lane 4, Fig. 3.4 B) when compared to the long DIG-tail labelled probe (lane 1-3, Fig.3.4 A) in a 6% polyacrylamide gel. However, when the concentration of the gel was increased to 12% polyacrylamide, the short DIG-tail labelled probe (lane 4, Fig.3.4 C) was detected with the same sensitivity as the long DIG-tail labelled probe (lane 1-3, Fig.3.4 C), migrating towards the bottom of the gel, as expected. Increasing the running time of the electrophoresis gel to 2.5 hours, improved even more the migration pattern of the probe, running more towards the bottom of the gel (Fig. 3.4 D). Finally, for further improvements, a larger polyacrylamide gel was run, using the long DIG-tail labelled probe as a control, and the results showed the most optimal conditions tested so far (Fig. 3.3 E). The probe migrated almost to the bottom of the gel, leaving no detectable smear in the top half of the gel. These optimized conditions were therefore used to run the EMSA gels.

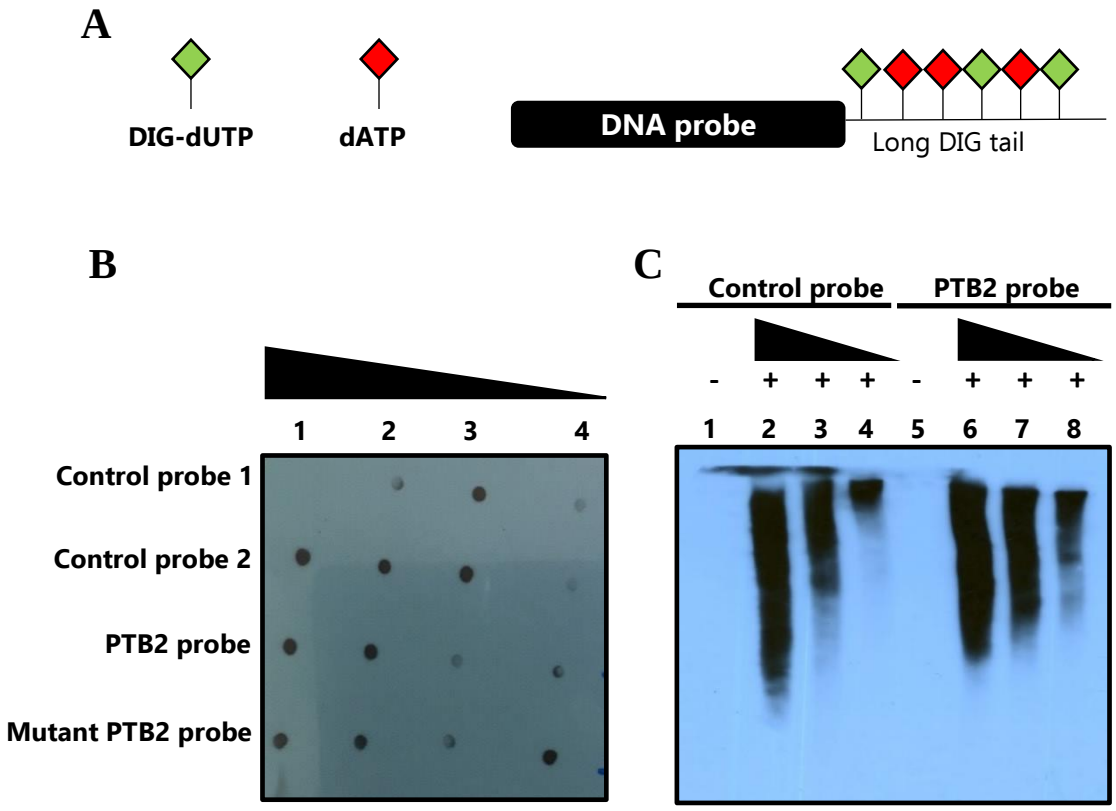

Figure 3.3: DNA probe labelling with a long DIG oligonucleotide tail. **(A)** Long DIG oligonucleotide tail was generated by incorporation of a mixture of DIG-dUTP/dATP nucleotides in the 3' of the DNA probe. **(B)** Dot blot was used to confirm the success in the probe labelling of the 4 probes used for standardization; control 1, control 2, PTB2 and PTB2 mutant probe, spotted in serial dilution as following: Spot 1: 1000 pg/µL; spot 2: 100 pg/µL; spot 3: 10 pg/µL; spot 4: 1 pg/µL. **(C)** Labelled probes were run in a native 6% polyacrylamide gel for 1 hour in serial dilutions.

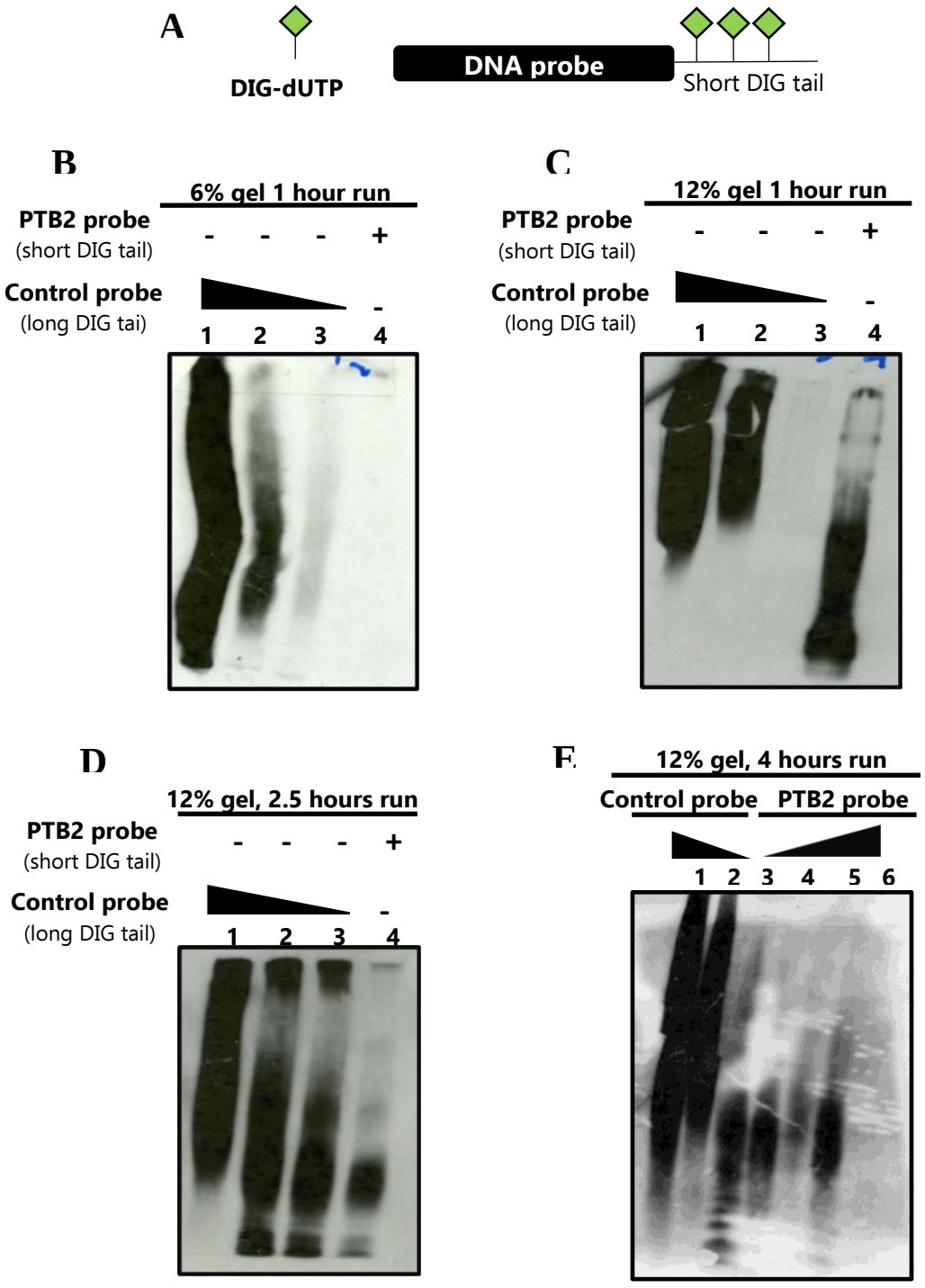

Figure 3.4: DNA probe labelling with a short DIG oligonucleotide tail. **(A)** Short DIG oligonucleotide tail was generated by incorporation of only DIG-dUTP nucleotides in the 3' of the DNA probe. **(B)** Labelled probe was run in a native 6% polyacrylamide gel for 1 hour (lane 4), using as control serial dilutions of 1:10, 1:100, and 1:1000 of the previously labelled long DIG tail PTB2 probe (lane 1-3). **(C)** Same samples were run in another native gel increasing the polyacrylamide concentration of the gel to 12%, and run for one hour, or 2.5 hours **(D)** Large 12% polyacrylamide gel run with a control probe labelled with a long DIG oligonucleotide tail (lane 1-2), or the PTB2 probe labelled with a short DIG-dUTP tail at increasing concentrations of the probe.

3.2.1.3 Determination of a negative control to ensure specificity of EMSA assay

One of the most important factors to determine the reliability of the EMSA assay is to ensure the specificity of the Protein-DNA interaction. TFs have binding affinities for their specific DNA sequences up to 10^6-fold higher than those for DNA in general (Phillips, 2008). Therefore, the determination of an appropriate negative control is key to ensure that any observed shift will be specific to the target probe of interest. For this purpose, an oligonucleotide with no P1BS was designed (random probe), and nonspecific DNA competitors were added to the binding reactions to enhance even more the specificity of the assay (salmon sperm and poly dI-dC). As expected, no shift was observed when PSR1 was incubated with the random probe in the presence of either of the nonspecific DNA competitors (lane 2 and 3, Fig.3.5), indicating that PSR1 does not bind to the random probe sequence. The stronger smeared pattern observed in the presence of salmon sperm (lane 2, Fig. 3.5) may be explain by nonspecific bindings that disappeared when poly dI-dC was use as a nonspecific DNA competitor (lane 3, Fig. 3.5). Overall, these results indicate that the designed random probe is a proper negative control for the EMSA assay, and that poly dI-dC is a better nonspecific DNA competitor that does not generates background signals of nonspecific binding.

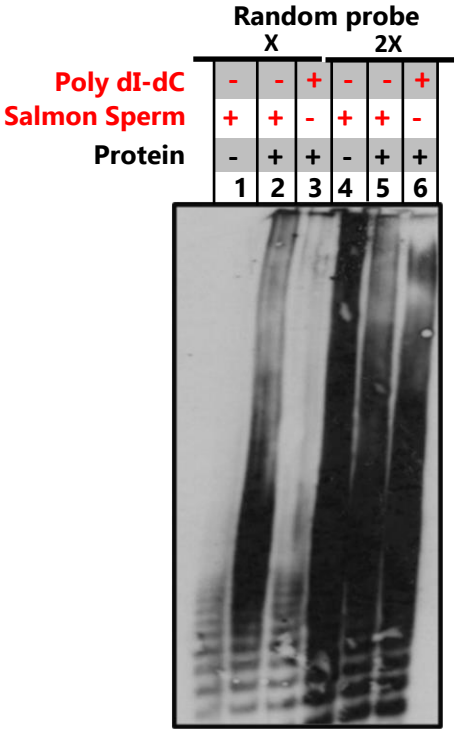

Figure 3.5: EMSA assay to ensure determine a negative control. EMSA assay of PSR1 TF and a randomly design probe (with no P1BS elements). PSR1 purified protein was incubated with the DIG labelled random probe in the presence of non-specific DNA competitors (salmon sperm and poly dI-dC) and a specific competitor. The reactions were loaded in two different concentrations in the gel, to ensure the intensity of the DIG signal (X and 2X).

3.2.2 EMSA assay for the determination of PSR1 putative target genes

3.2.2.1 *Putative PSR1 target involve in P-scavenging mechanisms*

The upregulation of a high affinity P transporter under P-starvation is one of the scavenging mechanisms known to be activated upon P-starvation in *C. reinhardtii* (Grossman and Aksoy, 2015). Bajhaiya *et al*. (2016b) reported high upregulation of the high affinity P transporter PTB2 under P-starvation conditions, but not in the *psr1* mutants under the same conditions. The inability of PTB2 to increase its expression levels when PSR1 is not active, suggests a key role of PSR1 in the regulation of this transporter. To determine whether PSR1 is directly regulating PTB2 expression through its TF activity, we tested the DNA binding ability of PSR1 to the promoter region of PTB2 by performing an EMSA assay. Considering the high homology of PSR1 with PHR1 TF in Arabidopsis, especially in their DNA binding domain, we assumed the PHR1 consensus binding sequence

(P1BS) GnATATnC as the PSR1 putative binding sequence. The DNA probe for the EMSA assay was designed from the promoter region sequence of the PTB2 gene, 3359 base pairs upstream from the translation origin, where the GgATATaC putative P1BS domain was found (Fig.3.6 A). The EMSA assay was performed using the previously optimized conditions. The results showed that the PSR1 TF produced a mobility shift with the PTB2 probe (lane 2 and 3, Fig. 3.6 B), whereas the probe alone did not (lane 1, Fig. 3.6 B), indicating the formation of a PSR1-PTB2 probe complex. The fact that the presence of either of the unspecific DNA competitors used (salmon sperm and poly dI-dC) did not affected the binding of PSR1 to the PTB2 probe (lanes 2 and 3, Fig. 3.6 B) verifies that the binding is sequence-specific. Moreover, the reduction in the shift intensity by the addition of unlabelled PTB2 probe (lane 5 and 6, Fig. 3.6 B), showed that this probe could specifically compete with the PTB2 labelled probe in binding to PSR1, confirming the specificity of the PSR1 binding. To further test the DNA binding specificity, a mutant PTB2 probe with a substitution of two nucleotides was used (Fig. 3.6 A). The shift observed with the PTB2 mutant probe, indicates that the substituted nucleotides were not essential for the PSR1 binding (lane 4 Fig. 3.6 C). However, the decrease in the band intensity of the PSR1-PTB2 mutant probe complex suggests that this binding may be weaker (reduced affinity).

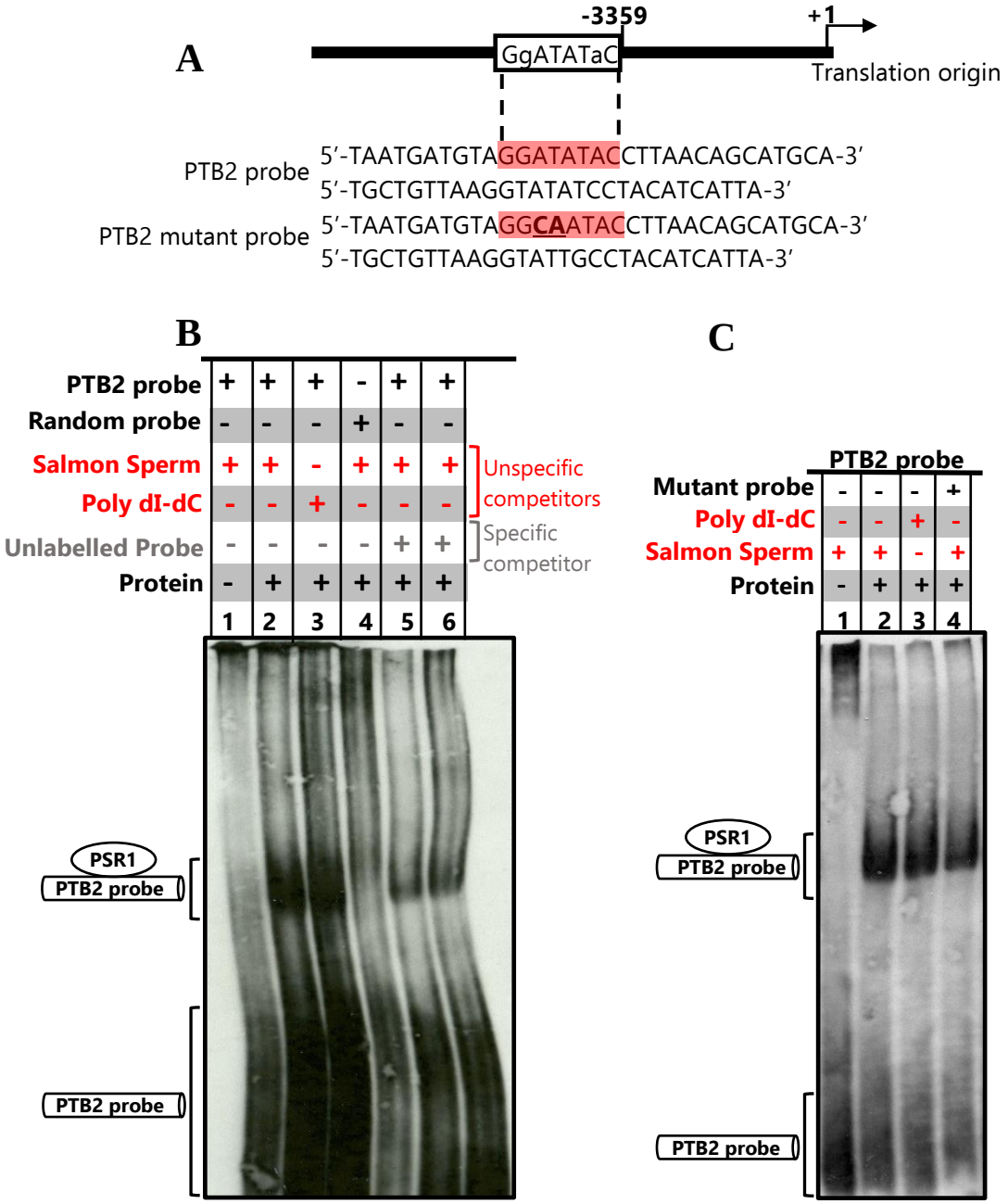

Figure 3.6: EMSA assay showing the binding of PSR1 to the PTB2 target sequence. **(A)** Diagram of the probes designed in the promoter region of the PTB2 gene that contained the P1BS putative binding site. Mutant PTB2 probe was design by the substitution of 2 nucleotides within the P1BS. **(B)** EMSA assay of PSR1 TF binding specifically to the PTB2 probe. PSR1 purified protein was incubated with the DIG labelled PTB2 probe in the presence of non-specific DNA competitors (salmon sperm, lane 2 and poly dI-dC, lane 3) and a specific competitor (PTB2 unlabelled probe, lane 5 and 6). The PTB2 unlabelled probe was used as a competitor at an excess molar ratio of 50:1 (lane 5) and 100:1 (lane 6) to the labelled probe. The previously standardized negative control (random probe) was run in parallel to verified specificity (lane 4). **(C)** EMSA assay of PSR1 TF binding to the mutant PTB2 probe. PSR1 protein was incubated either with the WT PTB2 probe (lane 2 and 3) or the mutant PTB2 (lane 4) and run in parallel in the EMSA gel.

3.2.2.1 Putative PSR1 targets related with lipid and starch biosynthesis

As mentioned earlier, starch and lipid biosynthesis are highly upregulated processes under P-starvation (Hu et al., 2008). Interestingly, transcripts of several genes involved in the biosynthesis of these molecules that are upregulated under P-starvation conditions, were found downregulated in *psr1* mutant lines (Bajhaiya *et al.*, 2016b). To determine whether PSR1 is directly regulating starch biosynthesis under P starvation, three genes of the starch metabolism were tested as putative targets for PSR1. The first one is the Soluble Starch Synthase 1 (SSS1), a key enzyme in the biosynthesis of starch, that participates in the amylopectin synthesis (Maddelein et al., 1994). A DNA probe was designed based on SSS1 promoter sequence 3833 base pairs upstream the translation origin, were the putative P1BS domain was found (Fig. 3.7 A). The EMSA was performed according to the previously standardized conditions. Results indicated that PSR1 can physically bind to the promoter region of the SSS1 gene in the P1BS domain shown by the shift in the migration of the probe in the EMSA gel in the presence of both nonspecific DNA competitors (lane 2 and 3, Fig. 3.7 B). The complex was not eliminated when an excess of the SSS1 unlabelled probe was used. However, the addition of a molar ratio of 1:100 of the unlabelled specific competitor decreased the complex formation, observed as a decrease in the band intensity of the shift (lane 6, Fig. 3.7 B), suggesting the specificity of the PSR1-SSS1 probe binding. The fact that the complex was not eliminated and only a small decrease of the complex formation was observed in the presence of an excess of the specific competitor, may be due to experimental error (unlabelled probe should be added to the binding reaction before the labelled probe, and it was added at the same time). The absence of a shift in the reaction of the random probe with PSR1 confirmed the specificity of the PSR1-SSS1 probe complex (lane 4, Fig. 3.7 B). Another EMSA gel showed the same previous results plus results of the PSR1 binding reaction to the mutant SSS1 probe (Fig. 3.7 C), showing that the complex formation was dramatically decreased by the substitution of 2 nucleotides in the P1BS of the SSS1 probe (lane 4, Fig. 3.7 C). Problems in the transference of this gel to the nylon membrane may have masked the intensity of the PSR1 binding to the mutant probe, however, the comparison of band intensity with the other lanes in the same gel, makes the later assumption credible.

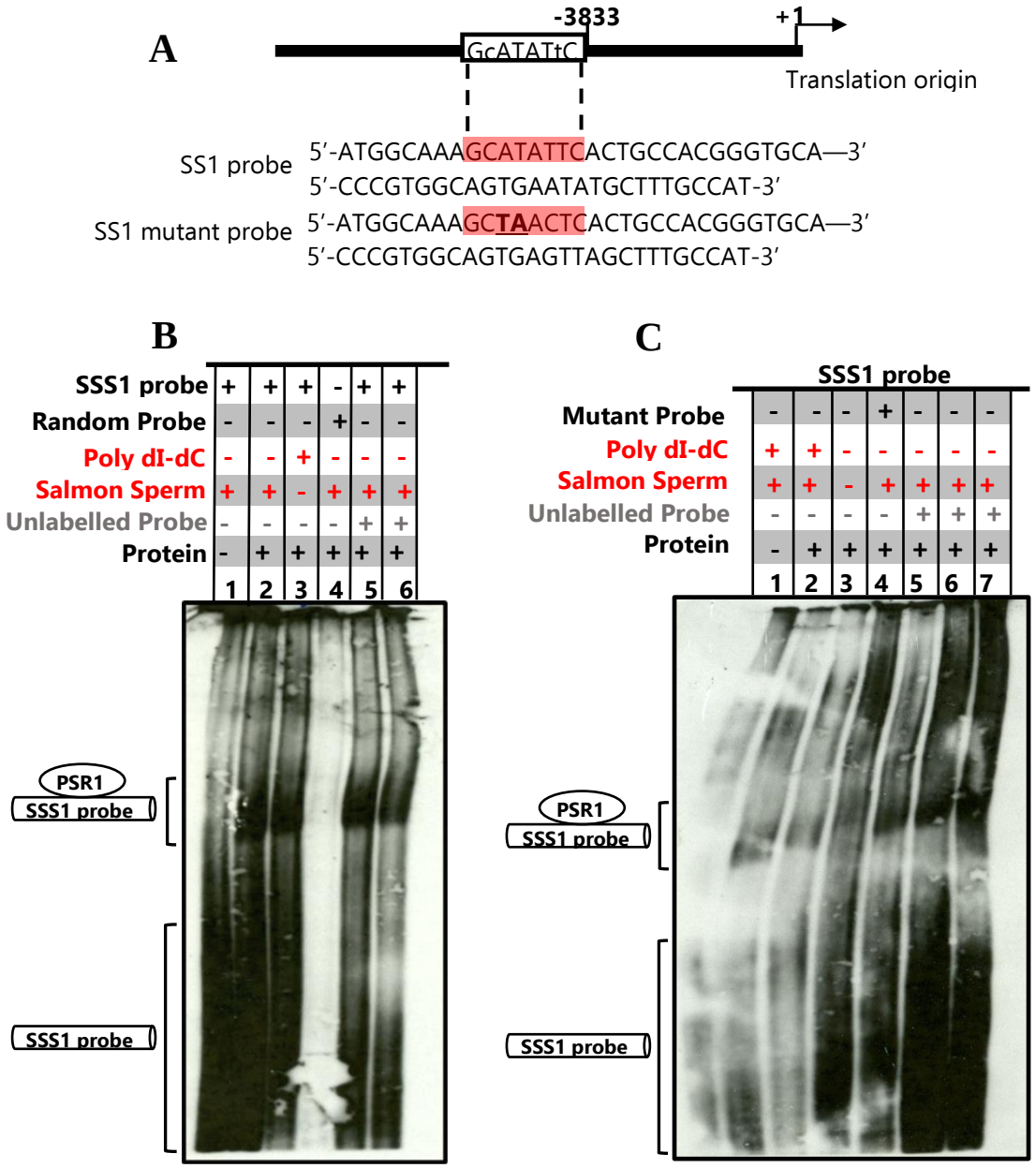

Figure 3.7: EMSA assay showing the binding of PSR1 to the SSS1 target sequence. EMSA assay showing the binding of PSR1 to the SSS1 target sequence. **(A)** Diagram of the probes designed in the promoter region of the SSS1 gene that contained the P1BS putative binding site. Mutant SSS1 probe was design by the substitution of 2 nucleotides within the P1BS. **(B)** EMSA assay of PSR1 TF binding specifically to the SSS1 probe. PSR1 purified protein was incubated with the DIG labelled SSS1 probe in the presence of non-specific DNA competitors (salmon sperm, lane 2 and poly dI-dC, lane 3) and a specific competitor (SSS1 unlabelled probe, lane 5 and 6) at an excess molar ratio of 50:1 (lane 5) and 100:1 (lane 6) to the labelled probe. The previously standardized negative control (random probe) was run in parallel to verified specificity (lane 4). **(C)** EMSA assay of PSR1 TF binding to the mutant SSS1 probe. PSR1 protein was incubated either with the WT SSS1 probe or the mutant SSS1 (lane 4) and run in parallel in the EMSA gel. Competition assay was performed by adding SSS1 unlabelled probe at an excess molar ratio of 50:1 (lane 5), 100:1 (lane 6) and 200 (lane 7) to the WT SSS1 labelled probe.

Similar results were obtained with the Soluble Starch Synthase 5, SSS5, another enzyme participating in the starch biosynthesis, specifically in the amylopectin synthesis (Ball and Morell, 2003). This gene also contained a putative P1BS domain in its promoter region from which the probe was designed (Fig. 3.8 A). This time an extra control was included to test whether the concentration of added protein was optimal. Ideally the optimal concentration of protein is the lowest that allows the specific formation of protein-DNA probe complex. Higher concentration of protein in the assay can produce unspecific bindings or present strong signal backgrounds, making it difficult to detect any shift produce by the protein-DNA complex. Results showed that PSR1 protein binds the SSS5 probe, at a minimal concentration of 20 nM, but not to the random probe at the same concentration, confirming the specificity of the binding (Fig. 3.8 B). The addition of an excess of unlabelled SSS5 probe did not help to show the specificity of the probe, creating a strong smear pattern (lane 6, Fig. 3.8 B). Increasing the concentration of the probe also increases the concentration of the buffers that came with it and this may have interfered with the normal migration of the probe producing this unwanted smear pattern. It is noteworthy that this was one of the first gel ran to test the EMSA assay, hence the protein concentration standardization. The protein concentration and the amount of unlabelled probe added were then optimize for the rest of the tests previously showed, considering the concentrations used in this gel (20 nM of PSR1 purified protein and lower concentrations of unlabelled probe).

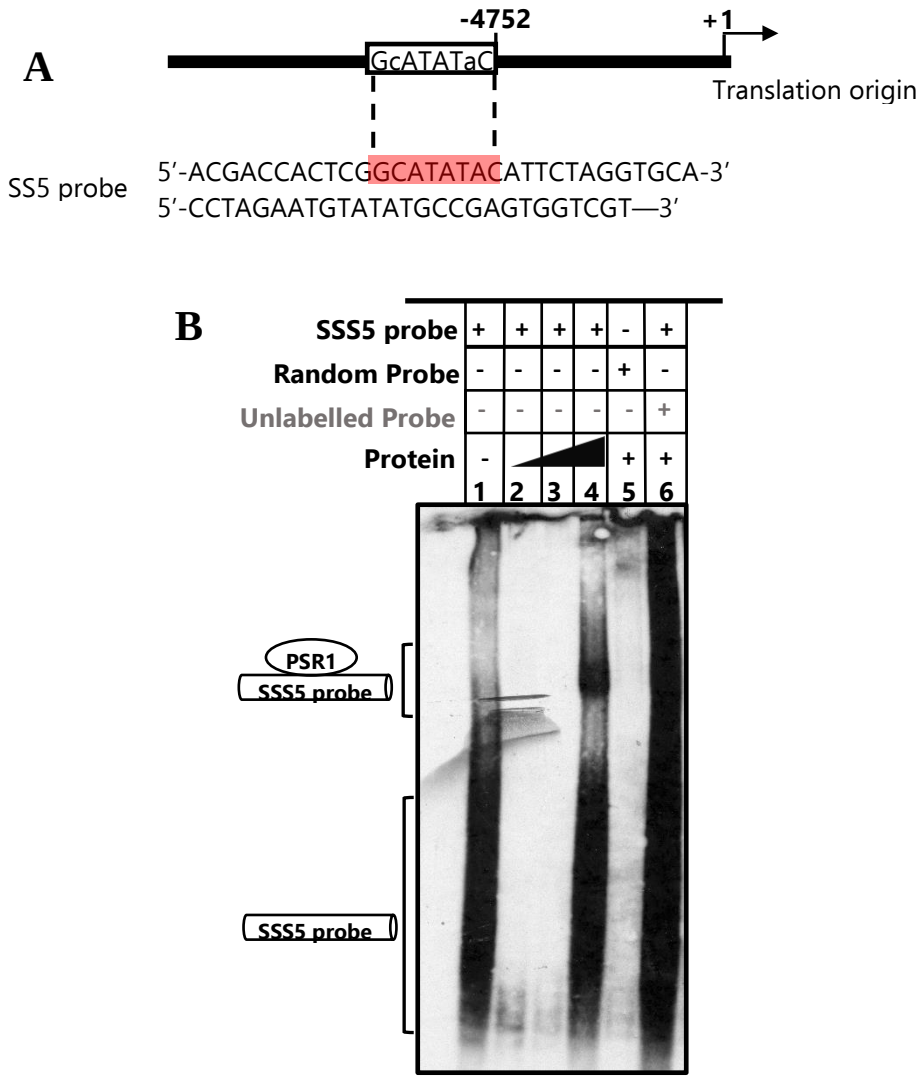

Figure 3.8: EMSA assay showing the binding of PSR1 to the SSS5 target sequence. **(A)** Diagram of the probes designed in the promoter region of the SSS1 gene that contained the P1BS putative binding site. **(B)** EMSA assay of PSR1 TF binding specifically to the SSS5 probe. Increasing concentrations of the PSR1 purified protein (5, 10 and 20 nM) were incubated with the DIG labelled SSS5 probe and a specific competitor (SSS5 unlabelled probe, lane 6) at an excess molar ratio of 300:1 (lane 6) to the labelled probe. The previously standardized negative control (random probe) was run in parallel to verified specificity (lane 5).

Contrary with the patterns seen with the other two starch biosynthesis enzymes previously tested (SSS1 and SSS5), SP2, a gene that encodes for the PhoB plastidial starch phosphorylase (Dauvillé *et al*., 2006), showed different results. The EMSA gel exhibited no specific shift in any of the lanes in which the PSR1 protein was incubated with the SP2 probe (Fig. 3.9 B). The binding reaction incubated with salmon sperm as a DNA competitor, showed a smear pattern (lane 2, Fig. 3.9 B) which was completely eliminated by the use of the poly dI-dC DNA competitor (lane

3, Fig. 3.9), indicating that the smear pattern could be explained by unspecific binding. Although the results of this gel were not as clean as expected and the noise background could be influencing the ability to detect a clear shift, it is important to mention that none of the repetitions made with this probe showed any clear shift in the migration of the probe, as seen previously (data not shown).

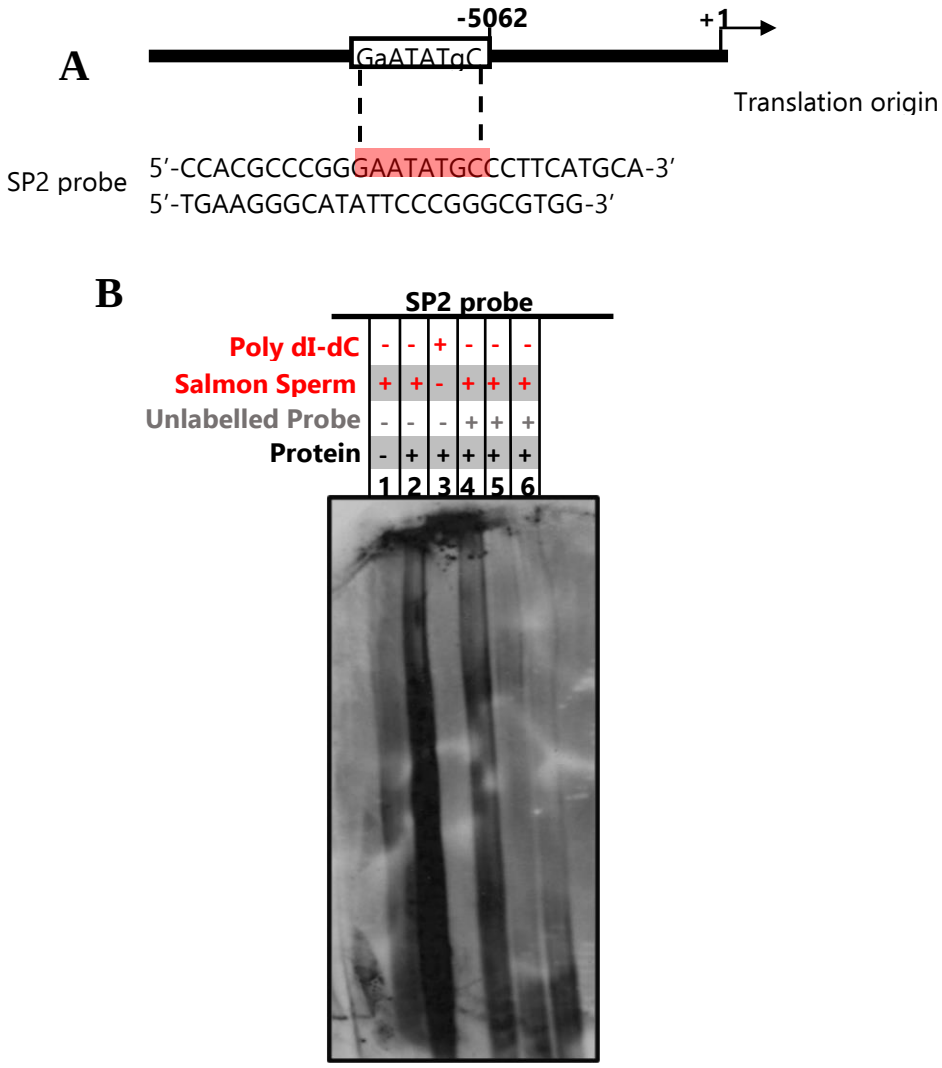

Figure 3.9: EMSA assay showing the binding of PSR1 to the SP2 target sequence. **(A)** Diagram of the probes designed in the promoter region of the SP2 gene that contained the P1BS putative binding site. **(B)** EMSA assay of PSR1 TF ant SP2 probe. PSR1 purified protein was incubated with the DIG labelled SP2 probe in the presence of non-specific DNA competitors and a specific competitor (SP2 unlabelled probe, lane 4-6) at an excess molar ratio of 50:1 (lane 5), 100:1 (lane 6) and 200:1 to the labelled probe.

To test whether PSR1 was binding the promoter region of a TAG biosynthesis gene, the EMSA assay was performed with a probe designed against the promoter region sequence of the glycerol-3-phosphate dehydrogenase 4 (GPD4) gene (Fig. 3.10 A). This gene encodes for an enzyme involved in reduction of dihydroxyacetone phosphate to glycerol 3-phosphate, a precursor of TAG biosynthesis (Driver *et al.*, 2017). The EMSA results indicated that the PSR1 proteins cause a mobility shift of the GPD4 labelled probe, in the presence of either of the DNA unspecific competitors (lane 2-3 Fig. 3.10 B), indicating the specific binding of GPD4 to the P1BS domain within the promoter of the GPD4 gene. The decrease in the intensity of the shift caused by the addition of an excess of unlabelled GPD4 probe, confirmed the specificity of the binding (lane 5, Fig. 3.10 B). The incubation of PSR1 with the mutant GPD4 probe in the binding reaction also produced a shift in the migration of the probe, with no apparent intensity differences with the WT probe, indicating that the mutations incorporated into the probe did not affect the affinity of the binding (lane 4, Fig. 3.10 B).

Finally, the EMSA reactions with PSR1 and PTB2, SSS1 and GDP4 probes were run in parallel in a polyacrylamide native gel to emphasise the specificity of the PSR1-probe binding. As previously observed, PSR1 produced a migration shift with all of the three probes in the presence of either of the unspecific DNA competitors (salmon sperm or poly dI-dC) (Fig. 3.10 C). By running in parallel the EMSA of the 3 different probes with PSR1, it is evident that the migration shift is produced at the same level in the gel, indicating that it is PSR1 causing the shift of the probes and not due to other potential unspecific binding events.

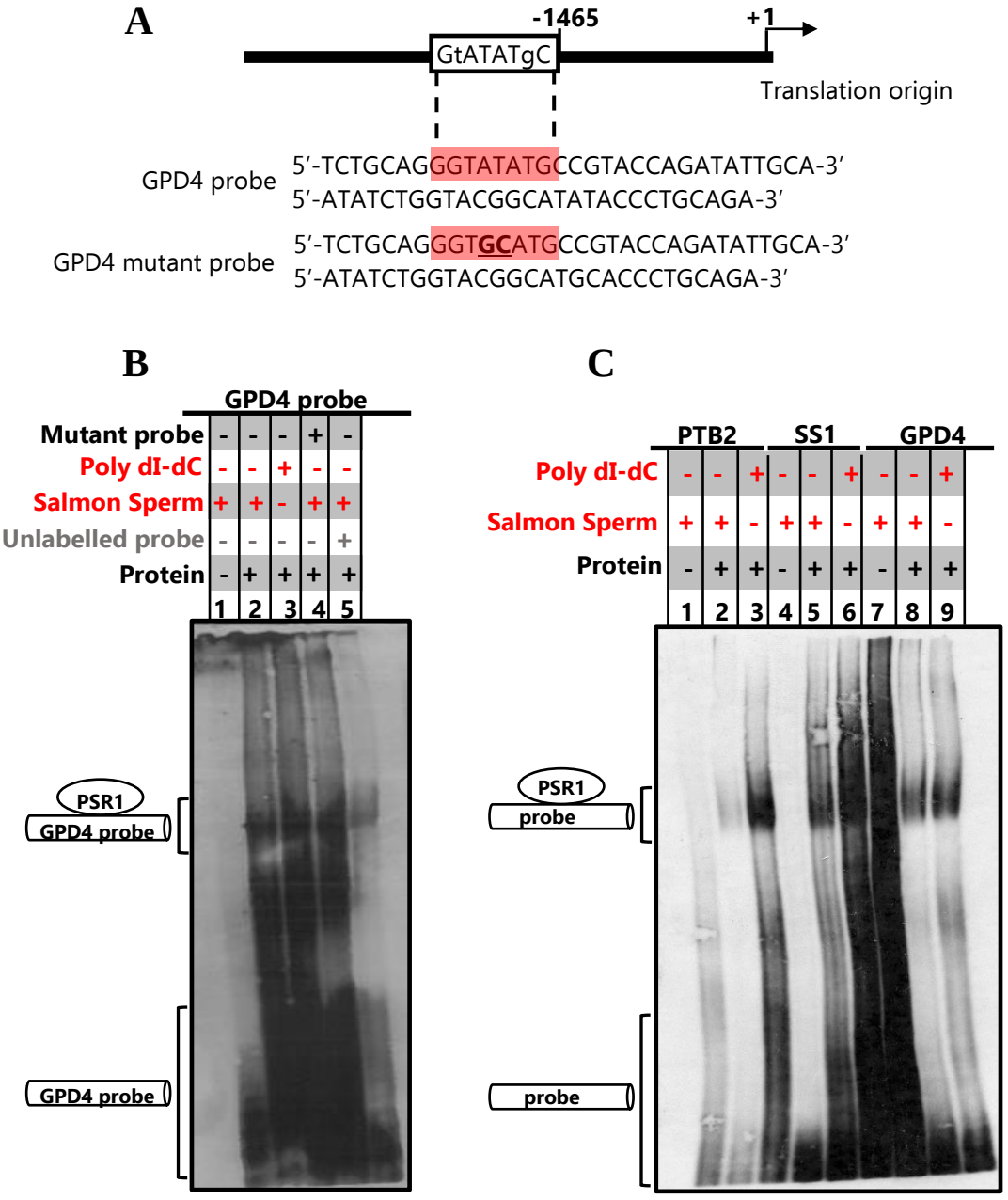

Figure 3.10: EMSA assay showing the binding of PSR1 to the GPD4 target sequence. **(A)** Diagram of the probes designed in the promoter region of the GPD4 gene that contained the P1BS putative binding site. Mutant GPD4 probe was design by the substitution of 2 nucleotides within the P1BS. **(B)** EMSA assay of PSR1 TF binding specifically to the GPD4 probe. PSR1 purified protein was incubated with the DIG labelled GPD4 probe in the presence of non-specific DNA competitors (salmon sperm and poly dI-dC) and a specific competitor (GDP4 unlabelled probe, lane 5) at an excess molar ratio of 200:1 to the labelled probe. Mutant GPD4 probe was run in parallel to ensure binding specificity (lane 4). **(C)** EMSA showing the specific binding of PSR1 to PTB2, SSS1 and GPD4 probes in the presence of either of the DNA unspecific competitors (salmon sperm and poly dI-dC).

3.3 Discussion

The mechanism underlying P starvation is well understood in plants, however, little is known about this process in microalgae. So far, experimental evidence has identified PSR1 TF as a key component in the activation of mechanisms involved in the acclimation of cells to P starvation. However, there is no information on how PSR1 is regulating these processes. A step forward was recently made by use of transcriptomic data reporting dramatic changes in gene expression in response to P starvation that was PSR1-dependent. Interestingly, besides from transcripts well known to be involved in the P starvation acclimation, transcripts involved in lipid and starch biosynthesis were found to be expressed in a PSR1 dependent manner (Bajhaiya *et al.*, 2016b). Thus, to further understand the role of PSR1 in the P starvation regulation mechanism, the binding function of PSR1 was validated. The first part of this work presented the challenges faced with the optimization of the EMSA assay as a tool to test the PSR1 TF activity. The second part reports the direct binding of PSR1 to the P1BS domain in the promoter of genes involved in P scavenging mechanisms, and in lipid and starch metabolism.

2.2.1 EMSA assay standardization

EMSA is a commonly used method for studying TF that was originally developed in 1981 and since then has evolved together with other molecular biology technologies (Jiang *et al.*, 2011). The detection and production of the TF protein are crucial for the performance of this assay. Bacterial protein expression has been the most widely used protein expression system for its versatility and low cost. However, it has its limitations when it comes to production of eukaryotic proteins. Some transcription factors require post-translational modifications for their activity that bacterial systems are not able to perform such a glycosylation and disulphide bonds (Sahdev *et al.*, 2008). PSR1 TF contains several putative post-translational modification sites; 12 phosphorylation sites, one glycosylation site and many myristylation sites (Wykoff *et al.*, 1999). Therefore a eukaryotic system of protein expression was the first option to consider production of PSR1. The ‘wheat germ cell free couple transcription/translation system’ was chosen for being recognised as one of the most advanced eukaryotic systems for protein expression. However, no successful results were obtained, and the protein was not detected either in the

soluble or insoluble fraction. The unsuccessful results may be due to the low yield reported by cell free systems, which are in part attributable to the low efficiency of translation initiation (Madin *et al.*, 2000).

One of the major drawbacks of the coupled transcription/translation system is that the RNA produced by a bacteriophage RNA polymerase is not 5'-capped. Uncapped RNAs are less efficient for translation than the capped counterpart. Weber *et al*. (1977), showed that uncapped transcripts are translated at very low efficiencies in wheat germ systems. Moreover, it has been reported that translation by wheat germ extracts is more cap-dependent than translation using other cell free extract systems, leading to lower yields in protein expression. Although different strategies have been implemented to overcome this with successful outcomes (Strenkowska *et al.*, 2016), for certain transcripts, capping of the RNA may be beneficial. Apparently, the effect of capping on translation efficiency is template-dependent (Andreev *et al.*, 2009). Other potential aspects behind low productivity of cell free systems that may have contributed to the failure of PSR1 expression is depletion of essential substrates needed for protein synthesis (Li *et al.*, 2017). During protein synthesis in growing cells, components like nucleotides, amino acids and the components required for regeneration of ATP are continuously being regenerated. However, in cell free systems, these components are added in set concentrations that can be inefficiently utilized causing depletion of key components in the process. Overall, even though the wheat germ cell free system has proven to be successful for purification of proteins from different organisms, it appears to be that the success of the protein expression still depends on the nature of the protein. Currently there is no data regarding PSR1 protein expression using cell free systems. Previous studies have reported the use of PHR1 expressed in the *E. coli* BL21(D3) for EMSA assays (Sun *et al.*, 2016). Based on these studies it was rational to hypothesise that even though eukaryotes translational modifications could be important for the function of the TF, it does not affect its ability to bind its DNA targets. Considering the homology between PHR1 and PSR1, especially in their DNA binding domains, it was then decided that using *E. coli* was a good alternative for the PSR1 expression.

Probe labelling was another crucial parameter of the EMSA assay that required optimization (Smith and Delbary-Gossart, 2001). ^{32}P has been traditionally used as a labelling method that produces strong signals, facilitating the detection of small amounts of biomolecules. However, the worries about the safety of

manipulating radioactive molecules have pushed the development of other alternatives such as fluorophores, biotin or digoxigenin (Holden and Tacon, 2011). This work presented the standardization of digoxigenin, a steroid molecule isolated from plants, for the 3' end labelling of DNA probes for EMSA assay. The advantage of the DIG end labelling using DIG tailed oligonucleotides is the increased sensitivity due to the incorporation of several small single-stranded DIG-nucleotides (Kass *et al*., 2000). Following the manufacturer's protocol, a long tail of DIG labelled nucleotides was added. However, even though the probe labelling was successful and with a very strong signal, the performance of the probe was not as expected when it was run on the EMSA native polyacrylamide gel. The long tail DIG label generated a smear migration pattern in the gel that did not improve by decreasing the probe concentration or by running the gel for longer times. These smear patterns are typical in gel electrophoresis that uses DNA samples that have been degraded for different reasons (nucleases, restriction enzymes, DNA samples from apoptotic cell), generating hundreds, or even thousands of DNA pieces of all fragment sizes (Suman *et al*., 2012). Possibly the addition of a mixture of DIG-dUTP and dATP to generate the long DIG label tail, produced a heterogeneous mixture of DIG-tail sizes that could explain the observed smear. However, the addition of only DIG-dUTP that only added up to 3 labelled nucleotides to the tail produced much better results, with the probe running almost to the end of the gel when the polyacrylamide concentration was increase to 12%. In conclusion, the most appropriate method for labelling probes with DIG for EMSA assay is to add a short tail of up to 3 DIG-dUTP nucleotides. This methodology will generate the desired probe migration and a strong enough signal to be detected by the anti-DIG antibody.

2.2.2 EMSA assay for the determination of PSR1 putative target genes

As outlined in the Introduction, *C. reinhardtii* has evolved specific mechanisms to respond to P-starvation that allows them to survive extended periods of depletion of this essential macronutrient. The PSR1 TF is the only regulator from this species identified to date as a key component for the execution of these specific responses. However, there are no studies about how PSR1 regulating these responses. The idea that microalgae may have simpler regulation mechanism upon P starvation, suggests that PSR1 may be one of the few or the only master regulator

coordinating P starvation responses, making it an attractive target for genetic manipulation. Transcription factor engineering (TFE) approach through the manipulation of TFs, aims to have a bigger impact than the manipulation of a single gene target, by regulating several sets of genes at the same time (Bajhaiya *et al*., 2017). However, the importance of PSR1 for the survival of cells under P starvation is so crucial that mutations of this TF that produce impaired function are highly deleterious. For example, it has been reported that *psr1* mutant cells, with loss of function of the TF, exhibit impaired growth and are much more sensitive to other stresses that could even lead them to death (high light exposure) (Shimogawara *et al.*, 1999). Therefore, knowing which are the key targets of PSR1would allow a better understanding of the PSR1 regulatory processes. This study showed for the first time the direct binding of PSR1 to targets involved in P scavenging and starch and lipid metabolism through binding to the PB1S motif in their promoter region.

Similar to other organisms, the most typical specific response to P-starvation includes the promotion of P acquisition, through the increased expression of high affinity P transporters and phosphatases (Grossman and Aksoy, 2015). One family of P transporters found highly activated upon P starvation are the PTB genes, a family of high affinity Na^+/P symporters. *C. reinhardtii* contains a total of 10 potential P transporters belonging to the PTB family. Genome base approaches have reported the upregulation of at least 7 of these genes under P starvation conditions but not in the *psr1* mutants (Moseley *et al*., 2006; Bajhaiya *et al.*, 2016b). Interestingly, the PTB transcripts showing a higher upregulation under P starvation contained P1BS in their promoter region, leading to the conclusion that PSR1 had a key role in the regulation of P transporter expression (Bajhaiya *et al.*, 2016b). To sustain this conclusion, the DNA binding of PSR1 to the P1BS domain in the promoter of the PTB2 was tested using the EMSA assay. Results confirmed the previous data, showing that PSR1 has a key role in the regulation of PTB2 expression, through the direct and specific binding to the P1BS in the promoter region, activating the transcription of this gene. Interestingly, the two transporters from the PTB family that are most upregulated under P starvation correspond to PTB2 and PTB4, and are the only ones from the family that present the P1BS in the promoter regions of their genes (Bajhaiya *et al.*, 2016b). Due to the strong binding specificity observed between PSR1 and the P1BS in the PTB2 gene, and PSR1-dependent expression reported by these two transporters, it is possible that PSR1 also binds to the P1BS in

the promoter region of PTB4. However, an EMSA assay using a probe design for the PTB4 promoter would be necessary to corroborate this assumption. Nevertheless, these results prove that at least one of the members of the PTB gene family, PTB2, is directly regulated by PSR1. This together with the fact that PTB2 is the high affinity transporter most upregulated under P starvation, suggests that PTB2 may be the transporter that contributes the most to the high affinity P uptake in P starved cells. Moreover, these findings reinforce the assumptions of PSR1 being a transcriptional activator. This is a theory that until now was only based in sequence analysis and the PSR1-dependent expression of a high number of transcripts under P starvation.

It is noteworthy that not all of the transcripts involved in the P stress response are positively regulated by PSR1. Two members of the family of PTA-type P transporters in *C. reinhardtii*, whose function is not fully determined, exhibited transcript levels that either remained unchanged or were more than >2 fold down-regulated under P starvation. However, even though these two PTA-type P transporters showed PSR1 dependent expression, they do not contain a P1BS domain in their promoters (Bajhaiya *et al.*, 2016b). Possibly, PSR1 may be activating directly its targets through its TF activity but indirectly repressing the expression of others. This duality of action of directly activating process but indirectly repressing others at the same time, has already been reported in PHR1 in Arabidopsis (Bustos *et al.*, 2010). Clearly, more experiments including a broader analysis of P starvation related genes are required to clarify the extent of the PSR1 regulation in the activation of P acclimation mechanisms.

The accumulation of storage molecules in the form of starch and lipids is another mechanism well known to be activated under P starvation in different organisms (Pant *et al.*, 2015; Y. Wang *et al.*, 2018). Accordingly, genes involved in starch and lipid biosynthesis are highly upregulated under P starvation (Bajhaiya *et al.*, 2016b). However, little is known about the regulation of these mechanisms under P starvation. Recent studies in Arabidopsis have identified PHR1 as a master regulator of the P starvation response also linked in the regulation of lipid accumulation (Pant *et al.*, 2015). On the other hand, studies in *C. reinhardtii* seem to be a step forward. PSR1 has been identified as key for the lipid and starch accumulation under P starvation. Psr1 mutant cells showed impaired accumulation of starch and lipids under P starved conditions. Moreover, a transcriptomic analysis of more than 17,000 transcripts under P starvation, identified clusters of starch and lipid

genes expressed in a PSR1 dependent manner (Bajhaiya *et al.*, 2016b). Based on these analyses, the remaining question was whether PSR1 was directly regulating these genes. To answer this question, starch and lipid transcripts showing the more dramatic upregulation under P starvation but not in the *psr1* mutants, were chosen and PSR1 binding ability to their promoter regions was tested using EMSA assay. Results showed that PSR1 binds specifically to the P1BS in the promoter region of the TAG biosynthesis enzyme glycerol-3-phosphate dehydrogenase 4 (GPD4). This enzyme has a key role in the TAG biosynthesis and its overexpression in *P. tricornutum* diatom showed increased accumulation of glycerol and neutral lipid (Yao *et al.*, 2014). The finding that PSR1 directly activates the transcription of this gene, evidence for the first time the direct role of a transcription factor in P stress response in lipid biosynthesis. Although it can be concluded that PSR1 is directly regulating TAG biosynthesis under P starvation conditions, more experiments are required to understand the extent of the control of this TF in this biosynthesis pathway. However, we know that, at least partially, this is through the transcriptional activation of GPD4. Similarly, EMSA assay showed that PSR1 binds specifically to the promoter region of the soluble starch synthase (SSS1, SSS5), key starch biosynthesis genes. This confirms the direct role of PSR1 in the regulation of the starch biosynthesis. On the contrary, PSR1 showed no binding to the P1BS in the promoter region of SP2, another enzyme reported to be key in the accumulation of storage starch. This may be explained by the fact that PSR1 is considered a transcriptional activator and SP2 has been found downregulated in cells that overexpress PSR1 under P starvation (Bajhaiya *et al.*, 2016b). Therefore it may not be surprising that PSR1 does not bind to the promoter of a gene that is being repressed. It is important to mention that even though this result confirms that PSR1 is not directly regulating SP2, it does not mean that it does not have a role in its regulation. It is possible that the reported increase in the starch content observed in cell overexpressing PSR1 activates a feedback regulation that represses the expression of genes involved in starch biosynthesis. PSR1 could be indirectly activating a feedback mechanism that repress genes involve in starch synthesis when there is an overaccumulation of starch. Given that PSR1 overexpressing cells accumulate much more lipids than the WT counterpart under P starvation, this would explain why WT cells overexpress SP2 and PSR1 overexpressing cell downregulated under the same P starvation condition. Similar scenarios have been identified before,

in which high-starch mutant *C. reinhardtii*, repress the expression of key starch biosynthesis genes in proportion with the amount of lipids accumulated (Koo *et al.*, 2017).

The importance of the P1BS element has been experimentally demonstrated in Arabidopsis as crucial for the transcription of several P starvation responsive genes (Bustos *et al.*, 2010). However, how important this element is for the PSR1 regulation under P starvation remains obscure. Studies of OsPHR2, a PHR1 homologue in *Oryza sativa* (rice) have demonstrated that variants of the P1BS elements, denominated as "P1BS-like", exhibit higher affinity for the OsPHR2 binding than to the P1BS (Ruan *et al.*, 2015). P1BS-like elements have also been reported in the promoters of P starvation responsive genes orthologous in different species (Ruan *et al.*, 2015). Single nucleotide mutations are common during evolution and the promoter regions of the P starvation responsive genes seems no exception. The adaptation to the availability of P in different types of environments may have pushed the evolution of the P1BS element to different variants. This could be the explanation of why PSR1 stills bind specifically to the mutant versions of the P1BS tested in this study. It may be possible that the P1BS evolved from a similar element in microalgae and the specific element to which PSR1 binds with more affinity is still not determined. Interestingly, studies on P1BS in Arabidopsis have reported that not only is the presence of this element necessary for the transcriptional activation of P starvation responsive genes but also the position of it within the promoter region (distance from the translation initiation site) and the number of P1BS elements (Sun *et al.*, 2016). Moreover, the sequences that flank the P1BS have also reported to be crucial for the function of this elements and not all the P1BS elements on a P starvation responsive gene are indispensable (Oropeza-Aburto *et al.*, 2012). Apparently, the functionality of the P1BS element seems complicated. However, there is no doubt about the importance of this element for the transcriptional activation P starvation responsive genes, Indeed, it can be noted that P1BS has been identified as an element specific for activation of genes under P starvation and not under any other stress (Bustos *et al.*, 2010).

Although this study provided a first step in the study of the P1BS element, confirming the binding of PSR1 to it, whether this binding is essential to activate the transcription of P starvation responsive genes is still not yet know. More experiments including a broader analysis of P starvation related genes and in vivo analysis of the

PSR1 binding activity, are required to clarify the role of P1BS in controlling P responses. The understanding of the cis regulatory elements through which PSR1 is regulating its downstream targets will give tools to TFE for the manipulation of key genes involve in P starvation response, for either improving P uptake or increasing the synthesis of lipids and starch storage molecules. The creation of tunable cis elements promoters that respond to P starvation have already shed some light for TFE. A recent investigation showed that the genetic manipulation of a P1BS element in the promoter region of a P starvation responsive gene, improved the P uptake in *Oryza sativa* (Ruan *et al.*, 2015). Although these technologies still have not been applied to improve the starch and lipid accumulation in microalgae, their success opens the horizons for the TFE to enhance the yield of carbon storage molecules for the biofuel industry.

Chapter 4

Analysis of the MYB transcription factor family in *Chlamydomonas reinhardtii* and prediction of potential roles

4.1 Introduction

Unicellular organisms are constantly being challenged by biotic and abiotic changes in the environment and adaptability is key to their survival. Regulation of gene expression provides a mechanism by which organisms respond to these changes, allowing them to adapt (Bleuven and Landry, 2016). In eukaryotes, the regulation of gene expression can occur at different levels ranging from the accessibility of chromatin to the post-translational modification of proteins (Cooper, 2000). However, most of the regulation of gene expression occurs at the level of transcription, where Transcription Factors (TFs) play a key role. These proteins recognize specific DNA motifs in gene regulatory regions and activate or repress their expression. Based on the similarities of their DNA binding domains, TFs have been classified into families (Stegmaier *et al.*, 2004). The MYB TFs constitute one of the largest families that are functionally diverse and ubiquitously expressed across eukaryotic organisms (Ambawat *et al.*, 2013). The name MYB comes from the first identified member of the family in the genome of an avian myeloblastosis virus (v-MYB) (Klempnauer *et al*., 1982). Evidence indicated that v-MYB may have originated from a vertebrate gene that mutated when it became part of the virus (Stracke *et al.*, 2001). This theory was later supported by the identification of three v-MYB-related proteins in vertebrates; c-MYB the proto-oncoprotein in humans and A-MYB and B-MYB (Nomura *et al.*, 1988; Lipsick, 1996). Since then, MYB domain proteins have been found in all eukaryotic organisms studied to date and they have been involved in a myriad of regulatory processes like responses to biotic and abiotic stresses, development, differentiation, metabolism, defence, etc. (Rubio *et al*., 2001; Liu *et al*., 2015; Liu *et al.*, 2017).

Members of the MYB TF family are characterized by their conserved DNA binding domain that consists of up to four imperfect repeats, each with approximately 52 amino acid residues containing three regularly spaced tryptophan (or hydrophobic) residues that together form a hydrophobic core (Ambawat *et al.*, 2013) (Fig. 4.1). NMR and X-ray crystallography has confirmed that each imperfect repeat encodes three α-helices, with the second and third helices forming a helix-turn-helix (HTH) structure necessary for DNA binding (Ogata *et al.*, 1996). It has been reported that only two MYB repeats are sufficient for the DNA binding activity, and that they are closely packed together in the major DNA groove (Sakura *et al.*,

1989). The third helix of each repeat is the DNA-recognition helix that identifies cis-elements in target genes and makes direct contact with DNA. In the plant kingdom, MYB proteins constitute one of the largest families of TFs with remarkably more members than those in fungi or animals (Riechmann *et al.*, 2000). For this reason, this TF family in plants has been the focus of much attention and several members have been characterized. The increasing number of sequenced plant genomes and development of databases such as Plant Transcription Factor Database (PlnTFDB) has permitted the accumulation of a large amount of data on MYB TFs in plants, providing a better understanding of the evolution and function of this large family of transcription factors. Putative MYB TFs have been identified in a variety of organisms such as *A. thaliana* (Dubos *et al.*, 2010), maize (Zea mays) (Grotewold *et al.*, 2000), rice (*Oryza sativa*) (Yang *et al.*, 2014) , grapevine (*Vitis vinifera* L.) (Holl *et al.*, 2013), petunia (*Petunia hybrida*) (Albert *et al.*, 2011), snapdragon (*Antirrhinum)* (Moyano *et al.*, 1996), poplar (*Populus tremuloides*) (McCarthy *et al.*, 2010) and apple (*Malus domestica*) (Wu *et al.*, 2017), giving new insights of the diversity, evolution and roles of MYBs. Consequently, the majority of what is known about MYB TFs relates to land plants.

Classification and functions

MYB protein can be classified into four types according to the number of adjacent imperfect repeats (one, two, three or four; Figure 4.1). The prototypic and first characterized MYB protein c-MYB contains three repeats, R1, R2 and R3, and repeats from other MYB proteins are named according to their similarity to R1, R2 or R3 of c-Myb (Stracke *et al.*, 2001; Dubos *et al.*, 2010). The retroviral version of c-MYB, v-MYB, its ortholog in mouse as well as most of the plant MYB only contains the R2 and R3 repeats, which have been shown to be sufficient for the DNA binding activity (Ogata *et al.*, 1996). The MYB domains of animals and plants are different in several aspects such as high variation of the amino acid sequences within the domain. However, they all tend to conserve the regular spaced tryptophan (W) residues, which are key for the helix-turn-helix structure that gives them the DNA-binding capacity (Figure 4.1) (Ogata *et al.*, 1992).

1R-MYB proteins are a type of MYB with only one repeat (or sometimes just a partial one). As a typical member of the MYB family, the majority of them contain

the three regularly spaced Trp (W) residues characteristic of MYB repeats. However, studies of MYB1R proteins in plants showed that either the first or the third W residue of the repeat is quite often replaced by other hydrophobic residues such as Ala (A) and Tyr (T) respectively (Rosinski and Atchley, 1998). Also, sequence analysis has revealed that the location of the MYB domain in 1R- MYB proteins is less conserved within members of the family, with this being located either in the N-terminal, middle region or C- terminal region of the protein (Du *et al.*, 2012). Similarly, sequences outside the MYB domain are quite divergent with non-conserved specific motifs. As a result, this type of MYB is considered the most variable one, and is usually confused with the MYB-related protein family, due to their high similarity (Safi *et al.*, 2017). Nevertheless, despite their divergent sequences, specific consensus sequences within the 1R-MYB domains have been found highly conserved, providing criteria for identifying these types of MYB proteins. Members of this family have been identified mainly with roles in circadian rhythm regulation (Schaffer *et al.*, 2001). Also, 1R-MYBs have been involved in regulation of anthocyanin production (AN2 in petunia flowers) (Quattrocchio *et al.*, 1993) and in development processes as the epidermal cell patterning (Lee and Schiefelbein, 1999).

The second type of MYBs are proteins that contain two MYB repeats, frequently referred as R2R3-MYB family, due to the presence of two MYB repeats similar to the R2 and R3 repeats of c-MYB. This type of MYB represents the most abundant class of MYB in plants, different to animals, where only a few members have been described (Ambawat *et al.*, 2013). Genome-wide studies have identified putative sets of around 157 R2R3-MYBs in the maize genome, 244 genes in soybean, 192 in *Populus* and around 168 in *A. thaliana* (Stracke *et al*., 2001; Wilkins *et al.*, 2009; Du *et al.*, 2012). The massive expansion of this TF family almost exclusively from plants, suggests that the evolution of this family may have helped the development of plant-specific processes. In contrast to 1R-MYBs, the R2R3 family possesses a highly conserved structure with two distinct regions: an N-terminal where the two MYB repeats are located and a diverse C-terminal modulator region responsible for the regulatory activity of the protein (containing transcriptional activation or repression domain) (Stracke *et al*., 2001a). Based on the recurrent presence of conserved amino acid motifs in the C-terminal region, the large

A. thaliana R2R3-MYB family is divided into 22 sub-groups (Stracke *et al*.,2001). This classification originally based on *A. thaliana* R2R3 proteins can also be applied to other plants, although sometimes the subgroups are expanded in number and size in other angiosperm species. The high level of conservation of these motifs and the fact that they appears to be only present in plant MYBs, makes R2R3 proteins easier to identify and classify (Du *et al.*, 2015). The *COLORED1 (C1)* gene of maize was the first R2R3 plant MYB gene identified and found to be responsible for the regulation of anthocyanin biosynthesis (Paz-Ares *et al.*, 1987). After two decades of the identification of this R2R3 MYB, many more have been characterized with diverse roles in plant-specific processes (Ambawat *et al.*, 2013). Interestingly, members of the same subgroups of R2R3 appear to have similar cooperative functions (Li *et al.*, 2016). Several members of the R2R3 MYB proteins have been identified with roles in the synthesis of Phenylpropanoid-derived compounds (Liu *et al*., 2015). Other R2R3 MYBs have been shown to regulate cell wall biosynthesis (McCarthy *et al.*, 2010). In addition to roles in secondary metabolism, R2R3 MYB proteins have been implicated in stress and defence responses to biotic and abiotic factors (Liu *et al.*, 2017), among other functions. In general, these functions appear to be broadly conserved for the R2R3 MYB proteins of the same subgroup in different angiosperms and appear to be involved predominantly in controlling plant-specific processes suggesting that R2R3-type MYB genes have, during evolution, contributed to plant speciation (Ambawat *et al.*, 2013).

The third type of MYB corresponds to proteins with three adjacent MYB repeats, R1R2R3 (3R-MYBs), with members found in most eukaryotic genomes, representing a conserved class of MYBs. In plants, this family is reduced in number compared to the R2R3 with only around five representative members. Studies based on c-MYB and v-MYB have confirmed that only R2 and R3 repeats are sufficient for the DNA-binding activity and R1, although not essential for the binding, might contribute to DNA-binding affinity. Functions of 3R-MYB proteins have mainly been related to cell cycle regulation (Feng *et al.*, 2017).

Finally, the last type of MYB protein is the smallest one and corresponds to the proteins containing four R1/R2-like MYB repeats (4R-MYB) in their amino acid sequences. Only one 4R-MYB protein is encoded in several plant genomes but little is known about the function of this class of MYB factors in plants (Ambawat *et al.*,

2013). Plants contain the four types of MYB, representing the taxon with the highest diversity of MYB proteins.

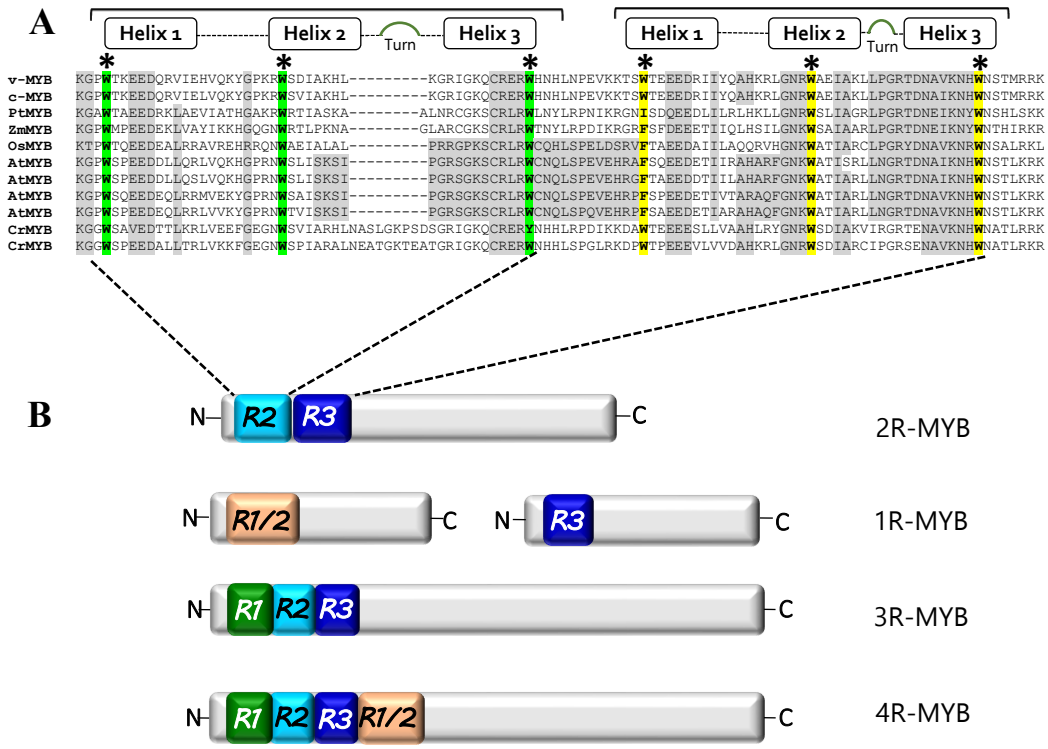

Figure 4.1: MYB DNA binding domain. (A) Sequence alignment of several representative MYB domains from plants and animals. The prototypical first discovered and characterized v-MYB and c-MYB were compared with plant MYB domains of maize (Zm), Populus (Pt), rice (Os), four Arabidopsis (At) and two *C. reinhardtii* MYB proteins. Asterisks indicate conserved Trp (W) residues. Grey highlighted areas indicate conserved motifs. The helix-turn-helix motif formed by the MYB domain is presented on the top. (B) Schematic representation of the four types of MYB proteins.

To date, genome-wide analyses of TFs in microalgae have been conducted, but these are based on bioinformatics predictions and no functional studies about the role of TFs in microalgae have been conducted (Hu *et al.*, 2014; Thiriet-Rupert *et al.*, 2016). Additionally, none of these studies focus on the MYB transcription family but more on a general landscape about the families of TFs in microalgae. On the other side, the MYB TF family in plants, in particular the R2R3 family has been extensively studied (Stracke *et al.*, 2001). Numerous genome-wide analyses of the R2R3 MB TF family in several plant species have been performed and the increasing availability of plant genome sequences has permitted a better understanding of the evolutionary relationships between plant MYBs. This, together with the increasing

amount of molecular data over the last fifteen years, has provided insights about their expression profiles, mechanism of control of their activities and the identification of direct targets of this family in plants. Interestingly, as mentioned before, similar MYB proteins in the same organism have been found to play similar and/or cooperative roles (Li *et al.*, 2016). Furthermore, MYB orthologs from several species have been reported to conserve their functions (Du *et al.*, 2012; Zhong *et al.*, 2015). This information, in combination with available genome sequences of the model organism *C. reinhardtii* and bioinformatics databases such as the PlnTFDB, provided valuable information to analyse the MYB gene family in *C. reinhardtii*. Therefore, one of the aims of this chapter was to perform a genome-wide study of the MYB gene family in *C. reinhardtii*. We identified 28 MYB related genes belonging to three of the families 1R, R2R3 and 3R-MYB. A phylogenetic tree combining the best characterized family of MYBs in plants, R2R3 MYB family, and *C. reinhardtii* predictive R2R3 MYB proteins was constructed to study their evolutionary relationships and putative functions based on comparative analysis with known functions of R2R3 MYBs in plants. Expression profile under different nutrient stresses known to increase the storage molecules as lipids and starch was performed to identify the potential role of selected MYBs in the regulation of these pathways. Emphasis was directed mainly to two genes found highly upregulated in transcriptomic profiles of *C. reinhardtii* under nitrogen and phosphorus starvation. This study represents the first comprehensive study of the MYB family in *C. reinhardtii* and provides information into the potential roles of MYB TF family in microalgae.

4.2 Results

4.2.1 Identification and classification of MYB-related proteins in *C. reinhardtii*

To identify MYB encoding genes, an architecture analysis using the SMART database was performed on the complete genome of *C. reinhardtii*. This database identifies conserved protein domains within proteins sequences using profile-hidden Markov models. Results showed twenty-eight hits, with protein fragments containing conserved MYB domains. These fragments were used as a query to perform a BLAST search against the *C. reinhardtii* Phytozome database for the corresponding protein sequence. Due to the high divergence of MYB genes and the high sequence variability of the domain, the identification of each hit had to be carefully confirmed. For this reason, to confirm the presence of the MYB domain, each matching sequence was searched against the Pfam database and afterwards manually checked for the presence of the domain. Twenty eight MYB-related genes were identified and classified into three groups based on the presence of one, two or three MYB repeats, respectively (Table 4.1). The analysis showed that 14 of the identified genes corresponded to the 1R-MYB type, 12 with two MYB repeats, and only two genes encoded for proteins with three MYB repeats. Due to the discrepancy of the obtained results with previews reports from literature that had identified TFs in the proteome of *C. reinhardtii* (Riaño-Pachón *et al.*, 2008; Thiriet-Rupert *et al.*, 2016), results were checked against the PlnTFDB. This database confirmed the information obtained with the SMART and Pfam databases of the 12 proteins containing two MYB domains and the two others containing three MYB domains. Moreover, these 14 proteins were the only proteins from *C. reinhardtii* recognized as MYB proteins by this database. However, results for the 14 other proteins identified by SMART and Pfam containing 1 MYB repeat, presented discrepancies. The PlnTFDB consider none of these 14 proteins as members of the MYB family, instead 10 were classified as members of the MYB-related protein family and four into the G2-like family. Sequence analysis revealed that even though DNA binding domains (DBD) of these three TF families are similar, there are differences that classified them within one family or the other (Safi *et al.*, 2017). In this study, the information obtained from the SMART and Pfam databases was complemented with the information obtained from the PlTFDB and later manually checked following the information gathered from

latest studies regarding these 3 TF families (Safi *et al.*, 2017). We finally identified 14 1R-MYB, 12 R2R3-MYB and two R3-MYB proteins.

In addition, transcript expression responses of the identified MYBs under phosphorus and nitrogen starvation were obtained from previous studies (Schmollinger *et al.*, 2014; Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b). The analysis revealed that at least 5 of them were differentially expressed in phosphorus and nitrogen stress (Table 4.1). Interestingly, 4 out of 5 genes upregulated under phosphorus starvation that were also shown to be upregulated in response to nitrogen starvation belonged to the R2R3 MYB family, suggesting an important role of this putative MYB TFs under stress conditions that appears not to be specific to the type of nutrient stress. However, the function of these genes under nutrient stress condition is still to be determined.

Table 4.1: Annotated MYB Transcription Factors from *C. reinhardtii* and expression profiles in phosphorus and nitrogen stress (fold change >2 as cut off values). N/I: No information. [a]Putative MYB domain TF identified by the Plant Transcription Factor database. [b] (Bajhaiya *et al.*, 2016b) .[c] (Schmollinger *et al.*, 2014). [d] (Gargouri *et al.*, 2015). [e](Ngan *et al.*, 2015). Genes highlighted in green will be the focus of later experiments.

Expression profile			
Type	**Gene code**	**Phosphorus Starvation**	**Nitrogen starvation**
	Cre02.g113450	Upregulated [b]	Unchanged [c,d]
	Cre02.g111200	Unchanged [b]	Unchanged [c]
	Cre02.g146629	Unchanged [b]	Unchanged [c]
	Cre06.g289900	N/I	Unchanged [c]
1R-MYB	Cre08.g364050	Unchanged [b]	Unchanged [c,d]
	Cre09.g397475	N/I	N/I
	Cre09.g391353	N/I	N/I
	Cre09.g399552	N/I	Upregulated [d]
	Cre09.g411600	Unchanged [b]	Unchanged [c]
	Cre10.g421021	Unchanged [b]	Unchanged [c]
	Cre14.g609350	Unchanged [b]	Unchanged [c,d]
	Cre14.g621172	Unchanged [b]	Upregulated [c]
	Cre14.g633789	Unchanged [b]	N/I
	Cre16.g686250	Unchanged [b]	Unchanged [c]
	Cre01.g034350 [a]	Upregulated [b]	Upregulated [c,d,e]
	Cre02.g103450 [a]	Upregulated [b]	Unchanged [c,d]
	Cre02.g108350 [a]	Unchanged [b]	Unchanged [c,d]
	Cre03.g144747 [a]	N/I	Unchanged [d]
	Cre03.g149201 [a]	N/I	N/I
2R-MYB	Cre03.g197100 [a]	Upregulated [b]	Upregulated [c,e]
	Cre03.g197350 [a]	Unchanged [b]	Unchanged [c]
	Cre07.g345350 [a]	Unchanged [b]	Unchanged [c,d]
	Cre09.g399067 [a]	N/I	Unchanged [d]
	Cre14.g621050 [a]	Unchanged [b]	Unchanged [c]
	Cre14.g632176 [a]	N/I	N/I
	Cre16.g677382 [a]	N/I	N/I
	Cre12.g522400 [a]	Upregulated [b]	Upregulated [c]
3R-MYB	Cre17.g714229 [a]	Unchanged [b]	Unchanged [c,d]

4.2.2 Phylogenetic analysis of R2R3-type MYB proteins

R2R3-type MYB is the biggest family of MYBs in plants and in the past decade they have been extensively studied (see 4.1 section). In *C. reinhardtii*, this family is smaller than in higher plants, with only 12 protein sequences identified (Table 4.1). To investigate the evolutionary relationship of this group of MYBs in *C. reinhardtii*, a phylogenetic analysis was carried out of the R2R3-type MYB family, which resulted in a tree with three main clusters (Fig. 4.2). Six putative MYB proteins were clustered together (group shaded in pink), five in the second cluster (group shaded in green) and only one MYB, Cre14.g632176, as a singleton in the tree. Sequence analysis showed that proteins within each cluster are quite divergent, with protein sequence identities ranging from 73% to 28%. 11 of the 12 studied R2R3-type MYBs, had the MYB domains located in the N-terminal of the protein, with the exception of Cre14.g632176 where the MYB domain is located towards the C-terminal, which could explain the distance from the other members in the phylogenetic tree (data not shown).

In *A. thaliana*, the R2R3-type MYB family is composed of 168 proteins with conserved MYB domains with several members functionally characterized *(Stracke et al*., 2001). Therefore, a phylogenetic analysis combining R2R3 proteins from *A. thaliana* and *C. reinhardtii* would help to understand the evolutionary relationships between the MYB families of these two organisms and would also suggest putative functions of the MYB proteins in *C. reinhardtii*. With this purpose, a phylogenetic tree of the R2R3-type MYB family in *A. thaliana* together with the R2R3 family in *C. reinhardtii* was constructed. Figure 4.3 shows the R2R3-type MYB family of *A. thaliana* divided in 25 subgroups according to the presence of conserved motifs in the C-terminal *(Stracke et al*., 2001). *C. reinhardtii* R2R3 proteins cluster together with *A. thaliana* proteins in one clade in the tree, in which subgroup 21, 22, 23 and 25 are found. The highly variable C-terminal region of R2R3 MYBs of *C. reinhardtii* was searched for conserved amino acid sequences using the MEME online tool. Interestingly, *C. reinhardtii* R2R3 MYBs do not contain any of the conserved motifs present in the *A. thaliana* subgroups that group close together in the phylogenetic tree. Instead, *C. reinhardtii* proteins contained other conserved motifs in their C-terminal region. From the 12 proteins studied, five conserved motifs were identified. The number of the conserved motifs in each MYB varied between 1 and 3. Two of

the five identified motifs were found in the C-terminal regions of eight of the twelve R2R3 MYB proteins studied. This indicates that most of the R2R3 MYB proteins in *C. reinhardtii* shared one or more identical motif outside of the MYB DBD, which suggests potential similarities in their functions. Besides the motifs with unknown function at the C-terminus identified using the MEME server, the Pfam identified coil-coiled domains in eight out of the twelve R2R3 MYB proteins from *C. reinhardtii* (Fig. 4.3).

To study the conservation of the DBD of these proteins, sequence logos were constructed (Fig. 4.4). Alignment analysis revealed that the R2R3-type MYB domains of the 12 *C. reinhardtii* proteins remarkably conserved the common features of all MYB domains: three regularly spaced W residues that together form a hydrophobic core with three α-helices. Moreover, the amino acids in between the W residues have been shown to have high variability within MYBs of different species (Williams and Grotewold, 1997), appeared to be highly conserved with repetitive motifs within the MYB domain. Furthermore, the R2 and R3 MYB repeats from *C. reinhardtii* MYB proteins also showed a high level of conservation in the MYB domains from the R2R3 family of *A. thaliana.* As in MYB-related genes from plants, where 13% to 65% of the R3 MYB domain have a substitution at either the first or the third W residue (Du *et al.*, 2013), in *C. reinhardtii* R3 MYB domain the first W residue was often substituted by a Phe (F) and the third by Tyr (Y). Taken together, these results reveal the evolutionary relationship of R2R3-type MYB proteins in *C. reinhardtii.*

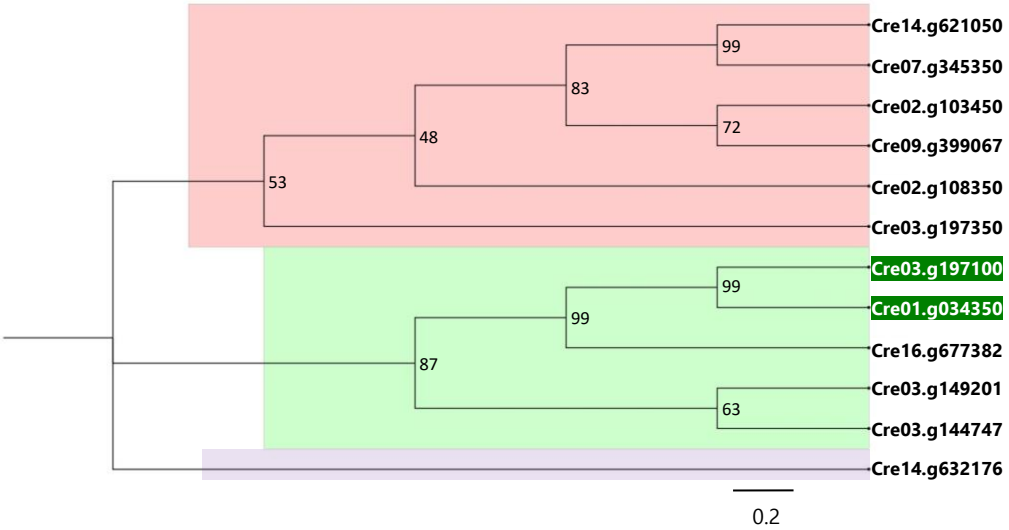

Figure 4.2: Phylogenetic tree of R2R3-type MYB family in *C. reinhardtii*. Multiple sequence alignment was performed by CLUSTAL OMEGA and the tree was constructed using the RaxML program with 1000 bootstrap replicates. Figtree was use to visualize the tree. Branch length scale bar indicates evolutionary distance of amino acid substitutions. Genes belonging to the R2R3 MYB protein family are identified by their Gene ID. Genes highlighted in green will be the focus of later experiments. The numbers at the inner nodes indicate bootstrap values. The scale bar (at the bottom of the tree) indicates 0.2 base amino acid substitutions per position.

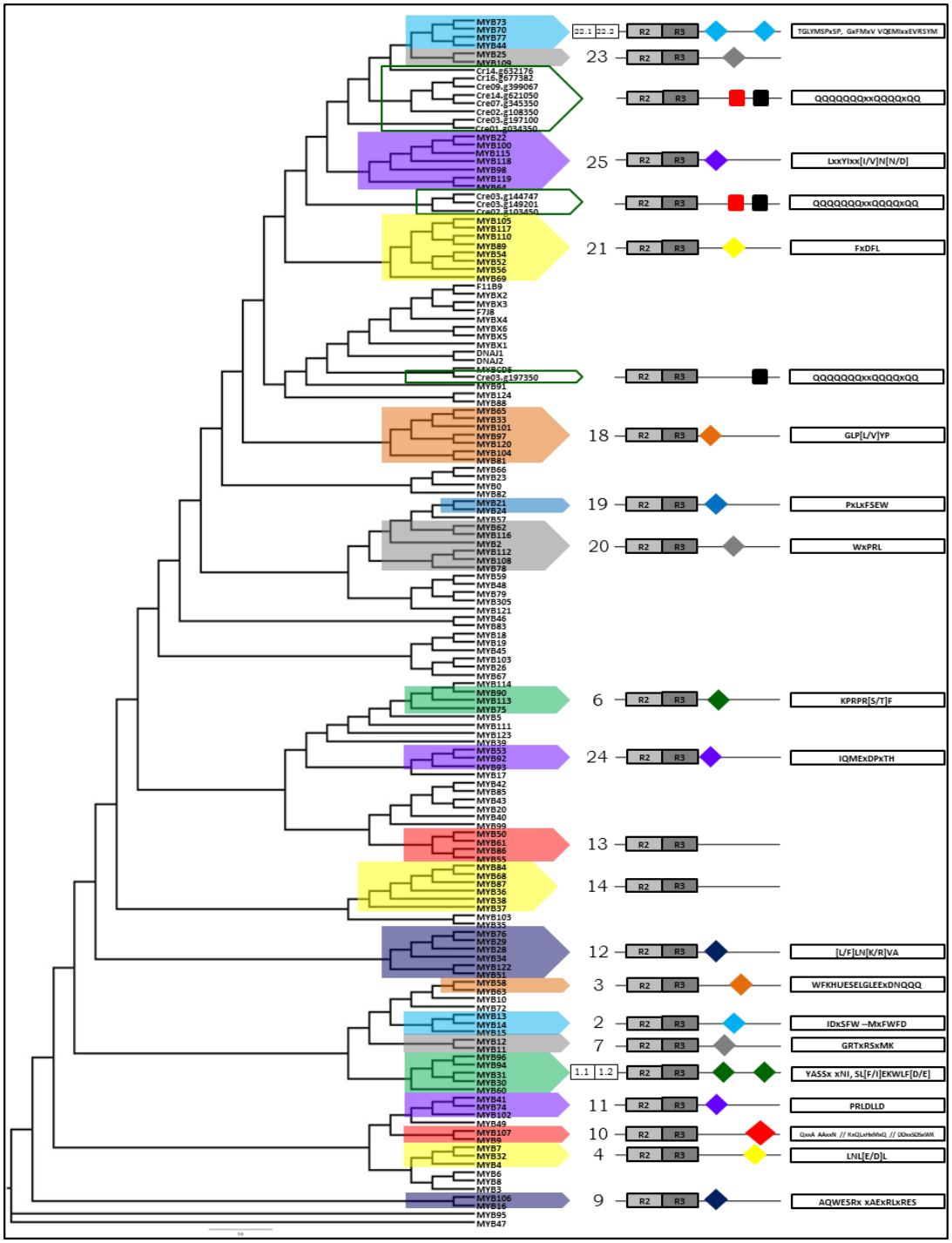

Figure 4.3: Phylogenetic relationship between R2R3 MYB proteins from *A. thaliana* and *C. reinhardtii*. The tree was constructed using the RaxML program and Figtree was used to visualize the tree. The subgroups were designated as previously reported and conserved C-terminal motifs are shown at the right of the tree (Stracke *et al*., 2001) and marked with colours to facilitate its identification. Green boxes represent *C. reinhardtii* proteins. Conserved motifs in *C. reinhardtii* proteins were detected using MEME using the complete amino acid sequences of the 12 previously identified R2R3 MYB proteins.

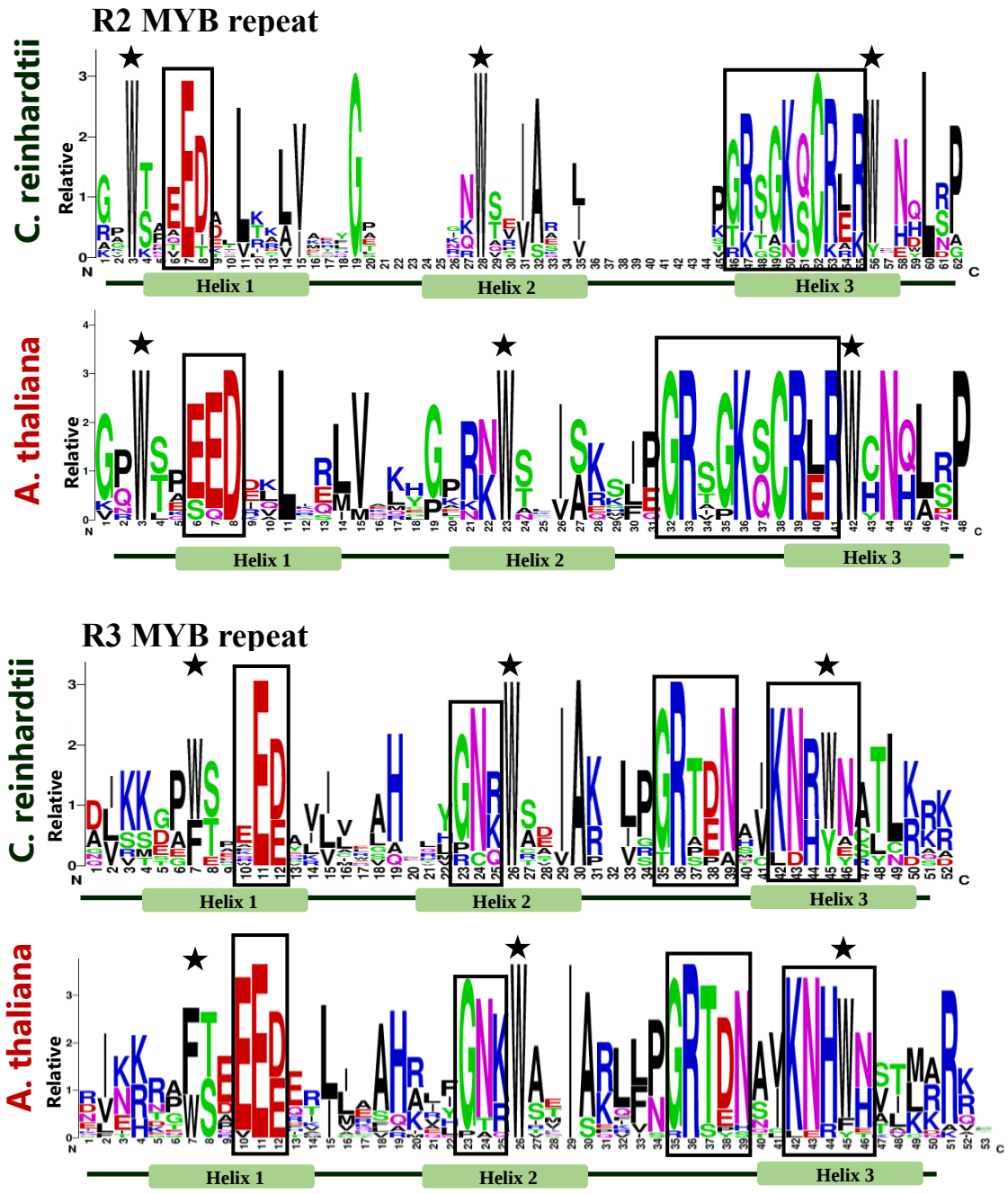

Figure 4.4: Sequence logos of the MYB domains of R2R3 MYB proteins from *A. thaliana* and *C. reinhardtii*. The sequence logo for *C. reinhardtii* was constructed using the twelve protein sequences previously identified as putative R2R3 MYBs. The sequence logo for *A. thaliana* was constructed using 35 protein sequences from R2R3 MYBs clustered together in the same clade of the phylogenetic tree with *C. reinhardtii* MYBs. Relative entropy is measured in bits; higher levels of entropy indicate higher levels of conservation of the residue. Stars indicate conserved Trp (W) amino acids. Black boxes indicate conserved motifs within the MYB domain. Conserved helices are shown below each logo.

4.2.4 Physiological parameters of *C. reinhardtii* under different stress

For a framework and physiological context for the study of the role of MYBs TFs in *C. reinhardtii*, physiological parameters were analysed. The aim was to obtain the optimal conditions that produced the most significant metabolic changes in starch and lipid accumulation to later determine a potential role of MYBs under these conditions. Cells were cultured under P, N and salt stress at various time points to assess growth, and chlorophyll, carotenoid, lipid and starch content using the different techniques described below.

Lipid and starch detection by FT-IR spectroscopy

Fourier-transform infrared (FT-IR) spectroscopy is a technique used to obtain an infrared spectrum of absorption or emission of a solid, liquid or gas. In metabolomics, FT-IR is used as a tool for the characterization of the complex building blocks of biological systems such as proteins, nucleic acids, lipids and carbohydrates (Baker *et al.*, 2014). The accuracy of the technique has led several research groups to use it as a tool for the identification and classification of microorganisms (Driver *et al.*, 2015). In microalgae, the potential of FT-IR spectroscopy has been tested in various earlier studies (Giordano *et al.*, 2001; Dean *et al.*, 2010; Driver *et al.*, 2015). However, only recently has it acquired more importance due to the progress in the use of chemometrics, using multivariate statistical tools such as the Principal Component analysis (PCA) for the extraction of more reliable information from FT-IR spectra. Compared to conventional biochemical analyses, FT-IR spectroscopy has shown many advantages like reliability, sensitivity, speed of measurement procedure and only a small amount of sample is required. In this study, FT-IR spectroscopy was used to measure the lipid and carbohydrate composition of *C. reinhardtii* cultured under Low P, Low N and 100 and 200 mM NaCl after 3, 5 and 7 days of growth.

Figure 4.5 A shows a typical mid infrared absorption spectrum in the region of 4000–950 cm^{-1}. The different peaks in the spectra had been attributed to specific molecular groups based on biochemical standards (Giordano *et al.*, 2001; Coates, 2006), with proteins showing peaks of absorption in the region of 1650 and 1540 cm^{-1}, carbohydrates in the spectral region from 1180 to 950 cm^{-1} and total lipids at 1740 cm^{-1}. The clustering in the spectra of the biological and technical replicas included in

the assay makes it difficult to determine differences in the absorption of the different samples from the spectra alone (Fig. 4.5 A). However, applying a PCA on the data, the differences become clearer. Figure 4.5 B shows that PCA resolved the spectra in three different major clusters along the PC1 axis, responsible for 93.1% of the variation, and two clusters along PC2, accounting for 3.3%. The type of stress appears to have caused the major differences in the samples, showing three big clusters of Low N at the left, Low P in the middle and the control (complete media) together with the salt stress at the right of the plot. The length of the treatment shows only a small effect on the differences of the samples. Two major clusters are distinguished along PC2; one is for the samples treated for three days under the before mentioned stresses and the other is samples after five and seven days of stress treatment.

For explaining the metabolic differences within the samples that produce this different clustering in the PCA plot, the PC loadings were analysed (Fig. 4.5 C). This plot explains what metabolic factors contributed to the differences among samples shown in the PCA plot. Figure 4.5 C shows that the major differences in the treatments cluster along PC1 was largely due to changes in the carbohydrate and lipid region of the spectra as determined by the sharp peaks at 1160-1036 cm^{-1} and 1740 cm^{-1}, respectively. The same scenario applies for PC2, which shows that the differences in the length of the treatments are attributed to differences in the lipid and carbohydrate content. Finally, considering the information provided by the PCA, a new spectrum was plotted showing only the spectra of the specific treatments that cause the major differences within the samples, averaging the results obtained by the days of treatment as well as technical and biological replicas (Fig. 4.5 D). The plot shows five spectra that correspond to samples treated in nutrient-replete media (control), Low P, Low N, 100 and 200 mM of NaCl, regardless of the days of treatment. The results correlated with the exhibit by the loading plots, showing clear differences in the lipids and carbohydrates among the samples. Low P and Low N showed the highest content of lipids and starch and the treatment with both concentration of NaCl produced a lower content of lipid and starch similar to the one produced by cells treated under control conditions (Fig. 4.5 D).

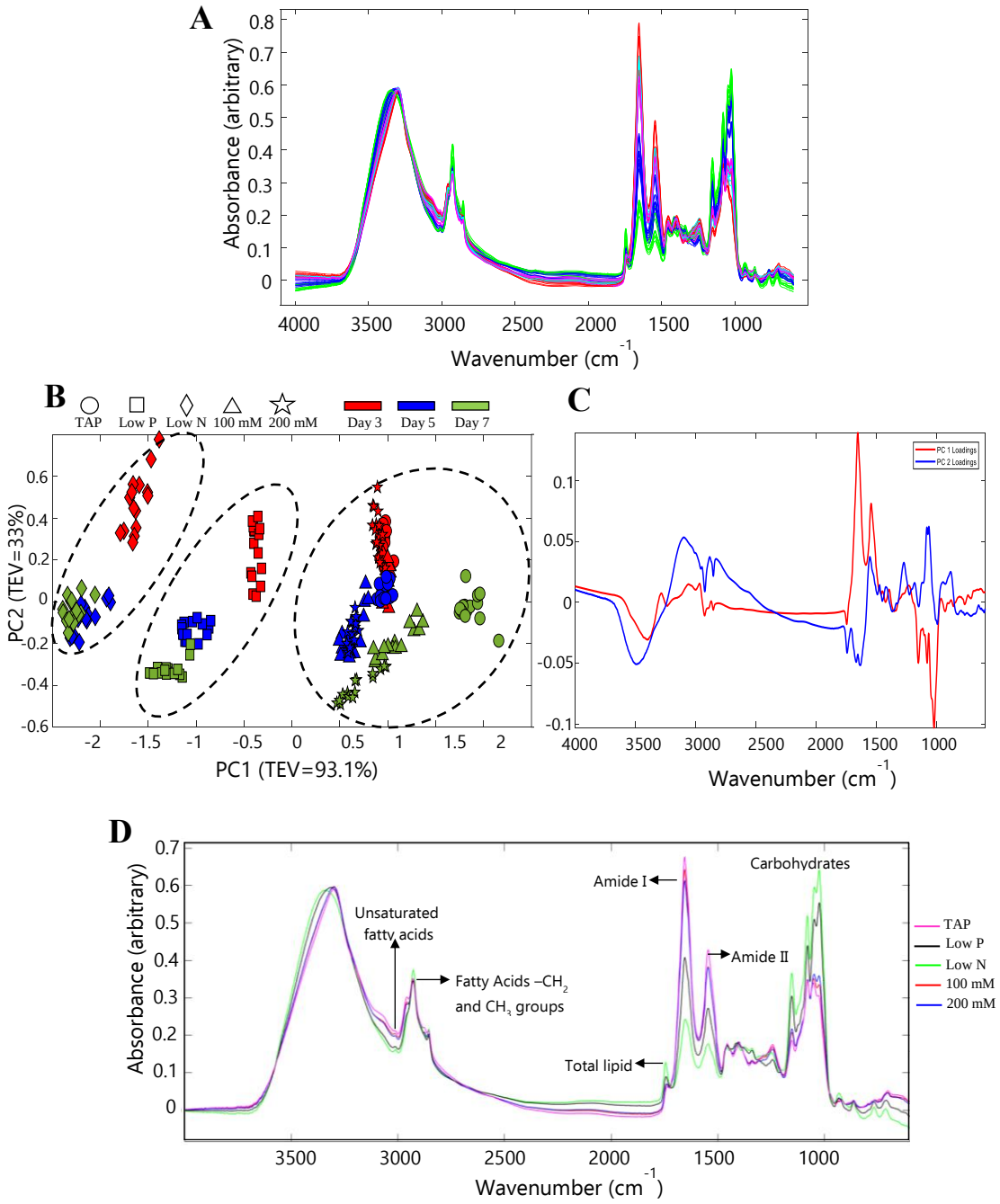

Figure 4.5: FT-IR spectroscopy data taken from *C. reinhardtii* grown under different stress conditions **(A)** EMSC corrected FT-IR spectra in the range of 4000-1000 cm^{-1} obtained from cell cultured with Standard TAP media, Low P, Low N and 100 and 200 mM NaCl media, during 3, 5 and 7 days (spectra super-imposed). **(B)** PCA analysis and loading plot **(C)** derived from the same experiment in the spectral range of 4000-1000 cm^{-1} on baseline corrected spectra. The type of treatment is represented by symbols and the days of the treatment by colours in the PCA plot. **(D)** FT-IR spectra derived from the same experiment using the average values from day 3, 5 and 7 from cells grown in Standard TAP, Low P, Low N and 100 and 200 mM NaCl media.

Biochemical determination of lipid content

Lipid content was quantified using a biochemical approach that measures the neutral lipid triacylglycerol (TAG) by using a fluorescence method based on the use of the Nile red reporter, a dye emitting a yellow fluorescence signal (around 580 nm) in the presence of neutral lipids. This time, the same time points used for FT-IR measurement was used, but shorter times were also explored. As shown in Figure 4.6 A, only after 12 hours of N-starvation, the lipid content becomes significantly higher than the control, increasing directly proportional with the course of time. However, P-starvation did not show any significant change when compared to the control condition at any of the tested time points within a 48-hour period. Longer exposure times to Low N continue increasing the lipid content of the cells until day 7 (Fig. 4.6 B). On the other hand, low P and 200 mM NaCl only produce a significantly higher accumulation of neutral lipid after 7 days of exposure to these conditions. 100 mM NaCl did not produce any significant lipid accumulation compared to control conditions at any of the time points tested. These results correlated with those previously obtained by FT-IR spectroscopy, which identify Low P and Low N as the conditions that produced highest lipid accumulation in the cells. However, in contrast with the results obtained with the FT-IR, Nile red measurements indicate that the exposure to 200 mM NaCl for 7 days also produced an important accumulation of lipids.

With the aim of obtaining more quantitative information from the FT-IR data and to facilitate the comparison with the results obtained by biochemical assays, the lipid:amide I ratios were calculated (Fig. 4.6 C). Previous studies have determined a strong correlation between lipid:amide I ratios and biochemical assays validating the FT-IR spectroscopy technique as an effective method for lipid determination in microalgae (Dean *et al.*, 2010; Dean and Pittman, 2015). The band heights for the total lipids obtained in the FT-IR plot are calculated and normalized as a ratio to the amide I band. These results further confirmed those observed in the PCA analysis, showing a gradual increase in the lipid content of cell grown under Low P and Low N condition with the course of time, with a faster rate of accumulation in the Low N treatment. Cells grown under standard TAP media and high salt concentration showed similar levels of lipid content that does not seem to increase with longer incubation times. Taken together, results from the Nile red assay and FT-IR

spectroscopy determining lipid content correlate together with exception of the results obtained with 200 mM NaCl.

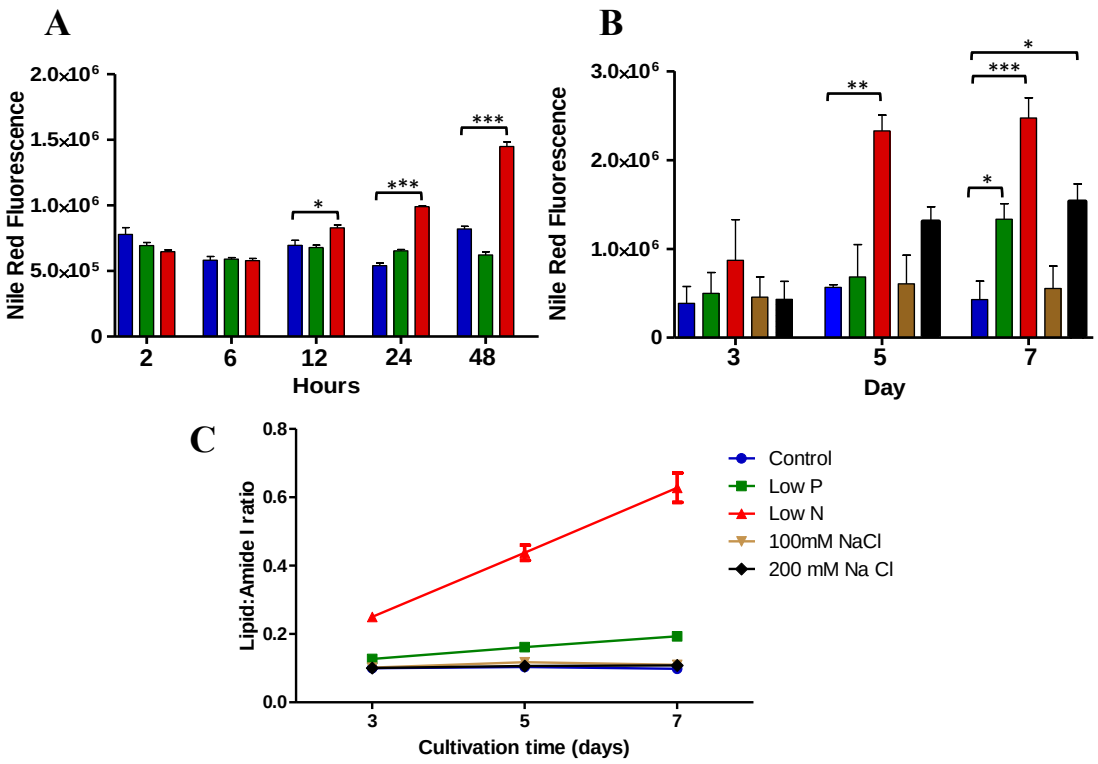

Figure 4.6: Lipid content of *C. reinhardtii* cells**. (A)** For short time stress analysis, cells were grown under standard TAP media, low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4) and low nitrogen (0.7 mM NH_4Cl) for 2, 6, 12, 24 and 48 hours. **(B)** For longer time stress analysis, cells were grown under standard TAP media, low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl concentration (100 and 200 mM) for 3, 5 and 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. Neutral lipid content was measured using Nile red and results were normalized against cell number using 680 nm optical density. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA). **(C)** Lipid:Amide I ratio derived from the previous FT-IR spectroscopy experiment obtained from cell cultured with Standard TAP media, Low P, Low N and 100 and 200 mM NaCl media, during 3, 5 and 7 days. Vertical bars indicate S.E.M of 3 FT-IR spectra from pooled data (derived from 3 biological replicas).

Biochemical determination of starch content

Starch content was measured using the alpha-amylase/amyloglucosidase activity assay (total starch assay kit). This assay consists in the enzymatic hydrolysis of starch to glucose and the subsequent peroxidase-catalysed oxidation reaction where glucose is oxidized to gluconic acid with quantitative production of hydrogen peroxide, which is detected by a colorimetric reaction. As with lipid content, short incubation times of cells under N-starvation only produce a significant amount of starch after 12 hours when compared to control condition, which continue increasing with the course of time (Fig. 4.7 A). However, there was no significant change in the starch content of cells under P-starvation at any of the time points tested within a 48 hour period of P-starvation (Fig. 4.7 A). Longer exposure times showed that after 5 days of culture under Low P or Low N conditions, the starch content is significantly higher than the one from cells grown under nutrient-replete conditions (control), which remains high until day 7 (Fig. 4.7 B). No difference in the starch content was observed in cells treated with either at 100 or 200 mM of NaCl at any of the time points tested compared to the control condition. In order to compare the results obtained with the previous data from FT-IR spectroscopy, the carbohydrate:amide I ratio was analysed. As with the lipid:amide I ratio, carbohydrate:amide I ratio has been reported to correlate with the starch content by using biochemical starch quantification assays (Dean *et al.*, 2010). Results showed a gradual increase in the carbohydrate content of cells grown under Low P and Low N (Fig. 4.7 C), and no significant change in cells grown under standard TAP media or under high salt concentrations (100 mM and 200 mM), correlating with the results obtained using the alpha-amylase/amyloglucosidase activity assay. However, FT-IR data presented as carbohydrate:amide I ratio, showed a clear difference between the carbohydrate content of cells grown under Low P and Low N media. With Low N stress causing higher levels of carbohydrate accumulation, a difference that is different from the results obtained with the alpha-amylase/amyloglucosidase activity assay.

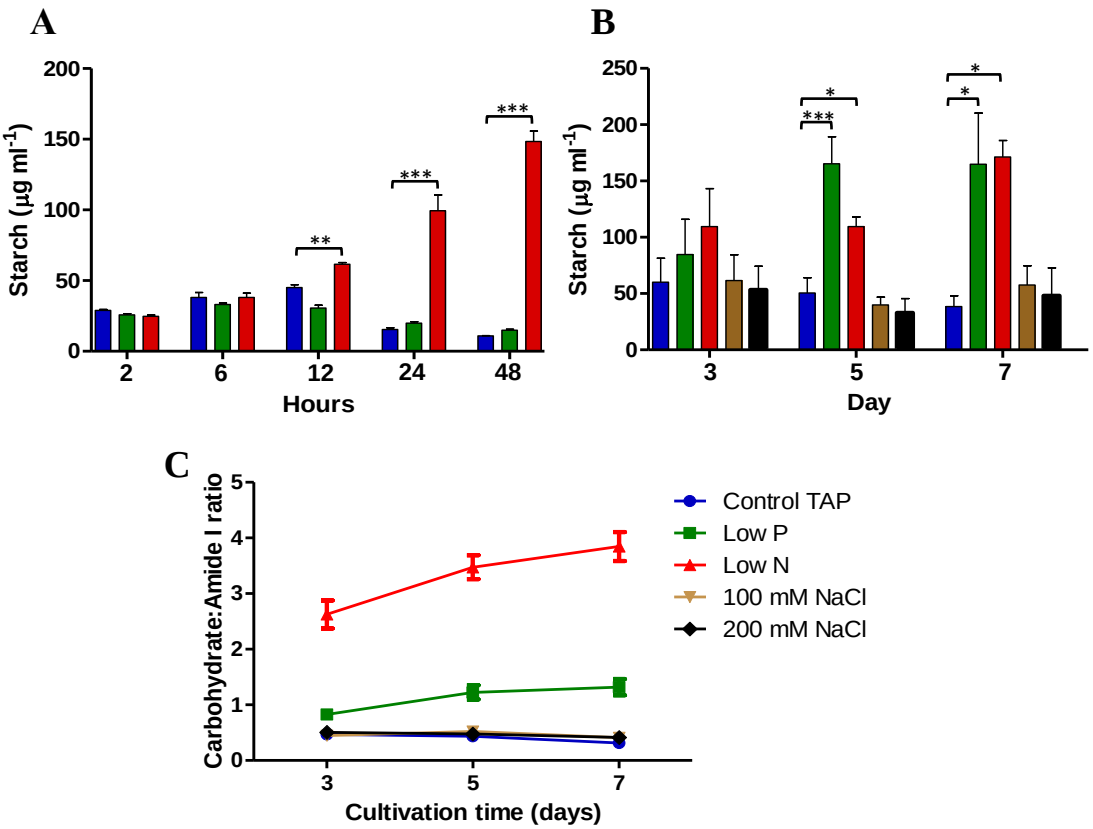

Figure 4.7: Starch content of *C. reinhardtii* cells**. (A)** For short time stress analysis, cells were grown under standard TAP media, low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4) and low nitrogen (0.7 mM NH_4Cl) for 2, 6, 12, 24 and 48 hours. **(B)** For longer time stress analysis, cells were grown under standard TAP media, low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl concentration (100 and 200 mM) for 3, 5 and 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA). **(C)** Carbohydrate:amide I ratio derived from the previous FT-IR spectroscopy experiment obtained from cell cultured with Standard TAP media, Low P, Low N and 100 and 200 mM NaCl media, during 3, 5 and 7 days. Vertical bars indicate S.E.M of 3 FT-IR spectra from pooled data (derived from 3 biological replicas).

Growth, chlorophyll, carotenoid and protein content

Cells grown in Standard TAP media (control), low phosphorus, low nitrogen and high-salt (100 and 200 mM of NaCl) media for 3, 5 and 7 days were tested for their growth rate, and total chlorophyll, carotenoid and protein content. For growth analysis, optical density values were taken at 680 nm to estimate cell density. As indicated in Figure 4.8 A, the algal growth under the control conditions exhibited fast and continuous growth with a clear lag, log and stationary phase. As expected, P, N and salt stress caused impaired growth compared to control, with 200 mM NaCl and Low N causing the highest reduction in the cell density. Accordingly, these conditions also had the biggest impact on chlorophyll (Fig. 4.8 B), carotenoid (Fig.

4.8 C) and protein content (Fig. 4.8 D), showing more than a two-fold decrease in their levels when grown under Low N conditions for more than 5 days or more than 7 days under 200 mM NaCl. Low P had no significant effect on carotenoid and protein content, and only showed a significant decrease on the chlorophyll levels after 7 days of growing under this condition. 100 mM NaCl had no effect on any of the parameters tested, showing similar levels of chlorophyll, carotenoids and protein to the control. In general, the most significant metabolic changes were observed after 7 days of treatment under either Low P, Low N or 200 mM of NaCl.

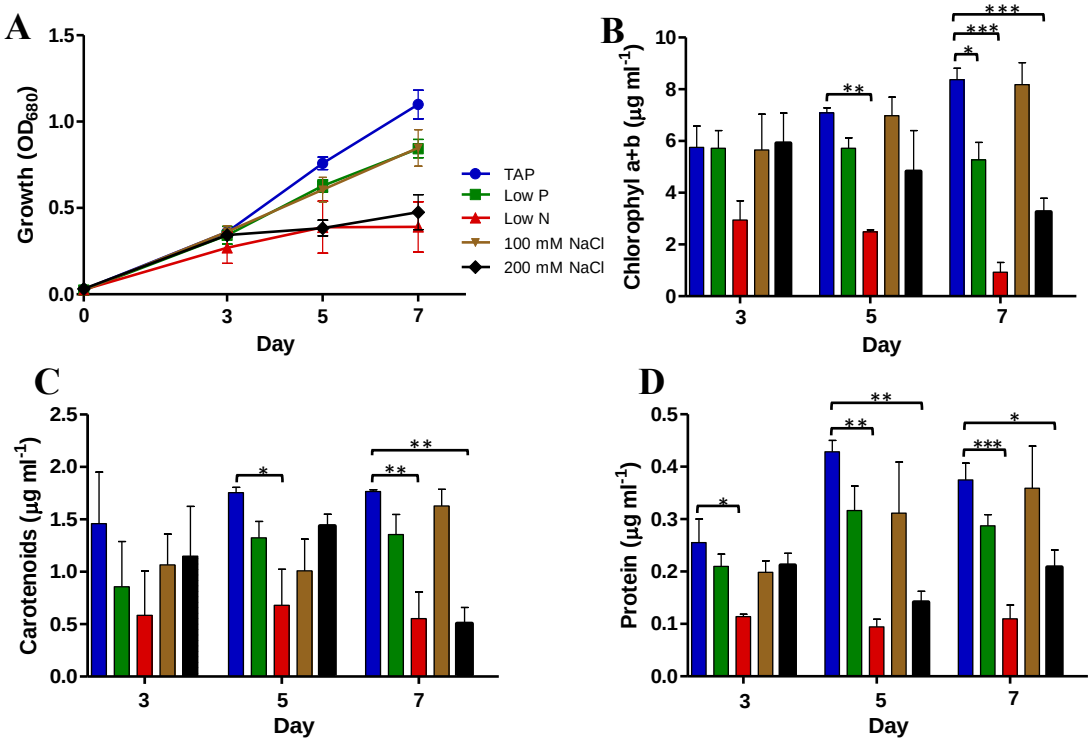

Figure 4.8: Growth, and total chlorophyll, carotenoid and protein content of *C. reinhardtii* under different stress conditions. Cells were grown under standard TAP media and with 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl (100 and 200 mM) media. For salt stress, cells were first cultivated in standard TAP media until day 3 and then different concentrations of salt were added. Growth (A), total chlorophyll (Chl a + b) (B), carotenoid (C) and protein (D) content were quantified at day 3, 5 and 7 in the different stress conditions. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA).

4.2.5 Expression profile of MYB in *C. reinhardtii* under different stress conditions

As mentioned in the introduction, several putative MYB TFs were found upregulated under P-starvation conditions, correlating with the high accumulation of lipids and starch. To study the expression profile of MYBs under the previously

tested stress conditions, two putative MYB genes showing the highest level of P starvation induced expression in RNA-Seq data (Bajhaiya *et al.*, 2016b), were chosen as targets for this study. Interestingly, these two genes were also found upregulated in Low N conditions (Table 4.1). To date, there is no study on the function of these genes, hence no specific descriptive name has been assigned, and they will be referred to as MYB1 and MYB2 based on decreasing rank of fold change in the RNA-Seq data (Bajhaiya *et al.*, 2016b).

Due to initial unsuccessful attempts to obtain MYB1 and MYB2 cDNA, 3 different primers to initiate reverse transcription were tested for cDNA synthesis from total RNA; oligo dT, random primer and a gene-specific primer. Total RNA was extracted from cells treated with Low P media after 3, 5 and 7 days, and cDNA was synthetized using the 3 different primers. Figure 4.9 A shows that in the three time points tested, the gene-specific primer was the most efficient for cDNA synthesis of MYB2, evidenced by the presence of amplification in the RT-PCR gel. Oligo dT and random primer, showed no clear amplification at the expected size. Figure 4.9 B shows the RT-PCR of another sample obtained after 7 days of growth in Low P conditions, with the random primer and gene-specific primer being the most efficient for the MYB2 cDNA synthesis. For ensuring the equality of cDNA in the PCR reactions and reliability of the results, the ribulose-bisphosphate carboxylase gene (*rbcl*) was amplified as an internal control gene (Fig. 4.9 B). Taken together, for later experiments, the gene-specific primer was chosen as the preferred method for MYB2 cDNA synthesis. Regarding MYB1, none of the three primers tested for the cDNA synthesis showed any amplification in the RT-PCR gels, after several tests with different sets of MYB1 PCR primers (data not shown).

Once the detection of the cDNA was standardized, the expression profile of MYB2 was detected under the stress conditions previously tested (Fig. 4.9 C). No apparent change was found on MYB2 mRNA levels under control or Low N conditions, whereas Low P and salt stress at the three tested concentrations (100, 150 and 200 mM) caused the upregulation of the gene.

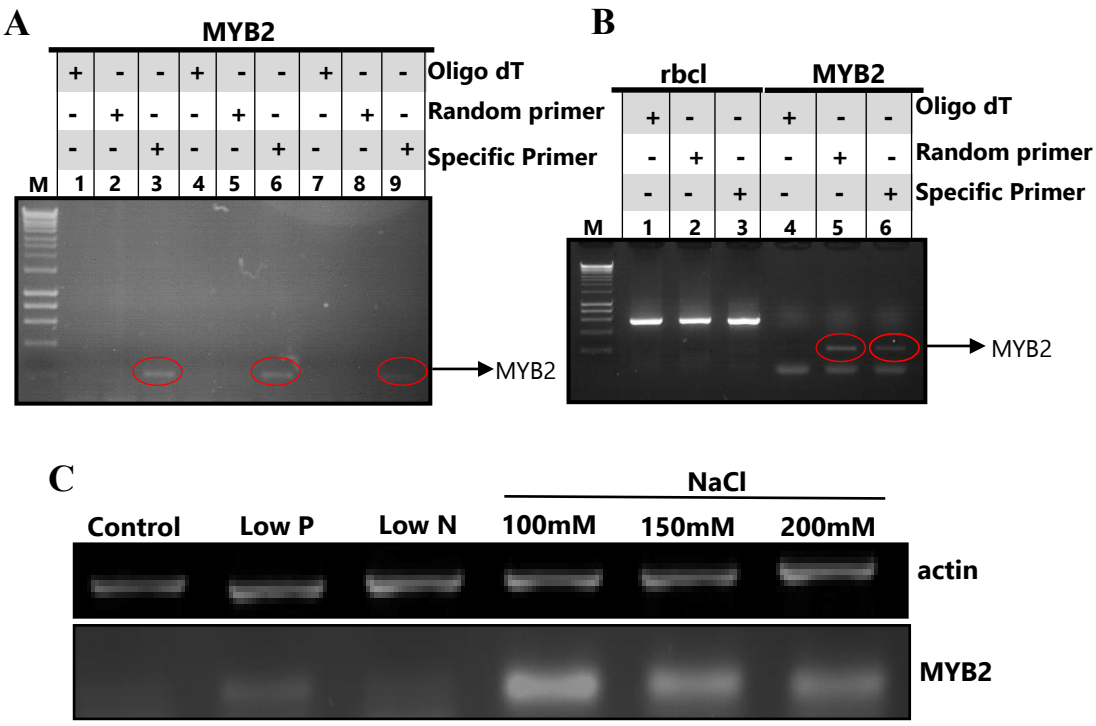

Figure 4.9: mRNA levels of MYB2 under different stress conditions. *C. reinhardtii* cells were grown under standard TAP media and with 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl (100 and 200 mM) media. For salt stress, cells were first cultivated in standard TAP media until day 3 and then different concentrations of salt were added. (A) Total RNA was extracted after 3, 5 and 7 days, converted to cDNA using the different primers; oligo dT, random primer and a MYB2 specific primer and subjected to PCR analysis. Red circles indicate amplification at the expected size. (B) Samples from day 7 are shown with the PCR amplification of the MYB2 gene, using cDNA synthetize with the three different primers, and the *rbcl* gene amplification as a loading control. (C) Expression profile of the MYB2 gene under standard TAP, Low P, Low N and salt stress, using actin gene amplification as a loading control.

4.3 Discussion

4.3.1 Identification of MYB proteins in *C. reinhardtii*

In order to identify members of the MYB TF superfamily in *C. reinhardtii*, candidate genes and their coding sequences were examined in different databases for the identification of MYB domains. As a comparison, we also analysed MYB proteins from *A. thaliana*, which have been extensively studied and with several of the members characterized, to gain insights into the potential functionalities of *C. reinhardtii* MYB proteins. The identification of MYB domains led to conflicting results. Proteins containing more than one MYB domain, belonging to either the R2R3- or 3R-MYB protein families were easily identified and the results from all databases used and the manual checks were in agreement in terms of the presence and position of the MYB domain. However, results from the identification of proteins with only one MYB domain, showed discrepancy. Databases such as Pfam and SMART, that use multiple sequence alignments and hidden Markov models (HMMs) for the identification of conserved protein domains, identified 28 proteins containing MYB domains. Accordingly, the platform Phytozome, a comparative hub for plant genome and gene family data and analysis that gathers information from other databases such as Pfam, confirmed the presence of MYB domains in these 28 proteins. On the contrary, the PlnTFDB, a portal that provides a comprehensive, high-quality and the most updated resource specifically dedicated to plant TFs, identified only 14 of these 28 proteins as belonging to the MYB TF family. The other 14 were classified as members of the MYB-related family or G2-like family.

Research papers and reviews in recent years have published conflicting results, with some classifying TFs such as PSR1 and PHR1 as members of the G2-like family, while others classify them as being in the MYB TF family or MYB-related family (V Rubio *et al.*, 2001; Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b; Safi *et al.*, 2017). After manually checking the sequences of these proteins, searching in different databases, the conclusion to these controversial classifications is that the DBD from the G2-like, MYB and MYB-related are very similar. This similarity may have led to the detection of false positives, explaining the differences between the number of MYB detected in *C. reinhardtii* in previous genome-wide studies about TF in microalgae (11 vs 46 MYB members) (Riaño-Pachón *et al.*, 2008; Hou *et al.*,

2014). However, despite their similarities, there are clear differences that have consistently conserved and grouped these proteins more closely together in phylogenetic analysis, with the G2-like, the MYB-related or MYB families. Safi *et al*. (2017) reviewed and clearly exposed the differences in the DBD of these three different families of TFs (Fig. 4.10). As previously mentioned, a MYB domain is a motif containing three regularly spaced W residues (with approx. 18 to 19 amino acids of spacing) with diversity in their amino acid sequences but also some conserve motifs within. Although this is the norm, some MYB domains have also been shown to have replaced either the first or the third W by other aliphatic amino acids as Tyr or Ala. MYB-related domains on the other hand, contain only two of the three conserved W of the typical MYB domain and highly conserved domains such as the SHAQK(Y/F) domain (Figure 4.10) or the LKDKW (N/T) domain. A G-2 like domain is a conserved sequence with only one of the three W from the MYB domain, and a highly conserved SHLQ motif. Due to the similarities in the DBD of these 3 families, they all form similar helix-turn-helix structures that confer them the DNA binding capacity. However, the differences across members of these families and the high conservation of the motifs, classified them into different phylogenetic groups.

The PlnTFDB currently presents the most updated information and classified proteins from plants and microalgae according to results from phylogenetic studies from different groups, acknowledging the differences between the domains of MYBs and G2-like proteins. However, this database does not acknowledge the differences between 1R-MYB proteins and MYB-related proteins and considers all proteins with 1 MYB repeat as belonging to the MYB-related family. A manual check of the 30 proteins considered as members of the MYB-related family by the PlnTFDB, evidenced that 15 of them were wrongly classified, as they contain 1 repeat of the conserved characteristic MYB domain, meaning that they should be classified as members of the MYB protein family. Taken as a whole, the current available information regarding MYB TF family in *C. reinhardtii* in databases is probably inaccurate and misleading. The classification of members of this family should be carefully done to avoid false positives and overestimation of the number of proteins in the family. It is worth noticing that the differences that have classified them as different families are merely based on phylogenetic analysis that grouped them

according to their amino acid sequences, suggesting different evolutionary origins. However, most of these proteins are still functionally uncharacterized and it should be considered that in the future, when more data of their functionalities is accumulated, the classification of these proteins in one or the other family could change if their function does or does not match with the predictive function of the members of the family. Nevertheless, the up-to-date information is mostly based on their sequences and the most accurate classifications should therefore acknowledge their differences that group them in different phylogenetic groups. In this study we have excluded false positives by manual inspection and identified a total of 28 proteins from the MYB family: 14 1R-MYBs, 12 R2R3-MYB and two R3-MYB proteins.

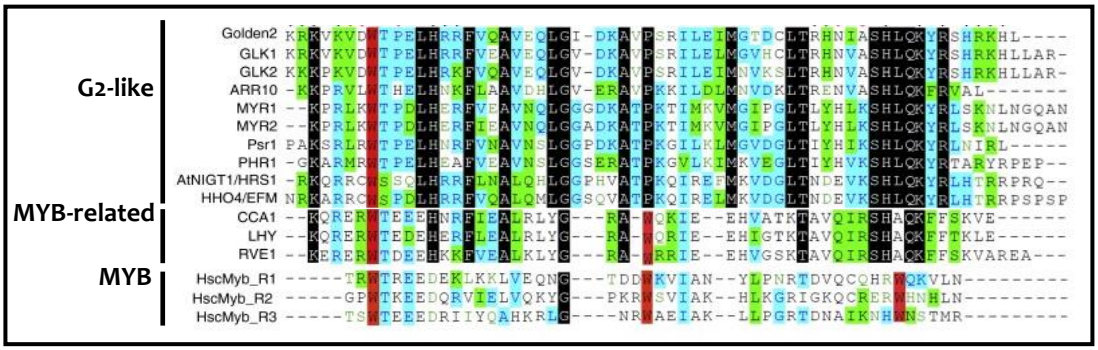

Figure 4.10 Similarities and differences between the G2-like, MYB-related and MYB domains. Sequence alignment of the G2-like, MYB-related and MYB domains. The three regularly spaced Try (W) residues are highlighted in red, and they represent the key difference to distinguish members of these families. Identical residues are highlighted in black, conserved residue in blue and less conserved in green. The MYB-related domain contains only two of the three regularly spaced W, and one conserved SHAQK motif. The G2-like members have only the first W. The two other W residues are replaced by P and L respectively, and they contain the conserved motif SHLQK. Figure adapted from Safi *et al.* (2017).

4.3.2 Phylogenetic analysis of *C. reinhardtii* MYB proteins

Besides classifying TFs into different families, the DBD confers TFs the ability and specificity to bind the DNA (Gonzalez, 2016). NMR and crystallography studies had provided insights into the conserved structure of the MYB domain, highlighting the importance of the conserved similarities of MYB domains to maintain the helix-turn-helix structure that confers on the protein the capacity to bind DNA (Ogata *et al.*, 1996). However, these studies do not address how the differences

among MYB domains affect this capacity and specificity of binding. As previously mentioned, the DNA binding of MYB in characterized by the action of two helices in the MYB domain that recognize their DNA target sequences and pack together to cooperatively bind the DNA. Studies have determined that the DNA sequence motifs recognized by plants and animal MYBs in their target genes are different (Williams and Grotewold, 1997). For example, the maize ZmP1 R2R3 MYB protein binds the consensus DNA sequence CC(T/A)ACC (Grotewold *et al.*, 1994), while vertebrate MYB proteins bind the (C/T)AACGG motif (Howe and Watson, 1991). The differences in their DNA binding domains have been shown to be responsible for their differential DNA-binding properties. In petunia, the single change of Leu71 to Glu in the MYB encoding gene PhMYB3, changes its specificity for the target DNA sequence from ANNC(G/C)GTTA to AGTTAGTTA (Solano *et al.*, 1995). However, the high variability within the MYB domain across species, and the large number of them, has made it difficult to identify specific residues involved in DNA-binding.

To study the similarities between R2R3 proteins in *C. reinhardtii* and *A. thaliana*, first a phylogenetic tree combining the genes of this family from both species was constructed. Results showed that all 12 members of the R2R3 family from *C. reinhardtii* clustered together in one major clade with members of the subgroup 21, 22, 23 and 25 of *A. thaliana*. Sequence logos were constructed with the sequences of R2R3 from *A. thaliana*, together with the 12 sequences of the R2R3 from *C. reinhardtii*. Results showed high similarities in R2 repeat as well as R3. Apart from the conserved W residues characteristic of MYB domains, several other motifs within the domain showed a high level of conservation between the member of the *A. thaliana* and *C. reinhardtii* genes that were closest together in the phylogenetic tree. Because no DNA-binding specificities for R2R3 MYB proteins from either *A. thaliana* or *C. reinhardtii* have been reported for members of these subgroups, it is unclear whether the similarities in DBD reflects similar DNA-binding preferences. Nevertheless, evidence has determined unique binding specificities of R2R3 MYBs family for particular subfamilies in *A. thaliana*, suggesting that similarities among groups clustering together in phylogenetic tress may have similar DNA-binding targets. Additionally, similarities in DBD may also indicate that the DNA-binding activities of these proteins may be regulated by similar mechanisms. Interestingly, while the majority of R2R3 MYB proteins in *A.*

thaliana contain two highly conserved Cys residues in the R2 repeat (between second and third W residue), members of subgroups 21, 22, 23 and 25 of *A. thaliana* and *C. reinhardtii* R2R3 cluster together in the phylogenetic tree, and do not conserve these two Cys residues, conserving only one of them within the R2 repeat. The presence of these two conserved Cys residues has been shown to form intramolecular disulphide bonds *in vitro* under oxidative conditions, preventing DNA binding (Heine *et al*., 2004). The absence of one of the Cys residues within the MYB domain may indicate that these proteins are not regulated by the redox status of the cells. However, the redox control of the DNA- binding activity of MYB proteins with only one conserved Cys residue in R3 repeat has also been reported (Myrset1 *et al.,* 1993), indicating an alternative mechanism for the formation of disulphide bonds. Other mechanisms that alter the DNA-binding activity of proteins and can potentially be regulating these proteins, include cysteine nitrosylation and phosphorylation of DNA-binding domain residue (Moyano *et al*., 1996; Serpa *et al.*, 2007).

Similarities within the DBD but also the presence of a conserved motif outside the MYB domain classified these proteins within different subgroups. Members of the same subgroups of R2R3 MYBs in plants have been shown to contain highly conserved motifs in the C-terminal region, which are specific to MYB proteins (Fig. 4.3). Although their importance is not fully characterized for all MYBs, evidence has shown that these conserved motifs are important for the regulation of the function of MYBs. Nevertheless, the results from this study showed that none of the R2R3 MYB proteins from *C. reinhardtii* share any common conserved motif with the members of the subgroups from *A. thaliana* clustered together in the same clade in the phylogenetic tree (Subgroups 21, 22, 23 and 25, Fig 4.3). Instead, they possess other conserved motifs that as conserved motifs in R2R3 proteins in plants could be crucial for their functions. Eight of the 12 R2R3 MYB proteins identified in *C. reinhardtii* contained a coil-coiled domain in the C-terminal region, shown to be important for the protein-protein interaction (Rubio *et al.*, 2001). The importance of coil-coiled domain in the sequences of several TFs for their regulation has been reported for several TFs, including MYB proteins. However, as mentioned previously, none of the R2R3 MYBs from *C. reinhardtii* has been functionally characterized so it is difficult to predict whether the presence of

different motifs in the C-terminal region between *A. thaliana* and *C. reinhardtii* is reflected the functionality of these proteins.

Based on the patterns observed for several MYBs in which members of the same group exerted similar functions, we could hypothesise that R2R3 MYB from *C. reinhardtii* in the same clade as subgroup 21, 22, 23 and 25 from *A. thaliana* could have similar functions. The majority of the report function for members of these subgroups has been related with response to abiotic stress such as dehydration, temperature, salt tolerance and disease resistance. Interestingly, one of the R2R3 MYBs from *C. reinhardtii* (MYB2) showed high expression when grow under high salt concentrations (Fig. 4.9), suggesting that it may be a conserved role of this protein in abiotic stress across microalgae and plant species. Taken together, the similarities of R2R3 proteins between *A. thaliana* and *C. reinhardtii* may suggest similar roles. However, it is not clear how their differences affect their functionalities. More experimental studies are required to determine the specific function of these proteins and determine whether proximity in the phylogenetic trees implicates similar functions.

4.3.3 Physiological parameters of *C. reinhardtii* under different stress

To set a framework for subsequent studies, the optimal conditions for determining the effects of the different stress on lipid and carbohydrate biosynthesis in *C. reinhardtii* were established. FT-IR spectroscopy and biochemical assays were performed and compared for the quantification of lipids and carbohydrates. As expected, the total lipid and carbohydrate content of N and P stress cells was higher, which is positive for a biofuel application, but appears not to be positive for the microalgae, as its growth stops and the cell duplication is hindered. These results are in agreement with most of the nutrient limitation studies carried out to date (Venkata Mohan and Devi, 2014; Goncalves *et al.*, 2016). Nutrient limitation conditions produce cell division arrest, and alter the cell's metabolism, directing the carbon flux mainly to the biosynthesis of storage molecules such as lipids and carbohydrates (Juergens *et al*., 2016).

This study determined that longer times of stress induction are ore optimal for the determination of lipid and starch accumulation and later studies will use 7 days as the time for stress-induced lipid and starch content determination. Furthermore, as

previously reported, N stress appears to have more effect on the lipid content and growth than P stress (Li *et al.*, 2014; Yodsuwan *et al*., 2017). The explanation for this may be that it has been shown that microalgae accumulate large storage of P in the form of polyphosphates which might be supplying the phosphorus demand under limited P conditions (Grossman and Aksoy, 2015). In addition, while P limitation does not affect the N uptake of the cells, the absence of nitrogen has been shown to have an effect on the phosphate uptake rate of the N-starved cells (Kamalanathan *et al.*, 2016). These factors could be triggering higher stress levels on N-starved cells compared to P-starved cells, leading to slower growth rates of the N-starved cells. The rearrangement of macromolecules under N-starvation could be mobilizing the carbon from proteins to neutral lipid synthesis, explaining the higher lipid content under N-starvation. In this study, results from both of the approaches used to measure starch and lipid, FT-IR and biochemical assays, were in accordance. Previous studies have reported high variability in the optimum staining conditions using Nile red as a biochemical assay to quantify lipid content across different species (Cirulis *et al.*, 2012). Physiological differences between microalgae species or strains such as structure and composition of cell wall and variation in lipid content, could explain this variability, and care must be taken in the standardization of this assay. However, the conditions used in this study for Nile red assay for the quantification of neutral lipids under P and N limitation appeared to be reliable. Results obtained by Nile red assay are in agreement with many other studies and correlate with the data obtained by FT-IR spectroscopy measurements. Similarly, starch measurements using a colorimetric enzymatic assay, correlated with data obtained from the FT-IR peaks. In conclusion, for cells under P- or N-starvation, either FT-IR or the biochemical assays tested could be use as reliable techniques for the lipid and starch quantification. However, this seems not to be true for every type of stress.

The lipid content of cells under high salt concentration determined by Nile red showed a discrepancy with the results obtained by FT-IR spectroscopy. While Nile red showed a significant increase in the neutral lipid content of cells treated with 200 mM of NaCl at day 7, FT-IR spectroscopy showed no significance increase, showing a lipid content similar to the cells cultured under standard TAP media. It has previously been reported that substances present in the culture media (meaning cell fragments, salts, precipitates of the dye, and other debris) can interfere with the

fluorescence reading of the Nile red stain. Hounslow *et al*. (2017), reported false positives when using Nile red for measuring Neutral lipids of salt stress cells. Apparently, high salt concentrations produce breakdown of cell walls, producing higher surface areas for the interaction between the cell lipids in cell fragments and the Nile red dye, leading to overestimations of the neutral lipid content. For this reason, FT-IR was considered a more reliable technique for measuring lipid content of the cells under high NaCl concentration. The lipid content obtained by FT-IR measurement of cells under salt stress is not surprising. This technique has already been proven as reliable for the macromolecular analysis of microorganisms under different type of stress, including P- and N-starvation and salt stress (Dean *et al.*, 2010; Driver *et al.*, 2015). Although salt stress has been proven effective in enhancing the lipid of microalgae cells, the concentration necessary for the induction of lipid accumulation seem to be very variable and species-dependent. Lower concentrations of NaCl in the media appear to be more appropriate for the induction of lipid accumulation in *C. reinhardtii* (Fan and Zheng, 2017). Altogether, these results highlight the necessity of the use of more than one method for the determination of lipid content in microalgae cells, due to the individual limitations of the existent lipid techniques (Hounslow *et al.*, 2017).

4.3.4 Expression profile of MYBs in *C. reinhardtii* under different stress conditions

Table 4.1 presented the expression profile of MYB proteins found in the literature from two different studies, one dedicated to transcript analysis of cells under P-starvation and the other focused on transcripts levels under N-starvation. Interestingly, both types of stress produced the upregulation of the same four R2R3 putative MYB proteins out of the 12 found in the *C. reinhardtii* genome. As an initial step to study the potential role of R2R3 MYB TFs in *C. reinhardtii,* the expression pattern of MYB1 and MYB2, two of the four proteins found upregulated under nutrient stress, was determined under the previously identified conditions.

Despite the crucial function of TFs in the control of gene expression in all living cells, they are frequently expressed at very low levels and only a small amount can activate a cascade of genes. For this reason, TFs can be difficult to study and their detection either at mRNA or protein level can be challenging (Milo and Phillips, 2015). In this study, the mRNA detection of MYB1 and MYB2 TFs

presented difficulties due to their low abundance. The choice of primer for the construction of the cDNA showed to be crucial for the detection of the mRNA. Although MYB1 mRNA detection was not successful, in spite of the different approaches used for its detection, MYB2 mRNA was successfully detected. Reverse gene-specific primer showed to be the best option to initiate reverse transcription and posterior PCR detection. Even though the use of reverse gene-specific primer can be complicated when more than one gene needs to be detected in the sample, it was the only plausible option for the detection of MYB2 mRNA levels. For subsequent studies on this gene (Chapter 5), reverse gene-specific primer will be used for the construction of the cDNA and housekeeping genes will be detected in a separate reaction using oligo dT.

Confirming the data of previous studies, RT-PCR showed the upregulation of MYB2 under Low P conditions (Fig. 4.8 C). However the upregulation of this gene under low N conditions was not very clear (Fig. 4.8 C). MYB2 showed the highest upregulation levels when exposed to concentrations of NaCl either of 100, 150 or 200 mM (Fig. 4.8 C). The identification of the overexpression of this TF under different stress conditions suggests a potential role of this protein in the regulation of responses against stress. As mentioned previously, in plants, R2R3 MYBs clustered within the same subgroups in phylogenetic trees exert similar functions, which have also been extrapolated to R2R3 orthologues in other species. Thus, it is logical to hypothesize that MYB2 can have similar functions to other similar R2R3 MYB proteins. MYB1 and MYB2 proteins are in the same clade in the phylogenetic tree (Fig. 4.2), and grouped together with R2R3 MYB proteins from *A. thaliana* that regulate mechanisms to respond to abiotic stress (Fig. 4.3). *AtMYB44*, *AtMYB70, AtMYB73*, and *AtMYB77* belong to R2R3 MYB subgroup 22, in the same clade as MYB1 and MYB2. These genes have been found upregulated under abiotic stresses such as wounding (Cheong *et al.*, 2002), white-light treatment (Ma *et al.*, 2005) and are transiently up-regulated by cold stress (Fowler and Thomashow, 2002). Interestingly, microarray analysis has identified that these genes are upregulated together by salt stress in *sos2* (salt overly sensitive2) mutants (Kamei *et al.*, 2005). These observations, together with the high expression levels of MYB2 under high salt concentrations suggest that it may be a conserved function of similar *C. reinhardtii* and *A. thaliana* R2R3 MYB proteins in response to abiotic stress, such as salt stress.

Chapter 5

Characterization of MYB2 transcription factor, a member of the MYB transcription factor family

5.1 Introduction

Microalgae have great potential as an alternative renewable source of energy, however improvements throughout the production process are needed before this can become commercially feasible. For the biofuel industry, one of the most important improvements would be the increase in the yield of macromolecules such as starch and lipids that can later be transformed into biofuels. Nutrient stresses have proven to be a strong inducer of the synthesis of this macromolecules with some species producing up to 80% of their dry weight in lipids when subjected to nutrient stress conditions (Spolaore *et al.*, 2006; Chisti, 2007). However, stress conditions also inhibit cell division leading to slow growth rates and lower biomass production, which is a major disadvantage for large-scale biofuel production (Hu *et al.*, 2008; Demirbas, 2011). Therefore, nutrient stress is not a feasible strategy to increase the production of these macromolecules, however, it has become a powerful tool to understand the regulatory networks controlling lipid and starch metabolism in microalgae. The changes in carbon flux upon nutrients stress leading to the accumulation of storage molecules such as starch and lipids is a complex process involving a series of changes in the levels of transcripts, proteins, and metabolites (Goncalves *et al.*, 2016). It has been observed that microalgae cells activate adaptation responses after just minutes of sensing the presence of a stress (López *et al.*, 2015; Park *et al.*, 2015). These rapid responses are in part the results of a tightly regulated network that involved transcription factors (TFs) and transcriptional regulators (TR) that act fast to change the expression of key genes involved in the acclimation responses (Gargouri *et al.*, 2015). Knowledge of the regulation of these metabolic pathways can lead to the identification of key regulators that can be targets for transcription factor engineering (TFE) to achieve high lipid and/or starch productivity while maintaining high biomass.

Genome-wide studies have represented a breakthrough in the understanding of the metabolic changes underlying abiotic stress in microalgae and have become a foundation for the identification of key regulators of these metabolic pathways. Although these analyses have contributed to the understanding of microalgae status under stress conditions, only a few TFs have been identified that have roles in the regulation of these responses. Currently, only four TFs have been functionally

implicated in the regulation of lipid and starch metabolism in microalgae under stress conditions (Table 5.1). One of them is the Nitrogen-Responsive Regulator-1 (NRR1) a putative TF belonging to the SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family, with a regulatory role in nitrogen (N) assimilation (Boyle et al., 2012; Gargouri et al., 2015). The mutation of NRR1 with loss of function of the protein, produced the loss of 50% of the triacylglycerol (TAG) accumulation ability of *C. reinhardtii*, highlighting the importance of this TF in the synthesis of this macromolecule (Boyle et al., 2012). NRR1 apparently is only involved in lipid accumulation under N starvation. Compromised hydrolysis of triacylglycerol (CHT7) is another putative TF that has been related with the lipid accumulation under N starvation. A CHT7 mutant with loss of function of the protein, showed an impaired in the degradation of lipid droplet protein MLDP and TAG, after resupply of N, suggesting a negative role in the lipid accumulation in *C. reinhardtii* (Tsai *et al.*, 2014).

More recently two others TFs had been added to the list, one of them is Rhythm of Chloroplast 40 (ROC40) and the other one is Phosphorus Starvation Response 1 (PSR1). A proteomic study identified the Rhythm of the Chloroplast 40 (ROC40), a MYB-related TF, as the most upregulated protein upon short-term N starvation. ROC40 mutant showed an inability to increase the accumulation of TAG upon N starvation, which was recovered when wild-type ROC40 was expressed (Goncalves *et al.*, 2016). PSR1 is a member of the G2-like TF family, key in the acclimation response of *C. reinhardtii* to phosphorus (P) starvation (Wykoff *et al.*, 1999). PSR1 overexpression has shown to increase the starch content of *C. reinhardtii* cells, which correlated with a higher expression of specific starch metabolism genes (Bajhaiya *et al.*, 2016b). Interestingly, PSR1 also appears to be regulating lipid and starch metabolism not only under P starvation but also under nitrogen (N) and (S) starvation, coordinating multiple lipid-induce stress responses (Ngan *et al.*, 2015).

Apart from the experimentally proven functions of these TFs, technological advances from different omics approaches (including transcriptomics, proteomics and metabolomics) in conjunction with in silico tools have revealed other TFs as potential regulators of microalgae lipid and starch metabolism. One of the major contributions of these recent published studies is the fact that the cellular response to

stress conditions cannot be explained by the action of individual TFs but by a group of them acting as a regulatory network orchestrating the metabolic changes observed in the cell in response to stress (Schmollinger *et al.*, 2014). Taking advantage of the sequenced genome and gene annotation studies of *Nannochloropsis* spp., Hu et al. (2014), used a computational strategy to identify the TFs involved in lipid related pathways. By identifying well-characterized orthologous with roles in lipids synthesis in plants (including WRINKLED1 and GmDofc), and co-expression analysis with well-known lipid synthesis genes, they identified 30 putative lipid-related TF genes. These TFs belonged to 11 different protein families, including MYB-related, NF-YC, AP2 and C3H TF families (Hu *et al.*, 2014).

In a more integrative study, the transcriptomic response to N starvation was correlated with the proteome and metabolome, and identified a series of putative TFs with roles in the regulation of this stress response (Schmollinger *et al.*, 2014). Interestingly, this study compared the stress response of the cell to P, N and Sulphur (S) starvation, and while the authors identified response networks that were specific for each stress, they also identified common TFs that were upregulated upon each of the three different stresses, suggesting a common mechanism for responding to micronutrient deficiencies. Among the common upregulated TFs is the R2R3-MYB protein Cre01.g034350, identified in the previous chapter as MYB1. Another study classified the responses of the cell upon N starvation at two stages; before TAG onset, in preparation for the start of lipid biosynthesis, and after the onset of TAG accumulation to allow the cell to acclimate to the stress (Gargouri *et al.*, 2015). It was revealed that some TFs function as hubs with synergistic roles in regulating these two stages and five major TFs were identified as the master regulator hubs. These five main hubs were also highly correlated the lipid and starch content of the cell, suggesting a potential role in the regulation of these pathways. Interestingly, this study also identified MYB1 (Cre01.g034350) as one of these hubs crucial for the regulation of stress response after the onset of TAG accumulation (Gargouri *et al.*, 2015).

In addition to the omics approaches mentioned above, one study that stands out by using a different approach was carried out by Ngan et al. (2015). By studying chromatin states, they determined the transcriptional profiles of the cells under stress conditions. It is well known that the DNA of actively expressed genes is less tightly compacted to allow the RNA polymerase to transcribe the genes. On the contrary,

DNA regions containing inactive genes are generally more condensed and associated with structural proteins known as histones (Zhang *et al*., 2015). Taking advantage of these changes, Ngan et al. (2015) constructed a genome-wide chromatin profile of microalgae cells under stress conditions to predict the transcriptional activity of genes. They identified 15 genes whose promoters changed from an inactive to active state upon N and S starvation, from which only three presented similar expression pattern under both stress conditions; these genes were PSR1, MYB1 and MYB2. The promoter of the NRR1 gene was found to be active only under N starvation, confirming the previously reported data and enhancing the reliability of this approach (Boyle *et al.*, 2012). This study highlighted the importance of using a combination of approaches to narrow down the identification of key regulators to a smaller number of high-confidence candidates for target for TFE. Finally, a recently published investigation in *C. reinhardtii* widened the alternatives for TFE by performing a genome-wide transcriptomic analysis of cells under salt stress (NaCl), which is another lipid-inducing stress condition. This study revealed important molecular cues for lipid accumulation under salt stress and identified the MYB family as the TF family with the largest number of upregulated genes (Wang *et al.*, 2018).

The studies described above have provided valuable insights into the regulatory networks in charge of displaying the stress responses in microalgae. Moreover, they have identified several TFs from different families as potential regulators of lipids and starch biosynthesis (Table 5.1). However, the actual role of these TFs in the synthesis of these macromolecules requires experimental validation. In the previous chapter, the attention was focussed on the MYB TF family and identified MYB1 and MYB2, as highly upregulated under P starvation. Interestingly, additional genome-wide studies have reported the upregulation of these genes under other stress conditions such as N and S starvation (Ngan *et al.*, 2015). With the aim of identifying potential targets for TFE to increase the yield of lipid and starch production in microalgae, this chapter focuses on the functional characterization of one of these TFs, MYB2 (Cre03.g197100), since RT-PCR amplification of MYB1 was unsuccessful. The role of MYB2 in lipid and starch metabolism was studied by comparing WT strain with MYB2 overexpression and knockdown lines under P, N and salt stress. The results obtained from this study will help to have a better understanding of the function of this TF and evaluate its potential as a target for TFE.

Table 5.1: Transcription factors involved in lipid and starch accumulation in *C. reinhardtii*.

Family	Name	Species	Method of identification	Effect on carbon metabolism	Role	Stress condition
SBP	**NRR1**	*C .reinhardtii*	Insertional gene disruption	50% decrease in TAG content	N assimilation and TAG accumulation	N^-
G2-like	**PSR1**	*C .reinhardtii*	Gene knockdown Gene overexpression	Decrease in starch and TAG content Increase in starch content	Adaptive responses to P starvation and biosynthesis of lipid and starch under stress	P^- N^- S^-
MYB-related	**ROC40**	*C .reinhardtii*	Insertional gene disruption	Decrease in TAG content	Control of Cyrcadian clock and TAG accumulation	N^-
CXC	**CHT7**	*C .reinhardtii*	Insertional gene disruption	Increase in TAG content after N resupply	Repressor of cellular quiescence and potentially involved in TAG turnover	N^-

5.2 Results

5.2.1 Generation of MYB2 overexpression lines

To investigate the role of the putative TF MYB2 in the lipid and starch metabolism of *C. reinhardtii*, MYB2 overexpression lines (MYB2-OE) were created. A plasmid harbouring the full length cDNA of MYB2 under the control of the hybrid HSP70A-RBCS2 promoter in the pCB740 plasmid was constructed (Fig. 5.1). This vector also contains the selectable marker Arg7 (arginine prototrophy) that allows the transformant cells to grow without exogenous arginine (Schroda, et al., 2000).

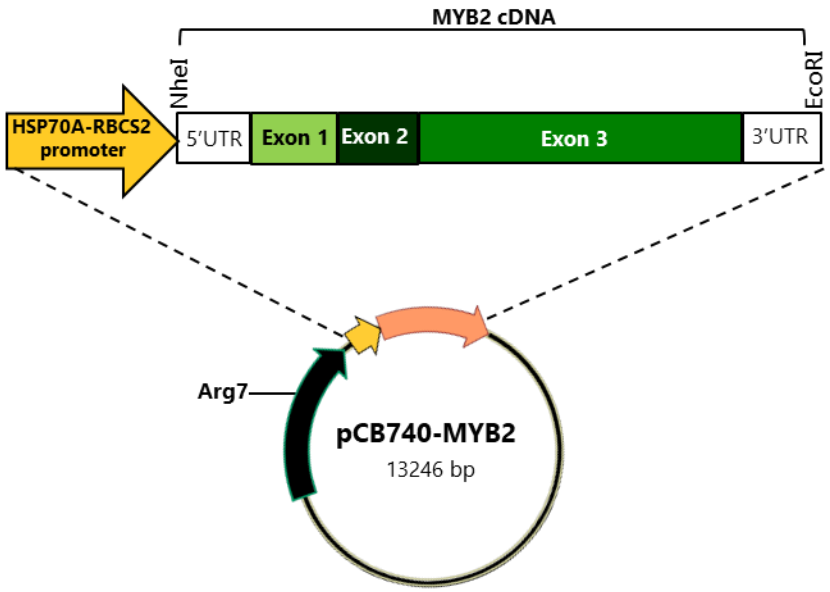

Figure 5.1: Schematic representation of the plasmid construct used for generating MYB2 overexpression lines. The plasmid use was the pCB740 originally containing the selectable marker Arg7 and the HSP70B gene under the control of the hybrid HSP70A–RBCS2 promoter (Schroda et al., 2000). To obtain the construct for the overexpression of MYB2, the HSP70B gene was replace with the full length cDNA sequence of the MYB2 gene by NheI and EcoRI restriction enzyme digestion generating the pCB740-MYB2 vector. M2F and M2R primers were used to confirm by genomic PCR the presence of the construct in the transformed lines.

The pCB740-MYB2 plasmid was introduced into the nuclear genome of the *C. reinhardtii* strains CC-4350 and CC-4351 by particle bombardment. Both of these strains are cell wall deficient and impaired in their capacity to produce arginine. The choice of these strains was made based on the ease of transformation and their ability to accumulate higher levels of carbon storage than the WT, possibly due to the increased sensitivity to stress caused by cell wall deficiency (Siaut *et al.*, 2011).

Transformant cells were selected by their ability to grow on plates without arginine and the transformation efficiency was determined by counting the colonies on the plate. As observed in Figure 5.2 A, CC-4351 strains was more efficient in the transformation of the pCB740-MYB2 showing a higher number of colonies on the plate, therefore this strain was the choice for performing the overexpression experiments. Genomic DNA PCR was performed to check the presence of the transgene within the successfully transformed colonies using the M2F and M2R primers. As shown in Figure 5.2 B, the MYB2 transgene was present in the three different tested colonies (OE4, OE5 and OE6) but absent in the control line (C-), which was transformed with the pCB740 empty plasmid (no MYB2 insert), indicating that pCB740-MYB2 had been successfully transferred to the transformants.

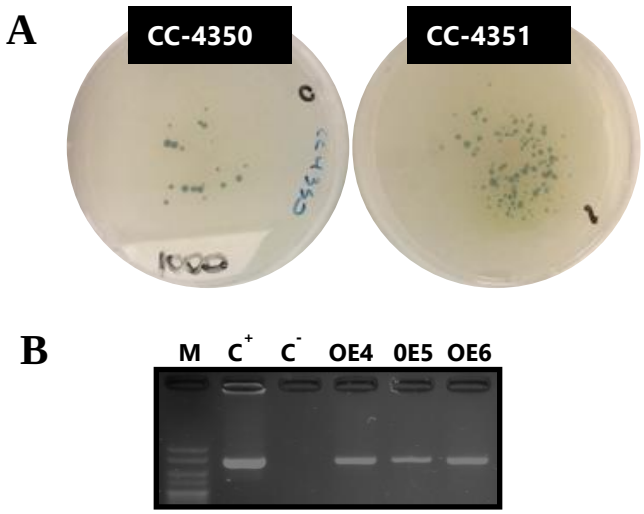

Figure 5.2: Confirmation of the success of pCB740-MYB2 transformation in *C. reinhardtii* (A) Arginine-requiring CC-4350 and CC-4351 (*cw15* background) colonies grown on selective Tris- acetate-phosphate (TAP) agar plates (without arginine). (B) PCR genotype analysis of the pCB740-MYB2 transformation lines. Genomic DNA was extracted from the control line (C^-, transformed with the pCB740 empty plasmid) and three individual MYB2-OE lines and the presence of the insert was confirmed by PCR amplification with the insertion specific primer set M2F and M2R. The pCB740-MYB2 plasmid was used as a template for positive control (C^+). Hypperladder V was used as a molecular marker to determine the amplified band size (expected size 384 bp).

It is important to mention that although the transformation of the pCB740-MYB2 plasmid was confirmed by PCR analysis (Fig. 5.2), the overexpression of MYB2 could not be confirmed. The level of expression of MYB2 in both, the control and transformant lines was low, with CT values above 35 in qPCR measurement (data not shown), generally considered not reliable values (Dotti et al., 2011). However, the expression of genes under the control of the hybrid promoter HSP70A-RBCS2 in the pCB740 plasmid has been proven highly efficient in *C. reinhardtii* with several genes (Molnar *et al.*, 2009; Bajhaiya *et al.*, 2016b; Driver *et al.*, 2017). Therefore, although the attempts to confirm the overexpression were not successful, the next sections show the characterization of the transformant lines assuming the likely overexpression of MYB2. In section 5.3, hypothesis about potential reasons for the inability to detect the overexpression of MYB2 and potential solutions are presented.

5.2.2 Growth analysis of the MYB2 overexpression lines

The growth of the MYB2-OE lines was analysed and compared with the control line under different stress conditions: standard TAP, Low P TAP, Low N TAP and salt stress (200 mM NaCl) TAP media. P and N limitation were achieved by reducing the concentration of phosphate and ammonium in the standard TAP media. Salt stress as achieved by administrating 200 mM NaCl after 3 days of culture into standard TAP media. Cellular growth was monitored by measuring optical density (OD_{680}) after 3, 5 and 7 days of culture. As shown in Figure 5.3 A, control cells grown in TAP and Low P showed similar growth rates as the MYB2-OE lines. However, in Low N and 200 mM NaCl media, the growth pattern of the MYB2-OE cells showed a decrease after 3 days when compared to control line. Interestingly, upon salt stress administration, OE lines behaved differently than control lines, forming biofilm-like aggregations at the bottom of the flask as shown in Figure 5.3 B. The decrease in the OD measurements of OE lines growth under salt stress when compared to the control line may be attributed to the fact that the majority of the cells were at the bottom of the flask and not in the supernatant where the OD was measured. Therefore, in this case lower OD may not be directly related with lower growth rates. It is worth noting that the growth analyses are the result of three independent experiments showing the same aggregation pattern of the OE lines under

salt stress, indicating that the overexpression of MYB2 may be causing this phenomenon.

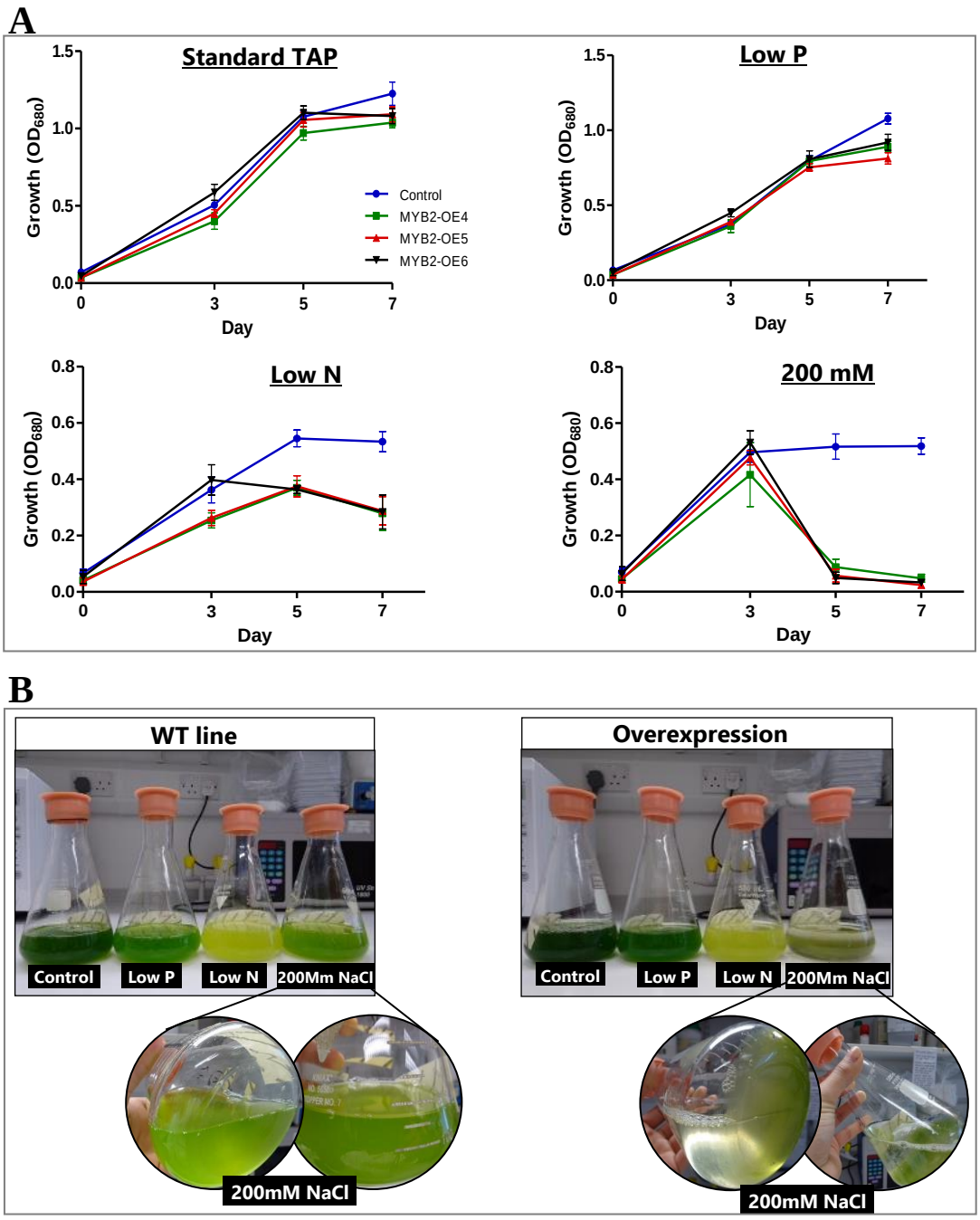

Figure 5.3: Growth analyses of the MYB2-OE lines. (A) Growth curves based on the optical density (OD_{680}) of control and MYB2-OE lines (lines 4, 5 and 6) grown in standard TAP media and 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl concentration (200 mM). For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. OD measurements were taken after day 3, 5 and 7 of culture. Data points represent the average of three independent experiments and error bars indicate the S.E.M. (B) Photographic images of the cultures taken at day 7. OE4 lines were taken as representative images form the OE lines.

5.2.3 Physiological characterization of MYB-OE lines

To evaluate the role of MYB2 in the carbon metabolism in *C. reinhardtii* and its potential as a target for TFE for the biofuel production, the carbon storage content of the MYB2-OE lines was analysed. In the previous chapter, the characterization of *C. reinhardtii* WT cells by FT-IR spectroscopy and biochemical assays determined that the lipid and starch content was significantly accumulated after 7 days of growth in Low P, Low N and 200 mM NaCl. Thus, the results presented below considered these conditions as the optimal for the characterisation of the OE lines.

Lipid and starch detection by FT-IR spectroscopy

MYB2-OE lines together with control lines cultured under the different stress conditions for 7 days were subjected to FT-IR spectroscopy analysis and screened for metabolic changes. Figure 5.4 A shows the representative FT-IR spectra obtained in the region of 4000–950 cm^{-1} from the control lines and the three MYB2-OE lines grown in standard TAP, low P, low N and 200 mM NaCl media, with their respective replicates. The PCA showed in Figure 5.4 B resolve the results from these spectra, showing clear differences between the samples. The majority of the variation within the samples was explained by PC1 accounting for 76% of this variation. Along the horizontal PC1 axis three major clusters were observed (Fig. 5.4 B, dotted ellipses) that correspond to the different treatments; one cluster grouped together the cells treated with low N conditions, the other cluster grouped together the cells treated with salt stress and low P, and the third cluster grouped together cells gown in standard TAP media. To understand what the principal components represented in the PC1 vs PC2 plot, the loading plots were analysed (Fig. 5.4 C). Positive loadings are associated with negative scores, while negative loadings are associated with positive scores in the PCA. Therefore, low N treated cells clustered towards the negative side of the PC1 axis as a result of having higher (negative red peaks on loading plot) carbohydrate (loading peak of C-O vibration at 1160, 1086, 1050 and 1036 cm^{-1}) and a higher total lipid content (loading peak of C═O of ester groups at 1745 cm^{-1}) as well as a reduction (positive red peak on loading plot) in protein (C-O of (amide I at 1655 cm−1 and amide II at 1545 cm−1) compared to the cells cultured in either low P, salt stress or standard TAP (Fig. 5.4 C). On the other hand PC2 explained most of the variability that is not explained by PC1, accounting for 33% of

the variance in the data. Along the vertical PC2 axis two major clusters are observed (Fig. 5.4 B, dotted squares), showing a clear difference between the OE and control lines treated with salt stress. The PC2 loading (Fig. 5.4 C, blue line) indicates that along the PC2 axis, MYB2-OE lines treated with high salt concentration (200 mM NaCl) clustered in the negative side of the PC2, indicating a higher amount of carbohydrates (negative blue peak) but lower amount of lipids (positive blue peak). Altogether, these results indicate that the major differences within the samples were given by the different macromolecular content of the cells induced by different stress conditions and by the MYB2 overexpression in lines treated under salt stress.

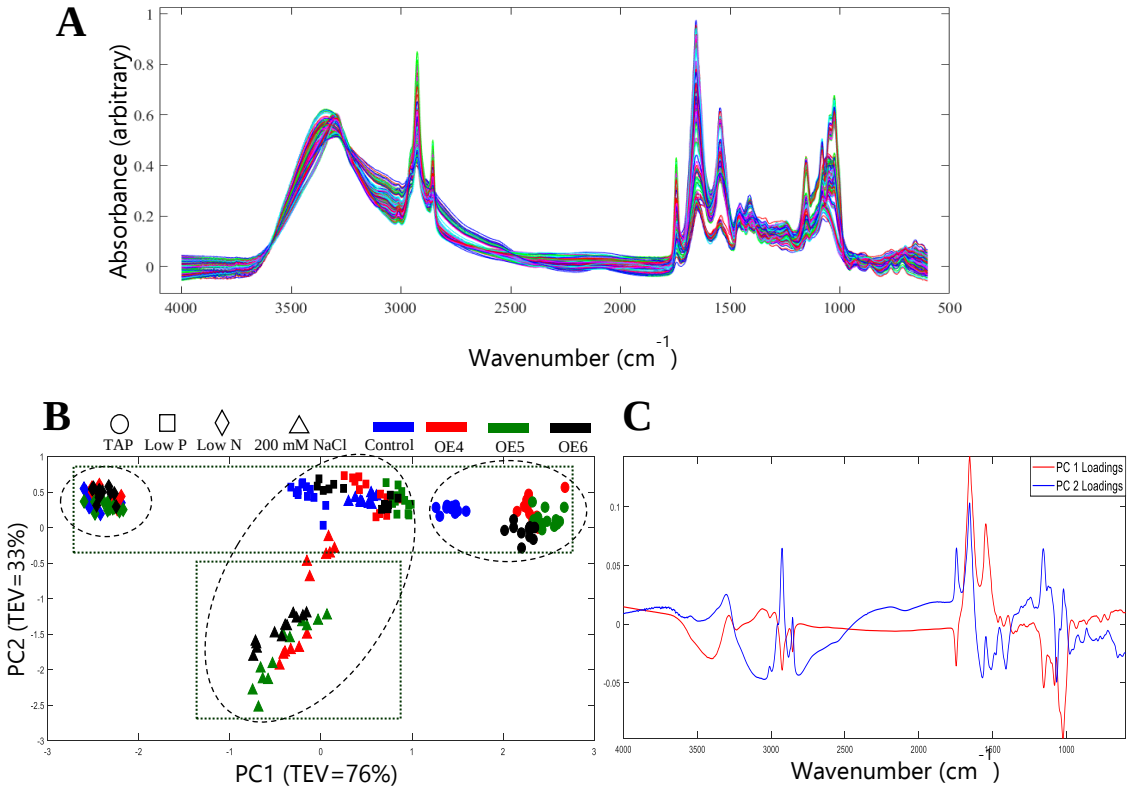

Figure 5.4: FT-IR spectroscopy data taken from *C. reinhardtii* grown under different stress conditions **(A)** EMSC corrected FT-IR spectra in the range of 4000-1000 cm^{-1} obtained from control and MYB2-OE lines cultured with standard TAP media, Low P, Low N and 200 mM NaCl media for 7 days (spectra super-imposed). **(B)** PCA analysis and loading plot **(C)** derived from the same experiment in the spectral range of 4000-1000 cm^{-1} on baseline corrected spectra. The type of treatment is represented by symbols and the different lines by colours in the PCA plot.

Lipid and starch detection by biochemical assays

In order to increase the reproducibility and accuracy of the measurements, the lipid content was also determined by using biochemical assays. The Nile red fluorescence, used to monitor the neutral lipid content, showed that the MYB2-OE lines had significantly higher lipid content than the control lines under low N and salt stress treatments (Fig. 5.5 A). Control cell showed no significant different in their lipid content compared to the OE lines grown in low P and standard TAP media. In addition to Nile red, the gravimetric SOXTEC system was used for the lipid quantification of the samples. This is an automated solvent extraction system that has been well established as a reliable technique to determine the lipid content of samples (Anderson, 2004). Figure 5.5 B showed that the SOXTEC quantification confirmed the results obtained by Nile red, indicating that under standard TAP or low P media conditions, control lines did not exhibit differences in their lipid content with the MYB2-OE lines, while under low N stress the three tested MYB2-OE lines contained significantly higher lipid levels. Unfortunately, SOXTEC could not be used to quantify the lipid content of the cells treated with salt stress condition as this method requires a large amount of microalgae biomass, and most of the biomass of OE lines treated with salt stress was attached to the bottom of the flask forming biofilm-like structures (Fig. 5.3B). To compare the results obtained by these biochemical assays with the FT-IR spectroscopy, the lipid:amide I ratios were calculated by measuring the band heights for the total lipids obtained in the FT-IR plot and normalized to the amide I peaks height. The lipid:amide I ratios confirmed the increase in the lipid content of the OE lines under low N conditions and salt stress, but also showed a slight increase in the lipid content of the OE lines gown in low P conditions compared to the control lines. Interestingly, Nile red as well as the lipid:amide I ratio showed a decrease in the lipid content of the MYB2-OE lines cultured in standard TAP compared to the control lines. These results suggest that *C. reinhardtii* cells respond differently to the MYB2 overexpression depending on the type of stress.

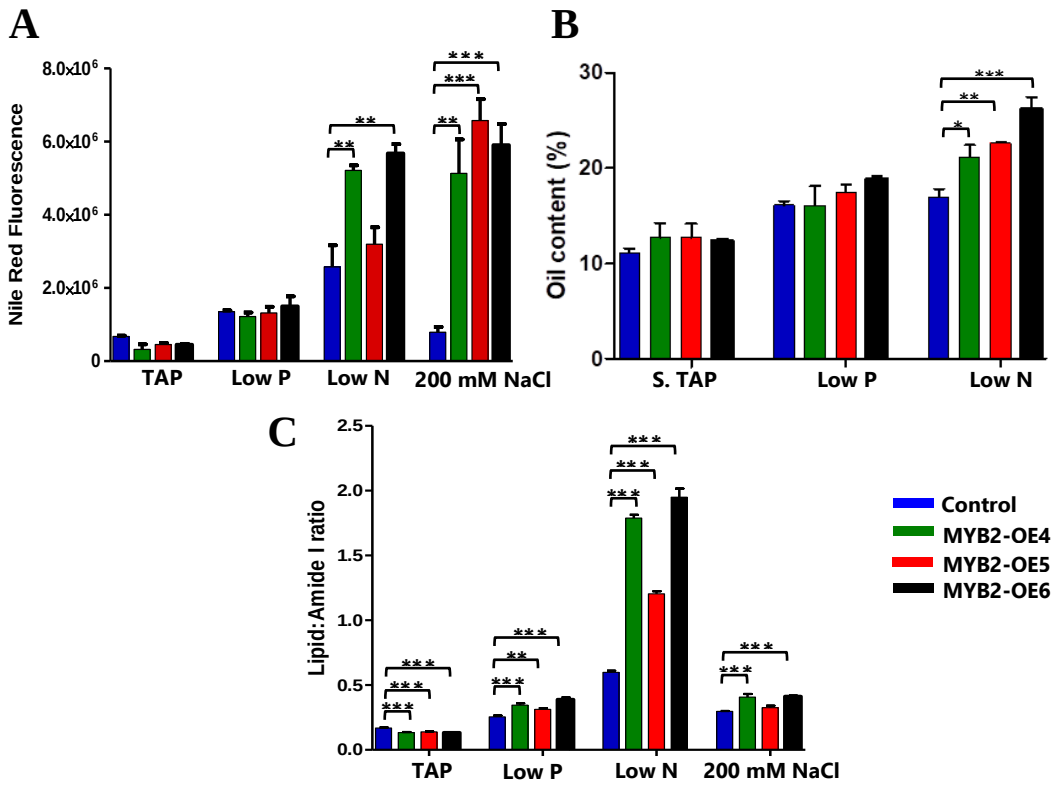

Figure 5.5: Lipid content of control and MYB2-OE *C. reinhardtii* cells**. (A)** Cells were grown in standard TAP media (S.TAP) and 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and 200 mM NaCl for 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. Neutral lipid content was measured using Nile red and results were normalized against cell number using 680 nm optical densities. (B) Lipid quantification by SOXTEC extraction utilising hexane. Results are expressed in percentage of dry weigh (C) Lipid:Amide I ratio derived from the previous FT-IR spectroscopy experiment. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA).

The starch content was also quantified biochemically by using the alpha-amylase/amyloglucosidase activity assay (total starch assay kit). Results showed that the carbohydrate content of OE lines was significantly higher only in salt stress cells when compared with control lines. Low P stress decreased the carbohydrate content of the OE lines while no significant difference was observed in OE lines treated with low N compared to control lines. The carbohydrate:amide I ratios obtained from the FT-IR spectroscopy assay, confirmed these results, showing an increase in the carbohydrate content only in the OE lines treated under salt stress when compared to control lines. The low P and low N conditions produce a decrease in the starch content of the MYB2-OE lines. Similarly to what is observed in the lipid content of OE lines grown in standard TAP, the starch assay kit as well as the

carbohydrate:amide I ratio indicated a decrease in the starch content of the OE lines cultured in standard TAP media.

In general, the changes observed in the lipid and starch content among the OE lines were consistent between the biochemical assays and the FT-IR spectroscopy. The most obvious phenotypes affecting the carbon storage metabolism were caused by MYB2 overexpression in cells grown under low N, exhibiting higher lipid accumulation levels (Fig. 5.5) and under salt stress showing higher carbohydrate accumulation compared to control lines (Fig. 5.6). Interestingly, while increasing their lipid content under low N conditions, OE lines decrease their carbohydrate content under the same stress conditions. These results suggest that the overexpression of MYB2 may be directing the carbon flux towards the lipid biosynthesis under low N stress.

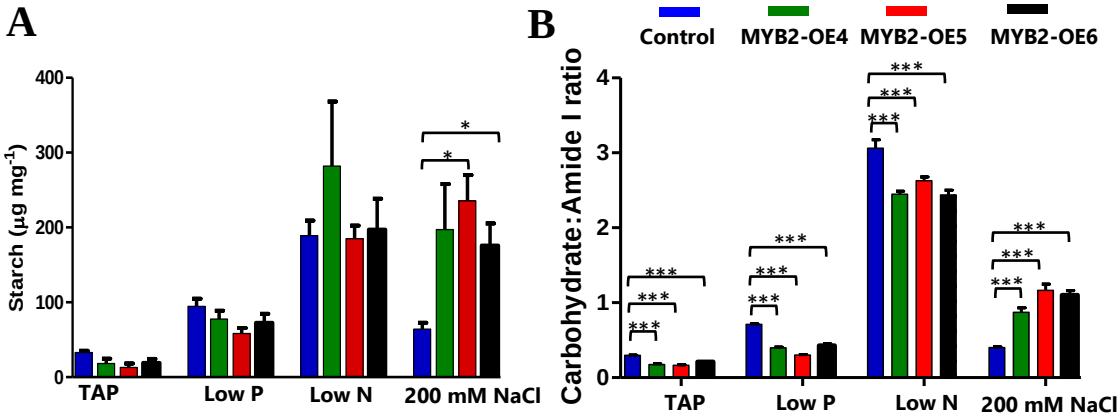

Figure 5.6: Starch content of control and MYB2-OE *C. reinhardtii* cells**. (A)** Cells were grown in standard TAP media and 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and 200 mM NaCl for 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. **(B)** Carbohydrate:Amide I ratio derived from the previous FT-IR spectroscopy experiment. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA).

The total chlorophyll and carotenoids content of the OE lines was also determined. As shown in Figure 5.7, the OE of MYB2 did not significantly affect the chlorophyll or carotenoid content of the cells grown in low N and salt stress media, showing similar levels to the control lines. However, OE lines exhibit a higher chlorophyll and carotenoid content when grown in standard TAP or low P media.

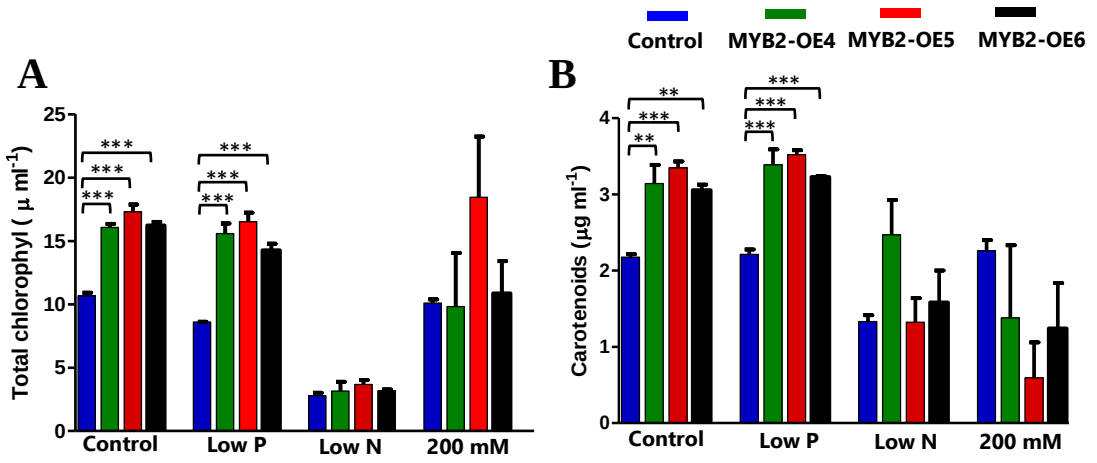

Figure 5.7: Chlorophyll and carotenoid content of control and MYB2-OE *C. reinhardtii* cells in different stress conditions. *C. reinhardtii* cells were grown in standard TAP media and with 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and high NaCl (100 and 200 mM) media. For salt stress, cells were first cultivated in standard TAP media until day 3 and then different concentrations of salt were added. Total chlorophyll (Chl a + b) (A) and carotenoids (B) were quantified at day 7 in the different stress conditions. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the standard TAP control condition (ANOVA).

5.2.4 Characterization of MYB2-KD lines

To further analyse the potential role of MYB2 in the lipid and starch metabolism, MYB2 knockdown lines (MYB2-KD) were generated and then characterized. These lines were previously constructed in my lab by using a MYB2 gene-specific amiRNA (Tan, Bajhaiya and Pittman, unpublished) but were not characterised. Figure 5.8 shows the RT-PCR based expression analysis of six MYB2-KD lines grown in low P media, previously shown to induce the MYB2 gene expression (Chapter 4). Four out of the six lines analysed showed significant knockdown of the MYB2 gene (lines 4, 8, 11 and 13) compared with the control lines (*cw15*). For further studies, lines 8, 11 and 13 were chosen due to the success of MYB2 silencing.

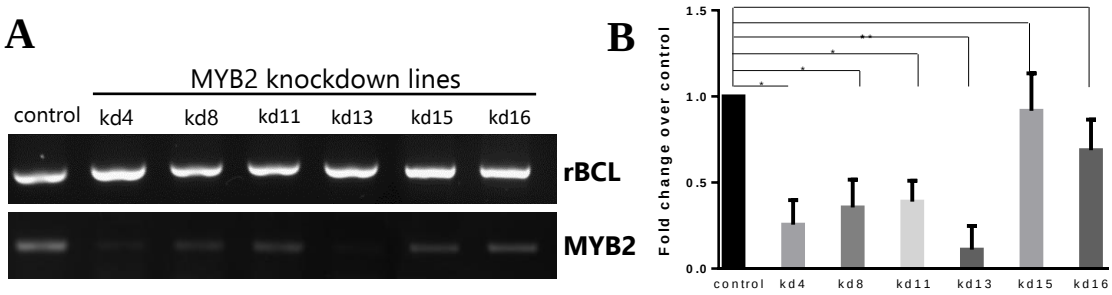

Figure 5.8: Characterization of MYB2-KD lines. (A) *C. reinhardtii* MYB2-KD cells together with control *cw15* cells were grown in low phosphorus TAP media. (A) Gel electrophoresis of MYB2 expression in transformed strain lines determined by RT-PCR. Six independent transformants were tested (line 4, 8, 11, 13, 15 and 16), with the rBCL gene as an internal control. Total RNA was extracted after 5 days of growth. (B) Densitometry analysis was performed using Image J gel analysis tool, with MYB2 expression normalized to rBCL and then the fold change in MYB2 over the control was plotted (Results from three independent experiments (n=3), *P < 0.01 , **P < 0.001 using Student´s t-test).

Physiological characterization of MYB2 KD lines

FT-IR spectroscopy was used to characterize the KD lines under low P, low N and high salt media at day 7 during stationary phase when the carbon storage metabolites begin to accumulate. As previously shown, the major differences observed in the FT-IR analysis of the lines were caused by the stress condition that they were subject to. To avoid the potential masking of differences within the lines in each treatment, this time the data set of each stress condition was analysed separately. The averaged FT-IR absorption spectra of the lines at each stress condition is shown in Figure 5.9. The areas of interest were focused in the total lipid region (1740 cm^{-1}) and the carbohydrate region (1300-950 cm^{-1}). MYB2-KD lines grown under standard TAP media showed no clear differences in the absorption spectrum when compared to the control lines, indicating similar biochemical composition (Fig. 5.9 A). However, upon stress, the differences within the lines become clear. The plots showed higher peaks in the lipid area of the three tested MYB2-KD lines (red, green and black lines) grown under low P, low N and high salt media, indicating a higher total lipid content of the KD cells compared to control lines (blue line). Regarding the carbohydrates, KD lines, appear to have higher carbohydrate contents upon low N and salt stress compared to the control line, with exception of KD line 13 in low N media which showed a lower carbohydrate content.

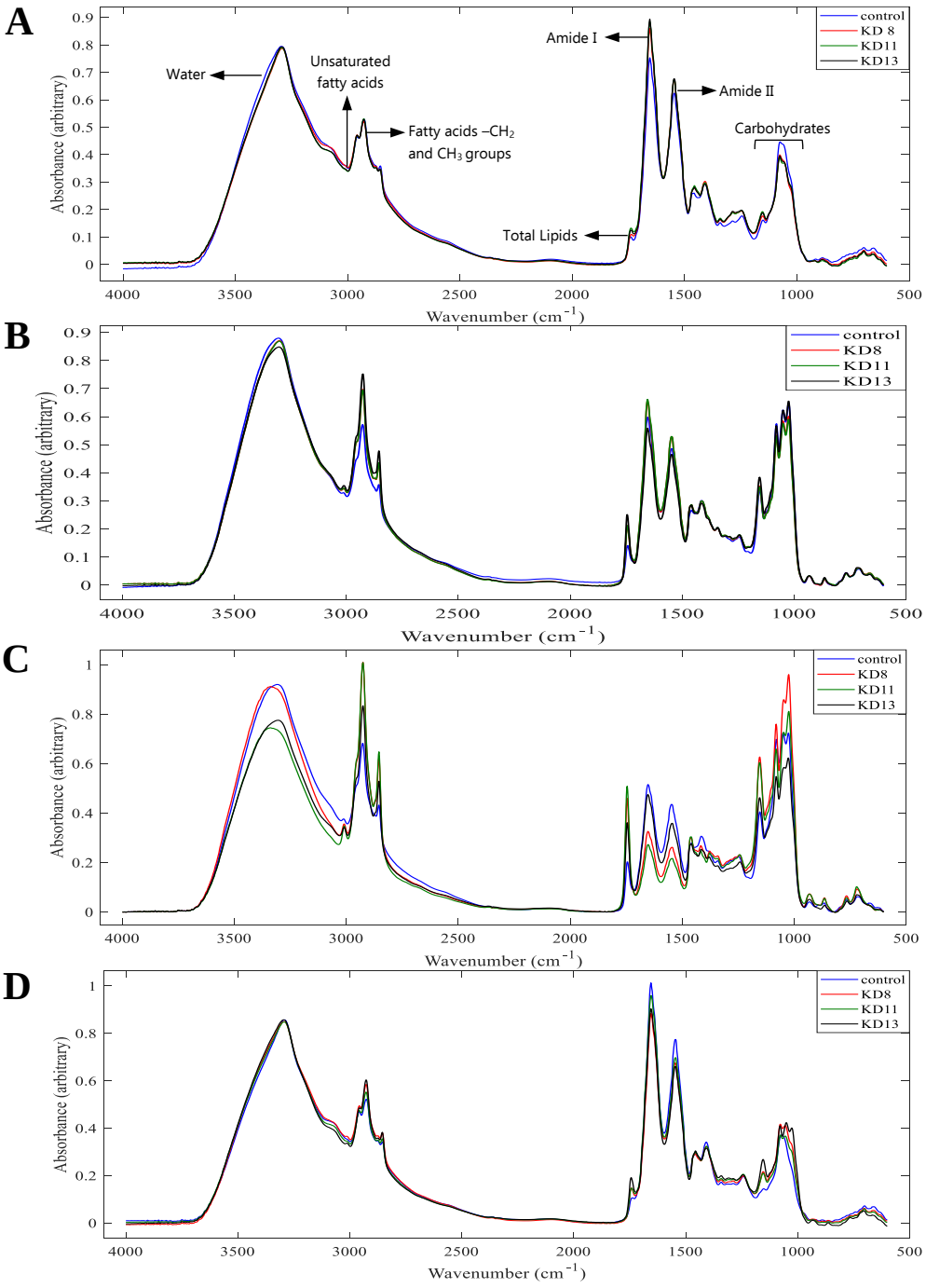

Figure 5.9: Figure 4.5: FT-IR spectroscopy data taken from MYB2 KD lines grown under different stress conditions. EMSC corrected FT-IR spectra in the range of 4000-1000 cm^{-1} obtained from average values from three biological replicas of MYB2-KD and control cells (*cw15*) grown in standard TAP (A), low P (B), low N (C) and 200 mM NaCl media (D).

As a further analysis of the FT-IR data, PCA was performed on the FT-IR spectra in the range of 4000–1000 cm^{-1} for the control and KD lines in each stress condition (Fig. 5.10). The score plots revealed that the spectral data collected from the MYB2-KD lines tend to cluster in a different group than the control lines in all the tested conditions. The segregation between the two groups in the PCA plots revealed that the biochemical composition of the KD lines has changed due to downregulation of MYB2 with the two PCs accounting for >90% of the total variation. The plots also showed that the MYB2-KD line 13 (black symbols) tends to cluster apart from KD lines 8 and 11 but also apart from the control line under stress conditions. One potential explanation for this is that KD line 13 has a different efficiency in the knockdown of the MYB2 gene, which could be inducing other metabolic changes. Low P treatment showed the clearest pattern of clustering of the KD lines along the positive side of the PC2 axis which are explained by the loading plots as a result of higher lipid content and lower carbohydrate content (Fig 5.10 B, positive blue peak at 1740 cm^{-1} and negative blue peak at 1160-1036 cm^{-1}, respectively). In contrast, a less clear pattern of clustering was observed in cells subject to Low N and salt stress, however, still a similar trend of clustering pattern can be observed along PC1 for both of the conditions between the KD lines and control line. KD lines subject to low N treatment tend to cluster to the positive side of PC1 as a result of higher lipid content and higher carbohydrate content (Fig 5.10 C, positive red peak at 1740 cm^{-1} and at 1160-1036 cm^{-1}, respectively). KD lines subjected to salt stress tend to cluster to the negative side of PC1 also as a result of higher lipid content and higher carbohydrate content (Fig 5.10 D, negative red peak at 1740 cm^{-1} and at 1160-1036 cm^{-1}, respectively). Altogether, these results demonstrate that there were significant metabolic differences between the MYB2-KD lines when compared to control lines.

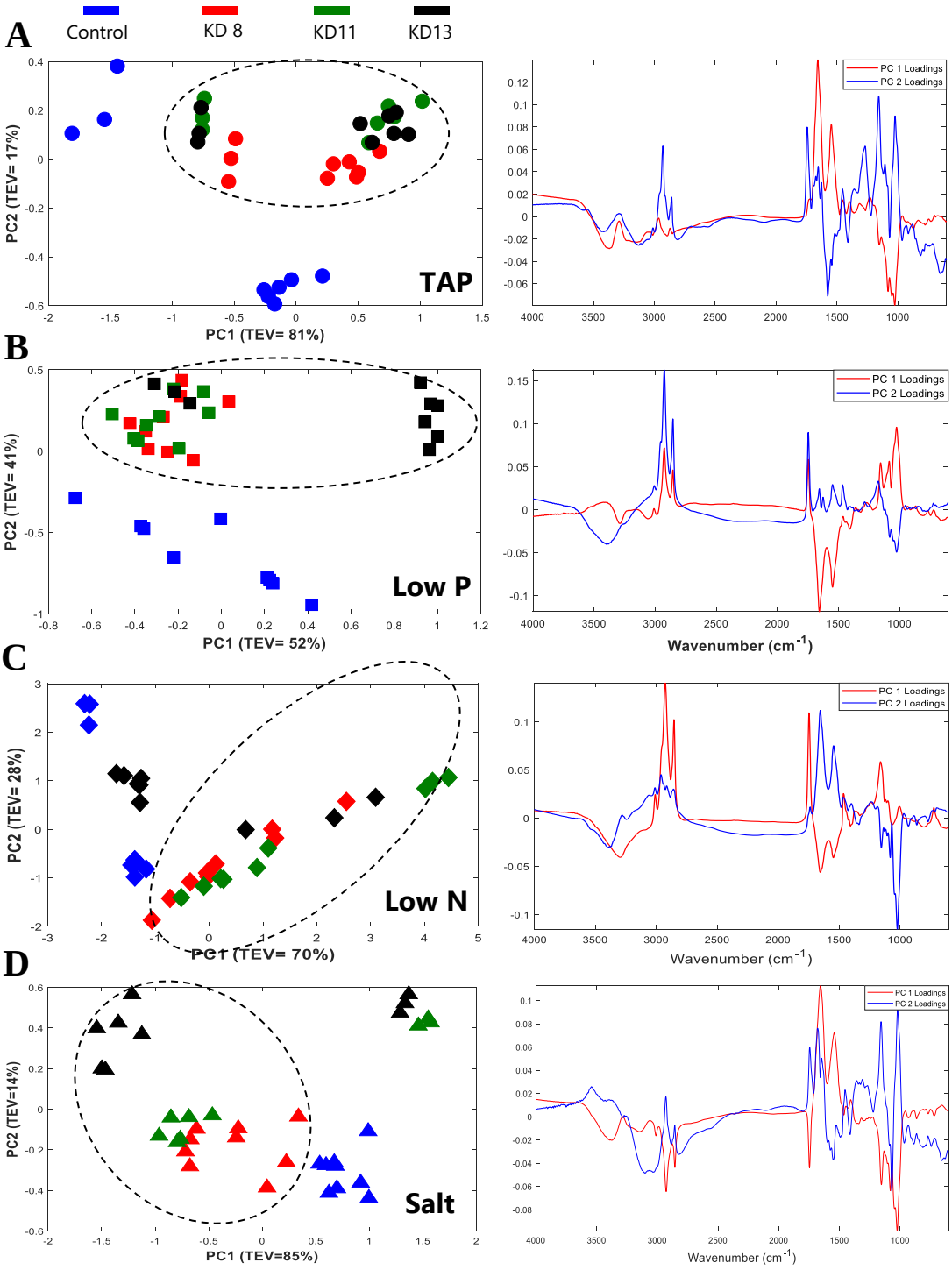

Figure 5.10: FT-IR spectroscopy analysis of MYB2-KD lines. Principal component analysis (PCA) scores (left panels) and PC loading plots (right panels) of FT-IR spectra from control lines (cw15) and MYB2 KD lines (A, B) and wild type (*cw15*) and PSR1 over-expression lines (C, D) under high P (A, C) and low P (B, D) conditions. Analysis was determined from three replicate spectra for each sample and treatment.

The Nile red quantification of neutral lipid showed an increase in the lipid content of KD lines treated under low N conditions but no significant increase was observed in any of the other stress (Fig. 5.11 A). Gravimetric quantification by SOXTEC confirmed these results showing a significant increase in the lipid content only of KD cells treated under low N conditions (Fig.5.11 B). These results are in accordance with the results obtained by the FT-IR spectroscopy. Lipid:amide I ratios showed the increase in the lipid content of the KD lines under low N conditions, however, it also reported higher lipid content of KD lines under low P and salt stress (Fig. 5.11 C). Nevertheless, it appears that the knockdown of MYB2 has the biggest impact on lipid content of cells treated under low N conditions.

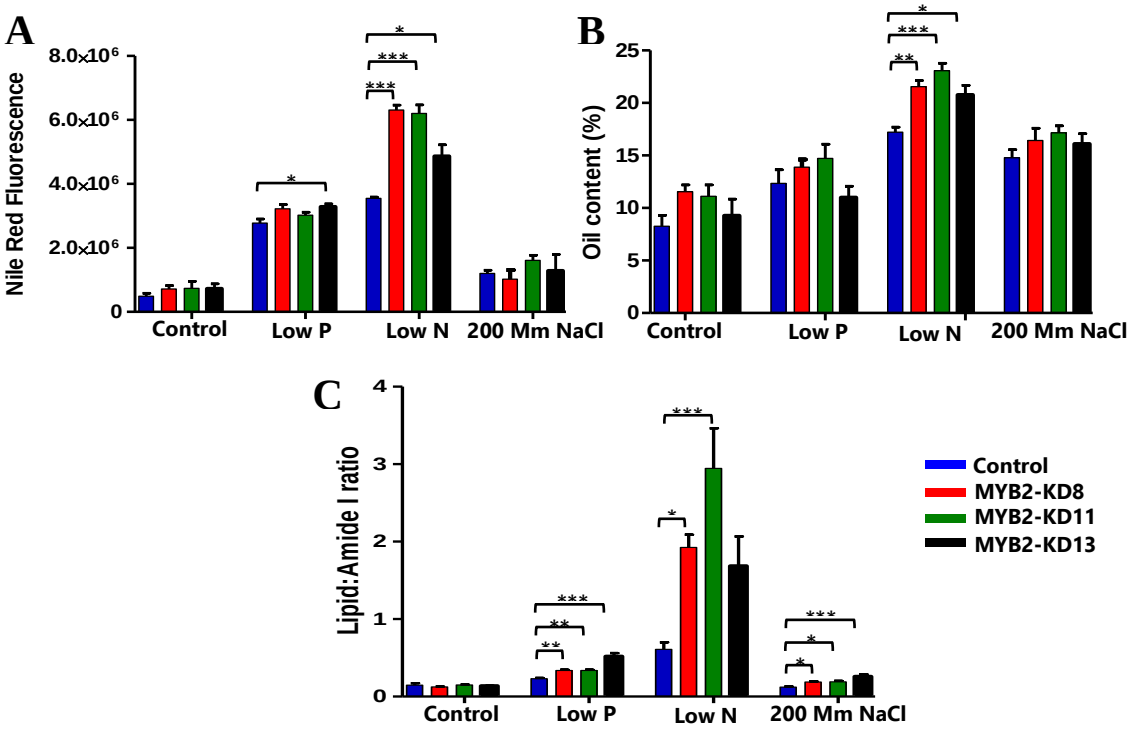

Figure 5.11: Lipid content of control and MYB2-KD *C. reinhardtii* cells**. (A)** Cells were grown in standard TAP media and 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and 200 mM NaCl for 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. Neutral lipid content was measured using Nile red and results were normalized against cell number using 680 nm optical densities. **(B)** Lipid quantification by SOXTEC extraction utilising hexane. Results are expressed in percentage of dry weigh **(C)** Lipid:Amide I ratio derived from the previous FT-IR spectroscopy experiment. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA).

Starch content was measured using the alpha-amylase/amyloglucosidase activity assay (total starch assay kit). The results indicate that the starch content was higher in the three tested OE lines under low N and salt stress conditions but significantly lower in low P when compared to control (Fig. 5.12 A). These results were corroborated through the analysis of the FT-IR data by calculating the carbohydrate:amide I ratios. This analysis also indicated that under low N conditions two of the three tested OE lines showed significantly higher levels of carbohydrates, and low P exhibit significantly lower levels of carbohydrates when compared to control. However, OE lines subjected to salt stress did not showed any differences I their starch content compared to the control line.

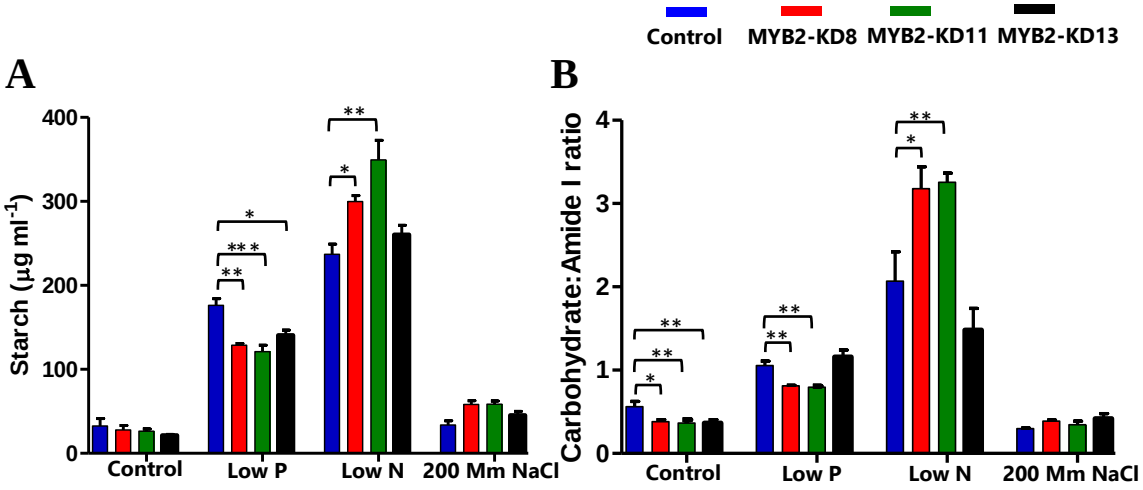

Figure 5.12: Starch content of control and MYB2-KD *C. reinhardtii* cells**. (A)** Cells were grown in standard TAP media and 3 different stress conditions: low phosphorus (0.01 mM K_2HPO_4/KH_2PO_4), low nitrogen (0.7 mM NH_4Cl) and 200 mM NaCl for 7 days. For salt stress, cells were first cultivated in standard TAP media for 3 days before the salt was added. **(B)** Carbohydrate:Amide I ratio derived from the previous FT-IR spectroscopy experiment. Error bars represent S.E.M from 3 independent experiments. *, **, *** ($p<0.05$, 0.01, 0.001) indicates significant difference from the Standard TAP control condition (ANOVA).

5.3 Discussion

Transcription factor engineering has emerged as a new strategy for a more efficient improvement in the lipid and starch biosynthesis in microalgae (Bajhaiya, Ziehe and Pittman, 2017). As the manipulation of individual enzymes within metabolic pathways has shown mixed success (Radakovits *et al.*, 2010; Driver *et al.*, 2017), a broader regulation of multiple enzymes mediated by TFs could provide a stronger response. Taking advantage of the available information from transcriptomic, proteomic and epigenomic studies, we have focused the attention in the characterization of MYB2, a MYB TF family member, highly upregulated under different stress conditions that are known to induce carbon storage metabolism. In this study we provide evidence that MYB2 is involved in the regulation of carbon storage metabolism under different stress conditions, however, the molecular mechanism underlying MYB2 regulation remain uncertain and appear to be complex and tightly regulated.

First, it is important to mention that this study did not show the levels of expression of the MYB2-OE lines due to challenges presented in its detection. Both control lines and OE lines showed the same levels of MYB2 expression (data not shown) which was an unexpected result. One possible explanation to this could be that the mRNA stability of MYB2 transcript is tightly regulated. As previously stated in Chapter 4, TFs are typically expressed at very low copy number per cell and only a small amount is often needed to activate a cascade of genes. This low expression can be the result of the low transcription rates of their promoters (weak promoters) and also from their RNA stability. Recent genome-wide studies have determined that upon stress, the microalgae cell experiences waves of genome-wide transcriptional events divided in different stages (Park *et al.*, 2015). Mathematical algorithm use for the analysis of RNA-Seq data of cells cultured under N starvation have revealed that there are at least 31 waves of transcriptional expression and that they are activated as early as 12 minutes after the stress induction and continue changing until 8 hours (López *et al.*, 2015). Based on the timing and the functional composition of the different transcriptional waves, they have been divided into two main stages: early response genes and late response genes. Within the early response genes we can find genes involved in N assimilation, acetate assimilation, the Calvin-Benson-Bassham cycle, starch accumulation and remodelling of lipid membranes. When the stress is

prolonged, another wave of differential gene expression dominates, with genes involved in metabolic changes with more specific functions related to the induction of genes directly involved in lipid metabolism (Gargouri *et al.*, 2015). This observation suggests that the changes in the metabolism towards the accumulation of carbon storage molecules are under tight transcriptional control. Moreover, studies have identified regulatory hubs that control the metabolic transition of the cells towards carbon storage metabolism (Gargouri *et al.*, 2015). Interestingly, some of the genes identified as regulatory hubs, exhibit oscillating expressions levels upon stress correlating well with various metabolic processes. MYB1 and MYB2 are two of the genes that presented this oscillatory expression pattern upon stress, suggesting that the regulatory hubs themselves are subjected to a tight regulation. It is possible that this regulation is modulated by posttranslational modifications such as riboswitches (Wachter *et al.*, 2007), or RNA-proteins interactions (Gutierrez, MacIntosh and Green, 1999). Therefore, if MYB2 oscillatory behaviour is controlled by post-transcriptional modifications, the fine mechanism of control activated upon stress would inactivate its activity when the TF is not required Thus, degradation of the overexpressed MYB2 through post-transcriptional modifications may have been the reason why we did not observed any significant difference between the OE lines and control lines after 7 days of stress. In support of this hypothesis, Bajhaiya et al. (2016b) reported a similar unexpected result when PSR1 TF was overexpressed in *C. reinhardtii* cells. The overexpression of this gene was detected only in low P conditions, where endogenously this TF is upregulated. However, the fact that the overexpression of this genes was under the control of a constitutive promotor, it would be expected that the overexpression could be detected under any condition. Moreover, the transcription inhibition by actinomycin D reduced the abundance of all transcripts tested with the exception of *PSR1*, suggesting that the levels of this TF were regulated post-transcriptionally. A time course analysis upon the exposure to the stress might be appropriate to confirm the overexpression of the gene.

An interesting and rather unexpected finding from this present study was observed with the growth of the overexpression lines. Compared to control lines, MYB2-OE lines formed aggregates at the bottom of the flasks upon salt addition. Although no further studies were performed to understand the origin of this behaviour, the fact that this occurred in all three of the tested OE lines and in all

biological replicas to the same extent, shows that this was not an isolated event but rather a consequence of MYB2 overexpression. As Figure 5.3 B shows, the majority of the cells were embedded in the biofilm-like structure attached to the bottom of the flask leading to a clear upper medium with very few cells. The capacity to form biofilms have been demonstrated on bacteria, actinomycetes, yeast, fungi and algae (Fattom and Shilo, 1984; Salehizadeh, Vossoughi and Alemzadeh, 2000; Zhanf *et al.*, 2002). This association to surfaces of microorganisms involves the secretion of different polymers of sugars denominated exopolysaccharide (EPS). It is thought that formation of biofilms is a way of protection of microorganism upon abiotic stress (Steele, Franklin and Underwood, 2014). Different stress conditions such as pH, light intensity, temperature and high salt have demonstrated to induce the biofilm formation of microorganisms (Kaplan *et al.*, 1987; Holzinger and Karsten, 2013). Diatoms, and microalgae species as *Dunaliella salina* and *Chlorella pyrenoidosa* have demonstrated to increases the production of EPS when expose to high salt concentrations, forming biofilms that enable them to cope with fluctuating salinity (Mishra and Jha, 2009). The composition of the EPS varies depend on several factors such as species and cultivation conditions, however, in general it is mainly composed from a mixture of different types of carbohydrates and proteins (Xiao and Zheng, 2016). Interestingly, MYB2 OE lines were shown to have considerably different macromolecular composition than the control lines, as observed in the FT-IR spectra PCA plot (Fig. 5.4), in part explained by their higher carbohydrate content. These results were confirmed by the biochemical assays in which MYB2-OE lines showed a higher carbohydrate content than the control. It is unclear whether this increase in carbohydrates corresponds to EPS and if this is the drive for the biofilm formation. It would be intriguing to determine the carbohydrate composition and if this phenomenon is a direct consequence of the overexpression of MYB2 or an indirect mechanism activated by the cells upon MYB2 overexpression. Although this was out of the scope of this investigation, it is worth mentioning it due to its potential application for the biofuel industry to reduce the harvesting cost and the potentially increase in the biomass productivity (Liu *et al.*, 2013; Gross, Mascarenhas and Wen, 2015).

The effect of the MYB2 overexpression in the carbon metabolism upon stress was another intriguing and unexpected finding. Both, overexpression and

downregulation of MYB2 showed similar phenotypes in the accumulation of carbon storage molecules. The most striking difference was observed upon N limitation, where both upregulation and downregulation of MYB2 showed an increase of more than 2-fold in the lipid content of the cells. P limitation and salt stress also produced an increase in the lipid production in both of the cell lines; however, this was not as striking as the cells grown in low N media. Regarding starch accumulation, OE lines grown in salt stress media showed the highest starch content with more than 2-fold increase when compared to the control. MYB2-KD also showed an increase in starch under salt stress conditions although the effects were moderate, only showing a slight increase. These results suggest that MYB2 plays a role in the carbon storage metabolism under stress conditions, however, it remains largely unclear what are the mechanism underlying this regulation but apparently are complex and tightly regulated.

There are two potential explanations that can account for the counterintuitive phenomenon of obtaining similar phenotype for overexpression and downregulation lines instead of opposite phenotypes as might be expected. Although speculative, it is conceivable that mechanisms of genetic robustness can be contrasting the loss of MYB2 in the MYB2-KD, leading to a similar phenotype than the overexpression lines. Genetic robustness can emerge from redundant genes where the loss of one gene is replaced by another one with overlapping functions and expression patterns, as previously reported for mutant lines in several organisms (Cohen *et al.*, 1987; Cadigan, Grossniklaust and Gehring, 1994; Rawls *et al.*, 1998; Danisman *et al.*, 2013). Consistent with this concept, Moyano et al. (1996) reported the redundancy of two MYB genes, MYB305 and MYB340 in *Antirrhinum* flowers that participate in flavonoid biosynthesis. Both of these TFs were able to regulate the expression of the same targets genes but MYB340 was a stronger activator than MYB305. However MYB305 more than MYB340 was able to bind to the target genes when both proteins were been expressed. This was explained by a competitive phenomenon in which MYB305 competed with MYB340 to reduce its effectiveness as a transcriptional activator when both TFs were expressed simultaneously (Moyano, Martínez-Garcia and Martin, 1996). This competition phenomenon of redundant TFs is not new and several examples have been reported in plants (Rudnik *et al.*, 2017). Moreover, the MYB TF family members show redundancy in maize. MYB TFs P

and C1 in maize are able to bind to the same targets in the anthocyanin biosynthesis (Grotewold *et al.*, 2000). Apparently, the similarity in their DNA binding domain is responsible for this redundancy in their function. Interestingly, MYB1 and MYB2 in *C. reinhardtii* both were found to be upregulated under stress conditions, present similar patterns of expression and share a 67% of similarity in their DNA binding domains. It is thus interesting to speculate that MYB1 and MYB2 could potentially have redundant competitive functions, and in the absence of MYB2, MYB1 could compensate its function. The increase in the content of carbons storage molecules in MYB2-OE lines could be given by the strong transcriptional activation produce by the OE of MYB2 in genes involved in carbon storage pathways. Now the same phenotype could be observed in MYB2-KD lines, whereby MYB1 in the absence of MYB2 acts as a stronger transcriptional activator of the same genes, thus leading to a similar phenotype than the OE lines.

Another possibility to consider, equally plausible, is that mechanism of genetic compensation can be activated upon the MYB2 KD. Different from genetic robustness, genetic compensation or transcriptional compensation occurs when the loss of function of a protein triggers the transcriptional activation of similar genes that can compensate the main function of the loss one. However, this activation occurs in similar but not equal genes, while compensating the function of the loss one, can also lead to different phenotypes (El-Brolosy and Stainier, 2017). If the loss of MYB2 is triggering mechanisms of genetic compensation, they could be strongly activating the transcription of genes involved in carbon storage metabolism to compensate the lack of MYB2 function. However, at the same time they could be activating other genes producing different phenotypes. If this is the case, this could explain why MYB2 OE and KD lines had similar phenotypes regarding the accumulation of carbon storage molecules yet only OE lines exhibit biofilm-like growth patterns. It appears to be common that genes involved in complex regulatory networks often have mechanisms of compensation to buffer the genetic or environmental changes that could affect their viability (Barabási and Oltvai, 2004). This study provides preliminary evidence of the role of MYB2 in carbon storage metabolism, however, whether MYB2 has a redundant function or if its loss is compensated by the activation of other genes remains unclear. Further experimental studies are needed in order to fully unveil the functions of this TF in the carbon

storage metabolism. Nevertheless, it is crucial to have awareness of these possibilities for the successful application of TFE.

This study also allowed us to observe that the role of MYB2 in carbon storage metabolism apparently is not a global response to macronutrient limitation but rather a stress-specific response. As shown in the results section, the most striking differences from MYB2 manipulation were observed under low N conditions, whereas low P and salt stress only showed a moderate response to MYB2 expression fluctuations. This is not surprising since previous work has shown that although stress conditions produce similar phenotypes in terms lipid and starch accumulation, the genetic response of the cells appears to be different. The comparison of the transcriptomes profiles of cells upon P, N and S revealed that the upregulated genes tend to be nutrient specific (Schmollinger *et al.*, 2014b). Moreover, gas chromatography-mass spectrometry-based metabolomics analysis indicated important differences in the metabolites composition of cells under P, N and S limitation (Bölling, Fiehn and Turpin, 2005). These observations suggest that the regulatory mechanisms activated under the different stress are different and induce different lipid and starch accumulation programs. This is an important point to consider when studying the role of regulators involved in the biosynthesis of carbon storage molecules, and appears to be that MYB2 plays a more important role in N limitation conditions.

An increasing amount of evidence indicates that MYB TFs play crucial regulatory roles in the stress response of cells but many of the functions of these proteins, especially in microalgae are largely unknown. In this study evidence has been provided about the importance of the MYB2 TF in *C. reinhardtii* in the accumulation of starch and lipids in response to abiotic stress. However, the downstream targets and its potential redundancy are still not determined. Certainly the regulation of MYB2 appears to be complex and its potential redundancy could provide a challenge for studies that aim to determine its role in more detail. In addition, other fundamental questions about MYB2, such its post-transcriptional modifications, need to be investigated. Thus, this study represents a first but important step toward gaining an understanding of the role of this TF in the starch and lipid biosynthesis in *C. reinhardtii*. This kind of understanding is essential as it

may hold a valuable potential for being engineered to enhance the production of carbon storage molecules for the biofuel industry.

Chapter 6

General Discussion

6.1 General Discussion

Carbon storage molecules (starch and lipids) from microalgal biomass have acquired substantial interest in recent years as a feedstock for biofuel production. The accumulation of these molecules in microalgae is a well-known response to stress (Goncalves *et al.*, 2016). While several abiotic stresses, including nutrient starvation or salinity, have been reported to increase the production of these carbon storage molecules in microalgae (Venkata Mohan and Devi, 2014; Zhu, Li and Hiltunen, 2016b), these stress conditions can significantly hinder their growth (Tan and Lee, 2016), leading to low feedstock productivity. This situation is a substantial problem for a microalgal-based bioenergy industry, which aims to maximize the production of biofuels in order to convert it into a cost effective eco-friendly alternative to fossil fuels. But currently the use of microalgal derived biofuel is not economically viable in contrast to conventional fuels, and thus microalgae biotechnology is currently focussed on alternative high-value products and applications, such as aquaculture, cosmetics and nutraceuticals rather than biofuels.

There are many different factors that still need to be addressed in order for microalgae biofuel to become a viable alternative to fossil fuels. Improvements to biomass is one key factor such as by generating an increase in the net lipid and starch yield (Hannon *et al.*, 2010). Genetic engineering has become an interesting approach to improve the carbon flux towards the production of lipid and starch in microalgae without the need of applying any condition (such as an extreme stress inducer) that may interfere with their growth (Radakovits *et al.*, 2010). The main requirement for genetic engineering to be successful is to have a deep understanding of the molecular mechanisms that coordinate carbon storage metabolism. Initially, genetic engineering approaches have focused on targeting specific key enzymes that mediate lipid and starch metabolism. However, the outcomes from such manipulations of individual enzymes have had a mixed success (Radakovits *et al.*, 2010; Driver, Bajhaiya and Pittman, 2014). Increasing evidence indicates that pathways leading to the accumulation of these storage molecules are part of a complex network, which is tightly regulated. Thus, the manipulation of individual enzymes may not be the most appropriate approach. For example, a recent attempt to manipulate TAG biosynthesis, through increased abundance of the glycerol-3-phosphate dehydrogenase enzyme, was able to drive more carbon flux into the Kennedy

pathway, but rather than increasing TAG concentration, only a precursor of TAG (phosphatidic acid) was increased, indicating an unexpected block in the TAG pathway (Driver *et al.*, 2017). The manipulation of several key enzymes simultaneously may instead provide stronger and more successful outcomes.

Transcription factors regulate the expression of pools of genes that share a high connectivity, such as genes encoding enzymes within a metabolic pathway (Alon, 2007). Therefore it was logical to hypothesise that by manipulating TFs that are key to the regulation of carbon storage biosynthesis, we would not only be manipulating individual genes, but a pool of them, leading to a stronger outcome in carbon storage metabolism. This is the goal of transcription factor engineering (TFE), which searches for key TFs as targets for manipulation to gain a significant impact on the yield of carbon storage molecules (Bajhaiya, Ziehe and Pittman, 2017). In this investigation, I have provided insights into the role of two TFs involved in the biosynthesis of carbon storage molecules in *C. reinhardtii*. In the first part of this report, I validated for the first time specific targets of PSR1, a TF previously shown to be involved in the biosynthesis of starch and TAG in microalgae (Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b). In the second part, I provided insights into the role of MYB2, a TF belonging to the MYB TF family, for which a role in carbon storage metabolism has not been previously reported before. Altogether, this report contributes towards the better understanding of the complex TF network regulating the carbon storage metabolism in microalgae.

PSR1 is a TF that is important in the response to phosphorus (P) starvation as it activates mechanisms that allow microalgae cells to survive and acclimate to P limitation (Wykoff *et al.*, 1999). In addition, it is one of three TFs in microalgae that has been functionally characterized with a role in lipid and starch metabolism under P starvation. Moreover, its role is not only limited to P stress, but apparently also linked with other nutrient stress. A recent study identified PSR1 as a pivotal switch that triggers lipid accumulation under nitrogen (N) limitation, and exhibited similar expression patterns in cells under sulphur limitation (S) (Ngan *et al.*, 2015). Together this information has revealed the importance of this TF for microalgae in the regulation of carbon metabolism. Therefore PSR1 is an attractive target for TFE. However, also due to its important role, its modification can produce undesired pleiotropic effects. For example, PSR1 mutants have exhibit defects in their

photosynthetic activity and are appear to be more sensitive to other stresses which could potentially compromise biomass production (Shimogawara *et al.*, 1999). In consequence, care must be taken when choosing a target for TFE. For this reason it is crucial to have a deep understanding of the role of this TF in carbon storage metabolism to be able to manipulate it in the most efficient manner.

In this investigation we explored a different angle of TFE, by identifying how PSR1 regulates its downstream targets. Due to its similarity with PHR1, a well know *A. thaliana* TF (Rubio *et al*., 2001), and the prediction of a DNA binding domain (DBD) within the PSR1 amino acid sequence, it was assumed that PSR1 was regulating its downstream targets directly through transcriptional activation. However, there was no experimental proof that PSR1 had the ability to bind to the promotor regions of genes and activate their transcription. Here I provided the first evidence that PSR1 is able to bind to the promoter region of four target genes. The assumption that a conserved PHR1 binding site (P1BS) sequence motif, GNATATNC, which was observed within the promoter regions of the four chosen *C. reinhardtii* target genes was indeed a putative element for PSR1 binding, was tested by an electrophoretic mobility assay (EMSA).

First, I tested the ability of PSR1 to bind to a protein that is directly linked to the P starvation response; PTB2, a high affinity Na^+/Pi transporter, which is highly upregulated in *C. reinhardtii* upon P starvation (Moseley, Chang and Grossman, 2006). PSR1 was able to bind to the putative P1BS motif within the PTB2 gene, confirming its ability to act as a TF activating the transcription of genes in response to P limitation. Likewise, PSR1 was demonstrated to bind the P1BS motif within the promotors of the glycerol-3-phosphate dehydrogenase 4 (GPD4) gene, involved in TAG biosynthesis (Driver *et al.*, 2017), and soluble starch synthase (SSS1, SSS5) enzymes involved in the starch biosynthesis (Ball and Morell, 2003). This gave the first insights into the mechanism underlying PSR1 regulation of starch and lipid biosynthesis. Although binding to the promoter regions of these genes is not enough alone to prove that PSR1 is activating their transcription, the fact that this enzyme have demonstrated to have a PSR1-dependent expression (Bajhaiya *et al.*, 2016b), makes the assumption of direct regulation stronger. In addition, the results suggested that PSR1 is not able to bind to the P1BS region in the promoter of the PhoB plastidial starch phosphorylase (SP2), another important enzyme in the starch

metabolism (Dauvillée *et al.*, 2006). Interestingly this enzyme was found downregulated in cells that overexpress PSR1 under P starvation (Bajhaiya et al., 2016b). This discovery reinforces the idea that PSR1 functions as a transcriptional activator and it may not be involved in the downregulation of genes.

In summary, this part of the study presented important evidence that further confirms the TF activity of PSR1 and its specificity of binding to the P1BS element in the promoter region of its target genes. This information is important due to the fact that TFE is not only limited to the modification of the TFs (trans-engineering), but also to the cis regulatory elements in their targets genes (cis-engineering) (Rushton *et al.*, 2002). This approach of TFE has arisen after the observation that constitutive expression of a TF is often accompanied by undesired pleotropic effects. The cis-TFE method may be able to produce tailored promoters that are optimized to facilitate tight control of the gene expression, by optimizing the copy number and spacing of the key *cis* elements in the promoter sequences of targets genes (Liu, Yuan and Stewart Jr, 2013). Thus, the knowledge of the cis regulatory elements of PSR1 is key for its potential as a target for TFE. So far, approximately 200 synthetic promoters have been successfully developed by engineering different cis stress-responsive elements or domains (Shrestha, Khan and Dey, 2018). This information gives light into the potential of PSR1 engineering for improving the yields of carbon storage molecules for the biofuel industry.

The second part of this investigation focussed on the MYB TF family, particularly the R2R3-type of MYB proteins. Members of this family have been found to be highly upregulated in different stress conditions that induce the accumulation of carbon storage molecules (Schmollinger *et al.*, 2014; Gargouri *et al.*, 2015; Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b). With the aim of further investigating a potential role of members of this family in lipid and starch metabolism in *C. reinhardtii*, I performed a phylogenetic analysis. Such analysis is an important first step to the functional characterization of proteins and provides useful information about sequence similarities and conserved domains in relation to the function of the proteins. Since there are very few functional studies of the MYB TF family in *C. reinhardtii*, the analysis was performed by comparing *C. reinhardtii* R2R3-MYB TFs with R2R3-MYBs from *A. thaliana* where several members have been functionally characterized (Stracke, Werber and Weisshaar, 2001; Mohamed *et*

al., 2017). Interestingly, the 12 R2R3-MYBs identified in the proteome of *C. reinhardtii* share high homology with a clade of R2R3-MYBs from *A. thaliana* that are involved in stress response such as dehydration, temperature, salt tolerance and disease resistance. This high homology was due to conservation in the DBD sequence, which could potentially mean that the *C. reinhardtii* and *A. thaliana* proteins bind to similar targets, thus exerting similar functions. In contrast, a conserved *A. thaliana* motif within the C-terminal region of the proteins is not shared with the *C. reinhardtii* proteins. Instead they possess other conserved motifs as a coiled-coil domain, which may be important for protein-protein interaction. It is unclear whether the similarities and differences between R2R3 MYBs from *A. thaliana* and *C. reinhardtii* are reflected in their functionalities. Nevertheless, it is logical to hypothesize that *C. reinhardtii* R2R3 MYBs could play important roles in stress response. In conclusion, the phylogenetic studies together with the transcriptomic data, place the *C. reinhardtii* R2R3-MYBs as a good target for regulators of carbon storage metabolism in response to stress.

To determine whether the potential role of R2R3-MYBs in stress response in *C. reinhardtii* was related with the metabolism of lipids and starch, we performed a functional analysis. The attention was focus on the characterization of MYB2. The choice of this TF as a potential regulator of lipid and starch biosynthesis was based on previous transcriptomic data highlighting MYB1 and MYB2 as highly upregulated under low P and low N conditions (Schmollinger *et al.*, 2014; Ngan *et al.*, 2015; Bajhaiya *et al.*, 2016b). Unfortunately due to experimental challenges, it was not possible to study further the role of MYB1. On the other hand, overexpression (OE) and knockdown (KD) lines for MYB2 were characterized. Results obtained were unexpected but interesting. Manipulation of MYB2, either through OE or KD induced an equivalent phenotype regarding the accumulation of carbon storage molecules. In both lines, the most striking phenotype was observed under low N conditions. All lines showed a significant increase of more than 2-fold in the lipid content of the cells when compared to control lines under low N treatment. P limitation and salt stress also produced an increase in the lipid production in both of the cell lines, however, it was moderate compare to low N treated cells. Starch levels also showed differences especially in OE lines under salt stress. These results suggest that MYB2 plays a role in the regulation of carbon

storage molecules; however, its regulation appears to be much more complex than expected. It is possible that mechanisms of genetic robustness or genetic compensation may be triggered by the loss of function of MYB2 and could explain the similar phenotypes in OE and KD lines (Cohen *et al.*, 1987; Cadigan, Grossniklaust and Gehring, 1994; El-Brolosy and Stainier, 2017). In support of this hypothesis, other members of the MYB TF family in other species show redundancy in their function (Moyano, Martínez-Garcia and Martin, 1996). Moreover, it is often a common situation that upon abiotic stress, more than one member of a TF family plays the same role in the response to stress (Guo *et al.*, 2015). Therefore it may be speculated that there are other MYB TFs that function redundantly with MYB2 in response to stress. Interestingly, MYB1 is upregulated under the same stress condition than MYB2 and shares a 67% of similarity within their DBD, suggesting that they could potentially share the same targets and function redundantly. Although this data is preliminary it is crucial for further studies on these TFs to consider the potential redundancies in their functions. Further experimental studies are needed to fully understand the functions of this TF in carbon storage metabolism. However we have provided the first evidence of a MYB TF involved in the carbon storage metabolism in *C. reinhardtii* that could potentially be used for a target of TFE.

In summary, the current study provided important evidence for the TFE of two TFs in *C. reinhardtii*. Regarding PSR1, future work needs to be done to understand the functional implications of the binding to the P1BS element. In addition, determining the exact sequence to which PSR1 binds with higher specificity could help for the design of stronger synthetic promoters than can trigger high accumulation of lipid and starch molecules. Regarding MYB2, further work is required to determine the potential redundancy of this TF and its potential as a target for TFE. However this study has provided new insights into the complex network of regulation of starch and lipid metabolism in microalgae that can potentially be manipulated for the benefits of the biofuel industry.

References

Acién, F. G., Molina, E., Reis, A., Torzillo, G., Zittelli, G. C., Sepúlveda, C., & Masojídek, J. (2017). Photobioreactors for the production of microalgae. *Microalgae-Based Biofuels and Bioproducts*, 1–44. https://doi.org/10.1016/B978-0-08-101023-5.00001-7

Al hattab, M. Al, Ghaly, A., & Hammoud, A. (2015). Microalgae Harvesting Methods for Industrial Production of Biodiesel: Critical Review and Comparative Analysis. *Journal of Fundamentals of Renewable Energy and Applications*, *05*(02), 1–26. https://doi.org/10.4172/2090-4541.1000154

Albert, N. W., Lewis, D. H., Zhang, H., Schwinn, K. E., Jameson, P. E., & Davies, K. M. (2011). Members of an R2R3-MYB transcription factor family in Petunia are developmentally and environmentally regulated to control complex floral and vegetative pigmentation patterning. *The Plant Journal*, *65*(5), 771–784. https://doi.org/10.1111/j.1365-313X.2010.04465.x

Alon, U. (2007). Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461. https://doi.org/10.1038/nrg2102

Alonso, D. L., Belarbi, E. H., Fernández-Sevilla, J. M., Rodríguez-Ruiz, J., & Grima, E. M. (2000). Acyl lipid composition variation related to culture age and nitrogen concentration in continuous culture of the microalga Phaeodactylum tricornutum. *Phytochemistry*, *54*(5), 461–471. https://doi.org/10.1016/S0031-9422(00)00084-4

Alves, C., & Cunha, C. (2012). Electrophoretic Mobility Shift Assay: Analyzing Protein - Nucleic Acid Interactions, Gel Electrophoresis - Advanced Techniques, Sameh Magdeldin, IntechOpen. DOI: 10.5772/37619. Retrieved from: https://www.intechopen.com/books/gel-electrophoresis-advanced-techniques/electrophoretic-mobility-shift-assay-analyzing-protein-nucleic-acid-interactions

Ambawat, S., Sharma, P., Yadav, N. R., & Yadav, R. C. (2013). MYB transcription factor genes as regulators for plant responses: an overview. *Physiology and Molecular Biology of Plants : An International Journal of Functional Plant Biology*, *19*(3), 307–321. https://doi.org/10.1007/s12298-013-0179-1

Anderson, S. (2004). Oil Extraction and Analysis, Critical Issues and Comparative Studies. Retrieved from

https://pdfs.semanticscholar.org/3d5f/6b6843020eb024d20e524b639174019afba9.pdf

Andreev, D. E., Dmitriev, S. E., Terenin, I. M., Prassolov, V. S., Merrick, W. C., & Shatsky, I. N. (2009). Differential contribution of the m 7 G-cap to the 5' end-dependent translation initiation of mammalian mRNAs. *Nucleic Acids Research*, *37*(18), 6135–6147. https://doi.org/10.1093/nar/gkp665

Baba, M., & Shiraiw, Y. (2013). Biosynthesis of Lipids and Hydrocarbons in Algae, Photosynthesis, Zvy Dubinsky, IntechOpen, DOI: 10.5772/56413. Retrieved from: https://www.intechopen.com/books/photosynthesis/biosynthesis-of-lipids-and-hydrocarbons-in-algae

Bajhaiya, A. K., Dean, A. P., Driver, T., Trivedi, D. K., Rattray, N. J. W., Allwood, J. W., Pittman, J. K. (2016a). High-throughput metabolic screening of microalgae genetic variation in response to nutrient limitation. *Metabolomics : Official Journal of the Metabolomic Society*, *12*, 9. https://doi.org/10.1007/s11306-015-0878-4

Bajhaiya, A. K., Dean, A. P., Zeef, L. A. H., Webster, R. E., & Pittman, J. K. (2016b). PSR1 Is a Global Transcriptional Regulator of Phosphorus Deficiency Responses and Carbon Storage Metabolism in Chlamydomonas reinhardtii. *Plant Physiology*, *170*(3), 1216–1234. https://doi.org/10.1104/pp.15.01907

Bajhaiya, A. K., Ziehe Moreira, J., & Pittman, J. K. (2017). Transcriptional Engineering of Microalgae: Prospects for High-Value Chemicals. *Trends in Biotechnology*, *35*(2), 95–99. https://doi.org/10.1016/j.tibtech.2016.06.001

Baker, M. J., Trevisan, J., Bassan, P., Bhargava, R., Butler, H. J., Dorling, K. M., Martin, F. L. (2014). Using Fourier transform IR spectroscopy to analyze biological materials. *Nature Protocols*, *9*(8), 1771–1791. https://doi.org/10.1038/nprot.2014.110

Balat, M., Balat, H., & Öz, C. (2008). Progress in bioethanol processing. *Progress in Energy and Combustion Science*, *34*(5), 551–573. https://doi.org/10.1016/j.pecs.2007.11.001

Ball, S. G., & Morell, M. K. (2003a). From bacterial Glycogen to Starch: Understanding the Biogenesis of the Plant Starch Granule. *Annual Review of Plant Biology*, *54*(1), 207–233. https://doi.org/10.1146/annurev.arplant.54.031902.134927

Ball, S. G., & Morell, M. K. (2003b). Understanding the Biogenesis of the Plant Starch Granule. *Annu. Rev. Plant Biol*, *54*, 207–240. https://doi.org/10.1146/annurev.arplant.54.031902.134927

Barabási, A.-L., & Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, *5*(2), 101–113. https://doi.org/10.1038/nrg1272

Bari, R., Datt Pant, B., Stitt, M., & Scheible, W.-R. D. (2006). PHO2, MicroRNA399, and PHR1 Define a Phosphate-Signaling Pathway in Plants. *Plant Physiology*, https://doi.org/10.1104/pp.106.079707

Bellou, S., Baeshen, M. N., Elazzazy, A. M., Aggeli, D., Sayegh, F., & Aggelis, G. (2014). Microalgal lipids biochemistry and biotechnological perspectives. *Biotechnology Advances*, *32*(8), 1476–1493. https://doi.org/10.1016/j.biotechadv.2014.10.003

Berg JM, Tymoczko JL, Stryer L. Biochemistry. (2002). 5th edition. New York: W H Freeman. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21154/

Bleuven, C., & Landry, C. R. (2016). Molecular and cellular bases of adaptation to a changing environment in microorganisms. *Proceedings. Biological Sciences*, *283*(1841). https://doi.org/10.1098/rspb.2016.1458

Bölling, C., Fiehn, O., & Turpin, D. H. (2005). Metabolite profiling of Chlamydomonas reinhardtii under nutrient deprivation. *Plant Physiology*, *139*(4), 1995–2005. https://doi.org/10.1104/pp.105.071589

Boyle, N. R., Page, M. D., Liu, B., Blaby, I. K., Casero, D., Kropat, J., Merchant, S. S. (2012). Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in Chlamydomonas. *The Journal of Biological Chemistry*, *287*(19), 15811–15825. https://doi.org/10.1074/jbc.M111.334052

Brennan, L., & Owende, P. (2010). Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. *Renewable and Sustainable Energy Reviews*, *14*(2), 557–577. https://doi.org/10.1016/j.rser.2009.10.009

Bustos, R., Castrillo, G., Linhares, F., Puga, M. I., Rubio, V., Pérez-Pérez, J., Paz-Ares, J. (2010). A Central Regulatory System Largely Controls Transcriptional Activation and Repression Responses to Phosphate Starvation in Arabidopsis. *PLoS Genetics*, *6*(9), e1001102. https://doi.org/10.1371/journal.pgen.1001102

Cadigan, K. M., Grossniklaust, U., & Gehring, W. J. (1994). Functional redundancy: The respective roles of the two sloppy paired genes in Drosophila segmentation (engaed/wngs/seMent polarity). *Proc. Nati. Acad. Sci. USA* (Vol. 91). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC44194/pdf/pnas01136-0082.pdf

Cagliari, A., Margis, R., Dos Santos Maraschin, F., Turchetto-Zolet, A. C., Loss, G., & Margis-Pinheiro, M. (2011). Biosynthesis of Triacylglycerols (TAGs) in plants and algae. *International Journal of Plant Biology*, *2*(1), 10. https://doi.org/10.4081/pb.2011.e10

Chen, G. Q., Jiang, Y., & Chen, F. (2008). Salt-induced alterations in lipid composition of diatom Nitzschia laevis (Bacillariophyceae) under heterotrophic culture condition. *Journal of Phycology*, *44*(5), 1309–1314. https://doi.org/10.1111/j.1529-8817.2008.00565.x

Cheng, C.-L., Lo, Y.-C., Lee, K.-S., Lee, D.-J., Lin, C.-Y., & Chang, J.-S. (2011). Biohydrogen production from lignocellulosic feedstock. *Bioresource Technology*, *102*(18), 8514–8523. https://doi.org/10.1016/j.biortech.2011.04.059

Cheng, D., & He, Q. (2014). Assessment of Environmental Stresses for Enhanced Microalgal Biofuel Production An Overview. *Frontiers in Energy Research*, *2*(July), 1–8. https://doi.org/10.3389/fenrg.2014.00026

Cheong, Y. H., Chang, H.-S., Gupta, R., Wang, X., Zhu, T., & Luan, S. (2002). Transcriptional Profiling Reveals Novel Interactions between Wounding, Pathogen, Abiotic Stress, and Hormonal Responses in Arabidopsis. *Plant Physiology*, 129(2):661-77 https://doi.org/10.1104/pp.002857

Chisti, Y. (2007). Biodiesel from microalgae. *Biotechnology Advances*, *25*(3), 294–306. https://doi.org/10.1016/J.BIOTECHADV.2007.02.001

Cirulis, J. T., Strasser, B. C., Scott, J. A., & Ross, G. M. (2012). Optimization of staining conditions for microalgae with three lipophilic dyes to reduce precipitation and fluorescence variability. *Cytometry Part A*, *81A*(7), 618–626.

https://doi.org/10.1002/cyto.a.22066

Coates, J. (2006). Interpretation of Infrared Spectra, A Practical Approach. In *Encyclopedia of Analytical Chemistry*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470027318.a5606

Cohen, R., Yokoi, T., Holland, J. P., Pepper, A. E., & Holland, M. J. (1987). Transcription of the Constitutively Expressed Yeast Enolase Gene ENOJ is mediated by Positive and Negative cis-Acting Regulatory Sequences. *Molecular and cell biology*, 7(8):2753-61. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC367892/pdf/molcellb00080-0111.pdf

Cooney, M., Young, G., & Nagle, N. (2009). Extraction of Bio-oils from Microalgae. *Separation & Purification Reviews*, *38*(4), 291–325. https://doi.org/10.1080/15422110903327919

Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Regulation of Transcription in Eukaryotes. Retrieve from: https://www.ncbi.nlm.nih.gov/books/NBK9904/

Costa, J. A. V., & de Morais, M. G. (2014). An Open Pond System for Microalgal Cultivation. *Biofuels from Algae*, 1–22. https://doi.org/10.1016/B978-0-444-59558-4.00001-2

Courchesne, N. M. D., Parisien, A., Wang, B., & Lan, C. Q. (2009). Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches. *Journal of Biotechnology*, *141*(1–2), 31–41. https://doi.org/10.1016/j.jbiotec.2009.02.018

d'Ippolito, G., Sardo, A., Paris, D., Vella, F., Adelfi, M., Botte, P., Fontana, A. (2015). Potential of lipid metabolism in marine diatoms for biofuel production. *Biotechnology for Biofuels*, *8*(1), 28. https://doi.org/10.1186/s13068-015-0212-4

Danisman, S., Van Dijk, A. D. J., Bimbo, A., Van Der Wal, F., Hennig, L., De Folter, S., Immink, R. G. H. (2013). Analysis of functional redundancies within the Arabidopsis TCP transcription factor family. *Journal of Experimental Botany*, *64*(18), 5673–5685. https://doi.org/10.1093/jxb/ert337

Dauvillée, D., Chochois, V., Steup, M., Haebel, S., Eckermann, N., Ritte, G., Ball, S. G. (2006). Plastidial phosphorylase is required for normal starch synthesis in *Chlamydomonas reinhardtii*. *The Plant Journal*, *48*(2), 274–285. https://doi.org/10.1111/j.1365-313X.2006.02870.x

Day, J. G., Benson, E. E., & Fleck, R. A. (1999). In vitro culture and conservation of microalgae: Applications for aquaculture, biotechnology and environmental research. *In Vitro Cellular & Developmental Biology - Plant*, *35*(2), 127–136. https://doi.org/10.1007/s11627-999-0022-0

Dean, A. P., Sigee, D. C., Estrada, B., & Pittman, J. K. (2010). Using FTIR spectroscopy for rapid determination of lipid accumulation in response to nitrogen limitation in freshwater microalgae. *Bioresource Technology*, *101*(12), 4499–4507. https://doi.org/10.1016/j.biortech.2010.01.065

Dean, A., & Pittman, J. (2015). Lipids from algae : novel applications and methods,. *Inform*, *26*, 14–17.

Delatte, T., Trevisan, M., Parker, M. L., & Zeeman, S. C. (2005). Arabidopsis mutants Atisa1 and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. *The Plant Journal*, *41*(6), 815–830. https://doi.org/10.1111/j.1365-313X.2005.02348.x

Demirbas, A., & Fatih Demirbas, M. (2011). Importance of algae oil as a source of biodiesel. *Energy Conversion and Management*, *52*(1), 163–170. https://doi.org/10.1016/J.ENCONMAN.2010.06.055

Dotti Isabella, Nardon Ermanno, P. D. and B. S. (2011). *G*uidelines for molecular analysis in archive tissues. Springer. Retrieved from https://books.google.co.uk/books?id=1IZPmlf29EEC&pg=PA128&lpg=PA128&dq=ct+greater+than+35+in+qpcr+are+considered+noise+background&source=bl&ots=QQMpe1zhGF&sig=ACfU3U3V5tPWrS3jGE6jZvRma4L3iTnfvQ&hl=en&sa=X&ved=2ahUKEwiyyZTs2oHhAhXu2OAKHQsOA2Q4ChDoATAIegQIA

Dragone, G., Fernandes, B. D., Abreu, A. P., Vicente, A. A., & Teixeira, J. A. (2011). Nutrient limitation as a strategy for increasing starch accumulation in microalgae. *Applied Energy*, *88*, 3331–3335.

https://doi.org/10.1016/j.apenergy.2011.03.012

Dragone, G., Fernandes, B., Vicente, A. A., & Teixeira, J. A. (2010). Third generation biofuels from microalgae. *Current research, technology and education topics in applied microbiology and microbial biotechnology*, 1355–1366.

Driver, T., Bajhaiya, A. K., Allwood, J. W., Goodacre, R., Pittman, J. K., & Dean, A. P. (2015). Metabolic responses of eukaryotic microalgae to environmental stress limit the ability of FT-IR spectroscopy for species identification. *Algal Research*, *11*, 148–155. https://doi.org/10.1016/J.ALGAL.2015.06.009

Driver, T., Bajhaiya, A., & Pittman, J. K. (2014). Potential of Bioenergy Production from Microalgae. *Current Sustainable/Renewable Energy Reports*, *1*(3), 94–103. https://doi.org/10.1007/s40518-014-0011-8

Driver, T., Trivedi, D. K., McIntosh, O. A., Dean, A. P., Goodacre, R., & Pittman, J. K. (2017). Two Glycerol-3-Phosphate Dehydrogenases from Chlamydomonas Have Distinct Roles in Lipid Metabolism. *Plant Physiology*, *174*(4), 2083–2097. https://doi.org/10.1104/pp.17.00491

Du, H., Feng, B.-R., Yang, S.-S., Huang, Y.-B., & Tang, Y.-X. (2012). The R2R3-MYB Transcription Factor Gene Family in Maize. *PLoS ONE*, *7*(6), e37463. https://doi.org/10.1371/journal.pone.0037463

Du, H., Liang, Z., Zhao, S., Nan, M.-G., Tran, L.-S. P., Lu, K., Li, J.-N. (2015). The Evolutionary History of R2R3-MYB Proteins Across 50 Eukaryotes: New Insights Into Subfamily Classification and Expansion. *Scientific Reports*, *5*, 11037. https://doi.org/10.1038/srep11037

Du, H., Wang, Y.-B., Xie, Y., Liang, Z., Jiang, S.-J., Zhang, S.-S., Tang, Y.-X. (2013). Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Research : An International Journal for Rapid Publication of Reports on Genes and Genomes*, *20*(5), 437–448. https://doi.org/10.1093/dnares/dst021

Du, H., Yang, S.-S., Liang, Z., Feng, B.-R., Liu, L., Huang, Y.-B., & Tang, Y.-X. (2012). Genome-wide analysis of the MYB transcription factor superfamily in soybean. *BMC Plant Biology*, *12*, 106. https://doi.org/10.1186/1471-2229-12-106

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., & Lepiniec, L. (2010). MYB transcription factors in Arabidopsis. *Trends in Plant Science*, *15*(10), 573–581. https://doi.org/10.1016/j.tplants.2010.06.005

Dunahay, T. G., Jaruis, E. E., & Roessler, P. G. (1995). Genetic transformation of the diatoms *Cyclotella Cryptzca. Journal of Phycology, 1012*, 1004–1012.

Eberlé, D., Hegarty, B., Bossard, P., Ferré, P., & Foufelle, F. (2004). SREBP transcription factors: master regulators of lipid homeostasis. *Biochimie*, *86*(11), 839–848. https://doi.org/10.1016/j.biochi.2004.09.018

El-Brolosy, M. A., & Stainier, D. Y. R. (2017). Genetic compensation: A phenomenon in search of mechanisms. *PLoS Genetics*, *13*(7), e1006780. https://doi.org/10.1371/journal.pgen.1006780

El-Dalatony, M., Salama, E.-S., Kurade, M., Hassan, S., Oh, S.-E., Kim, S., … Jeon, B.-H. (2017). Utilization of Microalgal Biofractions for Bioethanol, Higher Alcohols, and Biodiesel Production: A Review. *Energies*, *10*(12), 2110. https://doi.org/10.3390/en10122110

Fan, J., & Zheng, lvhong. (2017). Acclimation to NaCl and light stress of heterotrophic Chlamydomonas reinhardtii for lipid accumulation. *Journal of Bioscience and Bioengineering*, *124*(3), 302–308. https://doi.org/10.1016/J.JBIOSC.2017.04.009

Fattom, A., & Shilo, M. (1984). Phormidium J-1 bioflocculant: production and activity. *Archives of Microbiology*, *139*(4), 421–426. https://doi.org/10.1007/BF00408390

Feng, G., Burleigh, J. G., Braun, E. L., Mei, W., & Barbazuk, W. B. (2017). Evolution of the 3R-MYB Gene Family in Plants. *Genome Biology and Evolution*, *9*(4), 1013. https://doi.org/10.1093/gbe/evx056

Fowler, S., & Thomashow, M. F. (2002). Arabidopsis Transcriptome Profiling Indicates That Multiple Regulatory Pathways Are Activated during Cold Acclimation in Addition to the CBF Cold Response Pathway. *The Plant Cell*, *14*, 1675–1690. https://doi.org/10.1105/tpc.003483

Gargouri, M., Park, J.-J., Holguin, F. O., Kim, M.-J., Wang, H., Deshpande, R. R.,

Gang, D. R. (2015). Identification of regulatory network hubs that control lipid metabolism in Chlamydomonas reinhardtii. *Journal of Experimental Botany*, *66*(15), 4551–4566. https://doi.org/10.1093/jxb/erv217

Giordano, M., Kansiz, M., Heraud, P., Beardall, J., Wood, B., & McNaughton, D. (2001). Fourier Transform Infrared Spectroscopy as a novel tool to investigate changes in intracellular macromolecular pools in the marine microalgae *Chaetoceros Muelleri (Bacillariophyceae)*. *Journal of Phycology*, *37*(2), 271–279. https://doi.org/10.1046/j.1529-8817.2001.037002271.x

Goncalves, E. C., Koh, J., Zhu, N., Yoo, M.-J., Chen, S., Matsuo, T., Rathinasabapathi, B. (2016). Nitrogen starvation-induced accumulation of triacylglycerol in the green algae: evidence for a role for ROC40, a transcription factor involved in circadian rhythm. *The Plant Journal*, *85*(6), 743–757. https://doi.org/10.1111/tpj.13144

Goncalves, E. C., Wilkie, A. C., Kirst, M., & Rathinasabapathi, B. (2016). Metabolic regulation of triacylglycerol accumulation in the green algae: identification of potential targets for engineering to improve oil yield. *Plant Biotechnology Journal*, *14*(8), 1649–1660. https://doi.org/10.1111/pbi.12523

Gonzalez, D. H. (2016). Introduction to Transcription Factor Structure and Function. *Plant Transcription Factors*, 3–11. https://doi.org/10.1016/B978-0-12-800854-6.00001-4

Gross, M., Mascarenhas, V., & Wen, Z. (2015). Evaluating algal growth performance and water use efficiency of pilot-scale revolving algal biofilm (RAB) culture systems. *Biotechnology and Bioengineering*, *112*(10), 2040–2050. https://doi.org/10.1002/bit.25618

Grossman, A. R., & Aksoy, M. (2015). Algae in a Phosphorus-Limited landscape. *Annual Plant Reviews*, *48*, 337–374. https://doi.org/10.1002/9781118958841.ch12

Grotewold, E. (2008). Transcription factors for predictive plant metabolic engineering: are we there yet? *Current Opinion in Biotechnology*, *19*(2), 138–144. https://doi.org/10.1016/j.copbio.2008.02.002

Grotewold, E., Drummond, B. J., Bowen, B., & Peterson, T. (1994). The myb-homologous P gene controls phlobaphene pigmentation in maize floral organs by

directly activating a flavonoid biosynthetic gene subset. *Cell*, *76*(3), 543–553. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8313474

Grotewold, E., Sainz, M. B., Tagliani, L., Marcela Hernandez, J., Bowen, B., & Chandler, V. L. (2000). Identification of the residues in the Myb domain of maize C1 that specify the interaction with the bHLH cofactor R. *Plant Biology Biochemistry* (Vol. 97). Retrieved from www.pnas.orgcgidoi10.1073pnas.250379897

Guccione, A., Biondi, N., Sampietro, G., Rodolfi, L., Bassi, N., & Tredici, M. R. (2014). Chlorella for protein and biofuels: from strain selection to outdoor cultivation in a Green Wall Panel photobioreactor. *Biotechnology for Biofuels*, *7*, 84. https://doi.org/10.1186/1754-6834-7-84

Guiry, M. D. (2012). How many species of algae are there? *Journal of Phycology*, *48*(5), 1057–1063. https://doi.org/10.1111/j.1529-8817.2012.01222.x

Guo, M., Ruan, W., Li, C., Huang, F., Zeng, M., Liu, Y., Mo, X. (2015). Integrative Comparison of the Role of the PHOSPHATE RESPONSE1 Subfamily in Phosphate Signaling and Homeostasis in Rice. *Plant Physiology*, *168*(4), 1762–1776. https://doi.org/10.1104/pp.15.00736

Gutierrez, R. A., MacIntosh, G. C., & Green, P. J. (1999). Current perspectives on mRNA stability in plants: multiple levels and mechanism of control. *Trends in Plant Science*, 4(11):429-438. Retrieved from http://virtualplant.bio.puc.cl/Lab/doc/Gutierrez et al. (1999) TiPS.pdf

Han, S.-F., Jin, W.-B., Tu, R.-J., & Wu, W.-M. (2015). Biofuel production from microalgae as feedstock: current status and potential. *Critical Reviews in Biotechnology*, *35*(2), 255–268. https://doi.org/10.3109/07388551.2013.835301

Hannon, M., Gimpel, J., Tran, M., Rasala, B., & Mayfield, S. (2010). Biofuels from algae: challenges and potential. *Biofuels*, *1*(5), 763–784. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21833344

Harbers, M. (2014). Wheat germ systems for cell-free protein expression. *FEBS Letters*, *588*(17), 2762–2773. https://doi.org/10.1016/J.FEBSLET.2014.05.061

Harris, M. R. (1991). Non-Radoactive Techniques for the labelling of Nucleid Acids. *Biotech. Adv* (Vol. 9). Retrieved from https://ac.els-cdn.com/073497509190003E/1-

s2.0-073497509190003E-main.pdf?_tid=8c386b7d-c16c-4465-b312-6181b89086c8&acdnat=1538937799_0e994fef99cc31d82c0a84aa2eabba6b

Hattori, H., Imai, H., Kirai, N., Furuhama, K., Sato, O., Konishi, K., & Nakagawa, Y. (2007). Identification of a responsible promoter region and a key transcription factor, CCAAT/enhancer-binding protein epsilon, for up-regulation of PHGPx in HL60 cells stimulated with TNF alpha. *The Biochemical Journal*, *408*(2), 277–286. https://doi.org/10.1042/BJ20070245

Heine, G. F., Hernandez, J. M., & Grotewold, E. (2004). Two cysteines in plant R2R3 MYB domains participate in REDOX-dependent DNA binding. *The Journal of Biological Chemistry*, *279*(36), 37878–37885. https://doi.org/10.1074/jbc.M405166200

Hellman, L. M., & Fried, M. G. (2007). Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. *Nature Protocols*, *2*(8), 1849–1861. https://doi.org/10.1038/nprot.2007.249

Ho, D. P., Ngo, H. H., & Guo, W. (2014). A mini review on renewable sources for biofuel. *Bioresource Technology*, *169*, 742–749. https://doi.org/10.1016/J.BIORTECH.2014.07.022

Ho, S.-H., Chen, C.-Y., & Chang, J.-S. (2012). Effect of light intensity and nitrogen starvation on CO2 fixation and lipid/carbohydrate production of an indigenous microalga Scenedesmus obliquus CNW-N. *Bioresource Technology*, *113*, 244–252. https://doi.org/10.1016/J.BIORTECH.2011.11.133

Holden, N. S., & Tacon, C. E. (2011). Principles and problems of the electrophoretic mobility shift assay. *Journal of Pharmacological and Toxicological Methods*, *63*(1), 7–14. https://doi.org/10.1016/J.VASCN.2010.03.002

Holl, J., Vannozzi, A., Czemmel, S., D'Onofrio, C., Walker, A. R., Rausch, T., Bogs, J. (2013). The R2R3-MYB Transcription Factors MYB14 and MYB15 Regulate Stilbene Biosynthesis in Vitis vinifera. *The Plant Cell*, *25*(10), 4135–4149. https://doi.org/10.1105/tpc.113.117127

Holzinger, A., & Karsten, U. (2013). Desiccation stress and tolerance in green algae: consequences for ultrastructure, physiological and molecular mechanisms. *Frontiers in Plant Science*, *4*, 327. https://doi.org/10.3389/fpls.2013.00327

Hou, X.-J., Li, S.-B., Liu, S.-R., Hu, C.-G., & Zhang, J.-Z. (2014). Genome-wide classification and evolutionary and expression analyses of citrus MYB transcription factor families in sweet orange. *PloS One*, *9*(11), e112375. https://doi.org/10.1371/journal.pone.0112375

Hounslow, E., Noirel, J., Gilmour, D. J., & Wright, P. C. (2017). Lipid quantification techniques for screening oleaginous species of microalgae for biofuel production. *European Journal of Lipid Science and Technology*, *119*(2), 1–24. https://doi.org/10.1002/ejlt.201500469

Howe, K. M., & Watson, R. J. (1991). Nucleotide preferences in sequence-specific recognition of DNA by c-myb protein. *Nucleic Acids Research*, *19*(14), 3913–3919. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1861984

Hu, J., Wang, D., Li, J., Jing, G., Ning, K., & Xu, J. (2014). Genome-wide identification of transcription factors and transcription-factor binding sites in oleaginous microalgae Nannochloropsis. *Scientific Reports*. https://doi.org/10.1038/srep05454

Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., & Darzins, A. (2008). Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *The Plant Journal*, *54*(4), 621–639. https://doi.org/10.1111/j.1365-313X.2008.03492.x

IEA (2018). Renewables 2018: Analysis and Forecasts to 2023, IEA, Paris,https://doi.org/10.1787/re_mar-2018-en.

Ji, C.-F., Yu, X.-J., Chen, Z.-A., Xue, S., Legrand, J., & Zhang, W. (2011). Effects of nutrient deprivation on biochemical compositions and photo-hydrogen production of Tetraselmis subcordiformis. *International Journal of Hydrogen Energy*, *36*(10), 5817–5821. https://doi.org/10.1016/j.ijhydene.2010.12.138

Jiang, D., Jia, Y., & Jarrett, H. W. (2011). Transcription factor proteomics: identification by a novel gel mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography-electrospray-mass spectrometry analysis. *Journal of Chromatography*, *1218*(39), 7003–7015. https://doi.org/10.1016/j.chroma.2011.08.023

Juergens, M. T., Disbrow, B., & Shachar-Hill, Y. (2016). The Relationship of

Triacylglycerol and Starch Accumulation to Carbon and Energy Flows during Nutrient Deprivation in Chlamydomonas reinhardtii. *Plant Physiology*, 171(4):2445-57. https://doi.org/10.1104/pp.16.00761

Kamalanathan, M., Pierangelini, M., Shearman, L. A., Gleadow, R., & Beardall, J. (2016). Impacts of nitrogen and phosphorus starvation on the physiology of Chlamydomonas reinhardtii. *Journal of Applied Phycology*, *28*(3), 1509–1520. https://doi.org/10.1007/s10811-015-0726-y

Kamei, A., Seki, M., Umezawa, T., Ishida, J., Satou, M., Akiyama, K., Shinozaki, K. (2005). Analysis of gene expression profiles in Arabidopsis salt overly sensitive mutants sos2-1 and sos3-1. *Plant, Cell and Environment*, *28*(10), 1267–1275. https://doi.org/10.1111/j.1365-3040.2005.01363.x

Kamisaka, Y., Tomita, N., Kimura, K., Kainou, K., & Uemura, H. (2007). DGA1 (diacylglycerol acyltransferase gene) overexpression and leucine biosynthesis significantly increase lipid accumulation in the Deltasnf2 disruptant of Saccharomyces cerevisiae. *The Biochemical Journal*, *408*(1), 61–68. https://doi.org/10.1042/BJ20070449

Kaplan, D., Christiaen, D., Malis, S. (1987). *C*helating Properties of Extracellular Polysaccharides from Chlorella spp. *Applied and environmental Microbiology*, 53(12): 2953–2956. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC204228/pdf/aem00129-0265.pdf

Kapoore, R., Butler, T., Pandhal, J., & Vaidyanathan, S. (2018). Microwave-Assisted Extraction for Microalgae: From Biofuels to Biorefinery. *Biology*, *7*(1), 18. https://doi.org/10.3390/biology7010018

Kass, J., Artero, R., & Baylies, M. K. (2000). Non-radioactive electrophoretic mobility shift assay using digoxigenin-ddUTP labeled probes. *Dros. Inf. Serv.*, 83: 185-188. Retrieved from http://www.ou.edu/journals/dis/DIS83/Technique83/6tecKass/Kass.pdf

Khan, M. I., Shin, J. H., & Kim, J. D. (2018). The promising future of microalgae: current status, challenges, and optimization of a sustainable and renewable industry for biofuels, feed, and other products. *Microbial Cell Factories*, *17*(1), 36. https://doi.org/10.1186/s12934-018-0879-x

Klempnauer, K.-H., Gonda, T. J., & Michael Bishop, J. (1982). Nucleotide Sequence of tF Gene v-myb and Its Cellul; The Architecture of a Tran. *Cell*, 31(2 Pt 1):453-63.. Retrieved from https://www.cell.com/cell/pdf/0092-8674(82)90138-6.pdf?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2F0092867482901386%3Fshowall%3Dtrue

Koo, K. M., Jung, S., Lee, B. S., Kim, J.-B., Jo, Y. D., Choi, H.-I., Ahn, J.-W. (2017). The Mechanism of Starch Over-Accumulation in Chlamydomonas reinhardtii High-Starch Mutants Identified by Comparative Transcriptome Analysis. *Frontiers in Microbiology*, *8*, 858. https://doi.org/10.3389/fmicb.2017.00858

Kruskopf, M. M., & Plessis, S. Du. (2004). Induction of both acid and alkaline phosphatase activity in two green-algae (chlorophyceae) in low N and P concentrations. *Hydrobiologia*, 513(1-3), 59-70. Retrieved from https://link.springer.com/content/pdf/10.1023%2FB%3Ahydr.0000018166.15764.b0.pdf

Kumar, M., & Arvind, K. (2017). A Short Review on Biobutanol, a Second Generation Biofuel Production from Lignocellulosic Biomass. *Journal of Clean Energy Technologies*, 5(1), 27-30. https://doi.org/10.18178/jocet.2017.5.1.338

Lambers, H., & Plaxton, W. C. (2015). Phosphorus: back to the roots. In W. C. Plaxton, & H. Lambers (Eds.), *Annual plant reviews, Volume 48: Phosphorus metabolism in plants* (pp. 3– 22). Hoboken, New Jersey: Wiley-Blackwell.

Lee, M. M., & Schiefelbein, J. (1999). WEREWOLF, a MYB-Related Protein in Arabidopsis, Is a Position-Dependent Regulator of Epidermal Cell Patterning. *Cell*, *99*(5), 473–483. https://doi.org/10.1016/S0092-8674(00)81536-6

Lee, R. a., & Lavoie, J.-M. (2013). From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity. *Animal Frontiers*, *3*(2), 6–11. https://doi.org/10.2527/af.2013-0010

Lee, S. Y., Cho, J. M., Chang, Y. K., & Oh, Y.-K. (2017). Cell disruption and lipid extraction for microalgal biorefineries: A review. *Bioresource Technology*, *244*, 1317–1328. https://doi.org/10.1016/j.biortech.2017.06.038

Li, C., Yu, Y., Zhang, D., Liu, J., Ren, N., & Feng, Y. (2014). Combined effects of carbon, phosphorus and nitrogen on lipid accumulation of Chlorella vulgaris in

mixotrophic culture. *Journal of Chemical Technology and Biotechnology*, https://doi.org/10.1002/jctb.4623

Li, J., Zhang, C., Huang, P., Kuru, E., Forster-Benson, E. T. C., Li, T., & Church, G. M. (2017). Dissecting limiting factors of the Protein synthesis Using Recombinant Elements (PURE) system. *Translation (Austin, Tex.)*, *5*(1), e1327006. https://doi.org/10.1080/21690731.2017.1327006

Li, Y., Horsman, M., Wu, N., Lan, C. Q., & Dubois-Calero, N. (2008). Biofuels from microalgae. *Biotechnology Progress*, *24*(4), 815–820. https://doi.org/10.1021/bp070371k

Li, Z., Peng, R., Tian, Y., Han, H., Xu, J., & Yao, Q. (2016). Genome-Wide Identification and Analysis of the MYB Transcription Factor Superfamily in Solanum lycopersicum. *Plant Cell Physiology*, 57(8):1657-77. https://doi.org/10.1093/pcp/pcw091

Lin, Y., & Tanaka, S. (2006). Ethanol fermentation from biomass resources: current state and prospects. *Applied Microbiology and Biotechnology*, *69*(6), 627–642. https://doi.org/10.1007/s00253-005-0229-x

Ling, X., Weathers J., P., Xue-Rong, X., & Chun-Zhao, L. (2009). Microalgal bioreactors: Challenges and opportunities. *Engineering in Life Science*, 9(3), 178-189. https://doi.org/10.1002/elsc.200800111

Lipsick, J. S. (1996). One billion years of Myb. *Oncogene*, *13*(2), 223–235. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8710361

Liu, B., & Benning, C. (2013). Lipid metabolism in microalgae distinguishes itself. *Current Opinion in Biotechnology*, *24*(2), 300–309. https://doi.org/10.1016/j.copbio.2012.08.008

Liu, J., Osbourn, A., & Ma, P. (2015). MYB Transcription Factors as Regulators of Phenylpropanoid Metabolism in Plants. *Mol. Plant*, *8*, 689–708. https://doi.org/10.1016/j.molp.2015.03.012

Liu, J., Song, Y., & Qiu, W. (2017). Oleaginous microalgae Nannochloropsis as a new model for biofuel production: Review & analysis. *Renewable and Sustainable Energy Reviews*, *72*, 154–162.

https://doi.org/10.1016/J.RSER.2016.12.120

Liu, T., Wang, J., Hu, Q., Cheng, P., Ji, B., Liu, J., Wang, H. (2013). Attached cultivation technology of microalgae for efficient biomass feedstock production. *Bioresource Technology*, *127*, 216–222. https://doi.org/10.1016/J.BIORTECH.2012.09.100

Liu, W., Yuan, J. S., & Stewart Jr, C. N. (2013). Advanced genetic tools for plant biotechnology. *Nature Reviews Genetics*, *14*(11), 781–793. https://doi.org/10.1038/nrg3583

Liu, X., Yu, W., Zhang, X., Wang, G., Cao, F., & Cheng, H. (2017). Identification and expression analysis under abiotic stress of the R2R3-MYB genes in Ginkgo biloba L. *Physiology and Molecular Biology of Plants : An International Journal of Functional Plant Biology*, *23*(3), 503–516. https://doi.org/10.1007/s12298-017-0436-9

López, A., De Lomana, G., Schäuble, S., Valenzuela, J., Carter, W., Bilgin, D. D., Baliga, N. S. (2015). Transcriptional program for nitrogen starvation-induced lipid accumulation in Chlamydomonas reinhardtii. *Biotechnology for Biofuels*, *8*, 207. https://doi.org/10.1186/s13068-015-0391-z

Ma, L., Sun, N., Liu, X., Jiao, Y., Zhao, H., & Deng, X. W. (2005). Organ-specific expression of Arabidopsis genome during development. *Plant Physiology*, *138*(1), 80–91. https://doi.org/10.1104/pp.104.054783

Maddelein, M. L., Libessart, N., Bellanger, F., Delrue, B., D'Hulst, C., Van den Koornhuyse, N., Ball, S. (1994). Toward an understanding of the biogenesis of the starch granule. Determination of granule-bound and soluble starch synthase functions in amylopectin synthesis. *Journal of Biological Chemistry*, 269(40):25150-7.

Madin, K., Sawasaki, T., Ogasawara, T., & Endo, Y. (2000). A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(2), 559–564. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10639118

Maity, J. P., Bundschuh, J., Chen, C.-Y., & Bhattacharya, P. (2014). Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and

wastewater treatment: Present and future perspectives – A mini review. *Energy*, *78*, 104–113. https://doi.org/10.1016/J.ENERGY.2014.04.003

Markou, G., Angelidaki, I., & Georgakakis, D. (2012). Microalgal carbohydrates: an overview of the factors influencing carbohydrates production, and of main bioconversion technologies for production of biofuels. *Applied Microbiology and Biotechnology*, *96*(3), 631–645. https://doi.org/10.1007/s00253-012-4398-0

Markou, G., & Georgakakis, D. (2011). Cultivation of filamentous cyanobacteria (blue-green algae) in agro-industrial wastes and wastewaters: A review. *Applied Energy*, *88*(10), 3389–3401. https://doi.org/10.1016/j.apenergy.2010.12.042

Marrone, B. L., Lacey, R. E., Anderson, D. B., Bonner, J., Coons, J., Dale, T., Olivares, J. A. (2018). Review of the harvesting and extraction program within the National Alliance for Advanced Biofuels and Bioproducts. *Algal Research*, *33*, 470–485. https://doi.org/10.1016/J.ALGAL.2017.07.015

Martin, R. C., Vining, K., & Dombrowski, J. E. (2018). Genome-wide (ChIP-seq) identification of target genes regulated by BdbZIP10 during paraquat-induced oxidative stress. *BMC Plant Biology*, *18*(1), 1–12. https://doi.org/10.1186/s12870-018-1275-8

Mata, T. M., Martins, A. a., & Caetano, N. S. (2010). Microalgae for biodiesel production and other applications: A review. *Renewable and Sustainable Energy Reviews*, *14*(1), 217–232. https://doi.org/10.1016/j.rser.2009.07.020

McCarthy, R. L., Zhong, R., Fowler, S., Lyskowski, D., Piyasena, H., Carleton, K., Ye, Z.-H. (2010). The Poplar MYB Transcription Factors, PtrMYB3 and PtrMYB20, are Involved in the Regulation of Secondary Wall Biosynthesis. *Plant and Cell Physiology*, *51*(6), 1084–1090. https://doi.org/10.1093/pcp/pcq064

McNichol, J., MacDougall, K. M., Melanson, J. E., & McGinn, P. J. (2012). Suitability of Soxhlet Extraction to Quantify Microalgal Fatty Acids as Determined by Comparison with In Situ Transesterification. *Lipids*, *47*(2), 195–207. https://doi.org/10.1007/s11745-011-3624-3

Mehdi, S., Sohi, H., & Eghdami, A. (2014). Biodiesel production using marine microalgae Dunalialla salina. *J. Bio. & Env. Sci*, *2014*(2), 177–182. Retrieved from http://www.innspub.net

Milledge, J. J., & Heaven, S. (2013). A review of the harvesting of micro-algae for biofuel production. *Reviews in Environmental Science and Bio/Technology*, *12*(2), 165–178. https://doi.org/10.1007/s11157-012-9301-z

Miller, R., Wu, G., Deshpande, R. R., Vieler, A., Gärtner, K., Li, X., Benning, C. (2010). Changes in Transcript Abundance in Chlamydomonas reinhardtii following Nitrogen Deprivation Predict Diversion of Metabolism. https://doi.org/10.1104/pp.110.165159

Milo, R. & Phillips, R. (2015). Cell Biology by the numbers. Garland Science, p-171.

Mishra, A., & Jha, B. (2009). Isolation and characterization of extracellular polymeric substances from micro-algae Dunaliellasalina under salt stress. *Bioresource Technology*, *100*(13), 3382–3386. https://doi.org/10.1016/J.BIORTECH.2009.02.006

Misson, J., Raghothama, K. G., Jain, A., Jouhet, J., Block, M. A., Bligny, R., Thibaud, M.-C. (2005). A genome-wide transcriptional analysis using Arabidopsis thaliana Affymetrix gene chips determined plant responses to phosphate deprivation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(33), 11934–11939. https://doi.org/10.1073/pnas.0505266102

Mohamed, B. B., Aftab, B., Sarwar, M. B., Rashid, B., Ahmad, Z., Hassan, S., & Husnain, T. (2017). Identification and Characterization of the Diverse Stress-Responsive *R2R3*- RMYB Transcription Factor. *Hibiscus sabdariffa* L. *International Journal of Genomics*, *2017*, 1–12. https://doi.org/10.1155/2017/2763259

Mohr, A., & Raman, S. (2013). Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels. *Energy Policy*, *63*(100), 114–122. https://doi.org/10.1016/j.enpol.2013.08.033

Mohtasham, J. (2015). Review Article-Renewable Energies. *Energy Procedia*, *74*, 1289–1297. https://doi.org/10.1016/j.egypro.2015.07.774

Molnar, A., Bassett, A., Thuenemann, E., Schwach, F., Karkare, S., Ossowski, S., Baulcombe, D. (2009). Highly specific gene silencing by artificial microRNAs in the unicellular alga Chlamydomonas reinhardtii. *The Plant Journal*, *58*(1), 165–174. https://doi.org/10.1111/j.1365-313X.2008.03767.x

Moreno-Risueno, M. A., Martínez, M., Vicente-Carbajosa, J., & Carbonero, P. (2007). The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular Genetics and Genomics : MGG*, *277*(4), 379–390. https://doi.org/10.1007/s00438-006-0186-9

Moseley, J. L., Chang, C.-W., & Grossman, A. R. (2006). Genome-based approaches to understanding phosphorus deprivation responses and PSR1 control in Chlamydomonas reinhardtii. *Eukaryotic Cell*, *5*(1), 26–44. https://doi.org/10.1128/EC.5.1.26-44.2006

Moyano, E., Martínez-Garcia, J. F., & Martin2, C. (1996). Apparent Redundancy in myb Gene Function Provides Gearing for the Control of Flavonoid Biosynthesis in Antirrhinum Flowers. *The Plant Cell*, 8(9):1519-32. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC161295/pdf/081519.pdf

Mühlroth, A., Winge, P., Assimi, A. El, Jouhet, J., Maréchal, E., Hohmann-Marriott, M. F., Bones, A. M. (2017). Mechanisms of Phosphorus Acquisition and Lipid Class Remodeling under P Limitation in a Marine Microalga . *Plant Physiology*, 175(4):1543-1559. https://doi.org/10.1104/pp.17.00621

Myrset, A. H., Bostad, A., Jamin, N., Lirsac, P.-N., Toma, F., Gabrielsen, O. S., & Boye, E. (1993). *D*NA and redox state induced conformational changes in the DNA-binding domain of the Myb oncoprotein 4Corresponding author (in Norway). *The EMBO Journal*, 12(12):4625-33. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC413899/pdf/emboj00084-0167.pdf

Naik, S. N., Goud, V. V., Rout, P. K., & Dalai, A. K. (2010). Production of first and second generation biofuels: A comprehensive review. *Renewable and Sustainable Energy Reviews*, *14*(2), 578–597. https://doi.org/10.1016/J.RSER.2009.10.003

Ngan, C. Y., Wong, C.-H., Choi, C., Yoshinaga, Y., Louie, K., Jia, J., Wei, C.-L. (2015). Lineage-specific chromatin signatures reveal a regulator of lipid metabolism in microalgae. *Nature Plants*, *1*(8), 15107. https://doi.org/10.1038/nplants.2015.107

Nilsson, L., Müller, R., & Nielsen, T. H. (2010). Dissecting the plant transcriptome and the regulatory responses to phosphate deprivation. *Physiologia Plantarum*, *139*(2), 129–143. https://doi.org/10.1111/j.1399-3054.2010.01356.x

Nomura, N., Takahashi, M., Matsui, M., Ishii, S., Date, T., Sasamoto, S., & Ishizaki,

R. (1988). Isolation of human cDNA clones of myb-related genes, A-myb and b-myb. *Nucleid Acids Research*, 16(23):11075-89. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC338997/pdf/nar00165-0132.pdf

Norsker, N.-H., Barbosa, M. J., Vermuë, M. H., & Wijffels, R. H. (2011). Microalgal production — A close look at the economics. *Biotechnology Advances*, *29*(1), 24–27. https://doi.org/10.1016/j.biotechadv.2010.08.005

Odjadjare, E. C., Mutanda, T., & Olaniran, A. O. (2017). Potential biotechnological application of microalgae: a critical review. *Critical Reviews in Biotechnology*, *37*(1), 37–52. https://doi.org/10.3109/07388551.2015.1108956

Ogata, K., Hojo, H., Aimoto, S., Nakai, T., Nakamura, H., Sarai, A., Nishimura, Y. (1992). Solution structure of a DNA-binding unit of Myb: a helix-turn-helix-related motif with conserved tryptophans forming a hydrophobic core. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(14), 6428–6432. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1631139

Ogata, K., Kanei-Ishii, C., Sasaki, M., Hatanaka, H., Nagadoi, A., Enari, M., Sarai, A. (1996). The cavity in the hydrophobic core of Myb DNA-binding domain is reserved for DNA recognition and trans-activation. *Nature Structural Biology*, *3*(2), 178–187. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8564545

Oropeza-Aburto, A., Cruz-Ramírez, A., Acevedo-Hernández, G. J., Pérez-Torres, C.-A., Caballero-Pérez, J., & Herrera-Estrella, L. (2012). Functional analysis of the Arabidopsis PLDZ2 promoter reveals an evolutionarily conserved low-Pi-responsive transcriptional enhancer element. *Journal of Experimental Botany*, *63*(5), 2189–2202. https://doi.org/10.1093/jxb/err446

Owen, N. A., Inderwildi, O. R., & King, D. A. (2010). The status of conventional world oil reserves-Hype or cause for concern? *Energy Policy*, *38*, 4743–4749. https://doi.org/10.1016/j.enpol.2010.02.026

Owusu, P. A., & Asumadu-Sarkodie, S. (2016). A review of renewable energy sources, sustainability issues and climate change mitigation. *Cogent Engineering*, *3*(1). https://doi.org/10.1080/23311916.2016.1167990

Özçimen, D., & Inan, B. (2015). An Overview of Bioethanol Production From Algae. *Biofuels - Status and Perspective*. InTech. https://doi.org/10.5772/59305

Pandit, P. R., Fulekar, M. H., & Karuna, M. S. L. (2017). Effect of salinity stress on growth, lipid productivity, fatty acid composition, and biodiesel properties in Acutodesmus obliquus and Chlorella vulgaris. *Environmental Science and Pollution Research*, *24*(15), 13437–13451. https://doi.org/10.1007/s11356-017-8875-y

Pant, B. D., Burgos, A., Pant, P., Cuadros-Inostroza, A., Willmitzer, L., & Scheible, W.-R. (2015). The transcription factor PHR1 regulates lipid remodeling and triacylglycerol accumulation in Arabidopsis thaliana during phosphorus starvation. *Journal of Experimental Botany*, *66*(7), 1907–1918. https://doi.org/10.1093/jxb/eru535

Park, J.-J., Wang, H., Gargouri, M., Deshpande, R. R., Skepper, J. N., Holguin, F. O., Gang, D. R. (2015). The response of Chlamydomonas reinhardtii to nitrogen deprivation: a systems biology analysis. *The Plant Journal*, *81*(4), 611–624. https://doi.org/10.1111/tpj.12747

Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A., & Saedler, H. (1987). The regulatory c1 locus of Zea mays encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO Journal*, *6*(12), 3553–3558. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3428265

Posadas, E., Alcántara, C., García-Encina, P. A., Gouveia, L., Guieysse, B., Norvill, Z., Muñoz, R. (2017). Microalgae cultivation in wastewater. *Microalgae-Based Biofuels and Bioproducts*, 67–91. https://doi.org/10.1016/B978-0-08-101023-5.00003-0

Quattrocchio, F., Wing, J. F., Leppen, H., Mol, J., & Koes, R. E. (1993). Regulatory Genes Controlling Anthocyanin Pigmentation Are Functionally Conserved among Plant Species and Have Distinct Sets of Target Genes. *The Plant Cell*, *5*(11), 1497–1512. https://doi.org/10.1105/tpc.5.11.1497

Radakovits, R., Jinkerson, R. E., Darzins, A., & Posewitz, M. C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryotic Cell*, *9*(4), 486–501. https://doi.org/10.1128/EC.00364-09

Raven, J. A., & Beardall, J. (2003). Carbohydrate Metabolism and Respiration in Algae. *Photosynthesis in Algae*, 205–224. https://doi.org/10.1007/978-94-007-1038-

2_10

Rawls, A., Valdez, M. R., Zhang, W., Richardson, J., Klein, W. H., & Olson, E. N. (1998). Overlapping functions of the myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice. *Development*, 125(13):2349-58. Retrieved from https://pdfs.semanticscholar.org/5166/1111c3306d2e02ebbb9691d4c53ede40be20.pdf

Rehman, Z. U., & Anal, A. K. (2019). Enhanced lipid and starch productivity of microalga (Chlorococcum sp. TISTR 8583) with nitrogen limitation following effective pretreatments for biofuel production. *Biotechnology Reports*, *21*, e00298. https://doi.org/10.1016/j.btre.2018.e00298

Reik, A., Zhou, Y., Collingwood, T. N., Warfe, L., Bartsevich, V., Kong, Y., Gregory, P. D. (2007). Enhanced protein production by engineered zinc finger proteins. *Biotechnology and Bioengineering*, *97*(5), 1180–1189. https://doi.org/10.1002/bit.21304

REN21. (2017). *Renewable 2017 Global status Report*. Retrieved from www.ren21.net/gsr

Riaño-Pachón, D. M., Corrêa, L. G. G., Trejos-Espinosa, R., & Mueller-Roeber, B. (2008). Green transcription factors: a chlamydomonas overview. *Genetics*, *179*(1), 31–39. https://doi.org/10.1534/genetics.107.086090

Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C.-Z., Keddie, J., Yu, G.-L. (2000). Arabidopsis Transcription Factors: Genome-Wide Comparative Analysis Among Eukaryotes. *Science*, 290(5499):2105-10. Retrieved from http://science.sciencemag.org/

Riegman, R., Stolte, W., Noordeloos, A. A. M., & Slezak, D. (2000). Nutrient uptake and alkaline phosphatase (ec 3:1:3:1) activity of emiliania huxleyi (prymnesiophyceae) during growth under n and p limitation in continuous cultures. *Journal of Phycology*, *36*(1), 87–96. https://doi.org/10.1046/j.1529-8817.2000.99023.x

Ritchie Hannah and Roser Max. (2019). Energy Production & Changing Energy Sources. *European Review of Economic History*, *11*(2), 219–253. https://doi.org/10.1017/S1361491607001967

Robak, K., & Balcerek, M. (2018). Review of Second Generation Bioethanol Production from Residual Biomass. *Food Technology and Biotechnology*, *56*(2), 174–187. https://doi.org/10.17113/ftb.56.02.18.5428

Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., Biondi, N., Bonini, G., & Tredici, M. R. (2009a). Microalgae for oil: strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnology and Bioengineering*, *102*(1), 100–112. https://doi.org/10.1002/bit.22033

Rodolfi, L., Chini Zittelli, G., Bassi, N., Padovani, G., Biondi, N., Bonini, G., & Tredici, M. R. (2009b). Microalgae for oil: Strain selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor. *Biotechnology and Bioengineering*, *102*(1), 100–112. https://doi.org/10.1002/bit.22033

Roessler, P. G. (1990). Purification and Characterization of Acetyl-CoA Carboxylase from the Diatom Cyclotella cryptica1. *Plant Physiology*, 92(1):73-8.

Rosano, G. L., & Ceccarelli, E. A. (2014). Recombinant protein expression in Escherichia coli: advances and challenges. *Frontiers in Microbiology*, *5*, 172. https://doi.org/10.3389/fmicb.2014.00172

Rosinski, J. A., & Atchley, W. R. (1998). Molecular evolution of the Myb family of transcription factors: evidence for polyphyletic origin. *Journal of Molecular Evolution*, *46*(1), 74–83. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9419227

Ruan, W., Guo, M., Cai, L., Hu, H., Li, C., Liu, Y., Mo, X. (2015). Genetic manipulation of a high-affinity PHR1 target cis-element to improve phosphorous uptake in Oryza sativa L. *Plant Molecular Biology*, *87*(4–5), 429–440. https://doi.org/10.1007/s11103-015-0289-y

Rubio, V., Linhares, F., Solano, R., Martín, A. C., Iglesias, J., Leyva, A., & Paz-ares, J. (2001). A conserved MYB transcription factor involved in phosphate starvation signaling both in vascular plants and in unicellular algae. *Genes & Development*, *15*, 2122–2133. https://doi.org/10.1101/gad.204401.availability

Rudnik, R., Bulcha, J. T., Reifschneider, E., Ellersiek, U., & Baier, M. (2017). Specificity versus redundancy in the RAP2.4 transcription factor family of Arabidopsis thaliana: transcriptional regulation of genes for chloroplast peroxidases.

BMC Plant Biology, *17*(1), 144. https://doi.org/10.1186/s12870-017-1092-5

Rushton, P. J., Reinstädler, A., Lipka, V., Lippok, B., & Somssich, I. E. (2002). Synthetic plant promoters containing defined regulatory elements provide novel insights into pathogen- and wound-induced signaling. *The Plant Cell*, *14*(4), 749–762. https://doi.org/10.1105/TPC.010412

Safi, A., Medici, A., Szponarski, W., Ruffel, S., Lacombe, B., & Krouk, G. (2017). The world according to GARP transcription factors. *Current Opinion in Plant Biology*, *39*, 159–167. https://doi.org/10.1016/J.PBI.2017.07.006

Sahdev, S., Sunil, A. E., Khattar, K., Kulvinder, A. E., & Saini, S. (2008). Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, 307(1-2):249-64. https://doi.org/10.1007/s11010-007-9603-6

Sakura, H., Kanei-Ishii, C., Nagase, T., Nakagoshi, H., Gonda, T. J., & Ishii, S. (1989). Delineation of three functional domains of the transcriptional activator encoded by the c-myb protooncogene. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(15), 5758–5762. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2668947

Salehizadeh, H., Vossoughi, M., & Alemzadeh, I. (2000). Some investigations on bioflocculant producing bacteria. *Biochemical Engineering Journal*, 5(1), 39-44. Retrieved from https://ac.els-cdn.com/S1369703X99000662/1-s2.0-S1369703X99000662-main.pdf?_tid=092f3888-47f5-4425-8bb9-ad00a2955085&acdnat=1543075815_5412875f2f05722415913eb4f648ed49

Sanjaya, Durrett, T. P., Weise, S. E., & Benning, C. (2011). Increasing the energy density of vegetative tissues by diverting carbon from starch to oil biosynthesis in transgenic Arabidopsis. *Plant Biotechnology Journal*, *9*(8), 874–883. https://doi.org/10.1111/j.1467-7652.2011.00599.x

Schaffer, R., Landgraf, J., Accerbi, M., Simon, V., Larson, M., & Wisman, E. (2001). Microarray Analysis of Diurnal and Circadian-Regulated Genes in Arabidopsis. *The Plant Cell* (Vol. 13). Retrieved from www.plantcell.org

Schmollinger, S., Mühlhaus, T., Boyle, N. R., Blaby, I. K., Casero, D., Mettler, T., Merchant, S. S. (2014). Nitrogen-Sparing Mechanisms in Chlamydomonas Affect the

Transcriptome, the Proteome, and Photosynthetic Metabolism. *The Plant Cell*, *26*(4), 1410–1435. https://doi.org/10.1105/tpc.113.122523

Schroda, M., Blocker, D., & Beck, C. F. (2000). The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. *The Plant Journal*, *21*(2), 121–131. https://doi.org/10.1046/j.1365-313x.2000.00652.x

Scott, C. C., Vossio, S., Rougemont, J., & Gruenberg, J. (2018). TFAP2 transcription factors are regulators of lipid droplet biogenesis. *ELife*, *7*, 1–24. https://doi.org/10.7554/elife.36330

Scranton, M. A., Ostrand, J. T., Fields, F. J., & Mayfield, S. P. (2015). Chlamydomonas as a model for biofuels and bio-products production. *The Plant Journal : For Cell and Molecular Biology*, *82*(3), 523–531. https://doi.org/10.1111/tpj.12780

Serpa, V., Vernal, J., Lamattina, L., Grotewold, E., Cassia, R., & Terenzi, H. (2007). Inhibition of AtMYB2 DNA-binding by nitric oxide involves cysteine S-nitrosylation. *Biochemical and Biophysical Research Communications*, *361*(4), 1048–1053. https://doi.org/10.1016/J.BBRC.2007.07.133

Shamala, T. R., Ravishankar, G. A., Dayananda, C., Rao, A. R., & Sarada, R. (2006). Effect of salinity on growth of green alga Botryococcus braunii and its constituents. *Bioresource Technology*, *98*(3), 560–564. https://doi.org/10.1016/j.biortech.2006.02.007

Sharma, K. K., Schuhmann, H., & Schenk, P. M. (2012). High Lipid Induction in Microalgae for Biodiesel Production. *Energies*, *5*(12), 1532–1553. https://doi.org/10.3390/en5051532

Shimogawara, K., Wykoff, D. D., Usuda, H., & Grossman, A. R. (1999). Chlamydomonas reinhardtii Mutants Abnormal in Their Responses to Phosphorus Deprivation. *Plant Physiology*, 120(3):685-94. Retrieved from www.plantphysiol.org

Shrestha, A., Khan, A., & Dey, N. (2018). cis–trans Engineering: Advances and Perspectives on Customized Transcriptional Regulation in Plants. *Molecular Plant*, 11(7), 886–898. https://doi.org/10.1016/J.MOLP.2018.05.008

Siaut, M., Cuiné, S., Cagnon, C., Fessler, B., Nguyen, M., Carrier, P., Peltier, G.

(2011). Oil accumulation in the model green alga Chlamydomonas reinhardtii: characterization, variability between common laboratory strains and relationship with starch reserves. *BMC Biotechnology*, 11(1), 7. https://doi.org/10.1186/1472-6750-11-7

Sims, R. E. H., Mabee, W., Saddler, J. N., & Taylor, M. (2010). An overview of second generation biofuel technologies. *Bioresource Technology*, *101*(6), 1570–1580. https://doi.org/10.1016/J.BIORTECH.2009.11.046

Singh, M., & Das, K. C. (2014). Low Cost Nutrients for Algae Cultivation. In *Algal Biorefineries* (pp. 69–82). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-7494-0_3

Smith, M. F., & Delbary-Gossart, S. (2001). Electrophoretic mobility shift assay (EMSA). *Methods in Molecular Medicine*, 50:249-57. Retrieved from http://faculty.virginia.edu/smithlab/Protocols/EMSA book chapter.pdf

Solano, R., Nieto, C., Avila, J., Cañas, L., Diaz, I., & Paz-Ares, J. (1995). Dual DNA binding specificity of a petal epidermis-specific MYB transcription factor (MYB.Ph3) from Petunia hybrida. *The EMBO Journal*, *14*(8), 1773–1784. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7737128

Spolaore, P., Joannis-Cassan, C., Duran, E., & Isambert, A. (2006). Commercial applications of microalgae. *Journal of Bioscience and Bioengineering*, *101*(2), 87–96. https://doi.org/10.1263/jbb.101.87

Steele, D. J., Franklin, D. J., & Underwood, G. J. C. (2014). Protection of cells from salinity stress by extracellular polymeric substances in diatom biofilms. *Biofouling*, *30*(8), 987–998. https://doi.org/10.1080/08927014.2014.960859

Stegmaier, P., Kel, A. E., & Wingender, E. (2004). Systematic DNA-binding domain classification of transcription factors. *Genome Informatics. International Conference on Genome Informatics*, *15*(2), 276–286. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/15706513

Stracke, R., Werber, M., & Weisshaar, B. (2001). The R2R3-MYB gene family in Arabidopsis thaliana. *Current Opinion in Plant Biology*, *4*(5), 447–456. https://doi.org/10.1016/S1369-5266(00)00199-0

Strenkowska, M., Grzela, R., Majewski, M., Wnek, K., Kowalska, J., Lukaszewicz, M., Jemielity, J. (2016). Cap analogs modified with 1,2-dithiodiphosphate moiety protect mRNA from decapping and enhance its translational potential. *Nucleic Acids Research*, *44*(20), 9578–9590. https://doi.org/10.1093/nar/gkw896

Suman, S., Pandey, A., & Chandna, S. (2012). An improved non-enzymatic "DNA ladder assay" for more sensitive and early detection of apoptosis. *Cytotechnology*, *64*(1), 9–14. https://doi.org/10.1007/s10616-011-9395-0

Sun, L., Song, L., Zhang, Y., Zheng, Z., & Liu, D. (2016). Arabidopsis PHL2 and PHR1 Act Redundantly as the Key Components of the Central Regulatory System Controlling Transcriptional Responses to Phosphate Starvation. *Plant Physiology*, *170*(1), 499–514. https://doi.org/10.1104/pp.15.01336

Takagi, M., Karseno, & Yoshida, T. (2006). Effect of salt concentration on intracellular accumulation of lipids and triacylglyceride in marine microalgae Dunaliella cells. *Journal of Bioscience and Bioengineering*, *101*(3), 223–226. https://doi.org/10.1263/jbb.101.223

Tan, K. W. M., & Lee, Y. K. (2016). The dilemma for lipid productivity in green microalgae: importance of substrate provision in improving oil yield without sacrificing growth. *Biotechnology for Biofuels*, *9*, 255. https://doi.org/10.1186/s13068-016-0671-2

Thelen, J. J., & Ohlrogge, J. B. (2002). Metabolic engineering of fatty acid biosynthesis in plants. *Metabolic Engineering*, *4*(1), 12–21. https://doi.org/10.1006/mben.2001.0204

Thiriet-Rupert, S., Carrier, G., Chénais, B., Trottier, C., Bougaran, G., Cadoret, J.-P., Saint-Jean, B. (2016). Transcription factors in microalgae: genome-wide prediction and comparative analysis. *BMC Genomics*, *17*, 282. https://doi.org/10.1186/s12864-016-2610-9

Timilsina, G. R., & Shrestha, A. (2011). How much hope should we have for biofuels? *Energy*, *36*(4), 2055–2069. https://doi.org/10.1016/J.ENERGY.2010.08.023

Tsai, C.-H., Warakanont, J., Takeuchi, T., Sears, B. B., Moellering, E. R., & Benning, C. (2014). The protein Compromised Hydrolysis of Triacylglycerols 7

(CHT7) acts as a repressor of cellular quiescence in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(44), 15833–15838. https://doi.org/10.1073/pnas.1414567111

US EPA, O. (n.d.). Inventory of U.S. Greenhouse Gas Emissions and Sinks. Retrieved from https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks

Venkata Mohan, S., & Devi, M. P. (2014). Salinity stress induced lipid synthesis to harness biodiesel during dual mode cultivation of mixotrophic microalgae. *Bioresource Technology*, *165*, 288–294. https://doi.org/10.1016/J.BIORTECH.2014.02.103

Wachter, A., Tunc-Ozdemir, M., Grove, B. C., Green, P. J., Shintani, D. K., & Breaker, R. R. (2007). Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *The Plant Cell*, *19*(11), 3437–3450. https://doi.org/10.1105/tpc.107.053645

Wang, N., Qian, Z., Luo, M., Fan, S., Zhang, X., Zhang, L., Zhang, L. (2018). Identification of Salt Stress Responding Genes Using Transcriptome Analysis in Green Alga Chlamydomonas reinhardtii. *International Journal of Molecular Sciences*, *19*(11), 3359. https://doi.org/10.3390/ijms19113359

Wang, T., Li, Y., Ma, L., & Wu, C. (2011). Biomass to dimethyl ether by gasification/synthesis technology-an alternative biofuel production route. *Frontiers in Energy*, 5(3), 330-339. https://doi.org/10.1007/s11708-010-0121-y

Wang, Y., Zhang, S., Zhu, Z., Shen, H., Lin, X., Jin, X., Zhao, Z. K. (2018). Systems analysis of phosphate-limitation-induced lipid accumulation by the oleaginous yeast Rhodosporidium toruloides. *Biotechnology for Biofuels*, *11*(1), 148. https://doi.org/10.1186/s13068-018-1134-8

Wederell, E. D., Bilenky, M., Cullum, R., Thiessen, N., Dagpinar, M., Delaney, A., Hoodless, P. A. (2008). Global analysis of in vivo Foxa2-binding sites in mouse adult liver using massively parallel sequencing. *Nucleic Acids Research*, *36*(14), 4549–4564. https://doi.org/10.1093/nar/gkn382

Wilkins, O., Nahal, H., Foong, J., Provart, N. J., & Campbell, M. M. (2009). Expansion and Diversification of the Populus R2R3-MYB Family of Transcription

Factors. *Plant Physiology*, 149(2):981-93https://doi.org/10.1104/pp.108.132795

Williams, C. E., & Grotewold, E. (1997). Differences between plant and animal Myb domains are fundamental for DNA binding activity, and chimeric Myb domains have novel DNA binding specificities. *The Journal of Biological Chemistry*, *272*(1), 563–571. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8995298

Wu, R., Wang, Y., Wu, T., Xu, X., & Han, Z. (2017). MdMYB4, an R2R3-Type MYB Transcription Factor, Plays a Crucial Role in Cold and Salt Stress in Apple Calli. *Journal of the American Society for Horticultural Science*, *142*(3), 209–216. https://doi.org/10.21273/JASHS04030-17

Wykoff, D. D., Grossman, a. R., Weeks, D. P., Usuda, H., & Shimogawara, K. (1999). Psr1, a nuclear localized protein that regulates phosphorus metabolism in Chlamydomonas. *Proceedings of the National Academy of Sciences*, *96*(26), 15336–15341. https://doi.org/10.1073/pnas.96.26.15336

Wykoff, D. D., Grossman, A. R., Weeks, D. P., Usuda, H., & Shimogawara, K. (1999). Psr1, a nuclear localized protein that regulates phosphorus metabolism in Chlamydomonas. *PNAS*, (26) 15336-15341. Retrieved from www.pnas.org

Xiao, R., & Zheng, Y. (2016). Overview of microalgal extracellular polymeric substances (EPS) and their applications. *Biotechnology Advances*, *34*(7), 1225–1244. https://doi.org/10.1016/J.BIOTECHADV.2016.08.004

Yang, C., Li, D., Liu, X., Ji, C., Hao, L., Zhao, X., Zhu, L. (2014). OsMYB103L, an R2R3-MYB transcription factor, influences leaf rolling and mechanical strength in rice (Oryza sativa L.). *BMC Plant Biology*, *14*, 158. https://doi.org/10.1186/1471-2229-14-158

Yang, C., Li, R., Cui, C., Liu, S., Qiu, Q., Ding, Y., Zhang, B. (2016). Catalytic hydroprocessing of microalgae-derived biofuels: a review. *Green Chemistry*, *18*(13), 3684–3699. https://doi.org/10.1039/C6GC01239F

Yao, Y., Lu, Y., Peng, K.-T., Huang, T., Niu, Y.-F., Xie, W.-H., Li, H.-Y. (2014). Glycerol and neutral lipid production in the oleaginous marine diatom Phaeodactylum tricornutum promoted by overexpression of glycerol-3-phosphate dehydrogenase. *Biotechnology for Biofuels*, *7*(1), 110. https://doi.org/10.1186/1754-6834-7-110

Yodsuwan, N., Sawayama, S., & Sirisansaneeyakul, S. (2017). Effect of nitrogen concentration on growth, lipid production and fatty acid profiles of the marine diatom Phaeodactylum tricornutum. *Agriculture and Natural Resources*, *51*(3), 190–197. https://doi.org/10.1016/J.ANRES.2017.02.004

Zhanf, J., Liu. Z., Wang, S., & Jiang, P. (2002). Characterization of a bioflocculant produced by the marine myxobacterium Nannocystis sp. NU-2. *Applied Biotechnology and Microbiology*, 59(4-5):517-22. https://doi.org/10.1007/s00253-002-1023-7

Zhang, T., Cooper, S., & Brockdorff, N. (2015). The interplay of histone modifications - writers that read. *EMBO Reports*, *16*(11), 1467–1481. https://doi.org/10.15252/embr.201540945

Zhong, R., Yuan, Y., Spiekerman, J. J., Guley, J. T., Egbosiuba, J. C., & Ye, Z.-H. (2015). Functional Characterization of NAC and MYB Transcription Factors Involved in Regulation of Biomass Production in Switchgrass (Panicum virgatum). *PLOS ONE*, *10*(8), e0134611. https://doi.org/10.1371/journal.pone.0134611

Zhu, L. D., Li, Z. H., & Hiltunen, E. (2016a). Strategies for Lipid Production Improvement in Microalgae as a Biodiesel Feedstock. *BioMed Research International*, *2016*, 8792548. https://doi.org/10.1155/2016/8792548