UNDERSTANDING PHYSIOLOGICAL ADAPTATIONS, METABOLIC POTENTIAL AND ECOLOGY IN A NOVEL PHOTOAUTOTROPHIC ALGA FOR BIOFUEL PRODUCTION

by

Luisa Fernanda Corredor Arias

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Microbiology

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2019

DEDICATION

To Ben, the love of my life, partner in crime and my happy place. Your love, support, devotion and kindness are beyond measure.

To our little sunshine Lily. You turned my world into a wonderful rollercoaster of love and filled my life with joy and purpose.

Para mi mamá Magda, por el inmenso amor, la dedicación y la ternura que siempre me brinda.

Para mi papá Germán, quién despertó en mi la pasión por la ciencia y el conocimiento.

ACKNOWLEDGEMENTS

I would like to thank the Fulbright Scholarship Program for opening my mind to the world and its endless possibilities. It truly changed my life in the most positive ways.

I would like to thank my advisor, Dr. Matthew Fields, for giving me the chance to be part of his lab, the opportunities he provided for my professional development and for his support and advice for the last seven years of my life. I am very grateful to my committee members, Drs. Robin Gerlach, Mensur Dlakic and Abigail Richards for their support and always taking the time to help me, teach me and guide me through grad school and science. A special thank you to former and current members of the Fields lab for their support and friendship, coming to work was always great fun. I would also like to thank the members of the Center for Biofilm Engineering and the Microbiology and Immunology Department for all their assistance.

Finally, I would like to thank my husband Ben, for always keeping me strong and motivated, for his hopeful and contagious view of life and for always believing in me, even when I don't. My husband and my daughter Lily are my source of strength and encouragement for my personal and professional improvement every day. I would also like to thank my loving parents, Magda y Germán, for instilling in me the work ethic, commitment and perseverance that guides me through life, and my sister for her kind and loving soul. To my girls, Anali, Andrea, Erika, Dayane, Fede and Chiachi, I could not have done this without them. They are the family I found in the big sky country and will always be until we are old sipping wine and laughing without teeth. To Sandy.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF TABLES CONTINUED

LIST OF FIGURES

Figure Page

Chapter 2

LIST OF FIGURES CONTINUED

LIST OF FIGURES CONTINUED

Figure Page

LIST OF FIGURES CONTINUED

Figure Page

ABSTRACT

Commercial implementation of microalgal biomass as bio-oil/chemical feedstocks has been difficult to achieve, and challenges include water/nutrient sources, CO_2 delivery, and community dynamics of mixed cultures. We employed an integrated approach to the study of microalgal production systems to advance towards sustainable implementation of industrial microalgal biofuel production using a native alga (*Chlamydomonas*-like alga, PW95) isolated from Coal Bed Methane (CBM) production water. Our approach was based on the evaluation of PW95 physiological responses to combinations of growth constraints, the determination of its genomic and functional potential, phylogenetic relations and the implementation of an ecosystem view to algal biomass production. PW95 growth and lipid accumulation (biofuel potential) were ascertained in standardized media and CBM water through the evaluation of mixed effects of temperatures, nitrate levels, pH, and bicarbonate to elucidate interactions between multiple environmental variables and nutritional levels. The biofuel potential of PW95 ranges between 20–32% depending on culture conditions and our results suggest an important interaction between low nitrate levels, high temperature, and elevated pH for trade-offs between biomass and lipid production in the alga. Whole genome sequence was employed to predict biological and metabolic capacity in PW95, and the expression of these capabilities during growth in CBM water with the native microbial consortia was evaluated using RNA sequencing. genome determination and assembly resulted in a draft genome size of 92 Mbp with 14,000 genes predicted and 402 pathways mapped in the KEGG database. The gene complement of PW95 provided a glance into life in an oligotrophic environment (CBM water) and evidence of essential metabolic pathways for cell growth, survival and maintenance, also relevant for cultivation and value-added products generation. Microbial composition and shifts during growth were identified, as well as the algal phycosome During growth in CBM water, PW95 appeared to be supported by a native microbial consortium and differential expression analysis showed basic metabolic functions and adaptive physiological responses. Our findings build on previous knowledge for improved algal culturing for biomass and industry-valued products while exploring the biology of an organism with relevant impact in energy and water resource management.

CHAPTER ONE

INTRODUCTION

Microalgae are a highly ubiquitous and diverse group of photosynthetic organisms with a complex evolutionary history (Kim *et al.* 2014). They are critical for maintaining terrestrial and atmospheric conditions on a global scale (Grossman 2005), as their high photosynthetic and carbon fixation capacity make them important buffers against global warming via the fixation of CO_2 and production of O_2 (Tirichine and Bowler 2011). Some algae and diatoms can also be potential feedstocks of sustainable bioproducts, due to their ability to accumulate storage compounds such as lipids in the form of triacylglycerides (TAGs), starches and pigments under different environmental conditions and cultivations systems (Fields *et al.* 2014). In addition to serving as carbon storage compounds and performing biological functions, they have great commercial potential in the energy and nutraceutical industries (Grossman 2005; Fields *et al.* 2014). For instance, TAGs can be directly transesterified and converted to fatty acid methyl esters (FAMEs) (Georgianna and Mayfield 2012; Lohman *et al.* 2013; Fields *et al.* 2014; Hodgskiss *et al.* 2016; Singh, Upadhyay and Rai 2017) using direct *in situ* transesterification to efficiently convert the algal biomass to the full biofuel potential of the organism (Chen *et al.* 2018).

Microalgal biofuels are advantageous in comparison to conventional biofuels that originate from crops such as canola, soybean, palm oil or sugar cane (Hoekman *et al.* 2012). Among all photosynthetic organisms, microalgae are the most productive users of CO_2 and sunlight as larger levels of fixation can occur per land area than any other plant

crop. The associated high productivity and high growth rates are far superior to those of terrestrial plants, and biofuel production from algae does not compete with human food sources or arable land (Abo *et al.* 2019). Successful implementation of industrial-scale algal biofuel production has the capacity to generate significantly larger annual volumes of biodiesel per acre than current biodiesel crops (Hoekman *et al.* 2012), and taken together, these factors demonstrate the enormous potential of algae as a source of biofuel production. Furthermore, algae have the potential to grow in low quality water and wastewater for nutrient recycling, potentially eliminating the use of potable and clean water sources for algal production systems. However, the commercial implementation of algae as bio-oil/chemical feedstocks has been difficult to achieve due to a myriad of challenges (Fields *et al.* 2014). Important biological factors such as strain selection, growth physiology, culturing strategies, the quantity and the type of lipids that the strains accumulate (FAMEs) and the inverse relationship between biomass productivity and lipid content are parameters that can have an impact on production processes as well as economic viability of algae as a biofuel source (Talebi *et al.* 2013; Laurens, Chen-Glasser and McMillan 2017).

Factors that Influence Algal Growth and Lipid Accumulation

Algal growth and lipid accumulation are affected by physical factors (*i.e.,* light, temperature and pH) and chemical factors (*i.e.,* sources of carbon and nutrients). These factors can be modulated for better growth performance and lipid accumulation; however, the physiological responses to these modulations during cultivation can be species specific. Moreover, these parameters can vary significantly between geographic regions,

and matching indigenous strains with local growing conditions (*e.g.,* combinations of light, carbon, and temperatures) might improve algal production. Furthermore, the growth media in which the algae grow is an important parameter that affects the physiological outcomes; therefore, defining optimal conditions for growth and lipid production is a crucial focus for physiological studies in algal cultivation systems.

Physical Factors: Temperature, Light and pH

Light, temperature and pH are important factors for optimizing algal productivity. Photoautotrophic growth is by definition light-driven CO_2 assimilation, and in general, increases in light intensity show a proportional increase in growth and lipid production up to the optimal irradiance (Singh and Singh 2015; Abo *et al.* 2019), after which the growth declines, due to photoinhibition or light-induced damage of the photosystem II (Liu *et al.* 2019). It is known that phytoplankton in the photo-zone of the ocean occupy niches that match light intensity and wavelength to occupied water column depth, and freshwater algae likely exhibit similar environmental profiles to maximize energy conservation with biomass outputs. Previous works in freshwater green algae have shown physiological responses to different wavelengths of light (Hsia and Yang 2015; Singh and Singh 2015). Because outdoor cultivation is one of the most inexpensive ways to grow algae at larger-scales (and natural sunlight is the energy source under which green algae have evolved for millions of years), more work is needed to better understand growth and responses to natural sunlight in combination with other controllable parameters.

Temperature has also been shown to influence algal growth, as it affects cellular metabolic rates, enzymatic activities, respiratory and photosynthetic electron transport

and other important aspects, such as membrane fluidity and structure (Jiang and Gao 2004; Kurpan Nogueira *et al.* 2015). The optimum water temperature needed for cultivation of microalgae ranges from 15 to 30°C dependent upon the strain. Beyond this temperature range microalgal cell damage or death may occur (Singh and Singh 2015), unless the algal isolate can withstand extreme temperatures as is the case of extremophilic algae (Varshney *et al.* 2015).

Previous studies have demonstrated that the pH has significant impacts on biomass and lipid yields (Fields *et al.* 2014; Moll *et al.* 2014; Fakhry and El Maghraby 2015; Kurpan Nogueira *et al.* 2015; Lohman *et al.* 2015; Pancha *et al.* 2015). Microalgae require CO_2 for photosynthetic growth and the pH of the growth media determines carbon availability due to chemical speciation of CO_2 and its related ions in the liquid phase. Depending on the pH of the media, three different species of carbon may be present when CO_2 is dissolved in the water. CO_2 is prevalent at low pH, bicarbonate (HCO_3^-) at neutral pH, and carbonate (CO_3^{2-}) at high pH (Pedersen *et al.* 2018). Since algae can only use CO_2 and bicarbonate, a high pH will prevent algal photosynthesis, and in turn stop growth. The optimal external pH for algal growth usually ranges between pH 6 and 8.5 (Ihnken *et al.* 2014).

Chemical Factors: Carbon and Nitrogen

Inorganic carbon is vital for microalgal growth, and inorganic carbon (IC) fixation typically is carried out via C3 and/or C4 metabolism in microalgae but some species have evolved an inducible carbon concentrating mechanism (CCM) (Hopkinson, Meile and Shen 2013). There are two predominant forms of dissolved inorganic carbon (DIC),

bicarbonate (HCO_3^-) and CO_2 (Pörtner 2008). Microalgae can express a high affinity for DIC through a biophysical CCM, influenced by CO_2 or HCO_3^- availability and light supply (Reinfelder 2011; Mackinder 2018). Previous analysis of IC fluxes indicated that the major IC flux in the cell is the transport of IC into the chloroplast from the cytoplasm. IC accumulates in the chloroplast stroma and generates a CO_2 deficit in the cytoplasm, inducing CO_2 influx into the cell, which generally supplies most of the IC required for photosynthesis, although HCO_3^- transport also plays a role (Reinfelder 2011; Hopkinson, Meile and Shen 2013; Mackinder 2018). Higher rates of carbon fixation are attributed to the ability to uptake different DIC species (HCO_3^- and CO_2) from the media (Chen, Qiu and Shao 2006). Thereby, increasing the concentration of CO_2 at the active site of ribubulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), inhibiting the oxygenase activity of the enzyme, and reducing photorespiration and promoting photosynthesis and growth (Reinfelder 2011; Valenzuela *et al.* 2012; Mackinder 2018).

Nitrogen is indispensable for growth, as it is a major constituent of amino acids and nucleic acids, and most organisms across the three domains of life have evolved mechanisms to assess and respond to coupled acquisition of nitrogen and carbon. Carbon and nitrogen availability are continually assessed to ensure adequate ratios for biomass production, and in the event of unbalanced conditions, some organisms have the ability to store the excess substrate. For photoautotrophic organisms, nitrogen also plays a vital role, as a major component of chlorophyll (Chokshi *et al.* 2017). Numerous studies have demonstrated the ability of nitrogen-deprivation to ‘trigger’ or induce lipid accumulation in a variety of algae and diatoms (Fields *et al.* 2014), which is most likely a result of

photo-driven carbon-storage as a consequence of nitrogen-limitation (discussed below). Nitrate, nitrite, ammonium, and urea can be used as nitrogen sources for most green algae and diatoms (Eustance *et al.* 2013) but it has been suggested that algal cells could have a higher affinity for nitrate as opposed to nitrite and urea (Kothari *et al.* 2013; Hodgskiss 2015). Nitrate is also a common contaminant present in wastewater (Jia and Yuan 2016) hence the importance of understanding the physiological effects that nitrate has on algal cultivation for potential future applications in nutrient recycling coupled with the treatment and use of low-quality water using algae.

Nitrate Depletion and Bicarbonate Addition for Lipid Production

Microalgae mass culturing generally aims for high biomass yields combined with an intentional nutrient starvation aimed at increasing lipid contents (Fields *et al.* 2014; Ihnken *et al.* 2014). Although continuous nitrogen starvation increases the lipid and carbohydrate content of microalgae (Fields et al., 2014; Gardner et al., 2011; Huang et al., 2013; Jiang et al., 2012; Valenzuela et al., 2012; Yeesang & Cheirsilp, 2011), decreases in growth rate consequently causes an overall decline in productivity (Chokshi *et al.* 2017). Nitrogen starvation also lowers protein content and RuBisCO levels but activates biochemical responses that lead to an elevation of lipid bodies by a process that is not entirely understood (Ihnken *et al.* 2014). In practice, algae are grown in media with sufficient nitrate to rapidly stimulate higher concentrations of biomass during the exponential phase of growth, and then the overproduction of lipids often coincides with nutrient starvation that has been shown for many microalgal species under nutrient stress (Fields *et al.* 2014; Moll *et al.* 2014; Xia *et al.* 2014; Zhu *et al.* 2016).

An alternate way to stimulate lipid accumulation in some species of green algae and diatoms is the addition of different concentrations of bicarbonate before nitrate depletion (Gardner *et al.*, 2013; White *et al.*, 2013; Moll *et al.*, 2014). This strategy can potentially exploit the physical CCM that may be operating in certain species of algae, shuttling IC into lipid biosynthesis concurrently as the cells experience nitrogen starvation. The addition of bicarbonate at the time of nitrate depletion can change the metabolic activity of the culture, arresting cellular replication and beginning TAG accumulation (Gardner *et al.* 2013; White *et al.* 2013; Lohman *et al.* 2015). Gardner et al. previously reported that cellular replication was stopped in a chlorophyte but not in the diatom *Phaeodactylum tricornutum*, suggesting significant differences in the metabolic pathways employed for IC assimilation in eukaryotic photoautotrophs (Gardner *et al.* 2012).

Algal Growth in Standardized Media versus Low Quality Water

Algae represent a greatly heterogeneous and vastly diverse group of organisms characterized by very different physiological attributes and belonging to many different phyla. Consequently, different species of algae have very different growth requirements and responses to environmental stimuli and cues. In principle, microalgae can grow provided that adequate amounts of sunlight, water, carbon dioxide, and nutrients are available (Idenyi *et al.* 2016). In laboratory settings, these nutrients are provided in standardized media in the form of macronutrients (carbon, nitrogen, phosphorus), micronutrients (iron, manganese, molybdate, zinc, borate) and vitamins (typically B12 > thiamine > biotin) (Andersen 2005). However, even though standardized media provides

the optimal environment for growth, the use of laboratory-grade chemicals and water are unsustainable for biomass and lipid production at a commercial scale. An alternative to the use of standardized media is low quality water sources (CBM water; wastewater) and nutrients, although in these cases strain selection will be crucial to ensure adequate growth and production at the needed growth conditions.

The environmental conditions of a specific area can greatly influence microalgae populations and associated growth dynamics. Although many algal strains are available for the investigation of growth rates and bio-fuel production potential, the optimal strain in any given area will likely be highly dependent on the geographic location and associated weather (solar irradiance, temperature). The most logical approach then is to screen for highly productive strains with maximum lipid contents at selected locations and optimize the growth conditions for future large-scale cultivation (Idenyi *et al.* 2016) in similar conditions.

Coal Bed Methane Production Water as a Medium for Algal Growth

The Powder River Basin (PRB; northeastern Wyoming and southeastern Montana) has been determined to be one of the most promising sources of coal bed methane (CBM) in the United States (United States Geological Survey 2002) in the United States. The collection of the gas for use as an energy source generates large amounts (millions of gallons/well/day) of non-potable, low quality water that is stored in CBM water impoundments. Compared to standardized media (Bold's Basal Medium (BBM)), CBM water constitutes a minimal medium for growth but amended CBM water has been shown to be an adequate medium for algal biomass and lipid production in

CBM collected in Montana (Hodgskiss *et al.* 2016) and coal seam gas water in Australia (Harrington 2011). CBM water has very high levels of sodium, bicarbonate and total dissolved solids and low levels of nitrogen and phosphorus (Hodgskiss 2015). Due to the cost, environmental restrictions and logistical difficulty and costs of discharging the water to other locations, CBM water in the PRB is kept in unlined ponds for long-term storage. Water evaporation into the atmosphere and the resulting concentration of solids, sodium and other potential toxic hazards in the remaining water is an essential environmental consideration. Water infiltration into the soil could have undesired effects on local ground water and the economy (Hodgskiss *et al.* 2016; Grubert and Sanders 2018). Developing the capability to use low quality water impoundments such as these for the production of biofuel represents an additional benefit for the creation of renewable energy in the future, provided that an effective microalga can be identified that can survive and thrive in the stated conditions.

In the described studies, a novel green alga (strain PW95) has been isolated from a CBM water impoundment in the Powder River Basin (PRB), northeastern Wyoming (44° 52.613′N 106° 54.700′W; well FG-09). Growth of this alga was characterized for growth in standardized medium as well as CBM production water. PW95 has been observed to produce biomass and biofuel in monoculture and batch mode, using standardized media and low quality CBM water (Corredor *et al*., in review; Chapter 2, Hodgskiss *et al.* 2016) and, thus, represents an alternative approach for biomass and lipid production that could potentially alleviate the water burden and lower the cost of input into algal production systems.

Omics Approach to Understanding Algal Physiology

Whole genome sequencing (WGS) analysis consists of genome assembly, as well as structural and functional annotation and has been frequently employed to understand and predict the biochemical capacities and responses of model algal strains or production strains with potential to produce chemical feedstock products, such as biofuel (Hovde *et al.* 2018). Genome sequences provide a view into the functional potential of an individual organism and can inform phylogenetic analysis (phylogenomics; *i.e.*, concatenated protein trees) and functional genomics (*i.e.*, RNA-seq) to bring context and to transfer the knowledge from reference organisms into less well-characterized systems (Blaby-Haas and Merchant 2019). Despite algae being a highly heterogeneous and diverse group with very different physiological attributes and belonging to many different phyla (Prochnik *et al.* 2010; Chapman 2013; Idenyi *et al.* 2016), deep understanding of algal biology and physiology will advance our knowledge in primary productivity and nutrient cycling, as well as in algal production systems. WGS can reveal conserved genes and function among algae (comparative genomics) and can also reveal novel or unique genetic and functional potential, and currently over 100 algal genomes are publicly available or soon to be published (compared to >10,000 for bacteria).

The first draft genome of an alga was released in 2004 (Tanaka *et al.* 2004; Blaby-Haas and Merchant 2019) and the genome of the model alga *Chlamydomonas reinhardtii* was published in 2007 (Merchant *et al.* 2007). Although sequencing technology and computational methods for assembly have continued to improve, structural annotations can still be inaccurate even with finished genomes, and typically

algal research communities must invest significant resources to increase the number of evidence-based gene models both prior to and subsequent to publication of the genome sequence (Blaby-Haas and Merchant 2019). Historically, WGS has led to the development of algal genomic models that have provided insights into eco-physiological adaptation in environments, genome evolution, origin of multicellularity, the history of photosynthetic life on earth and even key animal functions such as the structure, assembly and function of eukaryotic flagella (Merchant *et al.* 2007; Tirichine and Bowler 2011; Kim *et al.* 2014).

RNA sequencing analysis typically aims to determine global gene expression in an organism or a cell. A RNA-seq method consists of mapping short reads to a reference genome (or to assemble them into contigs and align them to the reference genome), quantifying the number of reads aligning to individual genes (*i.e.*, number of mapped reads being indicative of expression levels) and finally determining the differential expression of a gene between 2 or more samples to identify how the transcript levels differ among samples of interest (Han *et al.* 2015).

RNA-seq is a next generation sequencing (NGS) tool frequently used for the quantitative assessment of RNA. Transcript identification and the quantification of differential gene expression can lead to a better understanding of physiological characteristics and allows the quantification of abundance levels or relative changes in gene transcript levels resulting from defined developmental stages or under specific treatment conditions (Han *et al.* 2015). RNA-seq studies have been used to determine many characteristics of microalgae. Transcriptomic studies of *Phaeodactylum*

tricornutum have shown that genes involved in nitrogen assimilation processes and the pentose phosphate pathway are up-regulated during nitrogen depletion (Valenzuela *et al.* 2012; Park *et al.* 2015). Genes encoding proteins that are essential for lipid accumulation are also up-regulated while genes for fatty acid degradation are down-regulated in *Chlamydomonas reinhardtii* (Park *et al.* 2015). Since RNA-seq can determine differential expression of all genes that are being transcribed in any given condition, it also has the capacity to reveal novel biochemical competencies under the tested growth conditions (Valenzuela *et al.* 2012).

Ecological Impacts on Algal Cultivation

Microalgae are often associated with diverse bacteria in their native aquatic ecosystems as well as in experimental culture systems dependent upon starting water and nutrients (Lakaniemi *et al.* 2012; Krohn-Molt *et al.* 2013). Polycultures (ecosystems) may be more stable and productive than monocultures and previous reports of natural and constructed algal assemblages, show that increased diversity can be associated with higher productivity (Stockenreiter *et al.* 2012; Beyter *et al.* 2016). Algal cultures include complex bacterial communities that, in experimental settings, are often considered to be bacterial contamination (Fulbright *et al.* 2018), but aside from playing a potential role in stimulating and supporting microalgal productivity, associated bacteria can help the algae perform critical tasks (Lian *et al.* 2018). Microbial communities can play a role in mediating biogeochemical cycling, micro-food web structure, stimulating algal growth by producing essential elements (*e.g.*, vitamins, phytohormones), exhibiting algicidal activity, and regulating algae-bacteria signaling such as quorum sensing (Adachi *et al.*

2003; Krohn-Molt *et al.* 2013; Lian *et al.* 2018; Zhou *et al.* 2018). The identification of microbial communities closely associated with microalgae in bioreactors can help understand potential algae–bacteria functional relationships for improved productivity in algal systems in the future (Lakaniemi *et al.* 2012; Krohn-Molt *et al.* 2013; Sambles *et al.* 2017). Most studies have surveyed bacterial communities in water treatment studies and facilities (Bell *et al.* 2019; Shetty *et al.* 2019) or algal blooms (Xiao *et al.* 2017; Zhou *et al.* 2018), but knowledge is lacking in the context of microalgal production systems, such as photobioreactors (PBRs). Moreover, basic research is lacking in systems implementing environmental isolates with growth and biofuel potential in low quality water. An ecological approach to the study of microalgal production systems can give insight into algal–bacterial interactions that could be explored for the establishment of highly productive and stable algal systems.

The novel green sporulating alga, PW95, isolated from a CBM water impoundment in the Powder River Basin was examined in the presented dissertation. PW95 shows promise for its ability to produce biomass and accumulate lipids for biofuel production in monoculture and batch mode in standardized media as well as low quality CBM water. However, at present little is known about the physiological, ecological, genomic, or transcriptomic characteristics of this alga. Understanding these characteristics will be key to assessing the true potential of PW95 as a viable biofuel producer in low quality water in the future.

Research Goals and Objectives

1. Determine the impact of different combinations of temperature, pH, and bicarbonate on PW95 growth and lipid production.

2. Predict biochemical capacity of PW95 via whole genome sequencing, de novo assembly and functional annotation.

3. Assess the potential contribution of differentially expressed genes to biomass production and lipid production in PW95 during biomass production in a photobioreactor system using low quality water.

4. Investigate the microbial community composition during algal biomass production and define the algal microcosm in a photobioreactor system using low quality water.

Objective 1 is described in Chapter Two where phylogenetic characterization (SSU rRNA gene and ITS) of a novel green microalga of a novel green microalga, isolate PW95, from low quality water was performed. Additionally, 24 combinations of treatments were tested to understand the physiological adaptations to limited resources and changing environmental conditions in a phototrophic green microalgal isolate during growth and lipid production over time.

Objective 2 is described in Chapter Three where high molecular weight DNA from PW95 was extracted for PacBio sequencing, *de novo* genome assembly and functional annotation.

Objective 3 is described in Chapter Four where RNA from PW95 was extracted for sequencing and differential expression analysis of key metabolic responses related to biomass and lipid production was performed. The cultures were grown in filtered and unfiltered CBM production water to elucidate potential impacts of a diverse bacterial community associated with low-quality water. Samples were collected at late-exponential and late-stationary phases to compare biomass and bio-lipid levels. Additionally, *de novo* transcriptome assembly was performed using the raw reads from the experimental conditions evaluated in the photobioreactor system to implement evidence-based genome annotation and contribute to the improvement of the functional annotation of PW95 (Chapter 3).

Objective 4 is described in Chapter Five where the microbial community composition in a photobioreactor system for algal biomass production was investigated. The source community was compared to the planktonic bacteria in the water as well as directly/indirectly associated with algal aggregates that formed over time during growth.

Chapter Six sums up the overall implications of the current work.

CHAPTER TWO

IMPACT OF TEMPERATURE, NITRATE CONCENTRATION, PH AND BICARBONATE ADDITION ON BIOMASS AND LIPID ACCUMULATION OF A SPORULATING GREEN ALGA

Contribution of Authors and Co-Authors

Manuscript in Chapter Two

Author: Luisa Corredor
Contributions: Experimental design, performed experiments, literature review, synthesized ideas, performed data and statistical analyses, wrote and edited the manuscript.

Co-Author: Elliot B. Barnhart
Contributions: Isolated the algal strain and edited the manuscript.

Co-Author: Al Parker
Contributions: Experimental design, performed statistical analyses and edited the manuscript.

Co-Author: Robin Gerlach
Contributions: Procured funding, synthesized ideas, reviewed and edited the manuscript.
Co-Author: Matthew W. Fields

Contributions: Principal Investigator, procured funding, experimental design, synthesized ideas, wrote and edited the manuscript.

Manuscript Information Page

Luisa Corredor, Elliot B. Barnhart, Al Parker, Robin Gerlach, Matthew W. Fields

Status of Manuscript:
____Prepared for submission to a peer-reviewed journal
__X_Officially submitted to a peer-reviewed journal
____Accepted by a peer-reviewed journal
____Published in a peer-reviewed journal

Date of submission: November 19th, 2019

Abstract

A novel sporulating alga, isolate PW95, isolated from Coal Bed Methane (CBM) production water was characterized morphologically, phylogenetically and physiologically with respect to lipid accumulation under varied growth conditions. Based on morphology and 18S SSU rRNA gene sequence, PW95 belongs to the order Chlamydomonadales, is similar to *Chlamydomonas*-like species and phylogenetically related to members of the genus *Neospongiococcum, Chlamydopodium* and *Chlorococcum,* also non-motile coccoid green algae. Comparisons of the culturing conditions were performed in terms of maximal biomass growth and lipid accumulation. A multivariate physiological analysis (24 treatment combinations) was used to assess the mixed effects of temperatures, nitrate levels, pH, and sodium bicarbonate (HCO_3^- source) on PW95 growth and lipid accumulation during batch experiments in a standardized growth medium. At high temperature (30°C) in combination with higher nitrate (2 mM), PW95 displayed a faster growth rate, as well as accumulation of chlorophyll. Fast growth was also observed at 25°C with 2 mM nitrate, and a pH buffer decreased the growth rate. Buffer addition (pH 8.5) had a negative impact on growth for most of the treatments, and bicarbonate addition had no significant effect on growth. In general, there was an inverse relationship between biomass accumulation (cells and chlorophyll) and lipid accumulation at the extremes of nitrate and temperature (low nitrate as well as low and high temperature). Saturated and unsaturated C16 (hexadecanoic) and C18 (octadecanoic) chains were the dominant FAMEs (Fatty Acids Methyl Esters) species at 20°C and 30°C. The treatment with the lowest cell density showed the highest percentage total FAME

content (32.83% weight of FAME/weight of biomass; w/w); however, at 30°C, PW95 showed almost as high lipid accumulation (27.15% and 25.43%) in treatments with the highest cell density. When total FAME levels were compared to the rate of cell or chlorophyll accumulation, chlorophyll accumulation rate was a better predictor of overall FAME content under unbuffered conditions (R=0.82). The results suggest an important interaction between low nitrate levels, temperature, and elevated pH for trade-offs between biomass and lipid production in PW95.

Introduction

Microalgae-based biofuel production is an advantageous process that can combine carbon dioxide uptake with the production of bio-oil in low quality water (CBM production water or wastewater) for nutrient recycling. Based on their growth efficiency, capacity to produce compounds for industrial applications, and ability to withstand different types of environmental stresses, Chlorophytes (green algae), have been the focus of major efforts towards microalgal biofuel sustainability (Hoekman *et al.* 2012; Talebi *et al.* 2013; Adeniyi, Azimov and Burluka 2018). Microalgae are promising feedstocks for biofuel and other valued products through the accumulation of triacylglycerides (TAGs) and other related compounds that can be directly transesterified and converted to FAMEs (Georgianna and Mayfield 2012; Lohman *et al.* 2013a; Fields *et al.* 2014; Hodgskiss *et al.* 2016; Singh, Upadhyay and Rai 2017). However, the commercial implementation of algae as bio-oil/chemical feedstocks has been difficult to achieve due to process challenges, including harvesting, oil extraction and conversion to fuel (Fields *et al.* 2014; Shurin *et al.* 2016). Important biological factors, such as strain

selection (growth physiology and potential to accumulate lipids), culturing strategies, the quantity and the type of FAMEs that the strains accumulate, and the inverse relationship between biomass productivity and lipid content (Talebi *et al.* 2013; Laurens, Chen-Glasser and McMillan 2017) are parameters that can have an impact on production processes as well as economic viability. A prospective contribution for these challenges is the exploitation of native microalgae exposed to different physical stresses *in situ* that can tolerate a variety of waters, nutrients, and weather conditions respective to different regions. Algal environmental isolates have the potential to lower culture costs and lipid production, through acclimation to fluctuating environmental conditions with low nutrient availability (Hoekman *et al.* 2012; Hodgskiss *et al.* 2016). Frequently, green microalgal strains from culture collections are used to explore maximization of biodiesel potential (Tasić *et al.* 2016); but native strains of algae can also be a viable option that most likely will respond more robustly to variable culture conditions (*e.g.*, temperature, sunlight, invasion during outdoor cultivation) (Zhou *et al.* 2011; Eustance *et al.* 2013; Fields *et al.* 2014; Hodgskiss *et al.* 2016).

The isolation of a novel green microalga, PW95, was previously reported from a Coal Bed Methane (CBM) production pond in the Powder River Basin (northeastern Wyoming and southeastern Montana). PW95 was shown to grow in supplemented CBM production water and produced elevated levels of lipids for potential biofuel production (Hodgskiss *et al.* 2016). Interestingly, PW95 did not accumulate significant levels of lipid when grown in a defined growth medium (Hodgskiss *et al.* 2016). Previous studies have demonstrated that the culture's pH, the use of a buffer, and nutrient limitation can have

significant impacts on biomass and lipid yields (Fields *et al.* 2014; Moll *et al.* 2014; Fakhry and El Maghraby 2015; Kurpan Nogueira *et al.* 2015; Lohman *et al.* 2015; Pancha *et al.* 2015); therefore, we assessed PW95 growth in a defined medium under various conditions. A multivariate physiological analysis using 24 different combinations of treatments elucidated synergistic or detrimental interactions between multiple environmental variables and nutritional factors. A high-concentration bicarbonate (HCO^-_3) trigger applied before the onset of nitrate depletion was tested to further optimize lipid production, as it has been shown to induce and/or accelerate TAG accumulation in green and marine algae (Gardner *et al.* 2012, 2013a, 2013b; Xia *et al.* 2014; Pedersen *et al.* 2018), and stimulate biomass growth (Gardner *et al.* 2013a; White *et al.* 2013; Lohman *et al.* 2015).

The aim of this study was to determine a promising combination of culture conditions for less input, low cost, sustainable biofuel production in a native algal isolate, the optimization of culture conditions for growth and lipid production to advance towards the use of native algal isolates from environmentally challenged field sites and low quality water sources for biofuel applications, and to determine the biofuel potential of PW95 by assessing the type and quantity of lipids that are being accumulated during nitrogen replete and deplete conditions. Results from this study give insight into the complex physiological responses of a green microalga to changing environmental conditions in the context of an algal biomass and lipid production system.

Materials and Methods

PW95 Isolation and Culture Maintenance

Water samples were collected from a CBM production water pond in the Powder River Basin, northeastern Wyoming (44° 52.613′N 106° 54.700′W) (average pH 8.4 and temperature = 15°C), and plated on Bold's Basal Medium (BBM) for growth at 20°C in a light incubator using a 14:10h light:dark cycle and constant shaking (125 RPM)(Hodgskiss *et al.* 2016). Isolated algal colonies were transferred to liquid BBM to stimulate biomass increase (Hodgskiss *et al.* 2016). To screen for bacterial contamination, cultures were analyzed via DNA sequence analysis (SSU rRNA gene sequences). The PCR amplicons were prepared for sequencing according to the "16S Metagenomics Sequencing Library Preparation" Illumina protocol for paired-end sequencing (Illumina, San Diego, CA, USA) and sequencing was carried out using an Illumina MiSeq v3 platform. The sequence reads were processed, and bacterial operational taxonomic units were identified with the MiSeq SOP pipeline of the MOTHUR software package (Kozich *et al.* 2013). Cultures of PW95 without bacteria were stored as stock cultures, maintained in liquid BBM at pH 8.4–8.5, and routinely checked for bacterial contamination using R2A agar plates incubated in the dark at 20°C to check for heterotrophic growth.

Morphological and Phylogenetic Characterization of PW95

Morphological characterization of PW95 was performed using electron (EM) and epifluorescent microscopy. Sample preparation for EM was performed using centrifuged cultures (5 mL at 14,500 x g), washed and resuspended in ultrapure deionized water

(dH_2O; 18.2 Ω). The cell suspension (30 µL) was placed on an EM chip, allowed to dry overnight, before the chip was sputter-coated with iridium the next day. Gross morphology of PW95 was characterized using Field Emission-Scanning Electron Microscopy (SUPRA 55VP; Zeiss). Sample preparation for epifluorescent microscopy was performed staining 1 mL of culture with 4 µL of Nile Red to observe lipid vacuoles (Cooksey *et al.* 1987) and cell mask orange for cell structure (as per manufacturer instructions; Thermo Fischer Scientific). Key intracellular markers were examined with transmitted light and fluorescence filters (B2A, TRITC, FITC) using an Epifluorescent Microscope (Nikon E800 Eclipse with fluorescent filters); images were captured using a Photometrics snap-MIO digital camera.

The phylogeny of PW95 was determined by amplification, sequencing (Sanger) and comparisons of the SSU rRNA gene sequences (18S) and internal transcribed sequences (ITS) against BLAST (Basic Local Alignment Search Tool (Altschul *et al.* 1990). Total genomic DNA was extracted using a Fast DNA SPIN Kit for Soil (MP Biomedicals). The 18S rDNA gene was amplified by PCR with the full-length primers 1-21F (5′-WACCTGGTTGTTGATCCTGCCAGT-3′) and 1741-1798R (5′-GATCCTTCYGCAGGTTCACCT-3′) as well as the internal primers NS3-F (5′-GCAAGTCTGGTGCCAGCAGCC-3′) and NS8-R (5′-TCCGCAGGTTCACCTACGGA-3′) (White *et al.* 1990). Both ITS sequences were amplified as a contig with the primers ITS-1F (5′-TCCGTAGGTGAACCTGCGG-3′) and ITS-4R (5′-TCCTCCGCTTATTGATATGC-3′) (White *et al.* 1990). PCR was performed using the following program: an initial denaturation at 94°C for 2 min; 30

cycles at 94°C for 30 s, 57°C for 1 min and 72°C for 1.15 min; and a final extension at 72°C for 7 min. The same program was used for the ITS gene amplification with an increase in annealing temperature to 60°C due to the high primer melting temperature (T_m). The size of each PCR amplicon was confirmed using agarose gel electrophoresis. Assembly and editing to build consensus sequences of both genes were carried out by comparing the direct and reverse chromatograms using Geneious R8 (https://www.geneious.com). The SSU rRNA gene sequence was trimmed according to sequence quality to a partial sequence (1060 bp out of 1798 bp); the ITS sequence did not require editing. Alignments for the 18S rRNA gene were conducted with the web BLASTN (Basic Local Alignment Search Tool-Nucleotide databases, https://blast.ncbi.nlm.nih.gov) and 47 DNA sequences from green algae were selected for phylogenetic analysis based on high percentages of identity (>90%) and query coverage. *Botryococcus braunii* sequences were selected as the rooted outgroup, after testing tree topology and resolution with other groups of algae. The multiple sequence alignment was generated with SSU-ALIGN v0.1.1 (http://eddylab.org/software/ssu-align/) (Nawrocki 2009) based on the conserved secondary structure and sequence of the SSU rRNA and TrimAl v1.2 (http://trimal.cgenomics.org) (Capella-Gutiérrez, Silla-Martínez and Gabaldón 2009) was used for automated trimming (gap threshold = 0.5). Bayesian Inference (BI) was used for phylogenetic analysis and the tree was constructed using MrBayes 3.2 (http://nbisweden.github.io/MrBayes/index.html) (Ronquist *et al.* 2012) with the following parameters: General Time Reversible (GTR) model with 8 gamma categories and 2 parallel runs, 4 chains, 2 million iterations and 0.25 burn-in fraction.

iTOL v4.4.2 (Interactive Tree of Life, https://itol.embl.de) (Letunic and Bork 2019) was used for tree viewing, editing and annotation. To construct a BI phylogenetic tree with the ITS consensus sequence, the same procedure was followed as described for the 18S sequence, except the alignment was constructed with Mafft v6 (BI parameters specified in the Supplementary Figure A1).

Experimental Design

A randomized factorial experimental design was used to determine the effect of temperature, nitrate concentration, buffer and bicarbonate addition on PW95 growth and lipid production. All the experiments were performed using two biological replicates and a total of 24 treatment groups were evaluated. Using Minitab (http://www.minitab.com), growth experiments were discerned by temperature (20°C, 25°C, 30°C) to evaluate the impact of 3 different treatments and 2 factors per treatment: Initial nitrate concentration in the media (0.5 mM and 2 mM), pH (with and without buffer) and bicarbonate trigger (with and without bicarbonate). Each treatment group has a label format consisting of all of the aforementioned factors, *e.g.*, 20°C_0.5 Nitrate_Buffer_No Trigger, which consists of the combination of temperature at 20°C, 0.5 mM of initial nitrate, no buffer and no bicarbonate trigger (Figure 1).

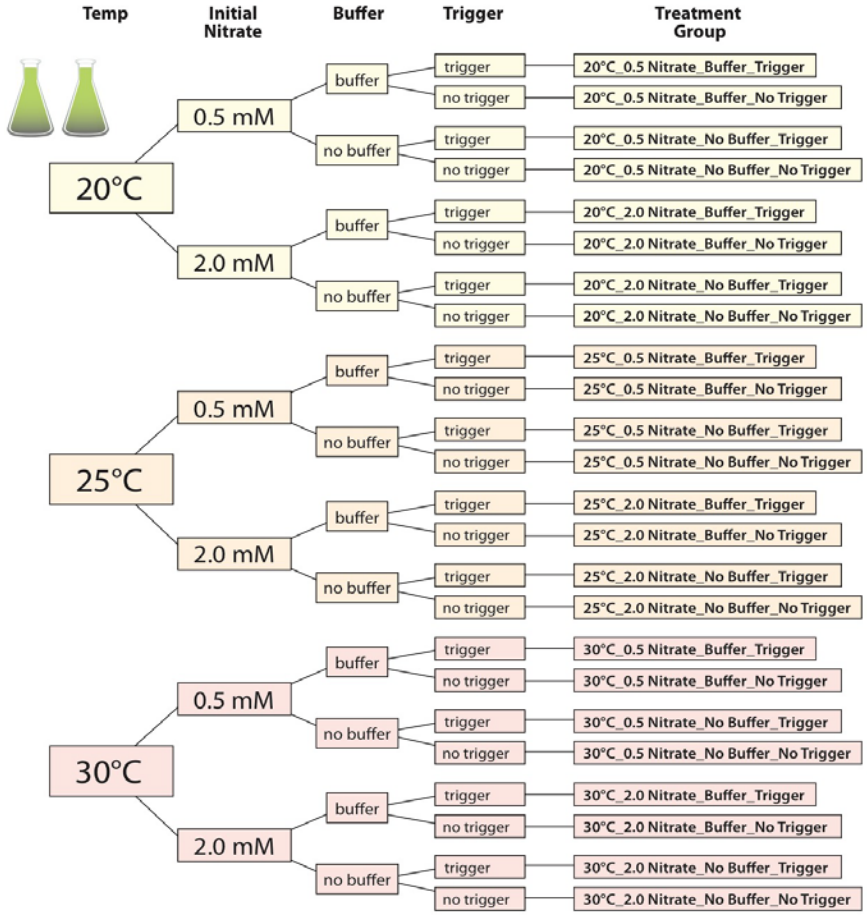


Figure 1. Experimental design for physiological characterization and growth optimization experiments in this study. Using a factorial design, 24 groups of treatments were evaluated and compared. All the experiments were performed using two independent biological replicates and 6 independent experiments.

Experimental Growth Conditions

PW95 stock cultures in exponential phase of growth, grown using identical concentrations of nitrate (0.5 or 2.0 mM) and environmental conditions as the experimental cultures, were used as initial inocula (10% V_T). PW95 batch experimental cultures were grown in duplicates, in conical flasks containing 200 mL of BBM with 0.5

or 2.0 mM nitrate. To test the effect of pH buffering, a mixture of analytical grade Trizma-Base and Trizma-HCl (Sigma-Aldrich, St Louis, MO, USA; Technical Bulletin No.106B) was added to provide a medium with an initial pH of 8.5, identical to the pH of the unbuffered control. To test the effect of bicarbonate addition, analytical grade sodium bicarbonate ($NaHCO_3$, Sigma-Aldrich, St Louis, MO, USA) was added at a final concentration of 50 mM immediately before nitrate depletion, as previously described (Lohman *et al.* 2015). Nitrate concentrations were measured with ion chromatography (IC; see below for details). Cultures were inoculated and grown under constant shaking (125 rpm) with gas exchange membranes and a 14:10h light/dark (L/D) cycle with a light intensity of 145 PAR, as measured with a photosynthetically active radiation (PAR) meter (LI-COR).

Physiological Response Analyses

Culture samples were collected every 24 hours before the end of the light cycle and subsequently used for the analysis of physiological parameters. Culture pH was measured using a standard pH meter (Hach, Sension+PH1 Portable pH meter). Growth performance was evaluated manually by cell counts in duplicate, using an optical hemocytometer (Reichert), and through chlorophyll concentration measurement. Chlorophyll was extracted using cold 100% methanol before pipetting an aliquot of the supernatant containing the extracted chlorophyll into a 96 well plate in triplicate. Chlorophyll fluorescence was measured using a Synergy H1 hybrid fluorometer/spectrophotometer reader (Bio-Tek) and Gen5 software was used to evaluate the fluorescence at 775 nm, 652.5 nm and 665 nm (Hodgskiss *et al.* 2016). Total

chlorophyll concentrations (Chl-a plus Chl-b) were calculated according to Ritchie (Ritchie 2006). Samples for water chemistry analysis were stored at −80°C and nitrate concentration changes were measured by ion chromatography using an AS22 Anion-Exchange Column (DIONEX ICS-1100) with a 4.5 mM NaHCO3/ 1.4 mM Na2CO3 buffer as eluent at a 1.2 mL flow rate with an ASRS 4 mm suppressor. A Chromeleon Chromatography Management System was used to determine the concentration of nitrate in each sample throughout the growth curve.

Lipid Content Analysis

Lipid content in the form of triacylglycerides (TAG) was estimated using Nile Red (NR) fluorescence as previously described (Cooksey *et al.* 1987); 4 µl of NR was added to 1 ml of culture, before 10 minutes incubation (optimal exposure time was determined using a time-course assay as reported by Chen et al (Chen *et al.* 2009)). Total NR fluorescence was determined at an excitation wavelength of 530 nm and an emission wavelength of 575 nm with a Synergy H1 hybrid fluorometer/spectrophotometer reader (Bio-Tek) and Gen5 software.

To assess the biodiesel potential of PW95, biomass was harvested and pelleted as described previously (Lohman *et al.* 2013b) before and after nitrate depletion or before and after bicarbonate trigger (dependent on treatment group), to examine the effect of nitrate and bicarbonate on PW95 lipid accumulation. Lyophilized biomass of PW95 was used to convert TAGs into a mixture of FAMEs or biofuel using direct *in situ* transesterification. FAMEs were quantified by gas chromatography–mass spectroscopy (GC-MS) as described previously (Lohman *et al.* 2013b; Pedersen *et al.* 2018).

Statistical Analysis

The statistical analysis was done using Minitab® Statistical Software v18 (http://www.minitab.com). A two way analysis of variance (ANOVA) using a Linear Mixed Effects Model (LMM) with a 95% significance level was fitted to all data for each response separately (log_{10}(cells/ml), pH, chlorophyll accumulation, NR fluorescence and nitrate depletion), to analyze the effects of temperature in combination with initial nitrate, pH buffer and bicarbonate trigger (fixed effects) and the interactions with time (days) and random effects (flask nested in experiment). Tukey's follow-up tests assessed all pairwise comparisons of the mean response among the treatment groups.

To determine the effect of the bicarbonate trigger, a second phase of analysis was conducted for each response separately (FAMEs, NR fluorescence, log_{10}(cells/ml), pH, chlorophyll accumulation, and nitrate depletion). An LMM, similar to that described above but with time (days) as a covariate and trigger as another fixed effect, was used for each of the 24 combinations of treatments to determine rate and mean differences in each response before and after the trigger. In cases when the linearity assumption (with respect to time) was violated, time (days) was included as a fixed effect in which case rates were not compared. All data are shown as the means and standard error of the mean (SEM) of two independent biological replicates (*i.e.*, flasks) and at least six independent experiments.

Results and Discussion

Morphological and Phylogenetic Characterization of PW95

PW95 is a unicellular green microalga. Vegetative cells of PW95 have thick cell walls, one nucleus and no flagella, as has been observed in a variety of culture conditions (Figure 2). Nile Red fluorescence can be used to demonstrate the accumulation of lipids (Figure 2B), and cells showed cell-to-cell variability. Small structures similar to cells can be observed in aggregates (Figure 2C & D) and based upon position and appearance the structures are aplanospores that are an asexual form of reproduction with thick cell walls and no flagella (Shubert 2003). Based upon microscopic observations, PW95 is a non-motile, coccoid alga that can be unicellular or colonial (aggregates) with the ability to produce spore-like structures (Figure 2E & F).

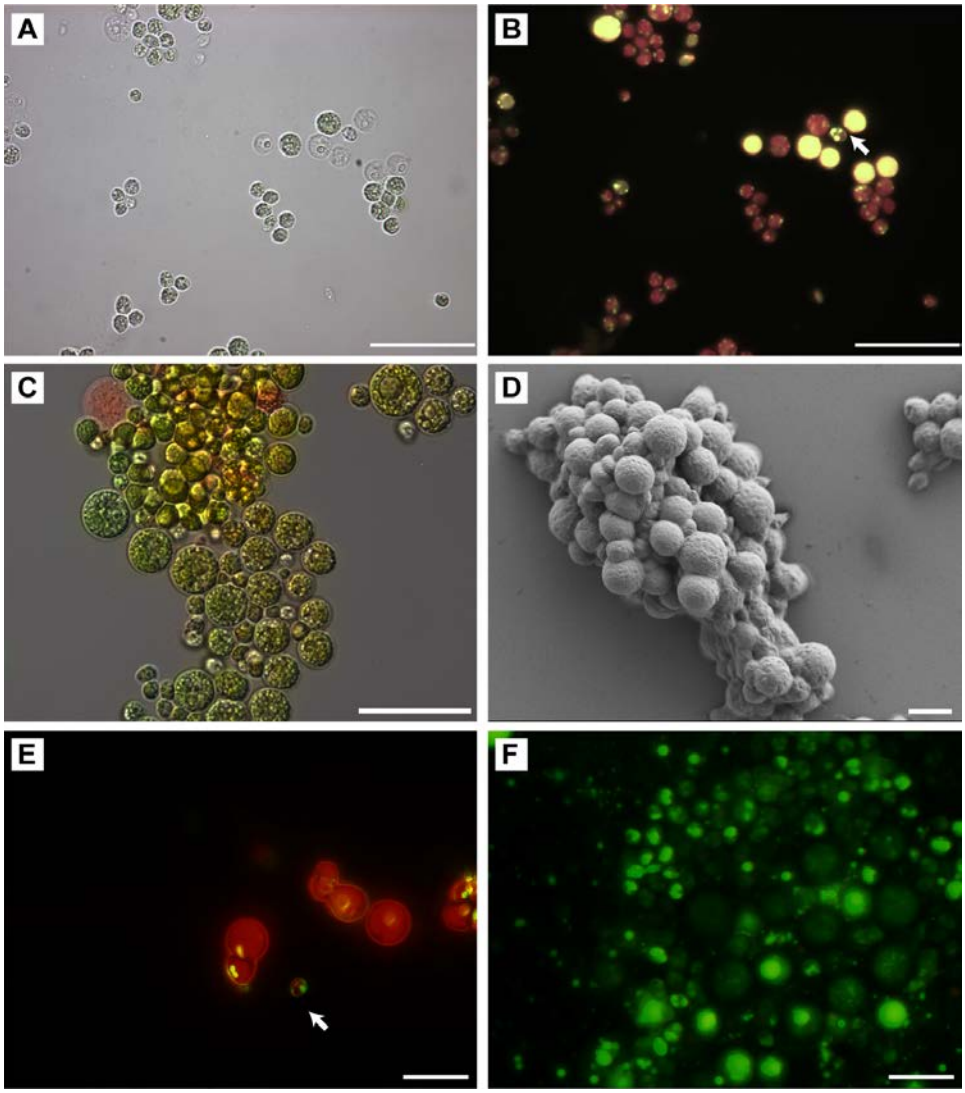


Figure 2. Morphological characterization of PW95 native alga cultured in Bold's medium. Transmitted light microscopy (A) and the corresponding epifluorescence microscopy image using Nile Red stain (B), show different levels of lipid accumulation inside the cells (shown by yellow fluorescence, white arrow). Epifluorescence microscopy (C) using Cell Mask Orange (CMO), a membrane-specific dye (green), shows vegetative mature and young cells of coccoid body shape with large nuclei and thick cell walls. Images B and C also show chlorophyll auto-florescence in red. Images A, B and C at 60x show a 35 μm scale bar. FE-SEM (D) image at a 318x magnification with 10 μm scale bar, showing a cell aggregate of PW95 and gross cell morphology. Epifluorescence microscopy using sybr green, nucleic acid specific dye (green/yellow) and chlorophyll auto-florescence in red, (E) shows vegetative mature and young cells with coccoid body accompanied by an aplanospore (white arrow), thick cell walls are observed in cells and spore at 20 μm scale. (F) Algal aggregate showing large vegetative cells and small aplanospores with different levels of sybr green intensity at 20 μm scale.

Bayesian phylogenetic inference using a 18S SSU rRNA sequence (Figure 3) and the ITS consensus sequence (Supplementary Figure A1), revealed that PW95 is closely related to Chlamydomonas-like species and members of the genus *Neospongiococcum*, *Chlamydopodium* and *Chlorococcum*, placing PW95 in the order Chlamydomonadales and grouped with non-motile coccoid green algae. Although *Chlorococcum* species show morphological similarities to PW95 vegetative cells, the SSU rRNA gene phylogenetic tree shows longer average branch lengths between *Chlorococcum* sequences and the PW95 sequence. Taxonomical reclassifications have been reported for *Neospongiococcum*, *Chlamydopodium* and *Chlorococcum* species (Floyd, Watanabe and Deason 1993; Maltsev and Konovalenko 2018) and demonstrate that algal phylogeny is complex and new isolates and sequences continue to improve the understanding of phylogenetic relatedness of eukaryotic photoautotrophs.

Figure 3. Phylogenetic relations of PW95 alga (indicated in bold) based on nuclear-encoded 18S SSU rDNA sequence comparisons. 47 sequences longer than 1000 bp were used for the analysis and the phylogenetic tree was constructed using Bayesian Inference. All Bayesian posterior probabilities are shown with colored dots on the branches. The global average standard deviation of split sequences was 0.008. *Botryococcus braunii* was chosen as outgroup (sequences shown as a collapsed clade).

PW95 Growth using Different Combinations of Temperature, Initial Nitrate Concentration, pH, Buffer Addition and Bicarbonate Trigger

Previous experiments at 35°C showed poor growth and morphological evidence of stress in PW95 (data not shown), indicating that optimal temperatures for PW95 should be in a lower range. The effects of temperature (20°C, 25°C, 30°C), initial nitrate

concentration (0.5 mM and 2 mM) and the addition of a buffer (buffer vs. no buffer) on PW95 growth and chlorophyll production were evaluated (Figures 4, 5 and 6 for 20°C, 25°C and 30°C, respectively). In an attempt to induce or optimize TAG accumulation, bicarbonate (trigger vs. no trigger) was added to different experimental batch cultures prior to nitrate depletion (Supplementary Figure A2) and compared to the unamended cultures to assess its impact on TAG accumulation (Figures 4, 5 and 6).

Cell density (Figure 4A & B, 5A & B, 6A & B) varied among treatment groups and the groups were statistically separated into group A, which showed the highest cell numbers and group B with the lowest. Group A is composed of the top performing treatments in terms of growth rate during exponential growth and maximum cell density: 30°C_0.5 Nitrate_No Buffer_No Trigger (0.8 d^{-1}, 5.30E+06 cells/ml), 30°C_2.0 Nitrate_No Buffer_No Trigger (1.38 d^{-1}, 6.54E+06 cells/ml) and 25°C_2.0 Nitrate_No Buffer_No Trigger (1.02 d^{-1}, 1.80E+06 cells/ml). Group B is composed of the treatment 25°C_0.5 Nitrate_Buffer_No Trigger (0.40 d^{-1}, 6.81E+05 cells/ml), that showed a spike in growth at day 2 but the overall performance was poor.

The treatment group 25°C_2.0 Nitrate_No Buffer_No Trigger also stands out showing a fast growth rate (1.02 d^{-1}) but the overall performance of the culture was poor, since growth declined after day 3 (Table 1, Figure 5B). The other tested growth conditions (shared grouping, A and B) had similar means with no statistically significant differences (Figures 4, 5 and 6 and Table 1). Interestingly, for both 25°C and 30°C, the increased growth rates were observed under no buffer conditions without bicarbonate trigger.

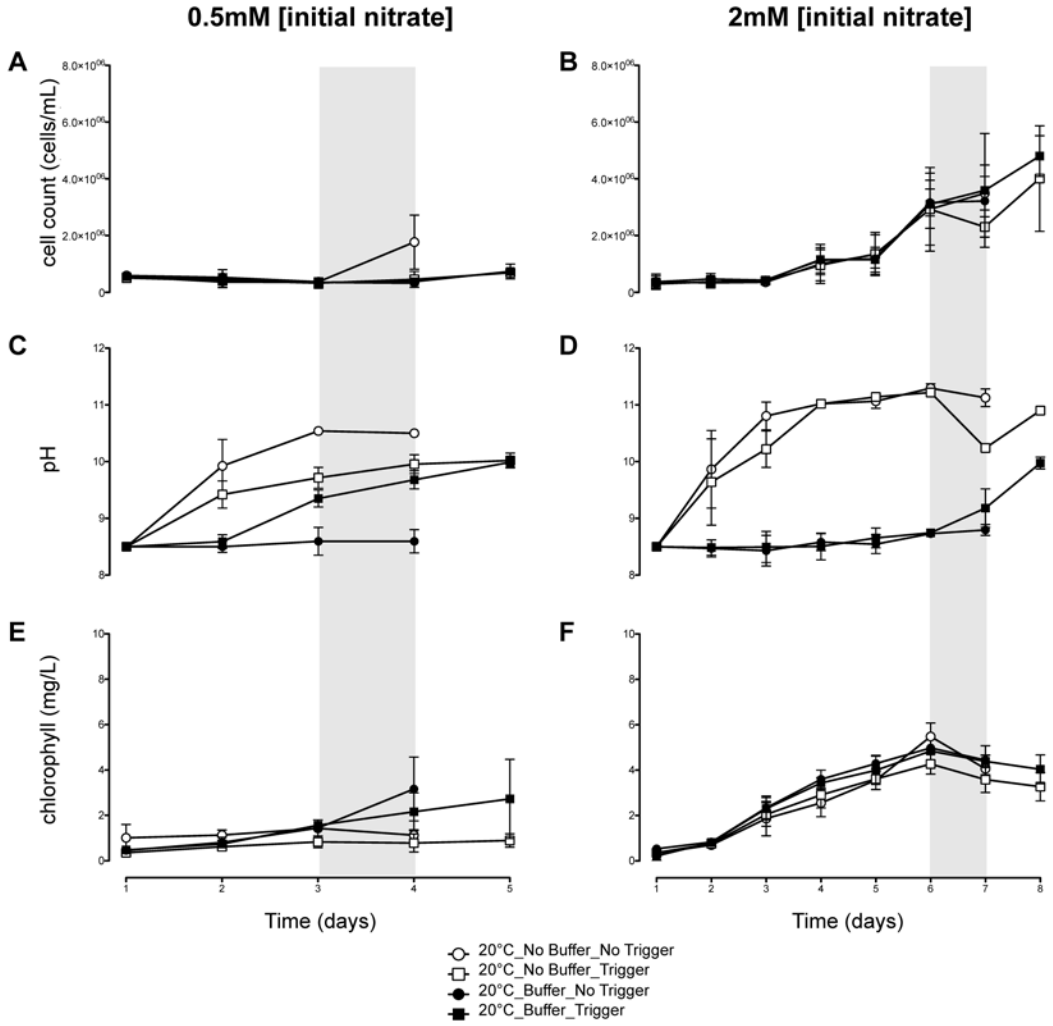


Figure 4. Comparison between the treatment groups evaluated at 20°C at two different concentrations of initial nitrate in terms of growth (cells/ml) (A & B), media pH (C & D) and chlorophyll production (E & F). Not all the cultures arrived at nitrate depletion at the same time point, so nitrate depletion is depicted as a time frame for nitrate depletion by the shaded region in all graphs. Bicarbonate trigger, if employed, was added prior to the onset of nitrate depletion. Filled symbols represent the combination of treatments that included the use of a buffer and the clear symbols represent treatments without buffer. Treatments with trigger are represented by squares and without trigger with circles. The mean of two independent replicates with the standard error of the mean (error bars) is displayed in each graph.

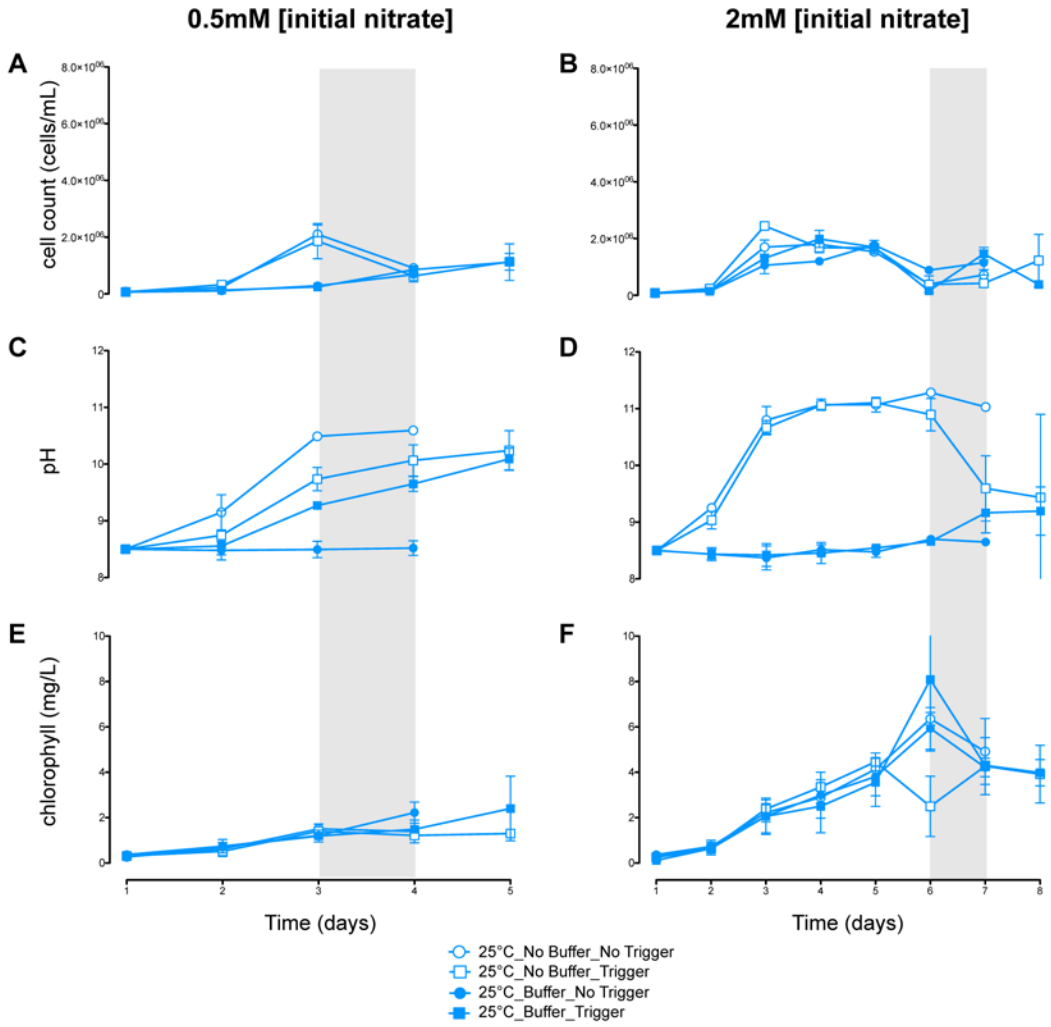


Figure 5. Comparison between the treatment groups evaluated at 25°C at two different concentrations of initial nitrate in terms of growth (cells/ml) (A & B), media pH (C & D) and chlorophyll production (E & F). Not all the cultures arrived at nitrate depletion at the same time point, so nitrate depletion is depicted as a time frame for nitrate depletion by the shaded region in all graphs. Bicarbonate trigger, if employed, was added prior to the onset of nitrate depletion. Filled symbols represent the combination of treatments that included the use of a buffer and the clear symbols represent treatments without buffer. Treatments with trigger are represented by squares and without trigger with circles. The mean of two independent replicates with the standard error of the mean (error bars) is displayed in each graph.

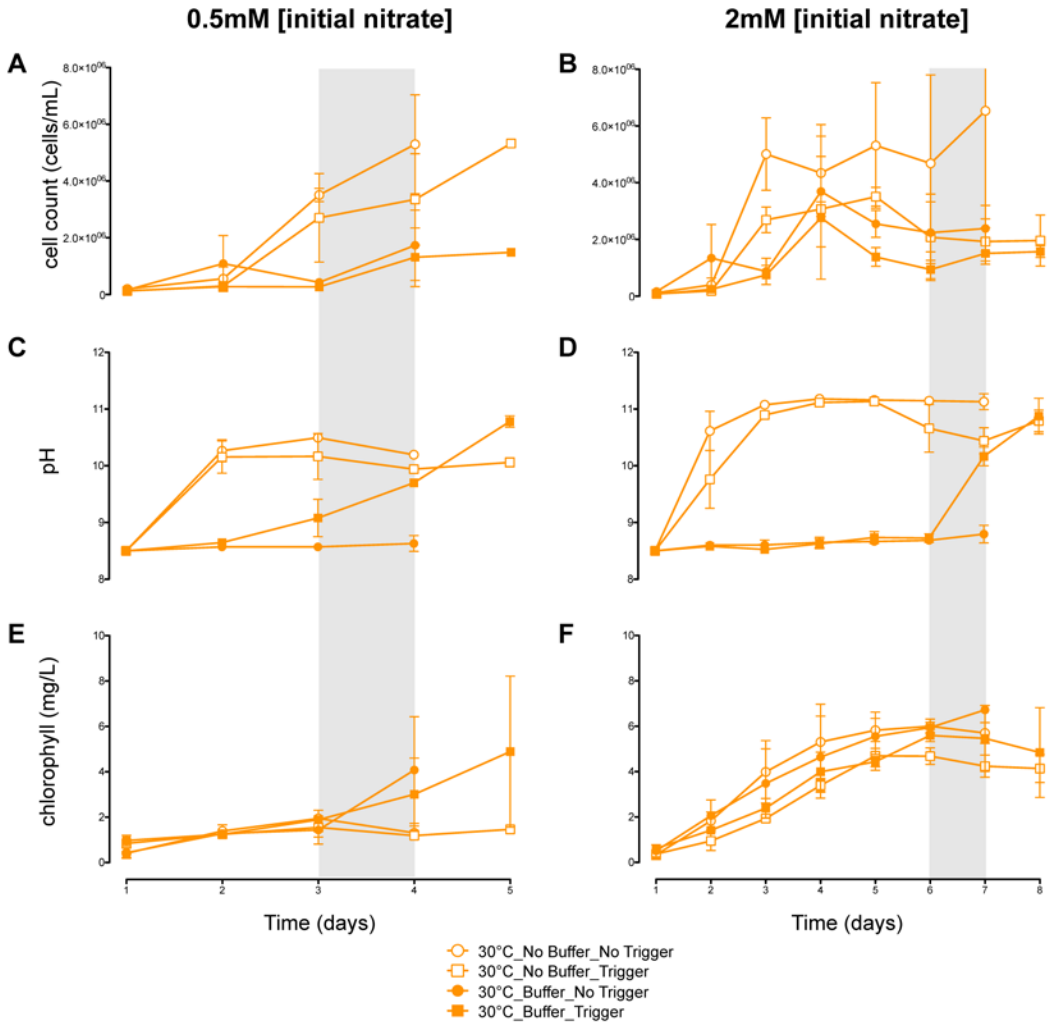


Figure 6. Comparison between the treatment groups evaluated at 30°C at two different concentrations of initial nitrate in terms of growth (cells/ml) (A & B), media pH (C & D) and chlorophyll production (E & F). Not all the cultures arrived at nitrate depletion at the same time point, so nitrate depletion is depicted as a time frame for nitrate depletion by the shaded region in all graphs. Bicarbonate trigger, if employed, was added prior to the onset of nitrate depletion. Filled symbols represent treatments that included the use of a buffer and the clear symbols represent treatments without buffer. Treatments with trigger are represented by squares and without trigger with circles. The mean of two independent replicates with the standard error of the mean (error bars) is displayed in each graph.

Table 1. Comparison of growth metrics between all the treatment groups evaluated in this study. Statistically significant ($p<0.05$) treatments according to the means of cell numbers (cell/ml) and the exponential rate of growth are highlighted in bold.

Temp (Celsius)	Initial Nitrate (mM)	pH Buffer	Bicarb Trigger	Growth rate (day -1) [a]	Doubling time (hours) [b]	Max cell number (cells/ml)	
20	**0.5**	**No**	**No**	**-0.15 ± 0.05**	**(n/a)**	**5.19E+05**	**± 8.8E+02**
			Yes	-0.09 ± 0.03	(n/a)	1.66E+05	± 8.3E+04
		Yes	**No**	**-0.18 ± 0.06**	**(n/a)**	**5.51E+05**	**± 3.4E+04**
			yes	-0.07 ± 0.02	(n/a)	7.4E+05	± 1.8E+05
	2	**No**	**No**	**0.45 ± 0.16**	**16.08**	**3.49E+06**	**± 6.7E+05**
			Yes	0.37 ± 0.13	19.59	4.01E+06	± 1.3E+06
		Yes	**No**	**0.41 ± 0.15**	**17.44**	**3.22E+06**	**± 9.0E+05**
			yes	0.43 ± 0.15	16.72	4.80E+06	± 5.1E+05
25	**0.5**	No	No	0.71 ± 0.51	10.11	2.10E+06	± 8.0E+05
			Yes	0.76 ± 0.27	9.51	1.86E+06	± 4.4E+05
		Yes	**No**	**0.40 ± 0.29**	**17.85**	**6.81E+05**	**± 9.3E+04**
			yes	0.27 ± 0.10	26.45	1.12E+06	± 4.6E+05
	2	**No**	**No**	**1.02 ± 0.36**	**7.10**	**1.80E+06**	**± 8.0E+04**
			Yes	0.43± 0.30	16.93	2.44E+06	± 5.3E+03
		Yes	**No**	**0.37 ± 0.39**	**19.54**	**1.78E+06**	**± 5.3E+04**
			yes	**0.68 ± 0.48**	**10.64**	**1.87E+06**	**± 2.1E+05**
30	**0.5**	**No**	**No**	**0.80 ± 0.28**	**9.04**	**5.30E+06**	**± 1.2E+06**
			Yes	0.96 ± 0.34	7.53	5.33E+06	± 3.5E+00
		Yes	No	0.83 ± 0.29	8.70	1.74E+06	± 8.8E+05
			yes	0.34 ± 0.12	21.50	1.31E+06	± 7.3E+05
	2	**No**	**No**	**1.38 ± 0.49**	**5.23**	**6.54E+06**	**± 2.9E+06**
			Yes	0.76 ± 0.54	9.53	3.51E+06	± 2.3E+05
		Yes	No	0.92 ± 0.32	7.88	3.69E+06	± 1.4E+06
			yes	0.57 ± 0.20	12.74	2.76E+06	± 1.5E+06

Values are reported as the mean of two independent replicates with the standard error for the growth rate and maximum cell number. n/a stands for not applicable since the growth rate has a negative value.

[a] Calculated using microbial growth rate equation (Taziki, Ahmadzadeh and A. Murry 2016).

[b] Exponential growth phase doubling time equation (Taziki, Ahmadzadeh and A. Murry 2016).

Overall, higher temperature appeared to have a positive impact on PW95 growth regardless of the amount of initial nitrate present in the media ($p<0.05$; Figures 4, 5 and 6). At 30°C, PW95 had a greater increase in cell numbers, a shorter lag phase (one day) and faster growth rates in comparison with treatment groups at 20°C and 25°C (Table 1). Han et al. 2013, observed optimum daytime temperature for *Chlorella pyrenoidosa* at 30°C for maximum biomass and lipid production when testing different levels of temperature during daytime and nighttime (Han *et al.* 2013). Temperature has been shown to influence algal growth, as it affects cellular metabolic rates, enzymatic activities, respiratory and photosynthetic electron transport and other important aspects, such as membrane fluidity and structure (Jiang and Gao 2004; Kurpan Nogueira *et al.* 2015). Nogueira et al. 2015 described higher growth rates at higher temperature when comparing growth in marine algae at 20°C and 30°C but reported no significant difference in overall cell density when exposed to different light intensity in combination with temperature (Kurpan Nogueira *et al.* 2015). A similar trend was observed in *Chlorella minutissima* under minimum irradiance, in which maximum growth rate increased from 0.12 d^{-1} at 10°C, to 0.66 d^{-1} at 30°C (Aleya, Dauta and Reynolds 2011). For PW95, higher temperature was a stimulating factor for faster growth rates and maximum cell density (Table 1). For example, in treatments with no buffer and no trigger with 0.5 mM of initial nitrate, the growth rate went from a negative value at 20°C to 0.8 d^{-1} at 30°C and from a max cell density of 5.19E+05 to 5.30E+06 cells/ml (No Buffer_No Trigger). The lower temperatures evaluated in this investigation (20°C and 25°C) did not show significant growth at 0.5 mM of nitrate (Figures 4 and 5). At 2 mM of nitrate, there

was a delayed exponential growth (starting at day 5) in all the treatments at 20°C (Figure 4B) and these results suggest an interdependence of nitrate utilization and temperature when nitrate and temperatures are lower. Optimal temperature for growth is species-specific, and PW95 accumulated cells and chlorophyll the fastest at 30°C. In addition, at 2 mM nitrate, a higher biomass yield (cells) was obtained at 30°C compared to 20°C (6.5 x 10^6 versus 3.5 x 10^6 cells/ml, respectively). In comparison, *Chlorella ellipsoidea* achieved maximum cell density at 15°C when cultivated in low salinity, even though a two-fold increase in growth rate was observed from 15°C to 25°C (Cho *et al.* 2007). As outdoor cultivation strategies are further investigated, daily temperature changes are likely during typical diurnal cycles at optimal light exposure. Further work is needed to delineate temperature optima for different algal species that may be mixed to minimize competition and maximize output per volume for different geographic regions.

Previous studies have shown that bicarbonate supplementation can enhance cell division, increase biomass and desired biochemical composition in marine algae (White *et al.* 2013) and chlorophytes such as *Scenedesmus sp.* (Pancha *et al.* 2015) and *Chlamydomonas reinhdartii* (Gardner *et al.* 2013). However, bicarbonate did not significantly affect PW95 growth and biomass accumulation under the tested growth conditions. The 2.0 mM nitrate conditions at 25°C with and without buffer did have higher cell counts (Figure 5B), but the comparison with no bicarbonate counterparts was not significantly different ($p>0.05$). The presence of bicarbonate after N depletion did not appear to impact biomass accumulation, and further work is needed to elucidate the

potential activation of carbon concentrating mechanisms in PW95 when bicarbonate is present in the media.

The pH was monitored daily to assess the changes in the media due to carbon fixation (Gardner *et al.* 2013b) and the effect of pH control with buffer addition on growth and lipid accumulation. As expected, culture pH increased for cultures with significant activity most likely due to nitrate utilization and photosynthesis (Figure 4C & D, 5C & D, 6C & D). The unbuffered culture pH increased to approximately 10 with 0.5 mM nitrate and 11 with 2.0 mM nitrate, and the buffered culture pH remained approximately 8.5 unless bicarbonate was added (Figures 4, 5, and 6). At each tested temperature, pH increased to between 9 and 10 during initial growth for both nitrate levels (0.5 and 2.0 mM) and approached approximately 10.5 as nitrate was depleted for the 0.5 mM nitrate treatments (Figure 4, 5 & 6, part C). The change in pH at 0.5 mM nitrate is similar to previous results when PW95 was grown in CBM water amended with macro- and micronutrients (Hodgskiss *et al.* 2016). The 2 mM nitrate treatments were at pH 11 as nitrate was depleted (Figure 4, 5 &6, part D), and the groups of treatments without bicarbonate trigger were significantly different ($p<0.05$) from the triggered treatments. These results demonstrate a buffering effect of bicarbonate when present as it has been used previously for pH control in algal cultures (Gardner *et al.* 2013b). The addition of bicarbonate was pre-nitrate depletion (shaded area in Figures 4, 5, and 6), and significant differences ($p<0.05$) among treatment groups with bicarbonate trigger were observed with a pH decline for unbuffered and a pH increase for the buffered conditions (Figure 4D, 5D, and 6D).

Gardner et al. 2013 (Gardner *et al.* 2013b) and Pedersen et al. 2018 (Pedersen *et al.* 2018), reported pH control as an effective strategy to maintain optimal growth of algal biomass during nitrogen replete conditions, but PW95 exhibited a different response. As a general trend, pH control using buffer (Trizma) addition to the media had an inhibitory effect on PW95 growth for most of the tested treatments combinations, and these results corresponded to a similar response reported by Mus et al. when *P. tricornutum* cultures were buffered with TRIS at pH 8.5 and 9.0 (Mus *et al.* 2013). The effect of bicarbonate addition during nitrogen depletion on buffered cultures has not been previously described, and the presented results suggest that buffer conditions could temper any possible benefit of bicarbonate addition. However, the response might be genus-, species- and even strain-dependent and further work is needed to better characterize carbon-concentrating mechanisms in different photoautotrophs.

Chlorophyll and biomass production (4E & F, 5E & F, 6E & F) are highly dependent on external nitrogen availability, and chlorophyll accumulation has been used previously to monitor culture health and as an alternative method for measuring biomass accumulation in different algae (Moll *et al.* 2014). In addition, rapid declines in chlorophyll have been associated with nitrogen starvation (Mus *et al.* 2013). As expected, PW95 had a higher chlorophyll production in cultures with 2 mM initial nitrate (2-6 mg/l chlorophyll) in comparison with the 0.5 mM cultures (1-4 mg/l chlorophyll), and similar results have been reported for PW95 in CBM production water (Hodgskiss *et al.* 2016) as well as for other algal cultures (Li *et al.* 2008; Valenzuela *et al.* 2012; Mus *et al.* 2013; Zhao *et al.* 2017). At 0.5 mM nitrate, the 30°C cultures had slightly higher chlorophyll

levels than the cultures at 20° and 25°C (Figure 6E vs. 4E & 5E). In addition, there was no statistical significance (*p>0.05*) that could be associated between differences in growth and chlorophyll levels (Figures 6B vs 6F), and these results suggested that there is not a detectable correlation between overall biomass and chlorophyll as has been previously reported for other alga (Ramaraj, Tsai and Chen 2013). Moreover, there was an inverse relationship between cell density and chlorophyll production, indicating that chlorophyll was not a good index for growth and cell production for PW95 under the tested conditions. For example, the two treatment groups with the best growth performance at 0.5 mM nitrate (30°C_0.5 Nitrate_No Buffer_No Trigger and 30°C_0.5 Nitrate_No Buffer_Trigger) had almost very low levels of chlorophyll production whereas the highest chlorophyll was observed in the buffered treatments (with or without trigger) but much less growth. The use of chlorophyll as an intracellular nitrogen pool can occur during N deprivation and has been previously reported in *Neochloris oleoabundans* (Li *et al.* 2008), *Phaeodactylum tricornutum* (Valenzuela *et al.* 2012) and in plants where chlorophyll degradation pathways have been discovered and are considered vital for leaf senescence and fruit ripening (Eckardt 2009). Additionally, low chlorophyll levels can be associated with an increase in antioxidant enzymes as a possible response to reactive oxygen species (ROS) after prolonged nitrogen starvation, as it was reported for *Chlorella sorokiniana* cultures (Zhang *et al.* 2013), but the physiological responses to oxidative stress during photosynthesis in PW95 are not known. At low initial nitrate (0.5 mM), an increase in chlorophyll production started during nitrate depletion (Day 3) for treatment groups with added buffer across temperatures, regardless of the

presence or absence of bicarbonate (Figures 4, 5 and 6). Although, this change was not found to be statistically significant ($p>0.05$), it can be argued that resources were being allocated as a reserve or 'persistent state' in response to the nitrogen (N) depletion perturbation. Green algae, such as *Chlamydomonas* and *Volvox*, produce morphologically distinct spores for this purpose (Ellegaard and Ribeiro 2018). PW95 is an environmental isolate that could possibly have a sporulating resting stage triggered by specific environmental conditions, such as, pH level and nutrient starvation. Further work is needed to better understand the role of spores in the life cycle of PW95 and how different sporulating green alga overall respond and survive nutrient deprivation in terms of biomass accumulation.

Lipid analysis and effect of sodium bicarbonate trigger on PW95 lipid production

Lipid accumulation in algae is frequently measured by monitoring TAG accumulation using Nile Red fluorescence or using direct transesterification of TAGs into FAMEs for GC-MS analysis to account for biofuel potential (Lohman *et al.* 2013a). NR fluorescence was used as a proxy for TAG accumulation under the tested growth conditions (Figure 7).

The start of significant TAG accumulation has been frequently correlated with the onset of nitrate depletion in green algae (Gardner *et al.* 2012; Eustance *et al.* 2013), diatoms (Valenzuela *et al.* 2012) and PW95 in amended low quality water (Hodgskiss *et al.* 2016). An increased and sustained production of lipids has been reported under nitrogen limitation, different temperatures, drastic fluctuations in pH and different combinations of these conditions (Moll *et al.* 2014; Xia *et al.* 2014). Previous studies

have shown that the addition of different concentrations of bicarbonate before nitrate depletion, can stimulate or increase lipid accumulation in some species of green algae and diatoms (White *et al.*, 2013; Gardner *et al.* 2013a; Moll *et al.* 2014). An increase in TAG accumulation and FAMEs was expected with nitrate depletion and with the addition of bicarbonate to PW95 cultures before the onset of nitrate depletion.

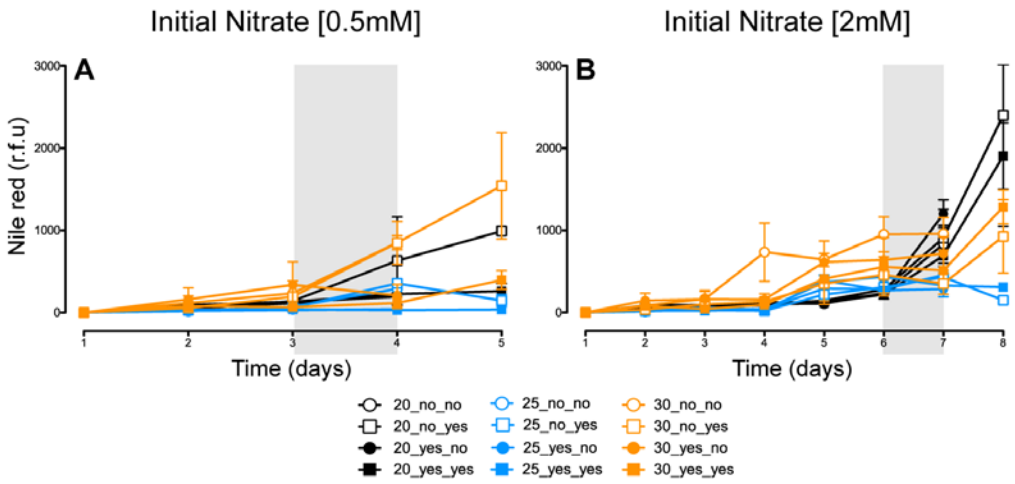


Figure 7. Daily lipid accumulation estimated by NR fluorescence during growth of all the treatment groups evaluated in this study at 0.5mM (A) or 2mM (B) initial nitrate concentration. Temperatures are differentiated by color: 20°C (black), 25°C (blue) and 30°C (orange). Filled symbols represent the treatments that included the use of a buffer and the clear symbols represent treatments without buffer. Treatments with trigger are represented by squares and without trigger with circles. The mean of two independent replicates with the standard error of the mean (error bars) is displayed in each graph.

Despite lower cell accumulation at 0.5 mM, an increase in NR fluorescence that started at the onset of nitrate depletion, and showed a maximum increase in fluorescence in the last day of the experiments curve for the following treatment groups: 30°C_0.5 Nitrate_No Buffer_No Trigger (800 rfu), 30°C_0.5 Nitrae_No Buffer_Trigger (1500 rfu)

and 20°C_0.5 Nitrate_No Buffer_ Trigger (1000 rfu). At 30°C both treatment groups showed a similar but not statistically significant trend of increase ($p>0.05$). In contrast, the 20°C treatment was statistically significant ($p<0.05$) and the increase in NR fluorescence could be attributed to the bicarbonate addition under unbuffered conditions (Figure 7A).

At 2 mM, before nitrate depletion (days 3–6) there was a slight increase in NR fluorescence as a function of time at 25°C and a significant increase of NR at 30°C (Figure 7B). The treatment group 30°C_2.0 Nitrate_No Buffer_No Trigger showed an increase in NR fluorescence during nitrate replete conditions but was not statistically significant ($p>0.05$). After nitrate depletion (day 6), there was an increase in NR fluorescence in all the treatment groups at 20°C and the triggered treatments at 30°C (Figure 7). These results indicate that there could be a relationship between temperature, nitrate limitation and possibly bicarbonate trigger with an increase of TAG accumulation in PW95. At 0.5 mM initial nitrate, the addition of buffer had no effect on TAG accumulation, in contrast to bicarbonate addition that appeared to increase TAG accumulation at 20°C and 30°C. On the other hand, treatment groups at 2 mM initial nitrate at 20°C and 30°C showed sudden increases in TAG accumulation in the last two days of the growth curve for the conditions that had both buffer and bicarbonate but the differences were not statistically significant ($p>0.05$) (Figure 7).

Direct *in situ* transesterification was used to obtain all fatty acids in the biomass and accurately represent the biofuel potential with no requirement of prior lipid extraction (Chen *et al.* 2018). The total FAME levels before and after nitrate depletion or before and

after trigger were determined for the different treatments (Figure 8A). A comparison of total FAMEs across temperatures and between the two concentrations of initial nitrate for all treatments showed that the highest total FAMEs content (~30%) was observed at 20°C without buffer or trigger at 0.5 mM nitrate (Figure 8). The lowest total FAME levels were observed for treatments at 25°C. When nitrate was increased to 2.0 mM, all treatments were similar in total FAME levels, but the levels were generally below 20% (Figure 8B). Among the treatment groups with the highest amount of lipids, *i.e.*, 20–30% (20°C_0.5 Nitrate_No Buffer_No trigger, 20°C_0.5 Nitrate_No Buffer_Trigger, 30°C_0.5 Nitrate_No Buffer_No trigger and 30°C_0.5 Nitrate_No Buffer_Trigger), the percentage of FAMEs was always higher before nitrate depletion.

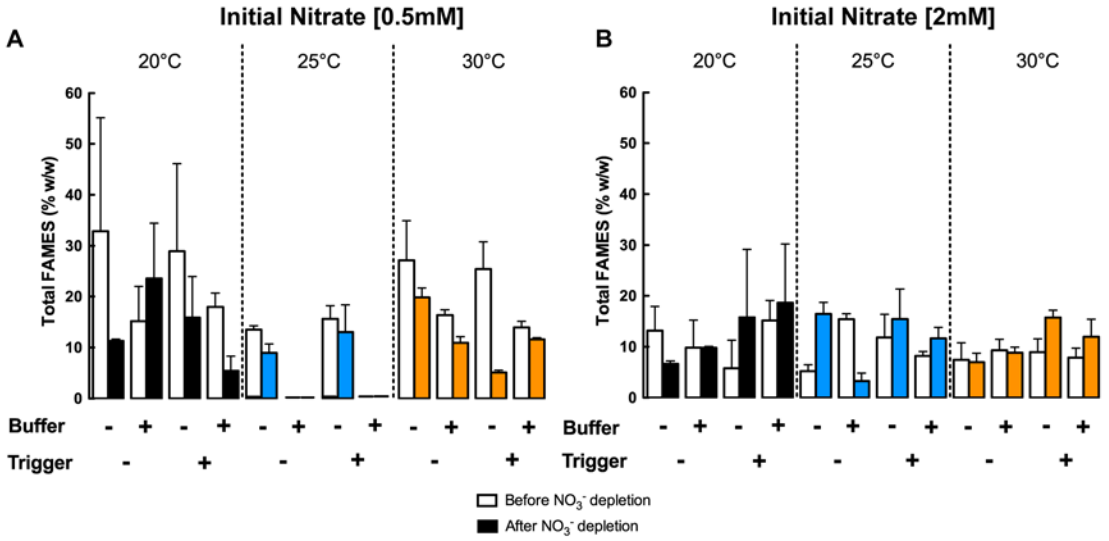


Figure 8. Percentage of total FAMEs before and after nitrate depletion and bicarbonate trigger. The concentration of nitrate in the samples was measured daily to predict the time of nitrate depletion for biomass collection and subsequent transesterification. For those treatments were bicarbonate trigger was added before the onset of nitrate depletion, biomass collection was done before the trigger and after the trigger for transesterification as well. The mean of two independent replicates with the standard error of the mean (error bars) is displayed in each graph.

These results indicate that in those treatments with bicarbonate trigger the percentage of FAMEs is lower after nitrate depletion/bicarbonate trigger. The bicarbonate trigger had no effect on lipid accumulation in PW95 and did not result in statistically significant differences ($p>0.05$) in any of the treatment groups. These results may suggest different carbon concentrating mechanism or that the typical biochemical and/or biophysical mechanisms operate uniquely in PW95. There was a strong correlation between high lipid production and low biomass productivity at 20°C and 0.5 mM initial nitrate. Cell density for the treatment groups 20°C_0.5 Nitrate_No Buffer_No trigger and 20°C_0.5 Nitrate_No Buffer_Trigger was low (Figure 4A) but they showed the highest percentage of total FAMEs, 32.83% and 28.95%, respectively (Figure 8A). These results suggest that cell replication (*i.e.*, biomass accumulation) was inhibited for the production of energy storage molecules and an inverse relationship exists between biomass and lipid accumulation that has been reported frequently in the literature (Laurens, Chen-Glasser and McMillan 2017). In contrast, at 30°C and 0.5 mM of nitrate there was also high lipid production in the treatments 30°C_0.5 Nitrate_No Buffer_No trigger (27.15%) and 30°C_0.5 Nitrate_No Buffer_Trigger (25.43%; Figure 8A). Both of these treatments showed the highest cell density during growth (Figures 4, 5 and 6), suggesting that the interaction between a low concentration of initial nitrate and higher temperature also play a role in biomass and lipid production in PW95.

At lower temperature, there is an inverse relationship between biomass and lipid accumulation, whereas at a higher temperature it can have an additive effect, resulting in more biomass and lipid accumulation, which is promising for biofuel applications. Some

of these treatment groups had bicarbonate addition (20°C_0.5 Nitrate_No buffer_Trigger and 30°C_0.5 Nitrate_No Buffer_Trigger) but the percentage of total FAMEs was significantly lower after the trigger, and these results indicate that the addition of bicarbonate had a negative impact on the total FAMEs detected in PW95 cultures that started with low nitrate concentration (0.5 mM). When paired with controlled pH strategies, the addition of a high concentration (50mM) of bicarbonate, such as the one used in this study, improved lipid content in *Nannochloropsis gaditana* (Pedersen *et al.* 2018). At higher nitrate (2 mM), none of the tested conditions had total FAMEs above 20% (Figure 8B) and total fatty acid contents in other green algae and diatoms range between 10 and 60% (Hoekman *et al.* 2012; Breuer *et al.* 2013; Lohman *et al.* 2015). In addition, the differences before and after the bicarbonate trigger were not statistically significant ($p>0.05$). Moreover, when total FAME levels were compared to the rate of cell or chlorophyll accumulation, chlorophyll accumulation rate was a better predictor of overall FAME content under unbuffered conditions (R=0.82). These results suggest an important interaction between low nitrate levels, temperature, and elevated pH for trade-offs between biomass and lipid production in PW95, and that chlorophyll accumulation may be an indicator of resource allocation during C:N imbalance (*i.e.*, N storage into chlorophyll as a pre-requisite to lipid accumulation during surplus C).

In this study, when the PW95 lipid profiles were compared between 20°C and 30°C for the 0.5 mM treatments before and after nitrate depletion or bicarbonate trigger, the dominant species of FAMEs were saturated and unsaturated C16 (hexadecanoic) and C18 (octadecanoic) FAMEs (Figure 9). Hodgskiss et al. determined that PW95 when

grown in CBM water amended with 0.5 mM of initial nitrate produced a higher amount of C:18 unsaturated acids and lower amounts of C:16 FAMEs in comparison with other reported algae (Hodgskiss *et al.* 2016). These results indicate that the FAME profile of PW95 can be manipulated with changes in culture conditions. Moreover, in comparison with other oleaginous algae, that contain significant amounts of lighter (C12–C14) and heavier (C20–C22) species, PW95 TAGs appear to be dominated by C16 and C18 fatty acids that are more common to vegetable oil fatty acids profiles from conventional biofuel producers (Hoekman *et al.* 2012). These results suggest that PW95 could be a potential source of C18/C16 feedstock.

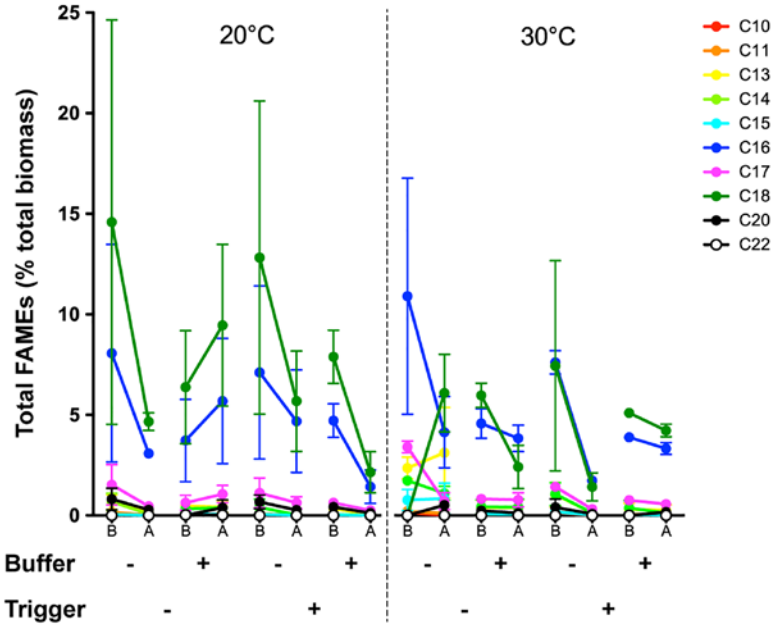


Figure 9. PW95 FAMEs profile comparison between 20°C and 30°C for treatments at 0.5 mM. The x-axis labels make a distinction between FAMEs before (designated B) and after nitrate depletion and bicarbonate trigger (designated A). Treatments that included the use of a buffer and the bicarbonate trigger are represented by the plus sign (+) and controls are represented by the negative sign (-). Bicarbonate trigger, if employed, was added prior to the onset of nitrate depletion. Error bars represent the standard error of the mean of two independent replicates.

Conclusions

PW95 is a novel, sporulating green alga isolated from CBM production water that has the capacity to accumulate lipids at different temperatures and culture conditions. Multivariate physiological analysis has proven useful to determine the optimal combination of temperature, nitrate levels, buffer, and bicarbonate treatments for PW95 growth and lipid production. As previously reported, PW95 accumulates lipids in amended low quality water at 20°C and the present study shows that PW95 can accumulate lipids, at 20°C and 30°C in standardized media with low concentration of initial nitrate (6X less than standard BBM), no buffer or bicarbonate addition. The bicarbonate trigger had a negative influence on lipid production for the treatments with the highest FAMEs percentages indicating that the success of this type of strategy to increase lipid accumulation is species-specific. In this case, the combination of stress conditions (nitrate depletion, buffer and bicarbonate addition) resulted in lower growth rates and biomass productivity. PW95 grew and produced lipids with minimal nitrate concentrations and under variable temperatures. FAMEs profiles can be varied by different culture conditions, but PW95 produced predominantly the FAMEs species optimal for biofuel applications (saturated and unsaturated C16 and C18).

Acknowledgments

The authors would like to thank the MSU Algal Research group for helpful discussion, collaboration and support. Also, the Center for Biofilm Engineering staff: Betsy Pitts for microscopy assistance. Ann Willis, Dr. Kristen Brileya and Sara

Altenburg for instrumentation and logistic assistance. Jill Story for graphics assistance and Dr. Ben Wheaton for editorial and graphics assistance. The authors would like to acknowledge the U.S. Geological Survey's Energy Resources Program which provided access to the field site where the water samples were collected for PW95 isolation. Funding was provided by the National Science Foundation under NSFCHE-1230632 and the Department of Energy under BETO. Partial support for L. Corredor was provided by the Fulbright Scholarship Program.

Disclaimer: All data generated or analyzed during this study are included in the main text or supplementary information of this publication. Any opinions, findings, conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the Department of Energy (DOE) or the National Science Foundation (NSF). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CHAPTER THREE

GENOMIC INSIGHTS INTO A SPORULATING, NON-MOTILE, OLIGOTROPHIC GREEN MICROALGA (PW95)

Contribution of Authors and Co-Authors

Manuscript in Chapter Three

Author: Luisa Corredor
Contributions: Experimental design, performed experiments, literature review, synthesized ideas, performed data and bioinformatical analyses, wrote and edited the manuscript.

Co-Author: Thiru Ramaraj
Contributions: Performed bioinformatical analyses and edited the manuscript.

Co-Author: Huyen Bui
Contributions: Performed bioinformatical analyses and edited the manuscript.

Co-Author: Mensur Dlakic
Contributions: Performed bioinformatical analyses and edited the manuscript.

Co-Author: Robin Gerlach
Contributions: Procured funding, synthesized ideas, reviewed and edited the manuscript

Co-Author: Matthew W. Fields
Contributions: Principal Investigator, procured funding, experimental design, synthesized ideas, wrote and edited the manuscript

Manuscript Information Page

Luisa Corredor, Thiru Ramaraj, Huyen Bui, Mensur Dlakic, Robin Gerlach, Matthew W. Fields

Status of Manuscript:
__X__Prepared for submission to a peer-reviewed journal
____Officially submitted to a peer-reviewed journal
____Accepted by a peer-reviewed journal
____Published in a peer-reviewed journal

Abstract

PW95 is a novel green microalga isolated from a coal bed methane (CBM) production water impoundment and is closely related to *Chlamydomonas reinhardtii*. PW95 is a sporulating alga that grows and accumulates lipids in the form of TAGs in a variety of physiological conditions in standardized media and CBM water. Due to its growth and lipid generation potential for biofuel in conjunction with the ability to grow in low quality CBM water, PW95 strain was selected for whole genome sequencing, assembly and annotation. Assembly of PacBio sequence reads with CANU resulted in a draft genome size of 92 Mbp. *Ab initio* gene prediction estimated between 10,000 and 14,000 genes and a total of 401 pathways and 58 modules in PW95 were mapped in the KEGG database. As expected for an oligotrophic alga, the PW95 gene complement contained predicted genes involved in nutrient acquisition through numerous membrane transporters, heat and cold resistance with temperature shock proteins, an alternative way to maintain cell structure and membrane fluidity via a novel squalene epoxidase for steroid biosynthesis, sporulation for persistence and survival and the absence of flagella. The investigation into the PW95 genome provides insight into the biological and biochemical traits of an environmental alga with the ability to grow in low quality water from the energy sector and offers a unique opportunity for an integrated approach to water resource management, energy production, and carbon dioxide utilization.

Introduction

Microalgae are a highly ubiquitous and diverse group of photosynthetic organisms with a complex evolutionary history that are critical for maintaining terrestrial and atmospheric conditions, as their high photosynthetic and carbon fixation capacity make them important contributors to the carbon and oxygen cycles (Grossman 2005; Tirichine and Bowler 2011; Kim *et al.* 2014). Microalgae are also potential feedstocks of sustainable bioproducts, due to the ability to accumulate storage compounds such as lipids, starch and pigments under different environmental and/or growth conditions and cultivations systems (Fields *et al.* 2014). Therefore, in addition to serving as carbon storage compounds and performing crucial biological functions, algae have commercial value that can be coupled to the recycling of low-quality water and nutrients (Grossman 2005; Fields *et al.* 2014).

The widespread use of genomic techniques has led to the development of algal genomic models that have provided insights into eco-physiological adaptation in environments, genome evolution, origin of multicellularity, the history of photosynthetic life on earth and even key animal functions such as the structure, assembly and function of eukaryotic flagella (Merchant *et al.* 2007; Tirichine and Bowler 2011; Kim *et al.* 2014). Well established algal genomic models such as *Chlamydomonas reinhardtii* (model green algae) and *Phaeodactylum tricornutum* (model diatom), are part of a list of 40 complete genomes publicly available and more than a hundred genome projects are in progress, including updates on existing versions of model organism genomes (JGI genome portal; https://genome.jgi.doe.gov/portal/).

Whole genome sequencing, assembly and annotation have been frequently employed to understand/predict the biochemical capacities and responses of algal strains with potential to produce industry valued-products, such as biofuel (Hovde *et al.* 2018). We have sequenced, assembled and annotated the genome of a novel sporulating, non-motile green microalga (PW95) isolated from coal bed methane (CBM) production water; due to the variety of physiological responses observed during changing cultivation conditions in standardized media and low quality water and its capacity to accumulate lipids for biofuel production (Corredor *et al.*, in review; Hodgskiss *et al.* 2016). PW95 has the potential to be a feedstock for biofuel (25–33% total FAMEs) while rapidly producing biomass in low quality water, offering a unique opportunity for an integrated approach to water resources management and energy production.

Results from this study shed light onto the genomic plasticity of a green microalga isolated from an oligotrophic environment and provide evidence of the genetic signatures for environmental responses, fundamental and unique biochemical capabilities and biological functions in a potential microalgal production strain.

Materials and Methods

Stock Culture Conditions and Biomass Collection for DNA Extraction

Monocultures of PW95 were grown and maintained in 500 mL of Bold's basal medium (BBM) at pH 8.4, in conical flasks (1000 mL) using rubber stoppers that allowed gas exchange. Cultures were kept in a light incubator at 20°C on a 14:10 light/dark (L/D) cycle (7872 lx, cool white fluorescent) and constant shaking (125 RPM).

To ensure clean, bacteria-free algal genomic DNA samples, antibiotic treatment was used to remove potential bacteria. Bacterial contamination was identified and eradicated using the following steps. DNA was extracted directly from stock cultures using a FastDNA Spin Kit for Soil (MP Biomedical) and the bacterial SSU rRNA gene was subsequently amplified using universal bacterial primers (8F and 529R) with Illumina overhang adapter sequences (Illumina, San Diego, CA, USA) in the following design: 8F (5′-Illumina forward overhang adaptor sequence-AGAGTTTGATCCTGGCTCAG-3′) and 529r (5′-Illumina reverse overhang adaptor sequence-CGCGGCTGCTGGCAC-3′) primer pairs (synthesized by Integrated DNA Technologies (Coralville, Iowa, USA)). The PCR amplicons were prepared for sequencing according to the "16S Metagenomics Sequencing Library Preparation" Illumina protocol for paired-end sequencing (Illumina, San Diego, CA, USA). DNA from each library was quantified with a fluorometric kit (QuantIT PicoGreen; Invitrogen), normalized and mixed with Illumina-generated PhiX control library (10% PhiX) and sequencing was carried out using an Illumina MiSeq v3 platform. The sequence reads were processed and identified with the MiSeq SOP pipeline of the MOTHUR software package (Kozich *et al.* 2013).

Stock cultures were then subjected to targeted antibiotic therapy (Ampicillin 0.5 mg/mL, Sigma) against the specific bacterial species (OTU) contaminating the culture. Antibiotic-treated cultures were confirmed to be bacteria-free using R2A agar plates incubated in the dark (for heterotrophic growth) and these bacteria-free, antibiotic-treated cultures of PW95 were used as the stock for DNA extraction, as follows.

A sample of culture (100 mL) was harvested for high molecular weight (HMW) genomic DNA extraction and biomass was concentrated via centrifugation at 4800 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA). Algal biomass pellets were washed with 100 mL of ultrapure (18.2 Ω) deionized water (dH_2O), frozen with liquid nitrogen and stored at -80°C until DNA extraction.

High Molecular Weight DNA Extraction

Genomic DNA was extracted using a customized high molecular weight (HMW) DNA extraction protocol for PW95. The homogenization step consisted of the careful grinding of algal pellets (twice) in liquid nitrogen in a pre-chilled mortar and pestle until a fine powder was obtained before resuspending in 1 mL of protoplasting stock solution, composed of 1 M sodium citrate, 2 M Sorbitol, 2 M EDTA (Fisher Scientific) (Lautenberger and Chen 1987), 14.3 M (2X) b-Mercaptoethanol (Sigma Aldrich) (Lever *et al.* 2015) and 4% Driselase (Sigma Aldrich) (Ramamoorthy *et al.* 2015). The cell suspension was incubated overnight at 37°C and constantly shaken at 120 rpm. After incubation, 1X volume of lysis buffer NDS (EDTA 0.5 M, pH 9.5 and lauroyl sarcosine sodium salt 1% (Sigma Aldrich) was added to the cell suspension and incubated overnight at 50°C and constantly shaken at 120 rpm. After incubation, potassium acetate (EMD Millipore Sigma, USA) was added at a final concentration of 0.2 mM (Dos Reis Falcão *et al.* 2008) and the cell suspension was incubated at room temperature for 1h, shaking at 120 rpm. The cell suspension was centrifuged at 4000 x g for 15 min, the supernatant was discarded, and the pellet was washed in 1 mL of ultrapure (18.2 Ω) dH_2O, centrifuged again at 4000 x g for 5 min and the supernatant was discarded. The

pellet was resuspended in solution BF1 from the DNeasy PowerBiofilm kit (Cat No./ID: 24000-50, Qiagen, USA) and carefully transferred to a clean 2 mL collection tube before DNA extraction and purification proceeded as per manufacturer's instructions. The recommended bead-beating step in the kit was not performed. Sample QC was performed using a Nanodrop One Microvolume UV-Vis Spectrophotometer (ThermoFisher, USA).

Genome Sequence and Assembly

The HMW genomic DNA was sequenced via PacBio at the National Center for Genome Resource (NCGR, Santa Fe, NM, USA) for long-read single-molecule sequencing. DNA libraries were constructed using the standard PacBio long insert library protocol (Pacific Biosciences, Menlo Park, CA) (Korlach *et al.* 2010) and sequenced on the RS II platform. Two rounds of PacBio sequencing were performed to improve genome coverage and the resulting data from both sequencing runs was used for *de novo* assembly of the long reads in CANU v.1.8 (Koren *et al.* 2017) genome assembler using default parameters (Table 1). CANU performs the assembly in three steps: correction, trimming and assembly. Reads are trimmed to high-quality sequences, removing regions such as Single Molecule Real Time (SMRT) adapters. Reads were assembled into contigs to generate consensus sequences to create alternate paths to scaffolding.

Assessment of the genome assembly completeness and quality was performed with Benchmarking Universal Orthologs (BUSCO v3) (Simão *et al.* 2015). The lineage dataset chlorophyta_odb10 (number of BUSCOs: 2168; https://busco.ezlab.org/) was used to run PW95 assembly in genome mode.

Transcriptome Sequencing and Assembly

Total RNA was extracted using a customized RNA extraction protocol for PW95 (Chapter 4). Library preparation, quality control and sequencing were performed at Genewiz (NJ, USA). RNA quality was determined using a 2100 Bioanalyzer system (Agilent Technologies, USA) and RNA library was constructed for mRNA sequencing via polyA selection. Following Illumina's standard protocols, HiSeq paired-end (2x150), high-output sequencing was performed. Paired-end reads for all samples were assembled using rnaSPAdes v3.12.0 (Bankevich *et al.* 2012) with *k*-mer value set to 105 base pairs and the remaining parameters set to default values.

Genome annotation

The draft genome assembly from CANU was annotated using Genome Sequence Annotation Server (GenSAS, v 6.0, https://www.gensas.org/) (Humann *et al.* 2019), an online integrated genome sequence annotation pipeline that provided structural and functional gene model predictions. The GenSAS platform contains three main modules: (i) Identifying Repeats and Masking; (ii) Structural Annotation; and (iii) Functional annotation. RepeatMasker (Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-3.0;* http://www.repeatmasker.org) and RepeatModeler (Smit, AFA, Hubley, R. *RepeatModeler Open-1.0*; http://www.repeatmasker.org) were used to identify and mask repeats *de novo* in the CANU genome assembly, before using theGenSAS "Masked Consensus" tool (Lee *et al.* 2011) that combines RepeatMasker and RepeatModeler results into a masked consensus genome assembly. Followed by repeat masking GenSAS uses a combination of tools, AUGUSTUS (evidence model *Chlamydomonas reinhardtii*)

(Stanke *et al.* 2008), GeneMarkES (Lomsadze, Burns and Borodovsky 2014), Genscan (Burge and Karlin 1997), GlimmerM (Pertea and Salzberg 2003) and SNAP (Li *et al.* 2007) for predicting features from the masked genome assembly. After prediction of gene models and other features, EVidenceModeler (EVM) (Haas *et al.* 2008) combines the *ab initio* gene predictions and protein and transcript alignments into weighted consensus gene models. Functional annotation analysis assigning function and gene product names to the consensus gene structures were performed by BLAST protein-versus-protein (blastp) (Altschul *et al.* 1990), which finds regions of local similarity between sequences against the plant reference sequence databases, InterProScan (Zdobnov and Apweiler 2001) and Pfam (Korlach *et al.* 2010). As a final step, GenSAS uses its built-in "Publish" tool (Lee *et al.* 2011) that creates final annotation results in GFF3 and FASTA format for public release.

Phylogenetic characterization of PW95: Concatenated Protein Tree

To perform a robust and informative phylogenetic analysis of PW95, a concatenated protein tree was constructed using a modified ezTree pipeline (Wu 2018). Protein sequences in fasta format from 17 publicly available algal genomes were downloaded from the JGI (https://genome.jgi.doe.gov) and the NCBI (ftp://ftp.ncbi.nlm.nih.gov/genomes/) genome portals. The pipeline was modified to skip the gene prediction step, and automatic identification of single-copy marker proteins was performed including PW95. Pfam v32.0 (El-Gebali *et al.* 2019) was used as a reference database to identify single-copy marker proteins shared by all genomes (Supplementary Table B1). These proteins were aligned individually with MAFFT-L-INS-i (Katoh *et al.*

2002), trimmed by trimAL (Capella-Gutiérrez, Silla-Martínez and Gabaldón 2009) using a 50% gap threshold, and concatenated to generate the final alignment file. IQ-TREE (Nguyen *et al.* 2015) was used to generate a maximum likelihood (ML) tree (100 replicates in ultrafast bootstrap approximation mode; Supplementary Figure B1), and its ModelFinder found the following model parameters to be optimal: LG substitution model with empirical residue frequencies, gamma-shaped rate variation (4 categories) with a proportion of invariable sites (LG+F+I+G4). Additional Bayesian phylogenetic inference (Figure 1) was performed with MrBayes v3.2.5 (Ronquist *et al.* 2012) using the same model parameters as in ML analysis (LG+F+I+G4), and with following run parameters: 2 parallel runs with 8 chains, 1 million generations and 0.25 burn-in fraction. The two runs achieved excellent convergence based on the potential scale reduction factors (none larger than 1.001) and the average standard deviation of split frequencies (<0.001). iTOL v4.4.2 (Interactive Tree of Life, https://itol.embl.de) was used for tree editing and annotation (Letunic and Bork 2019).

Metabolic Pathway Analysis

Ortholog assignment and pathway mapping were performed using the protein annotation file as query in the KEGG Automatic Annotation Server v2.1 (KAAS) (https://www.genome.jp/tools/kaas/) (Moriya *et al.* 2007). Functional annotation was performed using bi-directional best hit (BBH) by BLAST against the manually curated KEGG GENES against selected algal, plant and bacterial species (Supplementary Table B2). Gene lists annotated with KEGG Orthology (KO) identifiers (K numbers) were used

for metabolic pathway reconstruction in KEGG Mapper v4.0 (https://www.genome.jp/kegg/tool/map_pathway.html)(Kanehisa and Sato 2019).

Results and Discussion

PW95 is a novel green sporulating, non-motile alga isolated from a CBM production water impoundment in the Powder River Basin (northeastern Wyoming and southeastern Montana). PW95 is considered an industrially and environmentally relevant microalga due to fast growth rates and lipid generation potential for biofuel observed in standardized media (Chapter 2) and CBM water (Hodgskiss *et al.* 2016) in a variety of physiological conditions. Therefore, in this study, the genome of the PW95 strain has been sequenced and annotated to discover the genome encoded biochemical capacity of this organism for future cultivation strategies and industrial applications.

Phylogenetic Analysis and Description of PW95

A concatenated protein tree was generated for a robust and more complete understanding of the phylogenetic relationships of PW95 with other publicly available algae genomes for a robust better-informed organism description and genome annotation process (Figure 1). The phylogenetic analysis was performed using 30 single-copy proteins selected from PW95 genome and 17 eukaryotic photoautotrophs with publicly available genomes. The individual presumptive protein sequences were aligned and concatenated to produce a master alignment with 10,783 residues, as previously described (Wu 2018). PW95 grouped with *Chlamydomonas reinhardtii* and *Volvox carteri* (Figure 1) in the concatenated protein phylogenetic tree, showing a closer

phylogenetic relationship with *C. reinhardtii* (shorter branch distance).

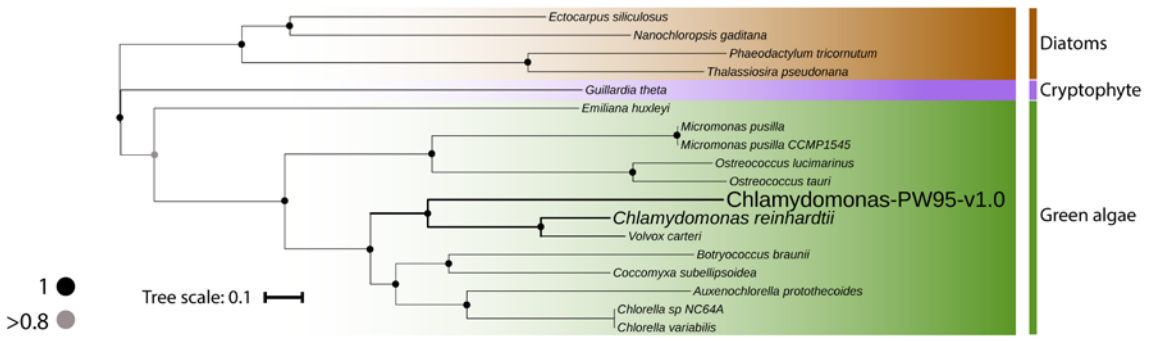


Figure 1. Phylogenetic relations of Chlamydomonas-PW95 green alga. Concatenated tree of single copy marker proteins aligned individually, trimmed and concatenated to generate a master alignment file. 17 publicly available algal genomes were included in the analysis. Bayesian phylogenetic inference with the model parameters: LG substitution model with empirical residue frequencies, gamma-shaped rate variation (4 categories) with a proportion of invariable sites (LG+F+I+G4). Average standard deviation of split frequencies (<0.001). Bayesian posterior probabilities are shown with filled circles in the branches.

PW95 Genome Assembly

De novo genome assembly was performed with CANU which specializes in assembling PacBio or Oxford Nanopore long-read sequences (Koren *et al.* 2017). CANU has been reported to generate accurate assembly of large genomic repeats (major proportion of nuclear DNA in eukaryotes (Biscotti, Olmo and Heslop-Harrison 2015)) from PacBio technology reads that have high error rates (10–15%) (Jayakumar and Sakakibara 2019). A total of 1,904,218 reads were generated from (SMRT) sequencing technology (PacBio) with an estimated 51X coverage of the PW95 genome. The PW95 draft genome assembly is 92 Mb, similar to the estimated genome sizes in *C. reinhardtii* (112 Mb) and *V. carteri* (120 Mb) (JGI genome portal; https://genome.jgi.doe.gov/portal/). The final draft assembly (v1.0; Table 1) consists of

4,402 contigs without gaps and reports high GC content (58.65%) similar to other sequenced algae (64% in *C. reinhardtii*; 56% in *Volvox carteri*; 67% in *Chlorella variabilis*; 53% in *Coccomyxa subellipsoidea*) (Hirooka *et al.* 2017).

Table 1. PW95 *de novo* genome assembly metrics. Assembly of the PacBio long reads was performed in CANU v.1.8 genome assembler using default parameters.

Assembly Metrics	PW95 *de novo* genome assembly
num_contigs	4,402
total genome length incl. gaps	91,951,860
avg_contig_size	20,888.70
Contig N50	29,477
Max contig size	164,018
Min contig size	1,420
GC (%)	58.65
Assembly completeness (%)	57

Genome contiguity is expressed as the contig N50, indicating that half of the genome sequence is in contigs larger than or equal to the N50 contig size. In principle, higher N50 translates into more complete genome assemblies and PW95 contig N50 (29,447) suggests that the assembly is still fragmented. Moreover, the number of scaffolds is equal to the number contigs which could be an indication that the assembler was unable to resolve repeats to perform contig scaffolding (Earl *et al.* 2011). Ambiguous genome reconstructions and fragmented assemblies arise when repeats are longer than the overlapping regions in the reads (used for contig assembly) (Phillippy, Schatz and Pop 2008).

Low coverage (51X) and high GC content (58.65%) can be an additional explanation for the fragmented assembly. PacBio instruments have the capacity to

produce reads of up to 14 kb; however, in reality most of the generated reads are smaller (800 bp on average). Therefore higher coverage is required to maximize the possibility of having enough sequencing depth and sufficient number of longer reads in order to accurately resolve genomic repeats and correctly assemble a genome (Nagarajan and Pop 2013), but this is not always possible given the high cost of sequencing an eukaryotic organism with ideal genome coverage. Moreover, rich GC regions can form secondary structures (hairpin loops) that are very stable and difficult to sequence (Blaby *et al.* 2014), hence high GC content is yet another challenge for genome sequencing and assembly.

Whole genome sequencing and assembly are processes that can always be improved by the addition of more sequencing data, using different assembly strategies and assemblers or by polishing the assembly (to detect and resolve misassemblies). These strategies can, however, add additional expenses and are not always necessary to answer the scientific questions of the project. The use of long reads in *de novo* assembly has intrinsic advantages, such as higher contiguity, fewer gaps and fewer errors (Jayakumar and Sakakibara 2019). For that reason, the assessment of the assembly completeness is a useful measure to move forward with the annotation. Genome completeness of the PW95 genome was estimated using BUSCO (Benchmarking Universal Single Copy Orthologs) and was determined that the PW95 genome assembly was sufficiently complete (57%) to proceed with annotation.

PW95 Genome Annotation

To achieve comprehensive functional annotation and pathway mapping of the PW95 genome, an assembled transcriptome (280,806 transcripts) was integrated into the annotation pipeline from a recently acquired transcriptomic data set of PW95 (Chapter 4). Extrinsic evidence, such as transcripts and protein databases, have been shown to significantly improve gene predictions when integrated with *ab initio* gene finders (Brejovńa *et al.* 2009).

Ab initio gene predictions of PW95 genomic features as reported by GeneMarkES, Genscan, GlimmerM and EVidenceModeler, showed that the number of genes, mRNA and proteins present in PW95 genome are similar to the reference green alga (and closest phylogenetic relative of PW95) *C. reinhardtii* (Table 2). None of these finders uses evidence models for the analysis. In contrast, AUGUSTUS showed the lowest numbers of features for PW95 and it can be attributed to using *C. reinhardtii* as the evidence model for the analysis, indicating that using a closely related organism gives a more specific (stringent) result.

Table 2. *Chlamydomonas* PW95 genome sequence annotation. Summary of genome features in comparison with the model green alga *Chlamydomonas reindhartii*.

Organism	Ab-initio Gene Predictors	genes	mRNA	proteins	Genome (Mbp)
PW95	AUGUSTUS	9,977	9,977	9,679	92
	GeneMarkES	16,404	16,404	15,759	
	Genscan	13,886	13,886	13,541	
	GlimmerM	14,064	14,064	14,058	
	EVidenceModeler (EVM)	12,483	12,483	11,355	
Chlamydomonas reinhardtii	JGI v4	14,474	14,422	19,526	111

After *ab initio* gene predictions, protein and transcript alignments were combined into weighted consensus gene models, followed by functional annotation. After the assignment of gene product names to the consensus gene structures, metabolic pathway reconstruction was performed using KEGG mapper. KEGG mapper identified a total of 402 pathways in PW95 (including complete and incomplete pathways) and 58 modules. A KEGG module consists of a segment of a biochemical pathway, *e.g.*, fatty acid (FA) initiation in the FA biosynthesis. A comprehensive list of fundamental complete biochemical pathways in PW95, shows the genetic signatures for carbohydrate metabolism and carbon storage, carbon fixation, nitrate assimilation, lipid metabolism, amino acid and nucleotide metabolism, all essential for cell growth, survival and maintenance, as well as relevant for cultivation and the production of value-added products such as lipids and starch (Table 3). Genome editing techniques of well understood, specific biochemical pathways can potentially enhance the accumulation of lipids for the biofuel industry (Sharma *et al.* 2018) and has been the target of research in *Chlamydomonas* species (Park, Nguyen and Jin 2019 and references therein).

Table 3. Fundamental biochemical pathways and modules in PW95. List of PW95 genes mapped to KEGG metabolic pathway and modules. KEGG mapper reported the pathways/modules that are complete in PW95 and the number of genes that mapped to the reference pathway.

Complete Pathway/Module KEGG in PW95	Number of genes
Carbohydrate metabolism	
Glycolysis (Embden-Meyerhof), glucose => pyruvate	
Glycolysis, core module involving three-carbon compounds	33
Gluconeogenesis, oxaloacetate => fructose-6P	
Pyruvate oxidation, pyruvate => acetyl-CoA	39
TCA (Citrate cycle), first carbon oxidation, oxaloacetate => 2-oxoglutarate	25
Pentose phosphate pathway, oxidative phase, glucose 6P => ribulose 5P	24

Pentose phosphate pathway, non-oxidative phase, fructose 6P	
Glycogen biosynthesis, glucose-1P => glycogen/starch	5
Starch and sucrose metabolism	28
Nucleotide sugar biosynthesis, glucose => UDP-glucose	3
Glyoxylate cycle, acetyl-CoA => succinate	38
Malonate semialdehyde pathway, propanoyl-CoA => acetyl-CoA	6
Propanoyl-CoA metabolism, propanoyl-CoA => succinyl-CoA	4
Energy metabolism	
Carbon fixation	
Reductive pentose phosphate cycle (Calvin-Benson), glyceraldehyde-3P => ribulose-5P	7
Photosynthesis	24
Photosynthesis - antenna proteins	6
Oxidative phosphorylation	65
Nitrogen metabolism	13
Assimilatory nitrate reduction, nitrate => ammonia	3
Lipid metabolism	
Fatty acid metabolism	
Fatty acid biosynthesis, initiation	
Fatty acid biosynthesis, elongation	27
β-Oxidation, acyl-CoA synthesis	
β-Oxidation	
Lipid metabolism	
Triacylglycerol biosynthesis	9
Acylglycerol degradation	2
Steroid biosynthesis	8
Sphingolipid metabolism	6
Glyceroplipid metabolism	23
Phosphatidylethanolamine (PE) biosynthesis, ethanolamine => PE	3
Amino acid metabolism	
Glycine, serine and threonine biosynthesis	30
Cysteine and methionine biosynthesis	40
Valine, leucine, isoleucine biosynthesis	11
Valine, leucine, isoleucine degradation	35
Histidine metabolism	14
Arginine and Proline biosynthesis	30
Tryptophan biosynthesis	15
Tyrosine biosynthesis and degradation	17
Phenylalanine, tyrosine and tryptophan biosynthesis	24

Nucleotide Metabolism	
Purine metabolism	56
Pyrimidine metabolism	35

Whole genome analysis can also lead to the discovery of novel or unique biochemical capacities, as is the case with PW95 genome (Table 4). A KEGG module involved in nitrate assimilation (nitrate transport), which is not present in any of the reference algal genomes in the KEGG database (*i.e.*, *Chlamydomonas reinhardtii*, *Ostreococcus spp.*, *Micromonas commoda*, *Guillardia theta*, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*), was found exclusively in PW95 (5 genes). PW95 is predicted to contain presumptive nitrate transporters that belong to both the MFS (major facilitator superfamily) and ABC superfamily (Saier *et al.* 1999; Davidson and Chen 2004). Typically, MFS transporters are specific for a class of solutes and function as uniporter, symporter, or antiporters driven by concentration gradients while ABC transporters have increased affinities at the expense of ATP. This could be correlated with the high rate of nitrate uptake observed in PW95 during cultivation at low initial nitrate concentrations (Chapter 2) and with differential expression over time of these transporters during autotrophic growth with nitrate as the sole source of nitrogen (Chapter 4).

Novel biochemical capacity was found in the PW95 genome with the squalene epoxidase gene (SQE). Despite SQE occurrence and conservation in animals, plants and fungi, it was only recently identified in a single diatom: *Phaeodactylum tricornutum* (Pollier *et al.* 2019). PW95 is phylogenetically characterized as a *Chlamydomonas*-like

alga but there is no report of this enzyme in *Chlamydomonas reinhardtii*, which has a fully annotated, manually curated genome and is a well-established model organism. *C. reinhardtii* has proven to be a powerful model for dissecting fundamental processes in biology not only in algae but has improved the understanding of these processes in plants and animals (Sasso *et al.* 2018). Therefore, this discovery suggests that, among algae, PW95 and *P. tricornutum* may have a unique capacity to perform steroid biosynthesis as an alternate way to potentially modulate the fluidity and flexibility of cell membranes as a maintenance/survival strategy. The potential impact on lipid accumulation and storage is not known and warrants further investigation.

Table 4. Genomic evidence of unique and novel biochemical capacity in PW95.

Signature Metabolic Capacity	Gene/Protein ID	KO	Number of genes in PW95
Nitrate assimilation unique to PW95			
MFS transporter, NNP family, nitrate/nitrite transporter	NRT, narK, nrtP, nasA	K02575	1
Nitrate/nitrite transport system substrate-binding protein, ABC transporter	nrtA, nasF, cynA	K15576	2
	nrtB, nasE, cynB	K15577	2
Novel alternative steroid biosynthesis in algae			
Squalene epoxidase	SQE, SQLE	K000100 K000909	1
Carbon concentrating mechanism			
Bicarbonate transport system/ ATP-binding protein	cmpC	K11952	1
Carbonic anhydrase	cah	K01674	1
Acetate assimilation			
Acetate kinase	ackA	K00925	1

Another unique aspect of photoautotrophic organisms is the ability to utilize a high affinity mechanism for dissolved inorganic carbon (DIC) uptake through a biophysical CCM (carbon concentrating mechanism), influenced by CO_2 or HCO_3^- availability and light supply (Reinfelder 2011; Mackinder 2018). Higher rates of carbon fixation are attributed to the ability to uptake different DIC species (HCO_3^- and CO_2) from the media (Chen, Qiu and Shao 2006) and thereby increasing the concentration of CO_2 at the active site of ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO). The CCM also helps to inhibit the oxygenase activity of the enzyme, thus reducing photorespiration and promoting photosynthesis and growth (Reinfelder 2011; Valenzuela *et al.* 2012; Mackinder 2018). Numerous genes and enzymes involved in inorganic carbon (IC) uptake, photosynthesis, and CCMs have been well studied in microalgae such as carbonic anhydrases and bicarbonate membrane transporters (Reinfelder 2011; Beardall and Raven 2017; Mackinder 2018), but there are still many uncharacterized proteins that could potentially play a role in IC fixation and the associated biochemical pathways responsible for physiological processes (*e.g.*, biomass and lipid accumulation).

Genomic evidence of PW95 potential capacity to activate and perform a CCM was found with the prediction of one carbonic anhydrase gene and a bicarbonate transporter gene (Table 4). Previous studies in PW95 using a bicarbonate addition for enhanced lipid accumulation, showed that bicarbonate did not stimulate growth and lipid production. The fact that PW95 did not respond to bicarbonate addition (Chapter 2; Corredor *et al.*, in review) suggests that any bicarbonate transporters present are high-affinity, low capacity but further investigation into PW95 response to increased

availability of IC and its relation to CCMs are necessary to better understand IC use during cultivation in this alga. CCMs are directly involved in the entry of IC into the cell and are under tight genetic and metabolic control and modulation of nutrient availability (Beardall and Giordano 2002; Li, Gao and Beardall 2012). Studies in *Chlamydomonas reinhardtii* showed a reduction of photosynthetic affinity for IC in N-sufficient cells versus N-limited cells (Beardall and Giordano 2002).

Interestingly, a gene involved in acetate assimilation, acetate kinase (ackA), was predicted in the annotation (Table 4). *C. reinhardtii* uses acetate as a sole carbon source during heterotrophic growth (Boyle, Sengupta and Morgan 2017), and given the close phylogenetic relationship to *C. reinhardtii*, PW95 might also have the capability to grow on acetate. Acetate assimilation depends also on an acetyl transferase (acetyl-P to acetyl-CoA) that was not found in the current version of the PW95 genome. This opens the possibility of PW95 cultivation in heterotrophic conditions, which has been reported as a strategy for wastewater treatment using green algae (Higgins *et al.* 2018) or as an alternative growth mode that has been shown to protect *C. reinhardtii* against photoinhibition during growth (Roach, Sedoud and Krieger-Liszkay 2013).

Interestingly, there were many similarities and differences between PW95 and *C. reinhardtii* genome annotations in the distribution of genes in functional categories (Figure 2). For both organisms the highest number of genes were distributed in pathways related to genetic information processes but thereafter differences emerged. For instance, in PW95, the third largest category was related to environmental information processing, a category that was not as highly represented in *C. reinhardtii*, while ‘protein families

involved in metabolism' was the third largest category in *C. reinhardtii* but did not appear in the top 10 categories for the PW95 genome. These differences may reflect the specialized environmental niche from which PW95 was isolated (Corredor *et al.*, in review; Chapters 2,5). Similarly, it has been shown that phytoplankton populations from more variable environments exhibit a higher level of genomic plasticity and diversity of environmental sensing systems (Schaum *et al.* 2013).

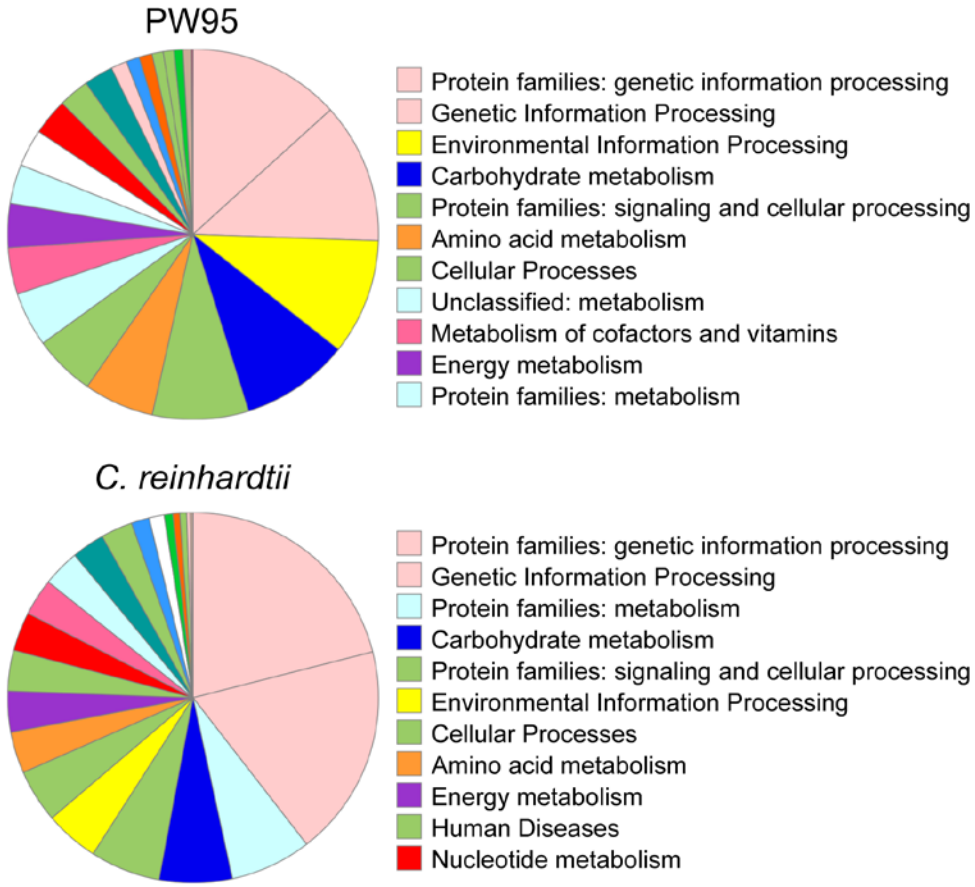


Figure 2. Comparison of PW95 and *Chlamydomonas reinhardtii* genomic functional categories. Distribution of annotated genes in both genome annotations as reported by GhostKOALA (KEGG Orthology and Links Annotation) functional characterization analysis. Protein sequence data in fasta format was submitted and searched against KEGG GENES database (nr database genus_prokaryotes+family_eukaryotes). Figure adapted from GhostKOALA report.

PW95 has more than 300 genes to encode for different transporters and could represent mechanisms to deal with oligotrophic conditions (Table 5). Transporters can be a defense mechanism to cellular perturbations due to change in environmental concentrations of solutes and metals, and macro and micronutrients (Blaby-Haas and Merchant 2012) (Tejada-Jiménez, Galván and Fernández 2011; Bonnot *et al.* 2017).

Table 5. Genes involved in the potential responses of PW95 to the environment stimuli.

Environmental Responses	Gene/Protein ID	Number of genes in PW95
Transporters		
nitrate/nitrite transport system	nrt	
ammonium transporter	amt	
urea transport system permease protein	urt	
Bicarbonate transport system	cmpC	
phosphonate transport system	phn,pst	
vitamin B12 transporter	btu,bio	
biotin transport system	bio	
thiamine transport system	thiP	
solute carrier families	sodium bicarbonate, zinc, (i) phosphate, sodium/calcium	>300
zinc transport system	znu	
iron transport system	afu	
manganese/iron transport system	sit, mnt	
molybdate transport system	mod	
multiple sugar transport system	mal,msm,frc	
multiple aminoacidic transport systems	liv,ABC.PA,pro,ABC.SS	
peptide/nickel transport system	ABC.PE	
ABC type transport system	ABC	
Temperature sensing		
heat shock protein	HSPA	8
cold shock protein	cspA	1

Response to photo-oxidation		
superoxide dismutase	SOD	2
Others		
starvation-inducible DNA-binding protein	dps	5
stress-induced-phosphoprotein	STIP	2

Transporters also aid in cell survival by scavenging for nutrients, sources of inorganic carbon or getting rid of the excess to maintain cellular homeostasis and avoid cytotoxicity (Blaby-Haas and Merchant 2012; Beardall and Raven 2017). PW95 also has a number of heat and cold shock protein genes that could potentially protect the cells from changes in temperatures and/or other environmental stresses. Sathasivam and Seu (2019) showed heat shock proteins in *Tetraselmis suecica* were involved in defense responses to environmental stresses, such as changes in water temperature and exposure to toxic substances (Sathasivam and Ki 2019). Lastly, the prediction of two superoxide dismutase (SOD) genes shows the potential activation of a protection mechanism against photo-oxidative stress in PW95 chloroplasts, most likely to prevent the damaging effects of reactive oxygen species (Zhang et al. 2013). Enzymatic antioxidants such as SOD are now frequently detected and associated with microalgal lipid accumulation cultivated under abiotic stress such as temperature, salinity, nitrogen, and light (Chokshi *et al.* 2017).

Genomic Evidence of Sporulation, Cell Motility and Sexual Cycle

We have discussed previously that PW95 appears to have the capacity to reproduce using spores and does not have flagella. However, its closest relatives (*C.*

reinhardtii and *V. carteri*) are bi-flagellated and can reproduce sexually or asexually. To investigate whether there was a genomic basis for these differences, the annotated genome was mined for evidence of genes related to these processes. Results are summarized in Table 6.

Table 6. Genomic evidence of sporulation, flagellar assembly and function and meiosis in PW95.

Biological Function/Structure	Gene/Protein ID	Number of genes in PW95
Sporulation		
Spore maturation protein B	spmB	1
DNA segregation ATPase FtsK/SpoIIIE	FtsK/SpoIIIE	2
Flagella		
Flagellin	fliC	5
Flagellum-specific ATP synthase	fliI	1
Intraflagellar transport protein	IFT46,140	2
Flagellar biosynthetic protein	flhA,B,F,Q,P,R,T	11
Flagellar P-ring protein precursor	flgI	3
Flagellar M-ring protein	fliF	1
Flagellar basal-body rod protein	flgG	1
Flagellar motor switch protein	fliM,NY,G	3
Flagellar system response regulator	ftcR	1
Flagellar hook protein	FlgE	1
Flagellar hook-associated protein	flgK,L	7
Pilus assembly protein	cpaB-F, flp	6
Tubulin alpha	TUBA	1
Tubulin beta	TUBB	1
Dynein heavy chain	DNAH	35
Dynein light intermediate chain	DNAI	2
Type IV pilus assembly protein	pilM,P,Q	3
Sexual cycle		
Meiosis associated recombinase	Rad51	1

Genomic evidence of spore formation in PW95 was found with the annotation of the spore maturation protein B (spmB; 1 gene) and the DNA segregation ATPase

(FtsK/SpoIIIE; 2 genes) (Table 6). The *spm*B gene product in *Bacillus subtillis* is associated with core dehydration during sporulation and not a spore structural component *per se* (Paredes-Sabja *et al.* 2008), and the *spo* related genes are typically involved in regulation of the sporulation process in bacteria. Therefore, despite observational evidence of spore-like structures during PW95 growth, the current genome annotation did not contain a full complement of a presumptive sporulation system. However, sporulation in photoautotrophic eukaryotes is not well understood, and future improvements in the assembly and annotation could potentially reveal many more genes associated with a presumptive sporulation process in PW95.

Neither spore-like structures nor vegetative PW95 cells show any phenotypic evidence of flagella. PW95 is assumed to be asexual and non-motile but more than 50 genes related to cilia/flagella assembly and function were identified in the PW95 genome (Table 6), and these results suggested that the capacity to assemble and use flagella exists and could be expressed in very specific, but yet unknown conditions. Evidence of flagella genes in other presumed asexual, non-motile algae has been demonstrated in the genome of two different species of *Chromochloris*, in which almost the entire set of motile flagella genes was revealed (Roth *et al.* 2017). Extracellular structures, such as flagella, have been shown to have other functions besides motility in bacteria (Clark *et al.* 2007) but additional roles in green alga have not been demonstrated.

Lastly, the PW95 genome also encodes the meiosis-associated Rad51 DNA recombinase homolog, also observed in the *Chlorella vulgaris* genome (Guarnieri *et al.* 2018), indicating the possibility of sexual mating. Novel gene combinations and loss of

deleterious alleles can occur during recombination in sexual reproduction and can be used as a breeding technique in the laboratory to improve biomass and bioproduct yields, as has been employed traditionally in plant breeding (Hovde *et al.* 2018). Further work is needed to better understand the reproductive cell cycle for unique green alga such as PW95.

Conclusions

The current study presents the draft genome assembly and annotation (v1.0) of a novel green alga isolated from coal-bed methane production water stored above ground in a high-plateau desert environment. Based on predicted protein sequences, a concatenation of 30 presumptive proteins from 17 photoautotrophic eukaryotes indicated that PW95 is most closely related to *C. reinhardtii*. From the current annotation, PW95 has a unique complement of nitrate transporters, novel steroid biosynthesis, a cadre of environmental sensing systems, and could be mixotrophic for acetate utilization. Although the PW95 genome is at a draft stage, the genome serves as a diverse reference for environmental green alga with the ability to grow in low-quality water for biomass and lipid production. Strategies for the improvement of the genome assembly will be implemented that can result in a less fragmented, higher quality assembly. Re-assembly and assembly polishing will be used to detect and resolve misassemblies and Illumina sequence data will be used for error-correction. For the annotation, the transcriptome will be quality filtered and incorporated into the pipeline as evidence. MAKER2 (genome curation and annotation pipeline) will be used in addition to GenSAS. Following the aforementioned improvements, an investigation into gene structure (introns and exons, CDS), telomeric

and repetitive sequences will expand our knowledge into the PW95 genome. As more of the gene complement, physiology and ecology of PW95 and other unicellular algae are investigated, relationships between gene content and function will inform the potential of environmental, industrial and biotechnological applications in novel and unique algae.

Acknowledgments

This work was supported by the National Science Foundation Sustainable Energy Pathways (SEP) Program under NSF CHE-1230632 and the Department of Energy PEAK program under DE-EE0008247. The authors would like to thank the Montana Bureau of Mines and Geology for providing access to the FG-09 well and the MSU CBM group for water sample collection. The authors also thank the MSU Algae Group for support and constructive exchange of ideas, Calvin Cicha for support with annotation process during his rotation in the Fields laboratory, Dr. Ben Wheaton for graphical and editorial assistance, and Dr. Chiachi Hwang for assistance with SSU amplicon sequencing and taxonomic analysis.

CHAPTER FOUR

TRANSCRIPTOMIC PROFILING OF *Chlamydomonas*-like PW95 CULTIVATED IN COAL BED METHANE PRODUCTION WATER WITH THE NATIVE MICROBIAL COMMUNITY

Contribution of Authors and Co-Authors

Manuscript in Chapter Four

Author: Luisa Corredor
Contributions: Experimental design, performed experiments, performed data analyses, literature review, synthesized ideas, wrote and edited the manuscript.

Co-Author: Huyen Bui
Contributions: Performed bioinformatical analyses and edited the manuscript.

Co-Author: Thiru Ramaraj
Contributions: Performed bioinformatical analyses and edited the manuscript.

Co-Author: Robin Gerlach
Contributions: Procured funding, synthesized ideas, reviewed and edited the manuscript.

Co-Author: Matthew W. Fields
Contributions: Principal Investigator, procured funding, experimental design, synthesized ideas, wrote and edited the manuscript.

Manuscript Information Page

Luisa Corredor, Huyen Bui, Thiru Ramaraj, Robin Gerlach, Matthew W. Fields

Status of Manuscript:
__X__Prepared for submission to a peer-reviewed journal
____Officially submitted to a peer-reviewed journal
____Accepted by a peer-reviewed journal
____Published in a peer-reviewed journal

Abstract

Differential gene expression analysis can lead to the discovery of genetic capabilities in green algae that can be exploited for the production of biomass, chemical feedstock products, and novel metabolic capacity. We isolated and characterized a novel sporulating green microalga from coal bed methane (CBM) production water, identified as *Chlamydomonas*-like species, designated PW95. PW95 was cultivated in photobioreactors (PBRs) with ambient air and artificial sunlight, using filtered and unfiltered CBM water. The physiological responses of PW95 in this system were investigated and global gene expression during growth and lipid production were characterized in the context of nitrate depletion in native CBM low quality water and in the presence of a diverse microbial community. In unfiltered water with the native microbial community, PW95 had a higher growth rate, cell density, and cell dry weight in comparison with filtered water. PW95 cultures grown in filtered and unfiltered water were harvested at late-exponential and late-stationary phases, and replicate gene expression samples could be grouped based upon growth phase and water condition. Principal component analysis, PC2, accounted for 30% of the data variability and differentiated water condition (filtered versus unfiltered), and PC1 accounted for 36% of the data variability and differentiated growth phase. The most differentially expressed genes were between exponential and stationary phases under the filtered water condition whereas fewer genes with statistically significant differential expression were detected (1.8x less) between growth phases for cells cultivated in unfiltered water. The fewest differences were observed when exponential-phase cultures were compared between

filtered and unfiltered water, but 5x more genes had significantly altered expression when stationary-phase cultures were compared between the filtered and unfiltered conditions. Genes with altered expression between exponential and stationary phases included sphingolipid metabolism, fatty acid metabolism, nitrate transporters, TCA cycle, glycolysis, and amino acid metabolism. The results suggest that the presence of a diverse, indigenous water microbial community impacts algal gene expression during cultivation, and the presence of a more diverse microbial community may modulate or dampen changes in algal gene expression particularly in exponential growth. Further work is needed to delineate the required systems for direct and/or indirect metabolic interactions between algae and associated microorganisms.

Introduction

Chlorophytes are unicellular green microalgae that play fundamental roles in primary productivity and nutrient cycling (Chapman 2013), as well as in the energy and nutraceutical industries (Fields *et al.* 2014). Renewed interest in green microalgae has been sparked by their ability to produce biomass and consume large amounts of atmospheric CO_2 while producing desirable compounds used in industry and medicine (Fields *et al.* 2014). Transcript identification and the quantification of gene expression can lead to the discovery of genetic capabilities in different organisms that can be exploited for numerous applications; photoautotrophic microorganisms such as green microalgae are no exception. They can be grown in low-quality water and can treat wastewater streams while also producing value-added products, such as lipids, starch and pigments (Hodgskiss *et al.* 2016; Singh, Upadhyay and Rai 2017). These applications

help to reduce water burden through nutrient and water recycling, while potentially providing potable water for human consumption; the utilization of large amounts of extracted water derived from energy-related water consumption (*e.g.*, coal bed methane (CBM) production water) and also contributing to economic feasibility (Corredor *et al.*, in review; Fields *et al.* 2014). Microalgae produce and accumulate lipids such as triacylglycerols (TAGs), usually under nutrient depleted conditions that can account for 20 to 50% of dry weight of the biomass depending on the strain, and fatty acid methyl esters (FAMEs) (derived from the transesterification of TAGs) that are key carbon and energy storage compounds desirable for the biofuel industry, making algae far more efficient for lipid production than any traditional food crop used for this purpose (Chisti and Yan 2011). Similarly, they can also produce other compounds of interest such as starch and carotenoids, that can have high nutraceutical value (Schnurr, Espie and Allen 2013).

To shed light on all these known desirable traits and associated basic physiological functions, we isolated and characterized a novel sporulating green microalga from CBM production water, identified as a unique *Chlamydomonas*-like species, designated PW95. We performed a physiological characterization of PW95 for biomass and lipid production in standardized media (Corredor *et al.* 2019, in review) and CBM water (Hodgskiss *et al.* 2016), sequenced and assembled the whole genome (Corredor *et al.* 2019, Chapter 3), and previous reports indicate that RNA-seq data is needed to improve functional gene annotation (Holt and Yandell 2011; Conesa *et al.* 2016). Hence, the current study is part of a continuous effort to gain knowledge of the

genetic and genomic basis that defines and controls the physiological behavior of this unique alga with implications for water management, biofuels, and other value-added products.

PW95 was cultivated in photobioreactors (PBRs) with artificial sunlight, using filtered and unfiltered supplemented CBM water (0.5 mM nitrate, 0.185 mM borate) at 30°C. The current study investigated the physiological responses of PW95 and characterized global gene expression during growth and lipid production in the context of nitrate depletion, establishing meaningful comparisons between different growth phases (late exponential and late stationary) and water types (filtered and unfiltered CBM water). Results from this study provide a better understanding of the underlying genetic capabilities of an environmental green microalga during algal biomass production in oligotrophic conditions. These results can be implemented into the design of improved cultivation strategies for PW95 and can be translated to other green microalgae for the same purpose. Additionally, the assembled transcripts (transcriptome) can be incorporated into the PW95 genome annotation pipeline.

Materials and Methods

Coal Bed Methane Production Water

PW95 (*Chlamydomonas*-like strain) was isolated from a CBM production pond in the Powder River Basin (northeastern Wyoming and southeastern Montana). CBM water samples were collected from the FG-09 well (Flowers-Goodale coal bed operated by the Montana Bureau of Mines and Geology) prior to performing the current investigation and

kept at 4°C until used in the PBRs for the present study. Supplementary Table C1 shows the water chemistry analysis report for the CBM water used in this study.

Experimental Culture Conditions

Stock PW95 cultures were grown in 500 mL of Bold's basal medium (BBM) in 1000 mL conical flasks with rubber stoppers that allowed gas exchange. Stock cultures were incubated at 30°C with artificial sunlight (white LED, 33 μmol photons $m^{-2}s^{-1}$, approx. 1435 lux output/tube) on a 14:10 light/dark (L/D) cycle and mixed constantly with a magnetic stirrer until the exponential phase of growth occurred. The biomass was concentrated via centrifugation at 4800 × g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA). Biomass pellets were resuspended in 250 mL ultrapure (18.2 Ω) deionized water (dH_2O) and used as inoculum for the PBRs (10% of the total volume).

Experiments were conducted in triplicate batch cultures using six glass tubes (70 × 500 mm) containing 1.25 L of CBM water as medium (initial pH 8.4), and amended with 0.5 mM sodium nitrate (Sigma-Aldrich, USA) and 0.185 mM boric acid (Sigma-Aldrich, USA), as previously described (Hodgskiss *et al.* 2016). Tubes 1 to 3 contained CBM water that had been filtered using a 0.2 μm bottle top PES filter (Corning, USA) and tubes 4 to 6 contained unfiltered CBM water. All experimental cultures (PBRs) were incubated side-by-side submersed in a controlled water bath maintained at 30 ± 1°C. Rubber stoppers, containing ports for aeration and sampling, were used to seal the tubes. Light (33 μmol photons m^{-2} s^{-1}, approx. 1435 lux output) was maintained on a 14:10 L/D cycle using a light bank containing eight white LED tubes (Waveform T5 PhotonTube™,

6500K, 95 CRI and R9>80) to simulate natural sunlight. Aeration (400 mL min^{-1}) was supplied by humidified compressed air bubbling through Nalgene bottles containing ultrapure (18.2 Ω) dH_2O and controlled using individual rotameters for each PBR tube (Cole-Parmer, USA). Tubes were constantly mixed using a magnetic stirrer.

RNA Sampling, Extraction and Sequencing

RNA was collected for RNA-seq at two different time-points (late exponential (E) and late stationary (S) phase of growth) in both filtered (F) and unfiltered (UF) water. A volume of a 100 mL of culture from each PBR were collected in sterile 50 mL Falcon tubes and concentrated via centrifugation at 4800 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA). Samples were washed with ultrapure (18.2 Ω) dH_2O and pooled. Biomass was pelleted by centrifugation at 14000 x g for 5 min (Thermo Scientific, Sorvall Legend XTR), snap frozen in liquid nitrogen and stored at -80°C for RNA extraction. Samples were designated E for late exponential and S for late stationary phases of growth between filtered (E-F1,2,3 and S-F1,2,3) and unfiltered (E-UF4,5,6 and S-UF4,5,6) CBM water.

Total RNA was extracted using a customized RNA extraction protocol for PW95 including a homogenization step, polysaccharide precipitation treatment and an extraction step prior to the use of the RNeasy PowerPlant Isolation kit (Cat No./ID: 13500-50, Qiagen, USA). The homogenization step consisted of grinding algae pellets (twice) in liquid nitrogen in a pre-chilled mortar and pestle until a fine powder was obtained before resuspending in 1 mL of extraction buffer TRIZOL (Invitrogen, USA), transferring the suspension to a bead-beating tube from the kit and proceeding with the polysaccharide

precipitation treatment step. Polysaccharide precipitation was achieved by adding potassium acetate (EMD Millipore Sigma, USA) at a final concentration of 0.2 mM (Dos Reis Falcão *et al.* 2008), vigorously shaking the tubes in the Fast prep-24 (MP Biomedicals, USA) at 6.5 m/s for 60 seconds, incubating at room temperature for 20 min and centrifuging tubes at 13,000 x g at 4°C for 10 min. To separate the supernatant phase that contains the RNA and a phase separation step was performed. The supernatant was transferred to a new tube and 200 μL of chloroform (Sigma-Aldrich, USA) were added; the tube was inverted by hand for 30 s and centrifuged at 13,000 x g at 4°C for 10 min. After centrifugation the aqueous phase (containing the RNA) was carefully removed and transferred to a clean 2 mL collection tube from the kit. The RNA was collected, washed and eluted using the RNeasy PowerPlant Isolation kit, as per the manufacturer's instructions with the addition of PR2 (patented inhibitor removal technology; https://www.qiagen.com/us/resources/resourcedetail?id=fe544355-a170-4e55-9c35-636a9e258cd4&lang=en) during step 6. Sample QC was performed using a Nanodrop One Microvolume UV-Vis Spectrophotometer (ThermoFisher, USA). Total RNA was not pooled but kept as the respective triplicates.

Library preparation, quality control and sequencing were performed at Genewiz (NJ, USA). RNA quality was determined using a 2100 Bioanalyzer system (Agilent Technologies, USA) and RNA library was constructed for mRNA sequencing via polyA selection. Following Illumina´s standard protocols, HiSeq paired-end (2x150), high output sequencing was performed.

Differential Expression Analysis

RNA-seq reads were aligned to the proprietary PW95 draft genome using the two-pass alignment mode of STAR 2.7.2b (Dobin *et al.* 2013) with a maximum intron size of 20 kb and customized parameters: "--outFilterScoreMinOverLread 0" (STAR default: 0.66), "--outFilterMatchNmin 30" (STAR default: 0), "--outFilterMismatchNmax 100" (STAR default: 10), "--outFilterMismatchNoverLmax 0.3" (STAR default: 0.3). The number of reads uniquely aligned to each locus was counted using HTSeq 0.11.1 (Anders, Pyl and Huber 2015) with feature type "exon". Differentially expressed genes (DEGs) between conditions of interest were detected using the DESeq2 package (version 1.24.0) (Love, Huber and Anders 2014). DEGs with a log2FC greater than 1 (in either direction) and a false discovery rate (FDR) adjusted *P*-value of less than 0.05 were considered differentially expressed for downstream analysis. Principal component analysis (PCA) was performed to assess sample clustering and variance.

Annotation and metabolic function assignation

Ortholog assignment and pathway mapping were performed using the PW95 predicted coding sequences as a query in the KEGG Automatic Annotation Server v2.1 (KAAS) (https://www.genome.jp/tools/kaas/) (Moriya *et al.* 2007). Functional annotation was performed using single-directional best hit (SBH) by BLAST against the manually curated KEGG GENES database against selected algal, plant and bacterial species (Supplementary Table C2). Gene lists annotated with KO identifiers (K numbers) were used to determine the metabolic function of DEGs and were mapped to metabolic

pathways. Genes without KO identifiers and unknown function were manually annotated with BLASTP (Altschul *et al.* 1990).

Results and Discussion

Physiological characterization of PW95 growth in CBM water

The physiological responses of PW95 during growth in amended CBM water (0.5 mM nitrate and 0.185 mM borate), artificial sunlight and 30°C in PBRs were monitored every 24 hours for 7 days, prior to the end of the light phase. A comparison between filtered and unfiltered CBM production water was performed with three replicates per type of water. Algal biomass samples for FAME analysis and RNA sequencing were collected during late exponential and late stationary phases of growth to evaluate the transcriptomic responses of PW9. As reported previously, batch cultures in standardized media with an initial nitrate level of 0.5 mM at 30°C had a positive impact on PW95 growth showing a greater increase in cell numbers, a shorter lag phase, faster growth rates and increased levels of lipids in comparison with 20°C and 25°C (Corredor *et al.*, in review). Moreover, PW95 was shown to grow in amended CBM production water (0.5 mM nitrate and 0.185 mM borate) and produced elevated levels of lipids in batch mode (Hodgskiss *et al.* 2016). Light has also been shown to be a positive stimulus that can increase photosynthetic rates and thereby growth and lipid production. The use of white LED lights for the indoor cultivation of *Chlorella sp.* demonstrated an increase in biomass and lipid production (Hsia and Yang 2015).

The unfiltered water had a higher mean response for growth rate, cell density, cell dry weight compared to cultures grown in filtered water (Chapter 5). Chlorophyll production was not significantly different between algae cultivated in filtered or unfiltered water, and pH profiles for both water conditions were also similar (Chapter 5). NR fluorescence was also similar for the two tested water conditions; however, NR fluorescence for algal cultures grown in filtered water was slightly higher (Chapter 5). The rate of increase in NR fluorescence was faster for algae cultivated in filtered water compared to unfiltered water. Detected FAMEs were similar between the two water conditions in exponential growth, but FAME levels were significantly higher for algal cultures grown in filtered water compared to unfiltered water (Chapter 5).

Global Transcriptional Profile of PW95 genes

The present study was designed to mimic oligotrophic environmental conditions in native water with low and high diversity indigenous microbial populations and determine the impacts on gene expression of sympatric microbial communities of PW95. To gain insight into the metabolic capabilities associated with the physiological responses of PW95 to cultivation in PBRs with CBM water during nitrogen starvation, the gene expression profile was quantified at late exponential and late stationary phases of growth (designated E and S) between filtered and unfiltered CBM water samples (designated F and UF). An average of 30 million reads per sample was obtained from RNA sequencing. From a total of 20594 genes reported in the genome reference (PW95 genome annotation), 9416 were included in the DESeq2 analysis and 510 genes were expressed at

statistically significant levels (*Padj<0.05*; Log2FC>1) at the two sampling points (5% average of uniquely mapped reads for each sample).

Principal component analysis of all the samples evaluated in this study showed that replicate samples within each group clustered together and samples corresponding to late stationary phase of growth showed greater variation (Figure 1A). Significant differentially expressed genes (DEGs) were identified and compared in two ways: 1) Exponential and stationary growth phases were compared *within* each water type (in filtered water: E-F vs S-F; and in unfiltered water: E-UF vs S-UF) to test the influence of growth stages on differential gene expression in each water type, and 2) Exponential and stationary growth phases were compared *between* water types (in the exponential phase: E-F vs E-UF; and in the stationary phase: S-F vs S-UF) to test the influence of water type on differential gene expression at different stages of growth. An overview of up- and down-regulated DEGs is shown in Figure 1B. The comparison that showed the lowest number of differentially expressed genes was exponential filtered vs. exponential unfiltered samples (E-F vs. E-UF) where 25 DEGs were identified, the majority of which were down-regulated. Stationary filtered vs. stationary unfiltered samples (S-F vs. S-UF) had 129 DEGs, the majority of which were up-regulated. Exponential unfiltered vs. stationary unfiltered (E-UF vs. S-UF) also had 129 DEGs, and these were up- and down-regulated in approximately equal proportion.

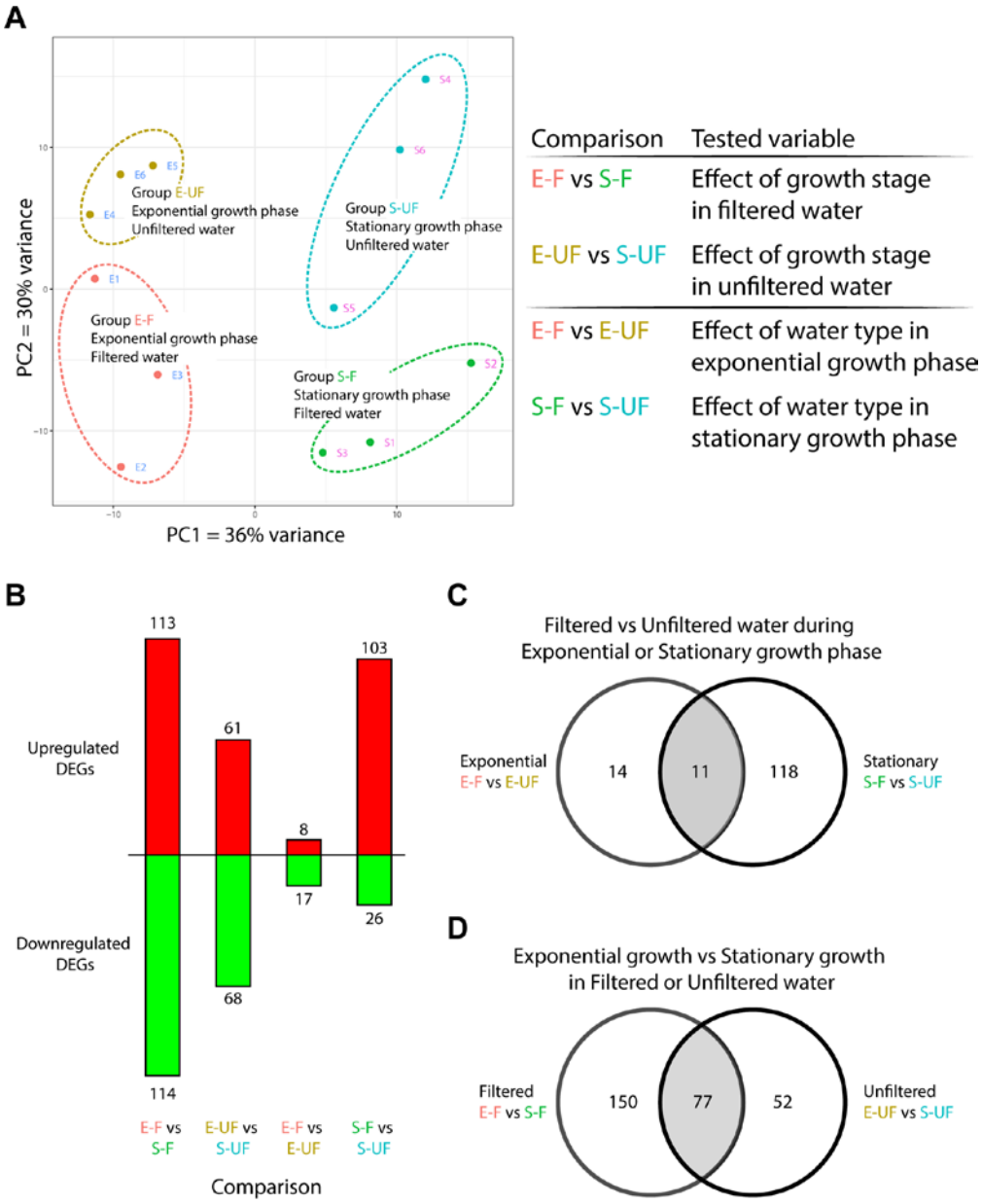


Figure 1. Global gene expression changes in PW95 during cultivation in amended coal bed methane production water at 30°C and artificial sunlight in PBRs. (A) PCA of expression data for the complete data set. Samples were designated E for late exponential and S for late stationary phases of growth in filtered (1,2,3 and unfiltered (4,5,6) CBM water. (B) Number of significant up-regulated and down-regulated genes detected in each comparison group (*Padj<0.05*; log2FC>1 in either direction). List of all DEGs with predicted function are available in the supplemental data (Supplementary Tables C3–C6). Venn diagrams showing unique and common elements of the transcriptomic responses between water types and between growth phases are shown in C and D, respectively.

Exponential filtered vs. stationary filtered (E-F vs S-F) had the highest number of DEGs (227); these were up- and down-regulated in approximately equal proportions. Greater variation in samples corresponding to late stationary phase of growth were found in the PCA (Figure 1A) and is reflected in the number of DEGs found in S-UF group comparisons. Lists of all DEGs with predicted function are available in the supplemental data (Supplementary Table C2–C6).

In order to determine PW95's base-line metabolic activity in this system, DEG datasets were overlaid to identify common elements regardless of the type of water or the growth phase. There were 11 DEGs common to both exponential and stationary phase time-points (Figure 1C) and there were 77 DEGs common to filtered and unfiltered water (Figure 1D). The overlapping DEGs in both data sets are involved in glycolysis, pyruvate metabolism, TCA cycle, Calvin-Benson cycle (phosphoglycerate kinase (PGK), glyceraldehyde-3-phosphate dehydrogenase (GAPA), carbon fixation via malate dehydrogenase (MDH1), oxidative phosphorylation (inorganic pyrophosphatase (*ppa*)) and amino acid metabolism (prolyl 4-hydroxylase (P4HA), 3-hydroxyisobutyrate dehydrogenase (*mmsB*).

A large proportion of overlapping DEGs was composed of non-assigned (NA) genes that make up the 63.6% of genes between the growth phases and 42.8% between the types of water. The genome reference can have a strong impact on the quantification and differential expression analysis (Conesa *et al.* 2016), incomplete annotations such as the current proprietary PW95 gene model could explain the high percentage of NA proteins present in this data set. An optimized version of PW95's genome annotation

could potentially identify more proteins in the DEGs (Zhao and Zhang 2015). For this same reason, pathway reconstruction based on DEGs was not possible (Supplementary Figures C1–C3). Instead, a subset of DEGs responsible for basic metabolic functions in PW95 was selected to present an overview of up- and down-regulation of genes involved in basic metabolic functions in this algal cultivation system (Figure 2), followed by an in-depth analysis of the up- and down-regulation of significant DEGs for each of the group comparisons in the following sections.

Steroids, sphingolipids and ceramides can be integral components of the cell membrane but also play roles in additional cellular processes, and the concentration of these lipids can change quickly depending on external stimuli (Luttgeharm, Kimberlin and Cahoon 2016). Sphingolipids (sphingosine lipids) have been detected in marine red algae membranes but the function is not known (Khotimchenko and Vas'kovsky 2004). A simple sphingolipid was identified in the green alga, *Ulva fasciata*, and was shown to have antiviral properties (Sharma, Garg and Chandra 1996). The down-regulation of genes annotated as magnesium or manganese-dependent protein phosphatases (PPM1L) involved in sphingolipid metabolism in exponential-phase cultures irrespective of water condition suggests a role during faster growth that is involved in sphingolipid and/or amino acid metabolism. Interestingly, a similar gene (PPM1L) in humans has been shown to be involved in adiposity (*i.e.*, lipid biosynthesis) (Merhi *et al.* 2015), with diferential expression in PW95 dependent upon lipid allocation. In addition, an annotated *fab*D (malonyltransferase) involved in fatty acid metabolism, was down-regulated in filtered water (exponential phase) but not unfiltered water (exponential-phase) (Figure 2).

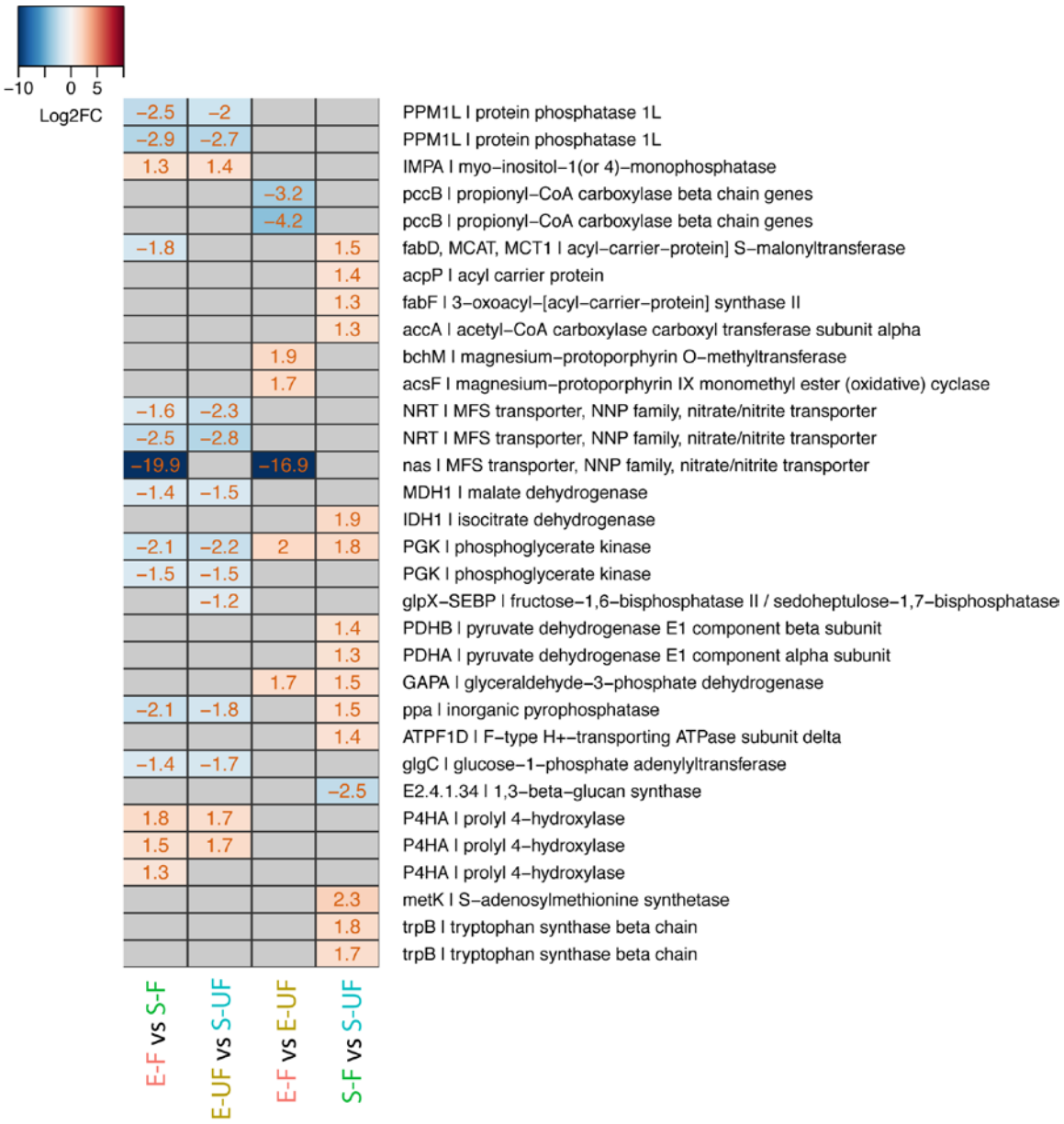


Figure 2. Overview of the changes in expression of significant DEGs in PW95 during biomass and lipid production in CBM water. DEGs representing metabolic pathways associated to fundamental processes in algae such as the carbon central metabolism, carbon fixation, nitrogen assimilation, photosynthesis amino acid metabolism and fatty acid biosynthesis. The heat map shows the level of up- (red) and down-regulation (blue) of DEGs in each group comparison for each of the selected DEGs and the pathway or pathways in which they are involved.

Interestingly, the same gene was up-regulated in stationary-phase (filtered water), most likely a result of lower expression in exponential-phase in filtered water versus unfiltered water. Following a similar trend, genes involved in fatty acid biosynthesis annotated as acyl-carrier-protein S-malonyltransferase (*fabD*), acyl carrier protein (*acpP*), 3-oxoacyl-acyl-carrier-protein synthase II (*fabF*) and acetyl-CoA carboxylase carboxyl transferase subunit alpha (*accA*) were up-regulated in filtered water (stationary-phase) compared to unfiltered water (stationary-phase). Correspondingly, two genes annotated as *pcc*B (involved in β-oxidation) were down-regulated in filtered water versus unfiltered water in exponential-phase growth (Figure 2). The down-regulation of genes involved in β-oxidation and then up-regulation of genes involved in fatty acid biosynthesis from exponential- to stationary-phase for algae grown in filtered but not unfiltered water corresponded to higher FAMEs for these samples.

Despite similar trends in chlorophyll levels for the two growth conditions, two genes annotated to be involved in chlorophyll biosynthesis (*bch*M and *acs*F) were up-regulated for filtered versus unfiltered condition and could suggest the role of a diverse bacterial community to cross-feed chlorophyll pre-cursors under different growth conditions. Superoxide dismutase (SOD2) was also up expressed and the increase of antioxidant enzymes after nitrogen depletion has been related to oxidative stress in *Chlorella sorokiniana* (Zhang *et al.* 2013). Declines in chlorophyll could be associated with the damaging effects of reactive oxygen species and the activation of a protection mechanism against photo-oxidative stress in the chloroplast, hence the up expression of superoxide dismutase (SOD2). Additionally, it could be argued that PW95 is acclimated

to fluorescent light at low to medium saturation (stock and experimental cultures) and the exposure to LED artificial sunlight triggered photo-oxidative stress. The type of light and saturation have been shown to play an important role in culture health and productivity (Hsia and Yang 2015; Cheloni and Slaveykova 2018).

Among the genes that did have assigned identities, the nitrate/nitrite transport system substrate-binding protein (*nrtA*) was the most down-regulated gene in this group, and because exogenous nitrate was completely depleted in the water at this sampling time the differentially expressed gene may represent a low-affinity, high rate transporter that would not be functional at low nitrate levels. Additional annotated transporters predicted to be involved in nitrate/nitrite transport were down-regulated in exponential versus stationary-phase, and an annotated *nas* (transporter) was strongly down-regulated even for filtered versus unfiltered condition. The differential expression for presumptive nitrate transporters corresponds to the N-source used for cultivation and could be a result of different transporters with different kinetic characteristics (rate and affinity) beneficial under different growth conditions. For example, the *nas* was not down-regulated for algal cultures grown in unfiltered water and could be linked to competition for nitrogen in more diverse microbial communities.

Differential Expression at Late Exponential Phase of Growth

The lowest number of DEGs (25) was observed during exponential phase of growth when comparing filtered water to unfiltered water (E-F vs. E-UF) and most of these DEGs were down-regulated (Supplementary Table C3). Late exponential phase of growth corresponds to the transition point between the exponential and stationary phases

for the tested growth conditions. Cultures received inorganic carbon from in-gassing of atmospheric air, and carbon fixation has been correlated with pH increase (Gardner *et al.* 2013). Presumptive genes annotated to be involved in TCA or carbon fixation (malate dehydrogenase, phosphoglycerate kinase, and sedoheptulose bisphosphatase) displayed down-regulation in exponential-phase samples irrespective of water condition (Figure 2). However, these same presumptive genes were no longer down-regulated or up-regulated in stationary-phase (filtered water condition when FAMEs were elevated). In addition, a presumptive glucan synthase was down-regulated in stationary-phase filtered water condition (elevated FAMEs) and suggests carbon-flow to lipids versus starch. Further work is needed to better understand carbon allocation in PW95 between lipid and starch.

Differential Expression at Late Stationary Phase in Filtered Water

A total of 129 significant DEGs were detected during late stationary phase of growth between filtered and unfiltered water (S-F vs. S-UF; Supplementary Table C4). The majority of the DEGs were up-regulated in the filtered water and were involved in photosynthesis, amino acid metabolism, oxidative phosphorylation, carbon metabolism via glycolysis and TCA cycle, oxidative phosphorylation and photosynthesis, Calvin cycle, FA biosynthesis, translation and cell division. Superoxide dismutase (SOD2) was up-regulated in this group as well (Supplementary Table C4). When exogenous nitrogen was depleted, chlorophyll content decreased suggesting a chlorophyll recycling strategy (utilization of intracellular nitrogen pool) may take place in PW95 to sustain metabolic activity and cell maintenance during cultivation. The use of intracellular nitrogen (from chlorophyll) resulted in an increased level of light-harvesting coupled with carbon

fixation, glycolysis and TCA to provide energy for amino acid synthesis (cysteine and methionine, glycine, serine and threonine metabolism), protein synthesis (7 ribosomal proteins genes with increased expression) and cell division (Supplementary Table C4). Although there was no apparent increase in cell numbers at late stationary phase, the levels of up-regulation may correspond to biomass turnover for an active algal culture with static biomass cell levels.

Differential Expression for Exponential-Phase in Filtered Water

The comparison between the exponential and stationary phases of algal growth in filtered water (E-F vs S-F) showed the highest number of significant DEGs (227), but up- and down-regulation was observed in equal proportions (Supplementary Table C5). The majority of up-regulated DEGs were housekeeping genes for DNA replication and chromosome associated proteins such as G2/mitotic-specific cyclin-B, DNA topoisomerase II, kinesin family member 11 (2 genes), structural maintenance of chromosome 4, ribonucleoside-diphosphate reductase subunit M1 (purine, pyrimidine metabolism). DNA repair and recombination proteins (DNA topoisomerase I, methyl-CpG-binding domain protein 4), processing of pre-mRNAs (heterogeneous nuclear ribonucleoprotein F/H), transfer RNA biogenesis (methionyl-tRNA synthetase, tRNA modification GTPase) and ribosome biogenesis proteins (ribosomal RNA-processing protein 8) and a nuclear chaperone involved in maturation and nuclear export of pre-60S ribosome subunits and expression regulation of genes related to plant growth and development in *Arabidopsis thaliana* (midasin (MDN1)) (Li *et al.* 2016). More than 30 down-regulated genes were ribosome proteins and proteins associated with the

spliceosome, the ubiquitin mediated proteolysis system, histones, messenger RNA biogenesis; potentially involved in maintaining a basal cellular state.

Transcripts involved in amino sugar metabolism were up-regulated at late stationary phase in the algae grown in filtered water. Three up-regulated prolyl 4-hydroxylase (P4HA) genes were observed, but the role of these presumptive proteins is unknown in PW95. This enzyme is responsible for catalyzing the post-translational formation of 4-hydroxyproline in -Xaa-Pro-Gly- sequences in proline-rich peptide sequences of plant (glycol)proteins and other proteins. Hydroxyprolines are important constituent of many plant cell wall glycoproteins, and *Chlamydomonas reinhardtii* cell wall structure is highly similar to plants (Imam *et al.* 1985; Domozych *et al.* 2012). Also related to cell structure and integrity, higher expression levels were observed for the sterol 14-alpha-demethylase (CYP51) involved sterol biosynthesis, as well as the squalene epoxidase (SQE), a novel alternative enzyme that catalyzes the first oxygenation reaction in the steroid biosynthetic pathway. Steroids modulate the fluidity and flexibility of cell membranes. SQE was recently discovered by Pollier et al. (2019) and despite how well conserved it is in animals, plants and fungi it has only been found in *P. tricornutum* (Pollier *et al.* 2019) and now it has shown to be presumptively transcribed in PW95. PW95 has been phylogenetically characterized as a *Chlamydomonas*-like alga and there is no report of such enzyme in *Chlamydomonas reinhardtii. C. reinhardtii* has a fully annotated, manually curated genome and is a well-established model organism. It has proven to be a powerful model for dissecting fundamental processes in biology not only in algae but has improved the understanding of these processes in plants and animals

(Sasso *et al.* 2018). Among algae, PW95 and *Phaeodactylum tricornutum* have a unique capacity to perform steroid biosynthesis in an alternate way to potentially modulate the fluidity and flexibility of cell membranes maintenance/survival strategy.

A presumptive urea-proton symporter (DUR3) and two solute carrier family 25 mitochondrial transporter proteins (UCP2_3, SLC25A32) that create proton leaks across the inner membrane (uncoupling oxidative phosphorylation from ATP synthesis) were down-regulated. Energy dissipation in the form of heat is a defense mechanism under stress conditions such as high temperature and light oversaturation and is known as non-photochemical quenching (Gerotto *et al.* 2012). The down-regulation of the two solute carrier family 25 transporter proteins might be an indication that PW95 was not stressed under the tested light and temperature conditions.

Differential Expression in Exponential-Phase in Unfiltered Water

The comparison between the exponential and stationary phases of algae grown in unfiltered water (E-UF vs. S-UF) showed 129 significant DEGs, also up- and down-regulated in equal proportions as observed previously for the filtered water differential expression (Supplementary Table C6). This was the only group comparison that revealed the expression of a carbon concentrating mechanism (CCM) in PW95, where the expression levels of a carbonic anhydrase (CAH) were higher at late exponential phase relative to late stationary growth. The up-regulation of the carbonic anhydrase suggests a potential role of biophysical CCM at nutrient depletion, increasing ambient level CO_2 flux to RubisCO by facilitating bicarbonate transport into the chloroplast (Valenzuela *et al.* 2012). This could explain why PW95 had a higher average biomass yield and in turn

better growth in the unfiltered water. The increased metabolic activity at late exponential phase was reflected in the up-regulation of DNA replication and chromosome associated proteins (G2/mitotic-specific cyclin-B, DNA topoisomerase II, kinesin family member 11 (2 genes), tubulin beta (TUBB), centrin-1 (CETN1)), transcription regulation (*rpaA* two-component system), protein processing and export (HSP20 family protein, chaperones), amino acid metabolism (prolyl 4-hydroxylase, 3-hydroxyisobutyrate dehydrogenase) purine and pyrimidine metabolism (ribonucleoside-diphosphate reductase subunit M1, urate oxidase (uaZ)) and phosphopantothenoylcysteine decarboxylase (PPCDC) responsible for a key decarboxylation step in the synthesis of coenzyme A (CoA). It was proposed by Avidan et al. (2015) proposed that high TAG levels in green algae during nitrogen starvation are dependent on the organism ability to allocate carbon flow towards acetyl-CoA biosynthesis, and were able to correlate increased levels of acetyl-CoA (acetylated CoA) to TAG accumulation in *Chlorella desiccata, Dunaliella tertiolecta* and *Chlamydomonas reinhardtii*. Malonyl-CoA and free CoA levels also increased, but to lesser extents (Avidan *et al.* 2015). PW95 lipid production in unfiltered water showed the lowest total FAMEs (13.42%). The up-regulation of phosphopantothenoylcysteine decarboxylase could be an indication that unacetylated (free) CoA was the predominant species and that carbon flow towards the synthesis of acetyl-CoA was poor at that time point, resulting in the observed low TAG accumulation and low FAMEs.

Conclusions

The present study was designed to mimic oligotrophic environmental in a native low-quality wastewater in photo-bioreactors. The CBM water contained a native

microbial community and PW95 was cultured in filtered and unfiltered CBM water. In unfiltered water, PW95 showed significantly higher average biomass concentration and growth rate in comparison with filtered water. Given that the filtered CBM water had a reduced microbial population due to filtering, PW95 cultivated in unfiltered CBM water might have established diversity-productivity interactions with the accompanying microbial consortia that had a positive impact on growth.

A correlation between low growth and high lipid production was observed in this system and differentially expressed genes could be correlated to this growth condition. PW95 showed less growth in filtered water but showed a significantly higher rate of TAG accumulation and total FAMEs (23%) at stationary phase. Total FAMEs produced by algae in filtered water at stationary phase was comparable to previous reports of CBM water algal cultivation in Australia (20–24%) and when grown in supplemented CBM water (from the same source; 20–27%). The FAMEs profile can be variable depending on culture conditions but PW95 could be a potential source of C18/C16 feedstock, since it produces predominantly those FAME species desirable for biofuel applications.

Although there was a high percentage of non-assigned (NA) significant DEGs (63.6% of genes between the growth phases and 42.8% between the types of water) and a low percentage of reads mapping back to the genome (5%), significant differential gene expression was observed, and meaningful comparisons were established to better understand the physiological responses of PW95 in a low-quality water with low and high microbial diversity.

A common trend observed in PW95 cultures both in standardized media and CBM water is the nitrogen starvation-induced chlorophyll degradation, the cells might have used the chlorophyll as a nitrogen intracellular pool, suggesting a cyclic chlorophyll recycling strategy could be taking place in PW95 for survival.

Novel biochemical capacity was discovered in PW95 (novel alternative pathway for steroid biosynthesis with DE of the squalene epoxidase gene) and insight into basic metabolic functions and adaptive physiological responses to environmental and intracellular cues in green microalgae was gained. Algal biomass production in CBM production water represents an alternative strategy for biomass and biofuel production that utilizes a low-quality water source for the Rocky Mountain north. Our findings build on previous knowledge for improved algal culturing for biomass and industry valued products while exploring the biology of an organism with relevant impact in energy and water resources management.

An inverse relationship has been observed between biomass production and lipid accumulation in PW95 in previous studies in different culture conditions (Corredor *et al.*, in review). In the current study, the low percentage of total FAMEs observed in the unfiltered water (faster growth) versus higher FAME (and slower growth) for filtered water displayed the same inverse relationship. The unfiltered water overall had better growth as well as lower FAMEs, and the unfiltered water condition had higher microbial diversity (Chapter 5). These outcomes could be associated with differentially expressed gene modes related to nitrate transport, central carbon metabolism, β-oxidation, and fatty acid biosynthesis. Further work is needed to delineate the underlying mechanism(s) of

improved biomass versus lipid accumulation of PW95 and the corresponding genotype-phenotype relationship in the presence of bacterial communities with high and low diversity.

Acknowledgments

This work was supported by the National Science Foundation Sustainable Energy Pathways (SEP) Program under NSF CHE-1230632 and the Department of Energy PEAK program under DE-EE0008247. The authors would like to thank the Montana Bureau of Mines and Geology for providing access to the FG-09 well and the MSU CBM group for water sample collection. The MSU Algae Group for support and constructive exchange of ideas. Dr. Ben Wheaton for graphical and editorial assistance.

CHAPTER FIVE

ASSOCIATIONS BETWEEN SYMPATRIC BACTERIAL GROUPS AND A NOVEL GREEN ALGA CULTIVATED IN COAL BED METHANE PRODUCTION WATER

Contribution of Authors and Co-Authors

Manuscript in Chapter Five

Author: Luisa Corredor
Contributions: Experimental design, performed experiments, literature review, synthesized ideas, performed data, statistical and bioinformatical analyses, wrote and edited the manuscript.

Author: Anna J. Zelaya
Contributions: Performed statistical analysis, assistance with bioinformatical analyses and edited the manuscript.

Co-Author: Robin Gerlach
Contributions: Procured funding, synthesized ideas, reviewed and edited the manuscript.
Co-Author: Matthew W. Fields

Contributions: Principal Investigator, experimental design, synthesized ideas, wrote and edited the manuscript.

Manuscript Information Page

Luisa Corredor, Anna J. Zelaya, Robin Gerlach, Matthew W. Fields

Status of Manuscript:
__X__Prepared for submission to a peer-reviewed journal
____Officially submitted to a peer-reviewed journal
____Accepted by a peer-reviewed journal
____Published in a peer-reviewed journal

Abstract

Algal biomass production in coal bed methane (CBM) production water represents a potentially useful strategy for cooperative biofuel energy production and water management. Algal polycultures may be more stable and productive than monocultures, and different algal species most likely form variable interactions with different microorganisms over time and space dependent upon environmental conditions. Culture conditions with controlled ecological parameters could enable higher and/or more stable biomass productivity and/or lipid accumulation for industry-valued algal products. The current study investigated the microbial communities associated with a *Chlamydomonas*-like green microalga (PW95) during biomass production in CBM production water using a closed system. The PW95 strain has been reported as a biofuel producer in CBM water and can serve as a model for predicting potential microbial interactions between an indigenous species and native microbial groups for an algal strain that naturally aggregates. While nitrate was consumed at similar rates in filtered and unfiltered CBM water, cultivation in filtered water resulted in elevated FAME content. Diversity and population dynamics were examined via paired-end sequencing of SSU rRNA gene sequences over time and space for CBM production water, the free planktonic fraction during algal growth, and the algal aggregate fraction at late exponential- and stationary-phases of growth in indoor photobioreactors. Both electron micrographs and fluorescent microscopy demonstrated microbial cells in close proximity to algal cells. Through these comparisons, enriched bacterial populations were identified in algal aggregates under nitrate-deplete conditions. Filtered versus unfiltered water had

the largest impact on bacterial populations associated to the algal aggregates or planktonic water phase in terms of species richness, diversity and evenness and for example, Chao species richness estimates were 4.5x and 3.8x higher for algal aggregates cultivated in unfiltered water under both growth phases, respectively. Under filtered water conditions, the algal aggregates maintained approximately 40 bacterial OTUs, whereas under unfiltered conditions the algal aggregates maintained approximately 180 bacterial OTUs. Sequences indicative of *Caulobacteraceae, Rhizobiaceae*, *Comamonadaceae, Erythrobacteraceae*, *Sphingomonadaceae* and *Phyllobacteriaceae* were the most abundant bacteria in close association with PW95 algal aggregates. PW95's microcosm is composed of the species *Truepera sp., Chelativorans sp., Mycoplana sp., Yonghaparkia sp., Roseomonas frigidaquae, Beijerinckiaceae* (unassigned), *Agrobacterium sp.,* Proteobacteria (unassigned) and Bacteria (unassigned). Results from this study give insight into algal-bacterial interactions and the potential role of bacterial communities in a particular ecological niche designed for algal biomass production.

Introduction

Water concerns are attracting more attention to water resource use prior to and following energy development for both traditional and non-traditional sectors. Energy-related water consumption (withdrawn from its source and not directly returned) accounts for about 10% of both total and freshwater water consumption in the United States (3.5 trillion gallons of freshwater per year) (Grubert and Sanders 2018). A major consumer of fresh water is coal bed methane (CBM) production, and the large amounts of extracted

water (millions of gallons/well/day) stored in impoundments can negatively impact the environment and the economy (Hodgskiss *et al.* 2016; Grubert and Sanders 2018). Therefore, energy and water needs present opportunities for the coordination of alternative energy production with treatment of different water streams, and coupling algal biomass production with low quality water is a unique opportunity for the cooperative management of water with both traditional and non-traditional and energy resources (Fields *et al.* 2014; Hodgskiss *et al.* 2016; Grubert and Sanders 2018).

Microalgae are often associated with diverse microbial communities in native aquatic ecosystems as well as in experimental and larger-scale cultivation schemes (Lakaniemi *et al.* 2012; Krohn-Molt *et al.* 2013), but few studies have shown a detailed phylogenetic characterization of the associated microorganisms in microalgal production systems (Krohn-Molt *et al.* 2013; Lian *et al.* 2018). Polycultures (ecosystems) may be more stable and productive than monocultures and previous reports of natural and constructed algal assemblages show increasing diversity associated with higher productivity (Stockenreiter *et al.* 2012; Beyter *et al.* 2016; Bell *et al.* 2019a).

Algal cultures include complex bacterial communities that are often considered to be bacterial contamination (Fulbright *et al.* 2018), but aside from playing a potential role in stimulating and supporting microalgal productivity, associated bacteria can help the algae perform critical tasks (Lian *et al.* 2018). Microbial communities can play a role in biogeochemical cycling, micro-food web structure, stimulating algal growth by producing essential nutrients (*e.g.*, vitamins, phytohormones), exhibiting algicidal activity, and regulate algae-bacteria signaling (such as quorum sensing) (Adachi *et al.*

2003; Krohn-Molt *et al.* 2013; Lian *et al.* 2018; Zhou *et al.* 2018). The identification of microbial communities closely associated with microalgae under different cultivation schemes can help understand potential algae-bacteria functional relationships for improved-productivity in algal systems in the future (Lakaniemi *et al.* 2012; Krohn-Molt *et al.* 2013; Fields *et al.* 2014; Sambles *et al.* 2017). Most studies have surveyed bacterial communities in water treatment studies and facilities (Bell *et al.* 2019b; Shetty *et al.* 2019) or algal blooms (Xiao *et al.* 2017; Zhou *et al.* 2018), but knowledge is lacking in the context of microalgal production systems, such as photobioreactors (PBRs). Moreover, basic research is lacking in systems implementing environmental isolates with growth and biofuel potential in low quality water, such as the current study.

The current study investigated the microbial community associated with a novel green sporulating alga (*Chlamydomonas*-like, designated PW95) isolated from a CBM water impoundment, which has been reported to produce biomass and biofuel in monoculture and batch mode, using either standardized media or low quality CBM water (Corredor *et al.*, in review; Hodgskiss *et al.* 2016). The aim of this investigation was to characterize the microbial community during algal biomass production in CBM water using a close system, to understand differences in species richness and diversity, to identify dominant and abundant species closely associated with the alga (phycosphere), and to determine if specific community assemblages or shifts take place during different phases of growth (exponential and stationary) or in different water types (filtered and unfiltered CBM water).

Materials and Methods

Coal Bed Methane Production Water

PW95 (*Chlamydomonas*-like strain) was isolated from a CBM production pond in the Powder River Basin (northeastern Wyoming and southeastern Montana). CBM water samples were collected from the FG-09 well (Flowers-Goodale coal bed operated by the Montana Bureau of Mines and Geology) in the Powder River Basin, prior to performing the current investigation and kept at 4°C until the photobioreactors were set-up. Supplementary Table D1 shows the water chemistry analysis report for the CBM used in this study.

Microscopy Analysis of PW95 Cultures

PW95 stock cultures were analyzed using confocal microscopy and electron microscopy (EM). Confocal microscopy (Leica TCS-SP5) was performed using 1 mL of culture incubated in the dark with equal volume of SYBR green (10X working solution, Invitrogen) for 10 minutes then filtered (PC membrane 0.2 μm, Whatman). The filter was washed with ultrapure (18.2 Ω) deionized water (dH_2O) and observed under the microscope. Field Emission-Scanning Electron Microscopy (SUPRA 55VP FE-SEM, Zeiss) was performed using 5 mL of a centrifuged culture (14,500 x g). The pellet was washed and resuspended in ultrapure (18.2 Ω) dH_2O. The cell suspension (30 μL) was placed on an EM chip, allowed to dry overnight, and the chip was sputter-coated with Iridium the next day (without critical point drying).

Experimental Culture Conditions

Cultures of PW95 were grown in 500 mL of Bold's basal medium (BBM) in 1000 mL conical flasks using rubber/filter stoppers that allowed gas exchange. Cultures were incubated at 30°C with artificial light (33 μmol photons m^{-2} s^{-1}, approx. 1435 lux output/tube) on a 14:10 light/dark (L/D) cycle and mixed constantly with a magnetic stirrer until the exponential phase of growth occurred. The biomass was concentrated via centrifugation at 4800 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA). Biomass pellets were resuspended in 250 mL of sterile nanopure water and used as inoculum for the PBR tubes (10% V_T).

Experiments were conducted in triplicate batch cultures using six glass tubes (70 × 500 mm) containing 1.25 L of CBM water as medium initially at pH 8.4, and amended with 0.5 mM of sodium nitrate (Sigma-Aldrich, St Louis, MO, USA) and 0.185 mM of boric acid (Sigma-Aldrich, St Louis, MO, USA) as previously described (Hodgskiss *et al.* 2016). Tubes 1 to 3 contained CBM water filtered using a 0.2 μm bottle top PES filter (Corning, USA) and tubes 4 to 6 contained unfiltered CBM water. All experimental cultures (PBR tubes) were incubated side-by-side submersed in a controlled water bath maintained at 30°C ± 1. Rubber stoppers, containing ports for aeration and sampling, were used to seal the tubes. Light (33 μmol photons m^{-2} s^{-1}, approx. 1435 lux output) was maintained on a 14:10 L/D cycle using a light bank containing eight T5 LED tubes (Waveform PhotonTube™, 6500K, 95 CRI and R9>80) to simulate natural sunlight. Aeration (400 mL min^{-1}) was supplied by humidified compressed air bubbling through Nalgene bottles containing ultrapure (18.2 Ω) dH_2O and controlled using individual

rotameters for each PBR tube (Cole-Parmer, USA). Tubes were constantly mixed using a magnetic stirrer. The experiment had a duration of 8 days the last samples was removed during the late stationary phase (N deplete conditions).

Cell and Lipid Content Analysis

Culture samples were collected every 24 hours before the end of the light cycle and subsequently used for the analysis of physiological parameters. Culture pH was measured using a standard pH meter (Hach, Sension+PH1 Portable pH meter). Growth was evaluated by optical density (OD) and dry cell weight (DCW). 1 mL of algal culture was collected and a subsample (200 μl) was used to measure OD at 750 nm in a Synergy H1 hybrid fluorometer/spectrophotometer reader (Bio-Tek). DCW was determined by filtering culture samples using GF/F Glass Microfiber filters (Whatman, USA). Samples were weighed after drying at 90°C for 48 h to ensure the water was evaporated. Chlorophyll was extracted using cold 100% methanol technique and its fluorescence assessed at 775 nm, 652.5 nm and 665 nm using a triplicate plate assay measured using a Synergy H1 hybrid fluorometer/spectrophotometer reader (Bio-Tek) (Hodgskiss *et al.* 2016). Total chlorophyll concentrations (Chl-a plus Chl-b) were calculated according to Ritchie (Ritchie 2006). Samples were centrifuged at 14000 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, USA) and supernatants for water chemistry analysis were stored at −80°C. Supernatants were used to track nitrate concentration changes by ion chromatography using an AS22 Anion-Exchange Column (DIONEX ICS-1100) with a 4.5 mM NaHCO3/ 1.4 mM Na2CO3 buffer as eluent at a 1.2 mL flow rate with an ASRS 4 mm suppressor. Chromeleon Chromatography Management System was

used to determine the concentration of nitrate in each sample throughout the growth curve.

Lipid content in the form of triacylglycerol (TAG) was measured using Nile Red (NR) fluorescence as previously described (Cooksey *et al.* 1987), by exposing 1 ml of culture to 4 µl of NR for 10 minutes (optimal exposure time was determined using a time course assay as previously reported (Chen *et al.* 2009)). Total NR fluorescence was determined at an excitation wavelength of 530 nm and an emission wavelength of 575 nm with a Synergy H1 hybrid fluorometer/spectrophotometer reader (Bio-Tek) and Gen5 software. Fatty Acid Methyl Esters (FAMEs) were quantified to assess biodiesel potential of PW95. Biomass was harvested and pelleted as described previously (Lohman *et al.* 2013). Harvesting was done at late exponential and stationary phase in order to compare PW95 lipid accumulation in two different phases of growth. Direct *in situ* transesterification was used on lyophilized biomass of PW95 (Pedersen *et al.* 2018) and FAMEs were quantified by gas chromatography–mass spectroscopy (GC-MS) as described previously (Lohman *et al.* 2013).

Statistical analyses of the physiological responses and the lipid production in PW95 were performed in Minitab® Statistical Software v18 (http://www.minitab.com). A Two-way analysis of variance (ANOVA) using a Linear Mixed Effects Model (LMM) with a 95% significance level was fitted to all data for each response separately (pH, OD750, DCW, chlorophyll accumulation, nitrate depletion, NR fluorescence and % of total FAMEs), to analyze the differences between filtered and unfiltered water and the

interaction with time (days). Tukey's follow-up tests assessed pairwise comparisons of the mean response among the evaluated conditions.

Microbial Community Analyses

Three types of samples for microbial community analysis were collected for this experiment. 1) Algal aggregates (samples designated AG) were collected and pelleted separately from the water, to identify closely associated microbial community members and to infer potential algal/bacterial/archaeal relationships. A sub-sample (100 mL) of algal culture from each PBR was collected in sterile Falcon tubes and allowed to settle (10 minutes) and the supernatant decanted (20 mL). The algal aggregates and associated microbial community were further concentrated via centrifugation at 4,800 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA) and the leftover supernatant was discarded. 2) The water (samples designated W) that was removed during the collection of algal aggregates was collected to identify planktonic bacteria/archaea independent from the alga. Collected water was filtered through an 8.0 μm MCE membrane (Merck Millipore, Ireland) to further discard any remaining algal biomass. Filtrate water samples were pelleted by centrifugation at 14,000 x g for 10 min (Thermo Scientific, Sorvall Legend XTR, Waltham, MA). 3) The filter membrane (0.2 μm PES membrane bottle top filter (Corning, USA)) used to initially filter the CBM water was collected to survey the native microbial communities present in the water prior to the inoculation of the alga. All samples were collected at two different time-points (late exponential and late stationary-phase of growth) and stored at -80°C for DNA extraction.

DNA Extraction, PCR and Sequencing

DNA was extracted from approximately 50 mg of biomass (algal aggregates or filtered water) or directly from the filter membrane (CBM water control sample or time zero) with a FastDNA Spin Kit for Soil (MP Biomedical). The bacterial and archaeal SSU rRNA genes were amplified using modified universal primers targeting the V4 region (Forward primer, 515F, 5'- GGACTACNVGGGTWTCTAAT-3'; reverse primer, 806R, 5'- GTGYCAGCMGCCGCGGTAA-3') (Apprill *et al.* 2015; Parada, Needham and Fuhrman 2016). PCR amplification was performed in triplicate with the following cycle program: an initial denaturation at 95°C for 5 min; 25 cycles at 98°C for 20 s, 58°C for 15 s and 72°C for 30 s; and a final extension at 72°C for 5 min. A 1% agarose gel in TAE buffer was used to assess the correct size of the PCR products. The PCR amplicons were prepared for sequencing according to the "16S Metagenomics Sequencing Library Preparation" Illumina protocol for paired-end sequencing (Illumina, San Diego, CA, USA). Following PCR clean-up, purification, and indexing, DNA concentrations were determined using NEBnext® Library quant kit for Illumina (NEB) for qPCR-based library quantitation. DNA concentrations were normalized and pooled with a 15% PhiX control library. Sequencing was performed on an in-house Illumina MiSeq™ using a 600-cycle Reagent kit v3 (MS-102-3003).

Sequence Data Analysis and Taxonomy Assignment

Bioinformatic analysis of the raw sequence data (demultiplexed reads) was processed using QIIME2 2019.4 (Bolyen *et al.* 2019). DADA2 was used for filtering, dereplication, chimera identification, and merging paired-end reads (Callahan *et al.*

2016). Taxonomy was assigned to amplicon sequence variants (ASVs) using the q2-feature-classifier (Bokulich *et al.* 2018) classify-sklearn naive Bayes taxonomy classifier against the Greengenes 13_8 database (99% OTUs reference sequences) (McDonald *et al.* 2012). Sequences classified as “Chloroplast”, “Mitochondria” and “Unassigned” were removed from the dataset, as well as singletons.

Data Analysis

Initial diagnostics for the representative ASVs in the microbial community analysis were done via q2-diversity after samples were rarefied (subsampled without replacement) using a sampling depth of 42812. At the specified sampling depth 1,070,300 (38.09%) features were retained in 25 (78.13%) samples. The α-diversity metrics were estimated using observed OTUs, Chao, Simpson, Shannon and Simpson evenness indices.

Analysis of variance (ANOVA) using a Bray-Curtis distance metric was performed with Adonis (999 permutations) using R statistical package VEGAN version 2.4-3 (Oksanen et al. 2013, https://CRAN.R-project.org/package=vegan) to generate non-metric multidimensional scaling (NMDS) plots using ggplot2 (Wickham 2016). ANOVA was performed to determine mean differences between sample type (aggregates vs. water), water type (filtered vs. unfiltered) and growth phase (exponential vs. stationary). All analyses included the filter membrane sample (designated CBMW) to assess all the mean differences of the samples against the native microbial communities present in the water before the inoculation of the alga.

Principal component analysis (PCA) in Canoco v4.5 was used to evaluate the interspecies correlations with the groups of samples. No data transformation was necessary, and species were centered and standardized (Lepš and Šmilauer 2003). Following the default parameters set by Segata et al. (Segata *et al.* 2011), linear discriminant analysis effect size (LEfSe) was performed to identify distinctive and abundant populations when comparing the algal aggregate samples between filtered water versus the unfiltered water at two different time-points throughout the growth curve (late exponential versus late stationary phase).

Results

Microscopic Observation of Bacteria in PW95 Cultures

Confocal microscopy of PW95 cultures show bacteria stained with SYBR green surrounding PW95 cells in red (chlorophyll autofluorescence). Electron microscopy showed algal vegetative cells with coccoid and rod-shaped bacteria in close association with the alga or dispersed in the area (Figure 1). Aggregate structures of algal and other microbial cells were approximately 20 μm x 15 μm and appeared to be densely packed (Figure 1).

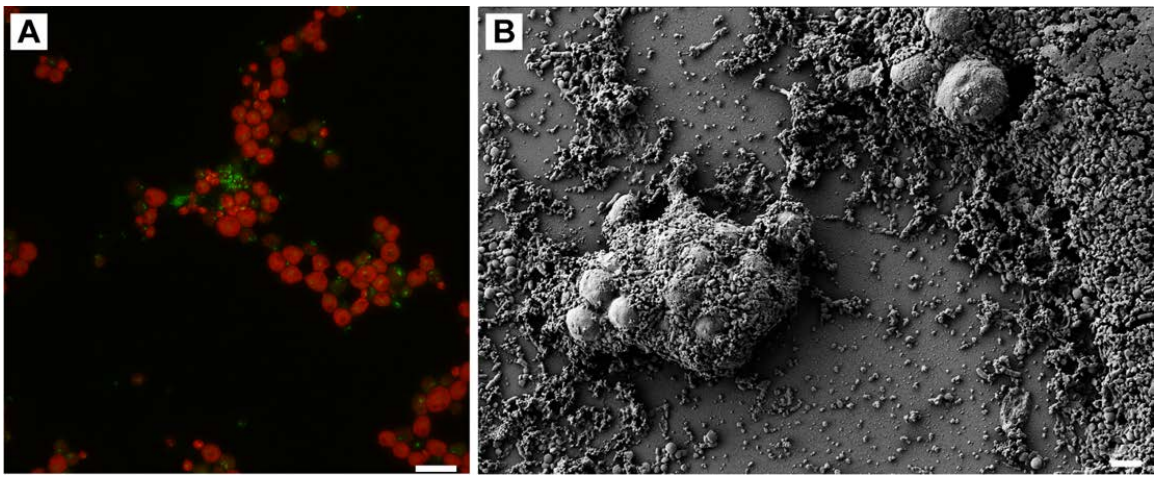


Figure 1. PW95 stock cultures with accompanying bacteria. (A) Confocal microscopy image at 20 μm scale and 100X magnification, showing bacterial cells in green fluorescence and PW95 cells displaying chlorophyll autofluorescence in red. Fluorescence was excited and collected using the following laser lines and emission parameters: Syto9, ex 488 nm (argon laser), em 500 to 550 nm (green signal) and autofluorescence of photosynthetic cells ex 633nm (HeNe laser), em 650 to 750 nm (red signal). (B) FE-SEM image at 5.54x magnification and 3 μm scale, showing gross morphology of algal vegetative cells with coccoid and rod-shaped bacteria in close association with the alga or dispersed in the area.

PW95 Growth

During the course of this study, cell density and biomass production were measured using optical density (OD^{750}) and dry weight (DW) (Figure 2A & B). Between filtered and unfiltered water, the average cell density and biomass production throughout the growth curve was significantly different (*p=0.00*) as well as the maximum growth rates based upon optical density (filtered=0.043 d^{-1} and unfiltered=0.059 d^{-1}) or cell dry weights (filtered=0.060 d^{-1} and unfiltered=0.098 d^{-1}) (*p=0.00*). The unfiltered water showed the highest mean response and the filtered water showed the lowest, placing them in statistically different groups (Tukey's, *p=0.00*).

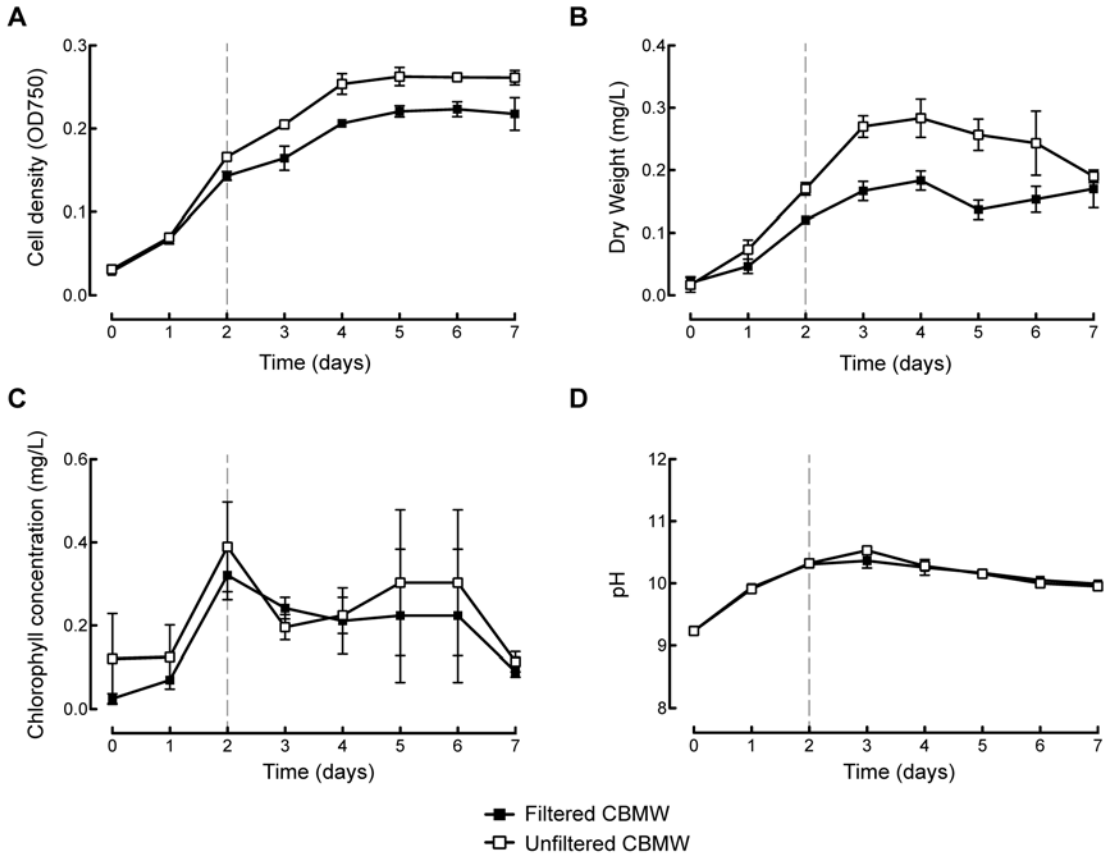


Figure 2. Physiological responses of PW95 during cultivation in amended coal bed methane production water (0.5 mM nitrate, 0.183 mM borate) at 30°C and artificial sunlight in PBRs. Growth is shown in terms of cell density (A) and biomass production (B). Chlorophyll accumulation (C) and pH (D) are indicators of carbon fixation. The time of nitrate depletion is indicated by the dashed line at day 2. Filled symbols represent the filtered water (F) and clear symbols represent unfiltered water (UF). Biomass collection for RNA sequencing was done at day 3 and day 7. The mean of three independent replicates with the standard deviation is displayed in each graph.

Chlorophyll production was used to assess culture health and as a qualitative measure of photosynthesis (Figure 2C) (Moll *et al.* 2014; Zhao *et al.* 2017). Statistical differences were not observed ($p>0.05$) in chlorophyll rate or average accumulation between filtered and unfiltered water, although chlorophyll accumulation was higher in

this study using CBM water at 30°C, artificial sunlight, ambient air in PBRs than reported previously for growth in CBM water at 20°C, fluorescent light and constant shaking for gas exchange in flasks (Hodgskiss *et al.* 2016). The average rate of chlorophyll increase was similar for both experimental conditions (filtered=0.25 d^{-1}, unfiltered=0.26 d^{-1}), and the peak of chlorophyll accumulation (filtered=0.32 mg/L, unfiltered=0.39 mg/L) was at day 2. After day 2 a decrease in chlorophyll accumulation, correlated to nitrate depletion, was observed in both water types. Culture pH was monitored daily to assess the changes in the media due to nitrate uptake, carbon fixation and photosynthesis (Figure 2D) (Gardner *et al.* 2013). There was no significant difference in pH between filtered and unfiltered water ($p>0.05$). As expected, the pH increased during exponential phase of growth (days 0–4) between 9 and 10.5 and plateaued entering stationary phase (day 4) until the last day (day 8) of sampling.

PW95 Lipid Production in CBM Water

Lipid accumulation in algae is frequently measured by monitoring TAG accumulation using Nile Red (NR) fluorescence. An increase in TAG accumulation was expected with the onset of nitrate depletion, and the NR fluorescence started to increase on day 1 and continued through late exponential phase (day 4) that corresponded to the maximum peak in NR fluorescence (Figure 3). After this time-point, a decrease in NR fluorescence was observed regardless of water type, though the decrease was more evident in unfiltered water by late stationary phase (day 7). Overall, NR fluorescence was lower than reported in previous studies of PW95 (Hodgskiss *et al.* 2016) and there was no difference in the average NR fluorescence between treatments, placing filtered and

unfiltered water in the same statistical group ($p>0.05$). Nonetheless, the filtered water had a higher rate of TAG accumulation that was statistically significant ($p=0.03$) and the NR fluorescence trended higher in filtered water (Figure 3).

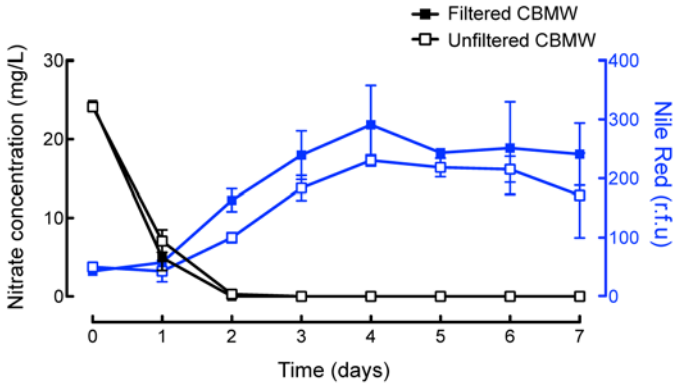


Figure 3. Characterization of lipid accumulation in PW95 showing increase in Nile Red fluorescence with respect to nitrate depletion. The left y-axis (black) shows the depletion of external nitrate concentration as a function of time (x-axis). The right y-axis (blue) shows Nile Red fluorescence intensity indicating the increase in TAG accumulation. Filled symbols represent the filtered water (F) and clear symbols represent unfiltered water (UF). The mean of three independent replicates with the standard deviation is displayed in each graph.

Total FAMEs detected during late-exponential phase of growth were similar between filtered and unfiltered water (15.4% and 16.0%, respectively) ($p>0.05$) (Figure 4). At late stationary phase of growth, filtered water showed the highest total FAMEs (23.1%) and unfiltered water showed the lowest total FAMEs (13.4%) (Figure 4); this observation can be correlated with the decline in NR fluorescence at day 7 in unfiltered water (Figure 3). The dominant species of FAMEs were saturated and unsaturated C16 (hexadecanoic) and C18 (octadecanoic) chain lengths; saturated and unsaturated C20

chain lengths were also present across all samples but in lower amounts than the C16 and C18 FAMEs (Supplementary Table D2).

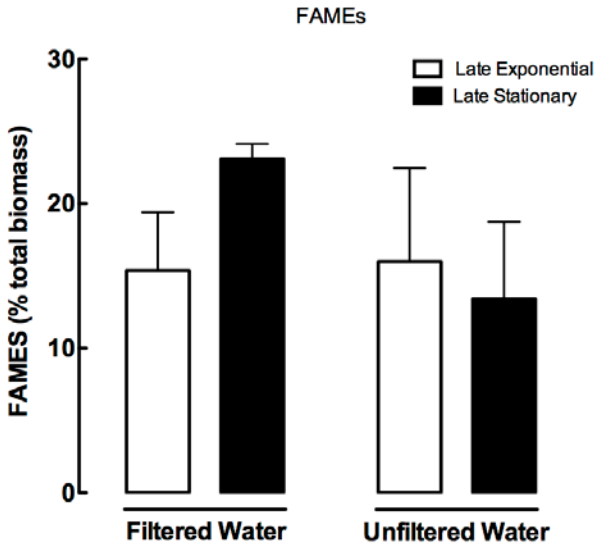


Figure 4. Lipid content of PW95 grown in CBM water at late exponential and late stationary phase of growth. Comparison of total FAMEs between filtered and unfiltered water. Biomass collection for GC-MS analysis and RNA sequencing was done at day 3 (late exponential) and day 7 (late stationary). The mean of three independent replicates with the standard deviation is displayed in each graph.

Microbial Diversity and Taxonomic Composition in the Algal Photobioreactors

Temporal samples (n=25) were sequenced and 1,261 taxonomic features were identified across the experimental conditions evaluated (type of water and type of sample) during both exponential and stationary phases of growth. A total of 1,217 (96.5%) features passed quality filtering and were used for the analysis. Unique taxa (n=537) were identified and classified by QIIME2. Almost all evaluated samples contained sequences corresponding to bacterial taxa only. Archaeal ASVs were found in

very low abundances (<0.001%) for just two samples: CBMW (native community prior to algal inoculation) and SW6 (unfiltered water taken during stationary phase of growth).

Table 1 shows initial diagnostics performed with QIIME2 consisting of pairwise comparisons within each experimental group (filtered vs. unfiltered water, algal aggregates vs. water and exponential vs. stationary phase) to assess species richness and evenness. Based on observed OTUs ($p= 0.000096$), Chao ($p= 0.000097$), Simpson ($p= 0.0080$) and Shannon ($p= 0.0005$) indices, overall richness was significantly higher by water type (unfiltered vs. filtered). The type of sample (aggregates vs. water) and the growth phase (exponential vs. stationary) had no significant effect on richness.

Table 1. Kruskal-Wallis test pairwise comparisons based on alpha diversity metrics between the evaluated conditions within the groups. Comparison of statistical differences between means using q-values (adjusted *p*-values) as reported by QIIME2. Statistical significance is designated with a star symbol (*).

Groups	Pairwise comparisons	Observed OTUs	Chao	Simpson	Shannon	Simpson evenness
Water type	Filtered vs. unfiltered	0.000096*	0.000097*	0.0080*	0.0005*	0.0005*
Sample type	Aggregates vs. water	0.5635	0.4529	0.9539	0.8624	0.6340
Growth phase	Exponential vs. Stationary	0.8624	0.8173	0.1088	0.1814	0.3570

As expected, unfiltered CBM water richness was higher than the filtered CBM water (Table 2) and that was confirmed by the higher number of OTUs present in unfiltered water samples and the corresponding richness indices per sample (Table 2).

The number of distinct features (observed OTUs) was close to the Chao index number for each sample, indicating that the sampled diversity was very similar to the probable species richness estimated for the samples. Shannon and Simpson indices confirm that the sample with the highest richness was the CBM water, representing the

native microbial community in the water followed by samples from unfiltered water, and filtered water samples had the lowest richness. On the other hand, diversity in each sample was measured using Simpson evenness index, were the CBM water has the lowest evenness index indicating higher diversity, followed by unfiltered water samples, with the filtered water samples ranking as the least diverse (highest evenness indices).

To further investigate these differences between groups of samples, NMDS plots were generated using Bray-Curtis dissimilarity (Figure 5). The analysis of variance revealed a significant effect of water type and sample type on community structure (Bray Curtis dissimilarity, $p=0.001$) in the samples. The water samples (CBM water, filtered and unfiltered water) cluster statistically differently from each other (R^2=0.36484), as well as the sample type (aggregates, water, CBM filter with native community) (R^2=0.28393).

Table 2. Alpha diversity indices and number of sequences analyzed per sample. Comparison between species richness and evenness between samples. The number of sequences analyzed post-filtering is shown followed by the alpha diversity metrics performed on rarefied data. Species richness was evaluated based on observed OTUs, Chao (probable species richness based on sampled diversity), Simpson and Shannon. Simpson evenness index was included as a measure of the relative abundance of different species.

ID	Sample type[a]	Water type[b]	Number of sequences	Observed OTUs	Chao	Shannon	Simpson	Simpson evenness
CBMW	Filter	CBM	84,260	322	329	5.59	0.96	0.07
E-AG1	AG	F	111,345	39	43	3.22	0.87	0.19
E-AG2	AG	F	114,012	39	43	3.09	0.86	0.19
E-AG3	AG	F	52,219	31	31	3.19	0.87	0.26
E-AG4	AG	UF	85,998	179	184	4.29	0.88	0.04
E-AG5	AG	UF	50,521	157	158	4.46	0.88	0.05
E-AG6	AG	UF	132,695	199	205	4.41	0.87	0.04
E-W1	Water	F	83,880	71	80	3.46	0.87	0.11
E-W2	Water	F	58,657	72	73	3.31	0.83	0.08
E-W3	Water	F	65,384	72	77	3.22	0.82	0.08
E-W4	Water	UF	184,614	170	191	3.78	0.89	0.05
E-W5	Water	UF	71,102	114	117	3.75	0.89	0.08
E-W6	Water	UF	274,904	172	189	3.76	0.89	0.05
S-AG1	AG	F	42,812	40	40	3.37	0.86	0.18
S-AG2	AG	F	97,715	51	52	3.41	0.86	0.14
S-AG3	AG	F	55,179	47	47	3.52	0.87	0.17
S-AG4	AG	UF	63,131	142	142	4.77	0.94	0.11
S-AG5	AG	UF	123,701	190	196	5.01	0.94	0.09
S-AG6	AG	UF	154,030	180	190	4.69	0.93	0.08
S-W1	Water	F	69,586	58	58	3.47	0.86	0.13
S-W2	Water	F	148,606	69	83	3.67	0.89	0.13
S-W3	Water	F	113,399	78	82	3.90	0.90	0.13
S-W4	Water	UF	229,383	113	135	3.49	0.87	0.07
S-W5	Water	UF	208,547	119	122	3.71	0.87	0.07
S-W6	Water	UF	52,098	595	598	5.62	0.94	0.03

[a]AG: algal aggregates

[b] F: filtered water, UF: unfiltered water

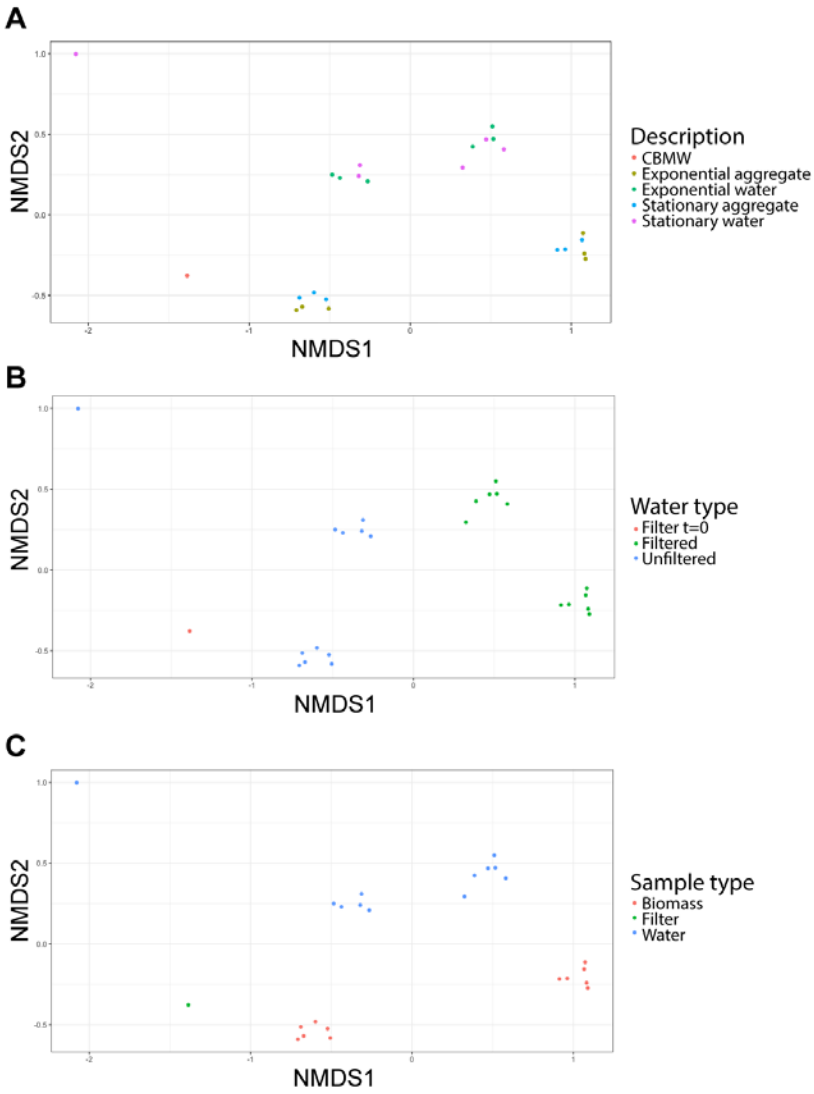


Figure 5. Non-metric multidimensional scaling (NMDS) plots using Bray-Curtis dissimilarity. Mean differences were determined between the groups: (A) Growth phase (exponential vs. stationary) using the sample description, (B) Water type (filtered vs. unfiltered), and (C) Sample type (aggregates vs. water). All analyses included the filter membrane sample (designated CBMW, filter or Filter t=0)) to assess all the mean differences of the samples against the native microbial communities present in the water before the inoculation of the alga. The NMDS reached a stable solution after 20 tries, with a stress value of 0.0447.

In contrast, there was no differential clustering observed between samples from exponential phase vs. stationary phase of growth, showing that there is no significant

difference in the community composition between these two sampling time-points. Principal component analysis (PCA) confirmed the statistically significant differences observed with the analysis of variance and helped identify the global community dynamics in the data set establishing specific correlations between samples and the associated bacterial taxa. From the total number of the unique ASVs (n=537) in the data set, 17 species were identified in the PCA as the dominant species that were driving the variability of the data (Figure 6). Most of the correlations were observed with water samples (designated W) from reactors that contained unfiltered water (samples W4–6), showing 13 species from the families *Cellulomonadaceae* (*Actinotalea sp.*), *Caulobacteraceae* (*Brevundimonas sp.*), *Cyclobacteriaceae* (*Algoriphagus alkaliphilus*), *Nocardiaceae* (*Rhodococcus sp.*), *Acetobacteraceae* (*Roseococcus sp.*), *Sphingomonadaceae* (*Blastomonas sp.*), *Microbacteriaceae* (*Microbacterium sp.*), *Xanthobacteraceae* (*Xanthobacter autotrophicus*), *Chitinophagaceae* (*Sediminibacterium sp.*), unassigned β-proteobacteria and unassigned Bacteria. Algal aggregate samples (designated AG) were closely related to each other but only showed correlation with three identified ASVs from the families *Comamonadaceae*, *Xantomonadaceae*, and *Hyphomicrobiaceae* (*Parvivaculum sp.*). For both aggregates and water communities in filtered water, *Erythrobacteraceae (Erythromicrobium sp.)* was a major ASV (Figure 6).

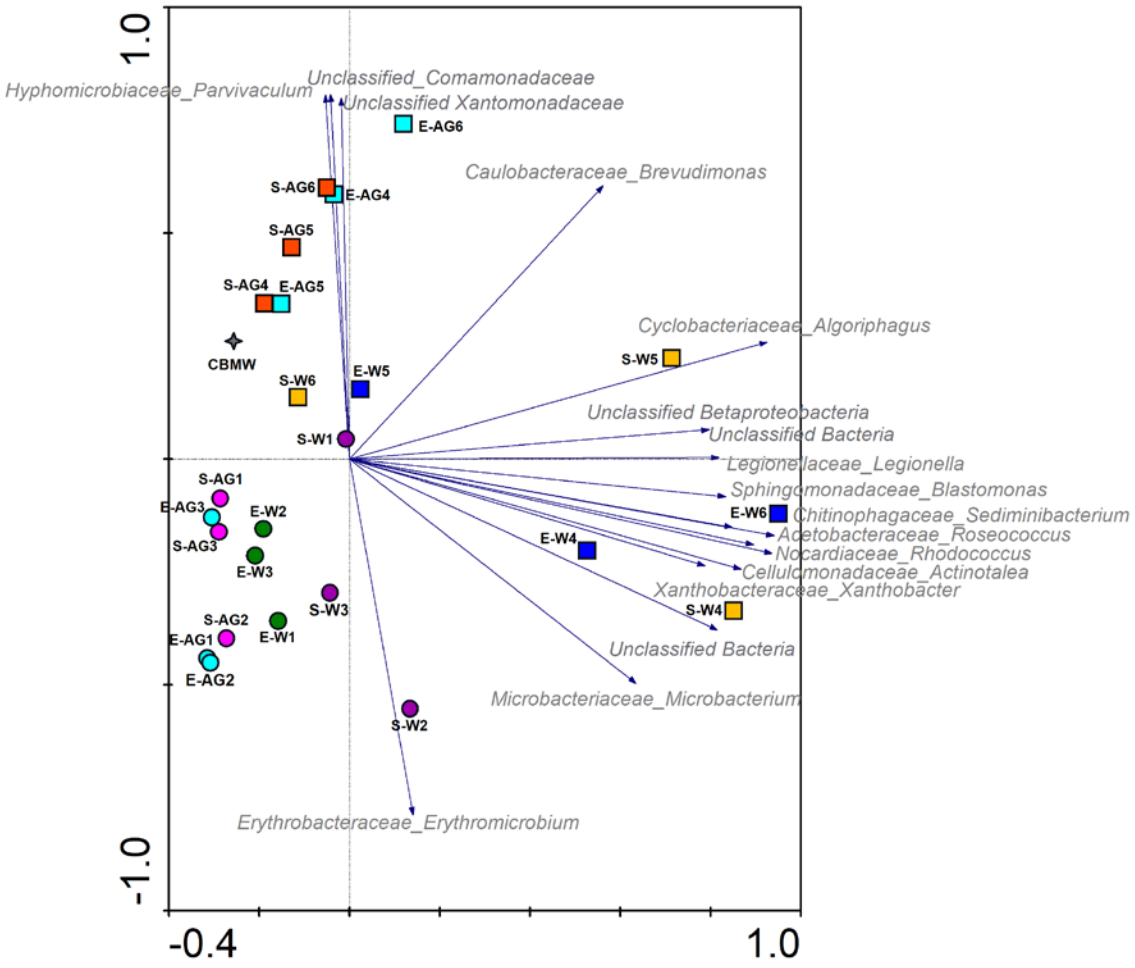


Figure 6. Principal components analysis (CANOCO 4.5 PCA) showing correlations among all the samples in this study and the sampled bacterial populations. The arrows indicate the direction of increasing variability for the taxa. Longer vectors signify a greater range of variation. Eigen value for axis 1 is 0.586 and for axis 2 is 0.143 Each color corresponds to each group of replicate samples (experiments were performed in triplicates). Algal aggregate samples (AG 1-6) are depicted in squares, water samples (W 1-6) in circles, the star corresponds to the CBM water sample (CBMW) prior to algal inoculation (native water community). Samples were taken in exponential phase of growth (letter E preceding sample ID) and stationary phase of growth (letter S preceding sample ID).

Algal Phycosphere: Microbial Community Structure in Algal Aggregates

In order to elucidate the microbial community structure and dynamics associated with the algal aggregates in a closed photobioreactor system, a high dimensional biomarker discovery algorithm (LEfSe) was implemented to identify the abundant bacterial taxa characterizing the differences between the algal aggregates samples in filtered water versus unfiltered water. From a total of 168 features with significant differential abundance, unique ASVs were detected in the algal aggregate samples from filtered (E-AG1,2,3 and S-AG1,2,3) versus unfiltered water (E-AG4,5,6 and S-AG4,5,6). ASVs that were considered exclusive to the aggregates in filtered water belong to the phyla Proteobacteria (unassigned_*Oxalobacteriaceae*, unassigned_Pseudomonadales, unassigned_Sphingomonadales, unassigned_*Moraxellaceae*, unassigned_*Rhizobiaceae*) and from the phyla Actinobacteria (unassigned_*Microthrixaceae* and unassigned_*Cellulomonadaceae*) (Figure 7, Supplementary Figure D1). In contrast, 77 ASVs were exclusive to aggregates in unfiltered water that included 33 taxa (42.9%) correspond to the α-proteobacteria. The remaining 44 taxa (57.14%) were almost evenly distributed between the phyla Verrucomicrobia, Planctomycetes, Chloroflexi, Bacteroidetes and Actinobacteria. The phyla Nitrospirae, Gemmatimonadete, Chlorobi, TM6, Armatimonadetes, NKB19 were present in fewer than 3 ASVs.

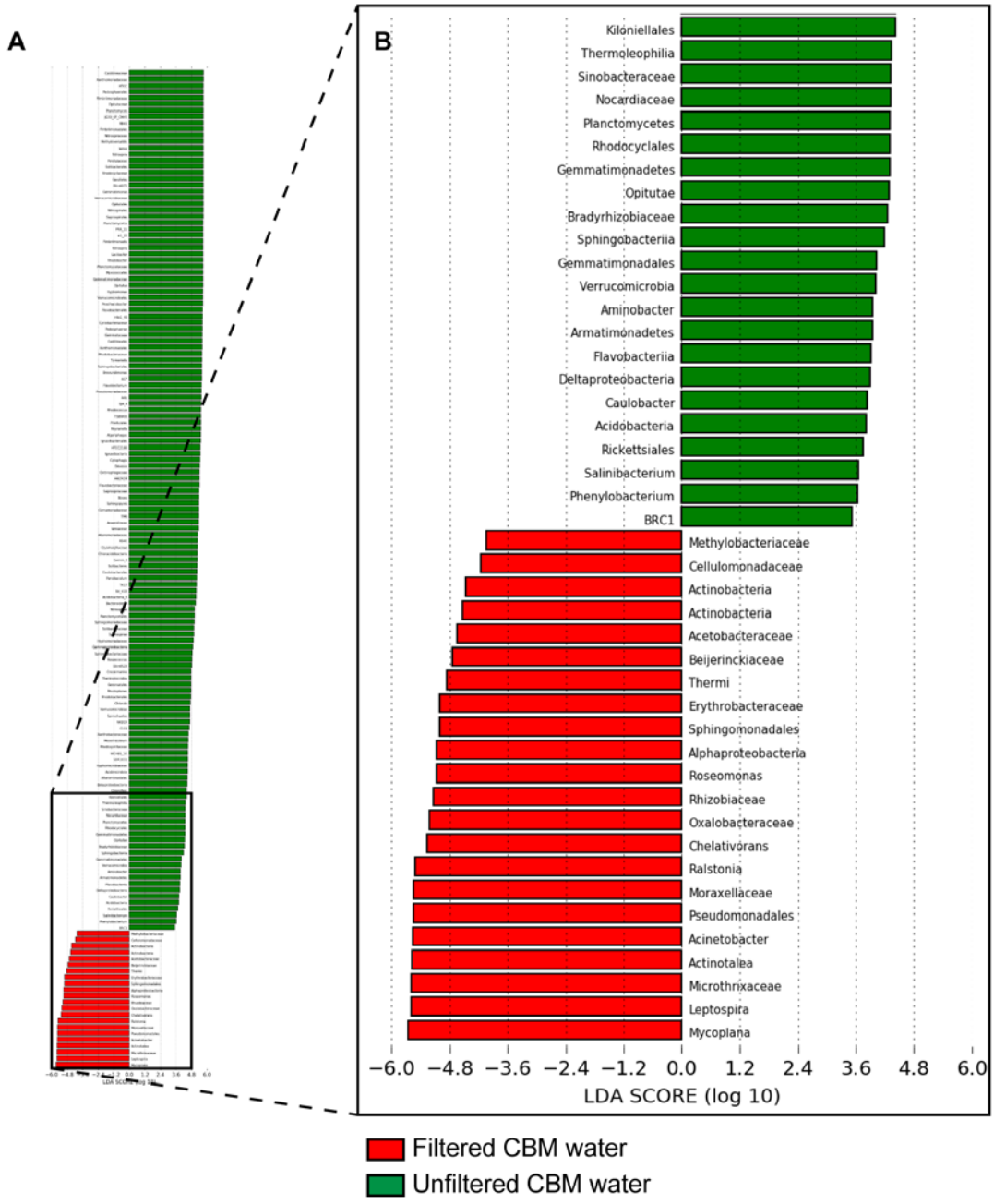


Figure 7. Comparison between algal aggregate samples versus the type of water (filtered and unfiltered) using LEfSe. High dimensional biomarker discovery algorithm (LEfSe) analysis output used to identify the abundant bacterial taxa characterizing the differences between the algal aggregate samples in filtered water versus unfiltered water. From a total of 168 features with significant differential abundance, unique ASVs were detected in the algal aggregate samples from each type of water.

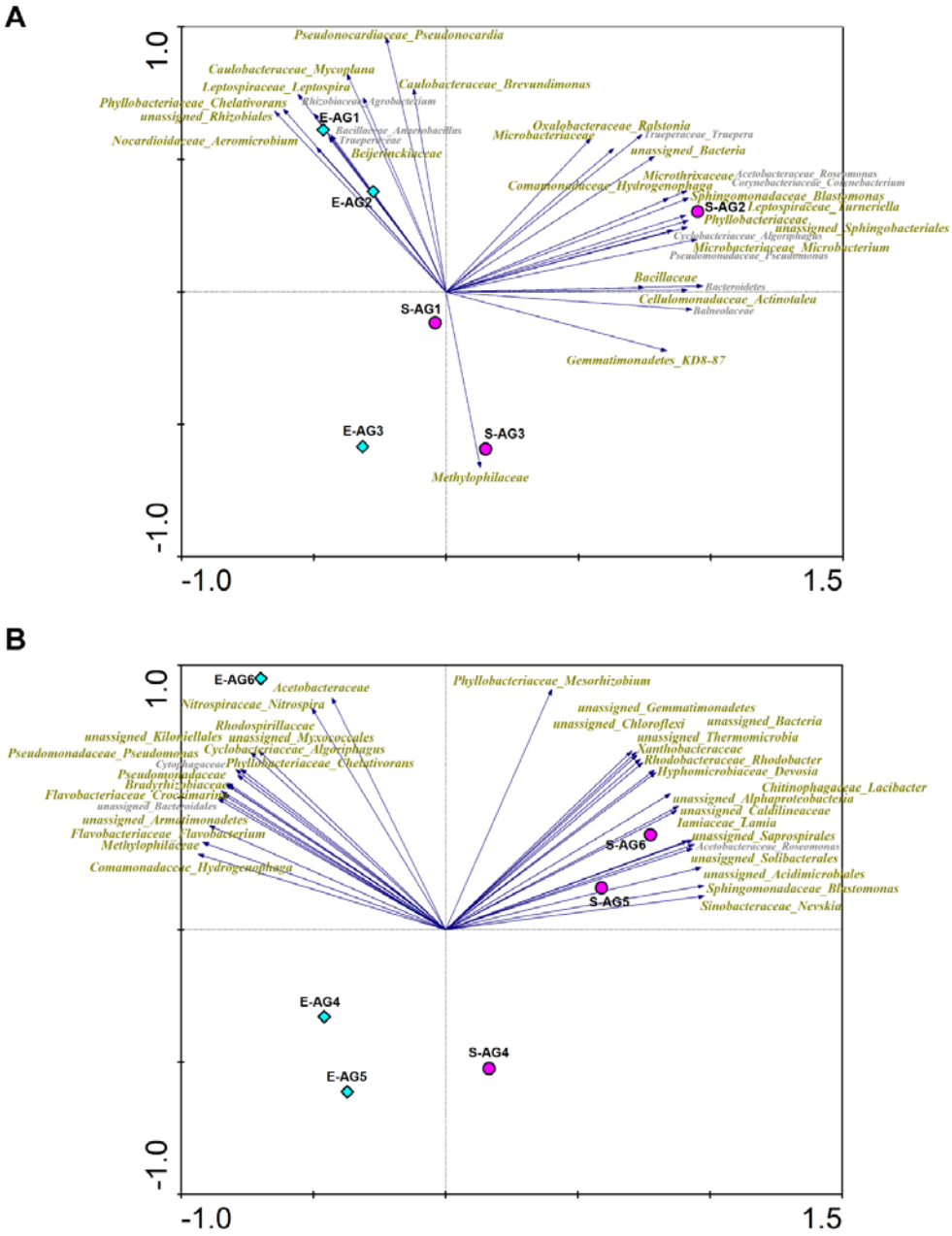


Figure 8. Principal components analysis (CANOCO 4.5 PCA) showing correlations among the algal aggregate samples (AG) and the sampled bacterial population from (A) filtered water only (PBR tubes 1-3) and (B) unfiltered water only (PBR tubes 4-6). Green diamonds correspond to replicate samples from late exponential phase of growth and magenta circles to replicate samples from late stationary phase. Bacterial taxa identified by both analyses (PCA and LEfSe) are indicated in brown colored font, taxa identified only by PCA analysis are indicated in grey. The arrows indicate the direction of increasing variability for the taxa. Longer vectors signify a greater range of variation. Filtered water only Eigen value for axis 1 is 0.379 and for axis 2 is 0.225. Unfiltered water only Eigen value for axis 1 is 0.440 and for axis 2 is 0.269.

The results from the LEfSe analysis were compared with PCA to further support the relationships observed between specific community members and the algal aggregates when grown in filtered or unfiltered water (Figure 8A). During growth in filtered water 33 bacterial taxa (Figure 8A) were identified by both analyses (PCA and LEfSe) and 35 bacterial taxa were identified during growth in unfiltered water (Figure 8B). Among them, 10 bacterial taxa were present across all aggregate samples regardless of the type of water (filtered and unfiltered water) and can be considered the core phycosome ASVs for the tested algal aggregates. The ASVs represented species in the *Phyllobacteriaceae* (*Chelativorans sp.*, *Mesorhizobium sp.*), *Pseudomonadaceae* (*Pseudomonas sp.*), *Comamonadaceae* (*Hydrogenophaga sp.*), *Sphingomonadaceae* (*Blastomonas sp.*), *Cyclobacteriaceae* (*Algoriphagus sp.*), *Acetobacteraceae* (*Roseomonas sp.*) and *Methylophilaceae* and also, to the phylum Gemmatimonadetes and an ASV corresponding to unassigned Bacteria. However, these taxa could be present in low relative abundances indicating that even though they are present, they are not important members of the system and are not establishing meaningful interactions with the alga.

Figure 9 shows 18 species that were identified as moderately to highly abundant (>1%) and common to all algal aggregate samples. The species present in the algal aggregates with ≥10% of relative abundance in at least one of the types of water (filtered versus unfiltered) were *Trueperaceae* (*Truepera sp.*), *Acetobacteraceae* (*Roseomonas frigidaquae*), *Erythrobacteraceae* (*Erythromicrobium sp.*), *Microbacteriaceae* (*Yonghaparkia sp.*), *Caulobacteraceae* (*Mycoplana sp.*), *Phyllobacteriaceae* (*Chelativorans sp.*). The species with relative abundances between 1% and 7% were:

unassigned *Balneolaceae*, *Rhizobiaceae* (*Agrobacterium sp.*), unassigned *Beijerinckiaceae*, unassigned Bacteria, unassigned_Proteobacteria, *Methylophilaceae* (*Methylotenera sp.*) and *Pseudonocardiaceae* (*Pseudonocardia sp.*). The abundant species unique to the filtered water condition were unassigned_Bacteroidetes (3%), and unique to the unfiltered water condition were *Caulobacteraceae* (*Brevundimonas sp.*) (29%), unassigned Gemmatimonadetes (phylum KD8-87) (5%), unassigned *Comamonadaceae* (2%), and *Comamonadaceae* (*Hydrogenophaga sp.*) (2%).

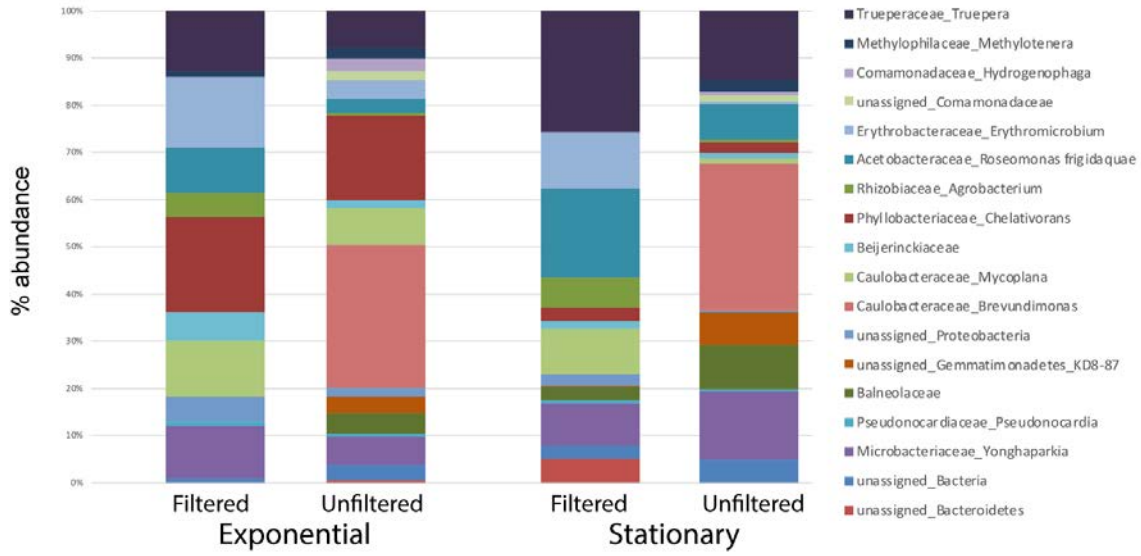


Figure 9. Bacterial species closely associated with algal aggregate samples with moderately to high average abundance (>1%). Comparison between type of water (filtered versus unfiltered) and phase of growth (exponential versus stationary) in aggregate samples.

Among these 18 bacterial taxa, shifts in abundance were observed between exponential- and stationary-phases of growth (Table 3, Figure 9) but 9 species remained highly abundant regardless of the water type (filtered versus unfiltered) or phase of growth. Trueperaceae (*Truepera sp.*), Phyllobacteriaceae (*Chelativorans sp.*),

Caulobacteraceae (*Mycoplana sp.*), Microbacteriaceae (*Yonghaparkia sp.*), Acetobacteraceae (*Roseomonas frigidaquae*), *Beijerinckiaceae* (unassigned), *Rhizobiaceae_Agrobacterium sp.*, Proteobacteria (unassigned) and Bacteria (unassigned) were closely associated with the alga irrespective of the culture condition and could be considered the predicted algal phycosphere under the tested cultivation system.

Table 3. Average relative abundances of species closely associated with algal aggregate samples. Species identified in 100% of algal aggregate samples. Shifts in relative abundance are observed in specific bacterial taxa between exponential and stationary phase of growth. Filtered water (F-W), unfiltered water (UF-W).

Phylum	Class	Order	Family_genus	Exponential Phase		Stationary phase	
				F -W	UF-W	F -W	UF-W
			unassigned_Bacteria	1%	3%	3%	5%
Actinobacteria	Actinobacteria	Actinomycetales	*Microbacteriaceae_Yonghaparkia*	11%	6%	9%	14%
			Pseudonocardiaceae_Pseudonocardia	1%	1%	1%	1%
Bacteroidetes	Rhodothermi	Rhodothermales	*unassigned_Bacteroidetes*	0%	0%	5%	0%
			Balneolaceae	0%	4%	3%	9%
Gemmatimonadetes	Gemmatimonadetes	KD8-87	unassigned_Gemmatimonadetes_KD8-87	0%	3%	0%	7%
Proteobacteria	Alphaproteobacteria		unassigned_Proteobacteria	5%	2%	3%	1%
		Caulobacterales	*Caulobacteraceae_Brevundimonas*	0%	30%	0%	29%
			Caulobacteraceae_Mycoplana	12%	8%	9%	1%
			Beijerinckiaceae	6%	2%	2%	1%
		Rhizobiales	*Phyllobacteriaceae_Chelativorans*	20%	18%	3%	2%
			Rhizobiaceae_Agrobacterium	5%	1%	6%	1%
		Rhodospirillales	*Acetobacteraceae_Roseomonas frigidaquae*	10%	3%	18%	7%
		Sphingomonadales	*Erythrobacteraceae_Erythromicrobium*	15%	4%	11%	0%
	Betaproteobacteria	Burkholderiales	unassigned_*Comamonadaceae*	0%	2%	0%	1%
			Comamonadaceae_Hydrogenophaga	0%	3%	0%	1%
		Methylophilales	*Methylophilaceae_Methylotenera*	1%	2%	0%	2%
Thermi	Deinococci	Deinococcales	*Trueperaceae_Truepera*	13%	8%	25%	14%

Discussion

Studies indicate that specific functional or phylogenetic types of bacteria establish mutualistic symbiotic relationships with algae, which support growth and survival in natural, artificial or extreme ecosystems (Ramanan *et al.* 2016). Diverse natural microalgal communities may be more productive than low diversity ones due to resource complementarity and facilitation that establishes diversity-productivity interactions between algae and the accompanying microbial consortia (Stockenreiter *et al.* 2012). Different modes of interaction have been shown but the exact nature of these interactions will depend on the specific microenvironment and the specific bacterial-algal consortia. Such types of interactions are the production and exchange of macronutrients (*i.e.*, carbon, nitrogen, and phosphorus) and micronutrients (*e.g.*, vitamins and phytohormones), degradation of complex molecules (polysaccharides, high molecular weight organic matter) and bacteria could potentially enable algal defense mechanisms, as has been demonstrated in plants (Krohn-Molt *et al.* 2013; Ramanan *et al.* 2015).

The unfiltered water contained the native microbial community in the CBM water, and the filtered water had a reduced microbial population due to filtering. A possible explanation for the improved performance of PW95 in unfiltered water is that diverse natural microalgal communities may be more productive than low diversity ones due to resource complementarity and facilitation (Stockenreiter *et al.* 2012). PW95 cultivated in unfiltered CBM water might have established diversity-productivity interactions with the accompanying microbial consortia that had a positive impact on PW95 growth.

Chlorophyll production was not significantly different between algae cultivated in filtered or unfiltered water. Rapid declines in chlorophyll have been associated with nitrogen starvation (Mus *et al.* 2013) but an increase in chlorophyll was observed during late stationary phase of growth (days 5 and 6). The use of chlorophyll as an intracellular nitrogen pool under N deprivation in green algae and diatoms has been suggested previously (Li *et al.* 2008; Valenzuela *et al.* 2012). Additionally, chlorophyll degradation pathways have been discovered in plants to sustain vital biological functions (Eckardt 2009) which could also be the case for green algae. Increases in pH during cultivation were similar for cultures cultivated in filtered or unfiltered water, and similar trends in pH increase were observed during PW95 growth in standardized medium and CBM water (Corredor *et al.*, in review; Hodgskiss *et al.* 2016).

Based upon NR fluorescence, algal cells cultivated in filtered water had slightly higher NR fluorescence and the rate of increase was also faster. Detected FAMEs from the different cultures showed similar levels in exponential growth, but FAME levels were higher in stationary-phase cells cultivated in unfiltered water compared to filtered water (1.7x). *Dunalliela tertiolecta* was reported to produce 20–24% total lipid when grown in CBM water from Australia (Aravinthan and Harrington 2014) and PW95 produced 20-27% FAMEs when grown in supplemented CBM water (from the same source) at 20°C (Hodgskiss *et al.* 2016). Similar to previous studies, the PW95 FAMEs were predominated by saturated and unsaturated C16/C18 fatty acid signatures with some C20 signatures (Corredor *et al.*, in review; Hodgskiss *et al.* 2016). The FAME profile can be variable depending on culture conditions but PW95 could be a potential source of

C18/C16 feedstock, and C16/C18 are desirable for biofuel applications (Hoekman *et al.* 2012). The increased rate and level of FAMEs for cultures grown in filtered water could be attributed to the lower average growth of PW95 in filtered water as has been previously demonstrated in standard growth medium (Corredor *et al.*, in review). An inverse relationship between biomass and lipid accumulation has been frequently reported in the literature for different green algae (Laurens, Chen-Glasser and McMillan 2017); however, the roles of different microbial populations that directly and/or indirectly contribute or hinder algal lipid accumulation are poorly understood.

The phylum-level characterization of bacterial structure and composition in this study revealed that Proteobacteria, Actinobacteria, and Bacteroidetes predominated the global taxonomic profile of the algal cultures. Actinobacteria have been shown to predominate in coal core samples from Flowers-Goodale coal bed from which the water samples used in this study were collected (Barnhart *et al.* 2016), and Proteobacteria have been reported as one of the dominant phyla in an algal pond ecosystem (Beyter *et al.* 2016). Bacteroidetes and Proteobacteria have been previously reported to predominate bacterial communities from all algae growth systems as well as the majority of the microorganisms in a mixed species biofilm associated with *Chlorella vulgaris* and *Scenedesmus obliquus* in a photobioreactor (Krohn-Molt *et al.* 2013; Fulbright *et al.* 2018). Proteobacteria and Bacteroidetes were the most dominant and prevalent phyla in the phycosphere of 11 algal strains isolated from a variety of ecosystems (Ramanan *et al.* 2015), and Bacteroidetes have been observed in organic aggregates in the ocean and are

important for the degradation of high-molecular weight organic matter (Krohn-Molt *et al.* 2013; Ramanan *et al.* 2015).

At the class level, Actinobacteria, α-Proteobacteria, β-Proteobacteria and γ-Proteobacteria predominated the global taxonomic profile of algal aggregates under the tested growth conditions. Actinobacteria have been correlated with hydrocarbon-containing environments such as oil and coal reservoirs (Barnhart *et al.* 2016) and may be more a consequence of source diversity (*i.e.*, CBM production water) rather than a specific algal interaction. Similar α-proteobacterial and γ-proteobacterial sequences have been shown to be the most abundant Proteobacteria in marine environments and in *Nannochloropsis* laboratory cultivations (Fulbright *et al.* 2018). Other studies have reported the families *Caulobacteraceae*, *Comamonadaceae*, *Xantomonadaceae* and the phylum Verrucomicrobia as the most abundant members of the microbial community associated with previously reported microalgal production systems (Beyter *et al.* 2016). These families and phyla were among the taxa identified in the PCA (Figure 6) as dominant species driving the variability of all the samples evaluated in this study and were correlated with unfiltered water. These observations validate the higher richness estimates observed for aggregate samples cultivated in unfiltered water (Table 2). Moreover, *Caulobacteraceae* and *Comamonadaceae* were observed as two of the most abundant bacterial groups (Table 3, Figure 9).

The most abundant bacterial sequence groups observed in close association with PW95 algal aggregates, that have been reported by other investigators as closely associated with, or attached to algal colonies or biomass are members of the families

Caulobacteraceae, *Rhizobiaceae*, *Comamonadaceae*, *Erythrobacteraceae*, *Sphingomonadaceae* and *Phyllobacteriaceae* (Lakaniemi *et al.* 2012; Krohn-Molt *et al.* 2013; Sambles *et al.* 2017).

Different α-proteobacteria have been implicated in plant growth promotion and have been designated plant growth promoting bacteria (PGPB) and the most notable are classified as Rhizobiales, Rhodospirillales and Sphingomonadales. These bacteria have the functional capabilities to be associated with green algae and could perform similar functions to that of plant-bacteria interactions (Kim *et al.* 2014; Ramanan *et al.* 2015). Highly abundant bacteria from the present study belong to Rhizobiales (*Phyllobacteriaceae_Chelativorans sp.*, *Rhizobiaceae_Agrobacterium sp.* and unassigned *Beijerinckiaceae*) and Sphingomonadales *(Erythrobacteraceae_Erythromicrobium sp.*). Species of the order Rhizobiales have been reported to sustain algal growth by producing vitamin B, can establish mutualistic interactions that have been reported to increase algal growth in model organisms such as *Chlamydomonas reinhardtii*, *Chlorella vulgaris*, *Scenedesmus sp.*, and *Botryococcus braunii* and are associated with endosymbiotic nitrogen fixation (Kim *et al.* 2014; Sambles *et al.* 2017; Fulbright *et al.* 2018). Interestingly, known *Erythromicrobium* have the ability to perform redox chemistry on tellurium (Yurkov, Jappe and Vermeglio 1996), and relatively high levels of tellurium have been associated with coal environments (Belzile and Chen 2017). The results showed that the *Erythromicrobium* sequences were an important group in filtered CBM production water, and these results suggested that this group was not completely removed with 0.2 μm filtration and/or selectively enriched in the presence of PW95 and could

provide protection against tellurium (or other heavy metals). Previous characterization of *Chelativorans* species has demonstrated a unique ability to utilize ethylenediaminetetraacetic acid (EDTA) as a source of C, N, and energy (Doronina *et al.* 2010). The metabolic interactions and/or role of potential *Chelativorans* is unknown but may play a role in turnover of unique N compounds such as diamines.

Shifts in abundance were observed between the late exponential and stationary phases of growth in algal aggregate samples possibly due to changes in algal biomass (Beyter *et al.* 2016) or oligotrophic conditions during the late stationary phase (*i.e.*, N deplete conditions; Table 3). Species that decreased in relative abundance during stationary phase were Proteobacteria (unassigned), *Caulobacteraceae* (*Mycoplana sp.*) *Beijerinckiaceae* (unassigned), *Phyllobacteriaceae Chelativorans sp.* and *Erythrobacteraceae (Erythromicrobium sp.).* Species that increased in relative abundance during stationary phase were *Trueperaceae* (*Truepera sp.*), *Comamonadaceae* (unassigned), *Comamonadaceae* (*Hydrogenophaga sp.*), *Acetobacteraceae* (Roseomonas *frigidaquae*), Gemmatimonadetes, *Pseudonocardiaceae (Pseudonocardia sp.)*, *Microbacteriaceae (Yonghaparkia* sp.) and Bacteroidetes (unassigned). The abundant species unique to filtered water during stationary phase was unassigned Bacteroidetes (3%), and unique to unfiltered water but present in both phases of growth were *Caulobacteraceae* (*Brevundimonas sp.*), unassigned Gemmatimonadetes (phylum KD8-87), unassigned *Comamonadaceae* and *Comamonadaceae* (*Hydrogenophaga sp.*). *Brevundimonas sp.* had the highest relative abundances for both late exponential phase (30%) and late stationary phase (29%), and these results suggested a role/interaction

irrespective of growth-related nitrogen condition.

The species that remained highly abundant regardless of the water type or phase of growth were considered universal and make up the algal phycosphere under the tested cultivation conditions. *Trueperaceae* (*Truepera sp.*) is a known alkaliphilic that uses a wide range of carbohydrates, organic acids or amino acids as carbon and energy sources with a respiratory metabolism, and extremely resistant to ionizing radiation (Ivanova *et al.* 2011). *Phyllobacteriaceae* (*Chelativorans sp.*) is known as *Mezorhizobium sp.* BNC1 (NCBI: txid266779) and this type of bacteria has been reported as a nitrogen-fixing species associated with green algae, as well as to produce vitamin B to support the alga growth for *Lobomonas rostrata* (Fulbright *et al.* 2018). *Caulobacteraceae* (*Mycoplana sp.*) are a group of aerobic soil bacteria with a reported ability to decompose aromatic compounds (Urakami *et al.* 1990). *Microbacteriaceae* (*Yonghaparkia sp.*) is a strictly aerobic alkaline soil bacterium and *Acetobacteraceae* (*Roseomonas frigidaquae*) is strictly aerobic as well but has been isolated from a variety of environmental samples such as drinking water distribution systems, freshwater sediment biofilm from a cooling tower, deep water marine invertebrates and a soil (Kim *et al.* 2009). *Beijerinckiaceae* (unassigned) *and Rhizobiaceae (Agrobacterium sp.)* are Rhizobiales part of a group of bacteria that specializes in symbiotic nitrogen fixation enabling them to thrive in habitats in which other potential sources of nitrogen are scarce (Wibberg *et al.* 2011; Marín and Arahal 2014). Moreover, species of *Beijerinckiaceae* include facultative methanotrophs (Tamas *et al.* 2014) showing the link between the origin of the CBM water used for this study and the association to the alga. Unassigned Bacteria and Proteobacteria sequences

were also determined to be part of PW95's phycosphere on the basis of abundance and persistence in algal aggregate samples and demonstrate the unknown diversity of even algal phycosome communities.

The presented data suggests that the unfiltered CBM water had higher species richness, and the algal aggregates had an established microbial consortium regardless of the type of water (filtered vs. unfiltered) used for growth. The different analyses performed during this study helped narrow down the bacterial diversity closely associated with the alga and 18 different species make up the algal microbial community under the tested cultivation scheme as a closed system with nitrate and CO_2 as the nitrogen and carbon sources, respectively. Within the enriched bacterial populations, 9 highly abundant taxa were universally present in all aggregate samples: *Truepera sp.*, *Chelativorans sp.* (also known as *Mezorhizobium sp.* BNC1), *Mycoplana sp.*, *Yonghaparkia sp.*, *Roseomonas frigidaquae*, *Beijerinckiaceae* (unassigned), *Agrobacterium sp.*, unassigned Bacteria and Proteobacteria. Interestingly, despite similar nitrate utilization for algal cultivation in filtered and unfiltered water, algal growth rate and biomass yield trended higher in unfiltered water. However, both NR fluorescence and FAME analysis indicated that algal cultivation in filtered water resulted in more accumulated lipids.

Conclusions

Bacterial taxonomic features were dominant in the PW95 algal cultivation system and archaeal taxa were rarely observed. Diversity indices for unfiltered water samples regardless of sample type and growth phase were higher, and as expected, the sample

representing the native microbial community in the water prior to algal inoculation (CBMW) had one of the highest richness indices and one the lowest evenness indices. Diversity declined during and after nitrate-depletion in the algal culture. In the unfiltered water as the starting point, the water and algal aggregate communities were different, and the aggregates were enriched in sequences indicative of *Parvivaculum*, *Comamonadaceae*, and *Xantomonadaceae*. For the cultures initiated in filtered water, the lack of diversity compared to the unfiltered water delineated the grouping with the exception of *Erythromicrobium* sequences. For PW95 under the tested growth conditions, filtered water had lower microbial diversity, slower algal growth, but higher lipid accumulation compared to cultivation in unfiltered water. Our findings build on previous knowledge of phylotypes that are commonly associated with algae and can be implemented for the enhanced production of biomass or biofuel in CBM water or other types of low-quality water. These results provide insights into specific algal-bacterial associations and potential functional relationships that sustain growth and productivity using a polyculture. The data suggest that eco-engineering of an algal phycosphere could be used to modulate the interplay between biomass accumulation and lipid accumulation, but further work is needed to elucidate specific functional relationships that may change upon repeated outdoor cultivations.

Acknowledgments

This work was supported by the National Science Foundation Sustainable Energy Pathways (SEP) Program under NSF CHE-1230632 and the Department of Energy PEAK program under DE-EE0008247. Partial support for L. Corredor was provided by

the Fulbright Scholarship Program. The authors would like to thank the Montana Bureau of Mines and Geology for providing access to the FG-09 well and the MSU CBM group for water sample collection. Dr. Rebecca Mueller and Hannah Goemann for assistance with sequencing. Dr. Al Parker for statistical analysis assistance. Dr. Federica Villa for the confocal microscopy image. The MSU Algae Group for support and constructive exchange of ideas. Dr. Ben Wheaton for graphical and editorial assistance.

CHAPTER SIX

EPILOGUE

Advances towards the successful implementation of industrial-scale algal biofuel production are dependent largely on the optimal combination of different factors such as the selection of robust strains (fast growth rates, high percentage of lipid accumulation and the right FAME profile), their growth physiology and the cultivation system (reactors or open ponds) of choice (Talebi *et al.* 2013; Fields *et al.* 2014; Laurens, Chen-Glasser and McMillan 2017). This thesis set out to propose an integrated approach to the study of algal production systems, based on the evaluation of the algal physiological responses to changing environmental conditions, the determination of the algal genomic and functional potential, its phylogenetic relations and the implementation of an ecosystem view to algal biomass production.

The studies outlined in the preceding chapters were based on the premise that PW95, a novel green sporulating microalga isolated from a CBM water impoundment, can be a potential feedstock for biofuel production using CBM low quality water (a by-product of methane gas extraction). A previous study demonstrated that PW95 produces biomass and biofuel in monoculture and batch mode, using CBM water, and that the lipid profile of PW95 is dominated by similar fatty acids (palmitic, oleic, linoleic and α-linoleic acid) to those produced by other plant-based sources of biofuel (Hodgskiss *et al.* 2016).

Algal environmental isolates have the potential to lower culture costs for lipid production, through acclimation to fluctuating environmental conditions with low

nutrient availability (Hoekman *et al.* 2012; Hodgskiss *et al.* 2016). Frequently, green microalgal strains from culture collections are used to explore maximization of biodiesel potential (Tasić *et al.* 2016); but native strains of algae can also be a viable option that most likely will respond more robustly to variable culture conditions (*e.g.*, temperature, sunlight, invasion) during outdoor cultivation in specific locations (Zhou *et al.* 2011; Eustance *et al.* 2013; Fields *et al.* 2014; Hodgskiss *et al.* 2016). Consequently, this thesis sought to characterize PW95 in order to assess its potential viability as a native alga for use in biofuel production.

Broadly, this thesis demonstrates four main findings. First, that PW95 is a *Chlamydomonas*-like alga. Second, that PW95 shows potential as a biofuel feedstock, accumulating hexadecanoic and octadecanoic fatty acids, although there are apparent trade-offs between biomass and lipid production in changing environmental conditions that must be considered. Third, the generation of the PW95 genome provides a deeper understanding of PW95's biological and biochemical capabilities. Finally, PW95 appeared to be supported by a native microbial consortium when grown in CBM water.

PW95, a *Chlamydomonas*-like Alga

Phylogenetic analysis using the 18S SSU rRNA and a concatenated protein tree identified PW95 as a *Chlamydomonas*-like alga. *Chlamydomonas reinhardtii,* the model green microalga, is motile and bi-flagellated, and widely used for understanding the structure, assembly and function of eukaryotic flagella (Merchant *et al.* 2007; Tirichine and Bowler 2011; Kim *et al.* 2014). PW95 is assumed to be non-motile but more than 50 genes related to cilia/flagella assembly and function were identified in the PW95 genome,

suggesting that the capacity to assemble and use flagella exists and could be expressed for very specific biological functions (sexual reproduction) or as yet unidentified, culture conditions. Another important phenotypic observation during the physiological characterization of PW95 was supported with the investigation into the genome. PW95 showed a sporulating phenotype under all the experimental conditions evaluated in this work and two genes involved in sporulation were predicted with the genome annotation (spore maturation protein B, DNA segregation ATPase FtsK/SpoIIIE), giving a strong indication that the spore-like structures observed in culture are most likely aplanospores. PW95 spores can be a way of reproduction or an expression of a persistent phenotype in response to oligotrophic growth conditions (nutrient depletion) in culture and the environment. The genetic similarity between PW95 and *C. reinhardtii*, and the phenotypic differences between these two chlorophytes may indicate that this unicellular green alga could be a cryptic morphotype of *Chlamydomonas spp*.

PW95 Biofuel Potential

The evaluation of the mixed effects of temperatures, nitrate levels, pH, and bicarbonate on PW95 growth and lipid accumulation elucidated synergistic and unfavorable interactions between multiple environmental variables and nutritional levels. In terms of growth, at higher temperatures (25°C and 30°C) in combination with higher initial nitrate (2 mM), PW95 displayed a faster growth rate, higher maximum cell numbers, as well as accumulation of chlorophyll. For most of the treatments the use of a pH buffer in the media had an inhibitory effect on growth. The use of a high concentration bicarbonate trigger had no significant effect on growth, indicating that

ambient levels of CO_2 were sufficient for carbon fixation and photosynthesis maintenance during nitrate starvation. The availability of a different form of DIC (bicarbonate) did not enhance cell division or increase biomass in PW95, suggesting that a carbon concentrating mechanism was not activated with the elevated levels of bicarbonate in the media. In terms of lipid accumulation, the highest total FAMEs content (33%) was observed at 20°C without buffer or trigger at 0.5 mM nitrate; however, at 30°C, PW95 showed almost as high lipid accumulation (27% and 25%). The lowest total FAME levels were observed for treatments at 25°C and when the initial nitrate concentration was 2.0 mM, all treatments were similar in total FAME, but the levels were generally below 20%. The results suggest there was an inverse relationship between biomass accumulation (cells and chlorophyll) and lipid accumulation at the extremes of nitrate and temperature (low nitrate as well as low and high temperature) and an important interaction between low nitrate levels, temperature, and elevated pH for trade-offs between biomass and lipid production in PW95. Taken together, the most promising combination of culture conditions for less input, lower cost, sustainable biofuel production using a native algal isolate is higher temperature (30°C), low initial nitrate concentration (0.5 mM), no buffer and no bicarbonate trigger. Additionally, in terms of the FAME profiles, PW95 can be a potential feedstock for biofuel production, because it accumulates predominantly saturated and unsaturated, C16 (hexadecanoic) and C18 (octadecanoic) fatty acid chains in low quality water and growth media. An interesting characteristic of lipid accumulation in this alga was that in treatments with the highest percentage of FAME, the fatty acids were always higher during nitrate replete conditions.

WGS Reveals Conserved and Novel Biochemical Capacities in PW95

Whole genome sequencing, assembly and annotation has been frequently employed to understand/predict the biochemical capacities and responses of algal strains with potential to produce industry valued-products, such as biofuel (Hovde *et al.* 2018). The present thesis is the first report of the complete genome sequence, assembly and annotation of an environmental alga isolated from low quality water derived from the production of alternative energy sources such as coal bed methane.

In an effort to advance towards the utilization of wastewater to generate biofuel, the first genome draft (v1.0) of *Chlamydomonas*-like PW95 alga served as the foundation for better understanding the biological and biochemical potential of this microalgal production strain. As further improvements are implemented in the assembly and annotation processes, a far more comprehensive view of PW95's physiological adaptations, predominant phenotype during cultivation, biology and biochemical capabilities will be generated. 57% of the genome is complete as estimated by BUSCO and conserved, complete pathways were annotated in the genome including carbohydrate metabolism and carbon storage, carbon fixation, nitrate assimilation, lipid metabolism, amino acid and nucleotide metabolism, all essential for cell growth, survival and maintenance, as well as relevant for cultivation and the production of value-added products such as lipids and starch.

The prediction of PW95's gene complement provided a glance at life in an oligotrophic environment (coal bed methane production water impoundment) with unique features: facilitated nutrient acquisition with a variety of membrane transporters (*i.e.*,

nitrate transporters specific to PW95), heat and cold resistance with temperature shock proteins, an alternative way for steroid biosynthesis to maintain cell structure and membrane fluidity (squalene epoxidase gene), a sporulating phenotype possibly contributing to persistence and survival that remained a dominant phenotype observed during cultivation and the absence of flagella, reallocating resources into metabolic maintenance instead of the assembly of a non-vital cellular structure that requires expensive energy investment, the potential for improved Ci fixation (CCM) and heterotrophic growth (acetate assimilation).

We also evaluated the physiological responses of PW95 during cultivation in photobioreactors (PBRs) with ambient air and artificial sunlight, using filtered and unfiltered supplemented CBM water (0.5 mM nitrate, 0.185 mM borate) at 30°C. In this system we characterized the microbial community structure and dynamics in close association with PW95 and the global gene expression during growth and lipid production in the context of nitrate depletion.

A correlation between lower growth and high lipid production was observed in this system as well and is consistent with the observations from the physiological responses of PW95 during growth in standardized media. The highest total FAME content (23%) was observed in filtered water, that showed significantly lower average biomass and rate of growth in comparison to unfiltered water. The unfiltered water contains the native microbial community in the CBM water, suggesting that algal polycultures with increasing diversity are indeed associated with higher productivity (Stockenreiter *et al.* 2012; Beyter *et al.* 2016; Bell *et al.* 2019). These observations were

further supported by the significant differential gene expression found at late exponential phase of growth where PW95's metabolic capacities were involved in active carbon and energy metabolism and resources were allocated towards growth, even at a stage were cells were starting to transition into the stationary phase. Although there was a high percentage of non-assigned (NA) significant DEGs (63.6% of genes between the growth phases and 42.8% between the types of water) and a low percentage of reads mapping to the genome (5%), significant differential gene expression was found for well conserved genes and functions. Nonetheless, a potentially novel biochemical capacity (squalene epoxidase; SQE) was discovered in PW95's genome and was found to be up-regulated at late stationary phase of growth regardless of the water type. This unique capacity to perform steroid biosynthesis represents an alternative way to potentially modulate the fluidity and flexibility of cell membranes maintenance/survival strategy. The metabolic capacity of PW95 was found to be versatile and adaptable to intracellular and environmental changes. For example, the down-regulation of nitrate transporters was observed during nitrate deplete conditions, and a slower metabolic rate was evident with the down regulation of genes involved in oxidative phosphorylation (inorganic pyrophosphatase (*ppa*)), carbon fixation (malate dehydrogenase (MDH1)), glycolysis, pyruvate metabolism, the TCA cycle (phosphoglycerate kinase (PGK)), and housekeeping functions (ribosome proteins, spliceosome, proteins involved in messenger RNA biogenesis).

PW95 Appeared to be Supported by a Native Microbial Consortium in CBM Water

While evaluating the physiological responses of PW95 during cultivation in photobioreactors (PBRs), we characterized the microbial community structure and dynamics in close association with PW95 during growth and lipid accumulation.

Under filtered water conditions, the algal aggregates maintained approximately a less diverse (40 bacterial OTUs) microbial consortium, whereas under unfiltered conditions the algal aggregates maintained a more diverse microbial community (180 bacterial OTUs). Interestingly, PW95 cultivated in unfiltered water had better growth but accumulated less lipids. Sequences indicative of *Caulobacteraceae*, *Rhizobiaceae*, *Comamonadaceae, Erythrobacteraceae*, *Sphingomonadaceae* and *Phyllobacteriaceae* were the most abundant bacteria in close association with PW95 algal aggregates; and PW95's phycosome was composed of the species *Truepera sp.*, *Chelativorans sp.*, *Mycoplana sp.*, *Yonghaparkia sp.*, *Roseomonas frigidaquae*, *Beijerinckiaceae* (unassigned), *Agrobacterium sp.*, Proteobacteria (unassigned) and Bacteria (unassigned). How these algal-bacterial interactions are established and maintained and how they benefit PW95 needs to be investigated for a better understanding of algal production systems.

Future Directions and Improvements

The physiological, genomic and functional studies outlined in this work could guide future experimental design and cultivation strategies using PW95. Aspects of PW95 biology, physiology and ecology worth exploring are: 1) the activation of carbon

concentrating mechanisms using different sources and concentrations of DIC during nitrate replete and deplete conditions, to evaluate whether the use of a bicarbonate trigger or elevated concentrations of CO_2 would enhance biomass or lipid production during growth in CBM water, 2) the evaluation of heterotrophic growth using acetate as an alternate way of biomass and lipid production and a way to avoid photo-oxidative damage, 3) to determine the mechanisms by which PW95 is potentially using chlorophyll as an intracellular nitrogen pool and its relation to biomass and lipid accumulation, 4) since the PW95 genome has the complete pathway for starch biosynthesis, it would be useful to determine whether starch is being accumulated during nitrate depletion and whether resources are being distributed into lipid and starch storage, which may be reflected in the low levels of FAMEs observed in certain growth states, 5) to investigate the nature of the algal-bacterial associations (algal phycosphere) observed during biomass production in CBM water and to identify symbiotic relationships that can translate into biomass and lipid production relationships, 6) induce flagella assembly and function in specific culture conditions, as well as spore formation for a deeper understanding of PW95's biology and potentially a more detailed morphological and taxonomic characterization of the alga.

PW95's genome v1.0 is currently in a draft state and continuous improvements towards the completion of the genome should be implemented in the assembly and annotation processes in order to obtain a far more comprehensive view of PW95's biology and metabolic capabilities in the context of biomass and lipid production. Genome projects are not end-points at publication and finished does not necessarily

means perfect. Telomere-to-telomere assembled chromosomes without gaps are considered finished, and only the relatively small nuclear genomes of *C. merolae* (red alga), *Micromonas commoda* RCC299 and *Ostreococcus lucimarinus* (prasinophytes) are algal genomes that equate to that state of completion (Blaby-Haas and Merchant 2019). CANU has been reported to generate accurate assembly of large genomic repeats but in reality, most of the generated PacBio reads are smaller and have high error rates (10–15%). Therefore, higher coverage is recommended to maximize the likelihood of having sufficient sequencing depth and numbers of longer reads in order to accurately resolve genomic repeats and correctly assemble a genome. However, since increasing coverage of a genome as big as PW95 can be very expensive, strategies for the improvement of the genome assembly that can result in a less fragmented, higher quality assembly are: re-assembly with modifications to default parameters in CANU, assembly polishing to detect and resolve misassemblies (*i.e.*, Quiver software) and rather than trying to improve assembly with a hybrid assembly approach (PacBio/Illumina data), use Illumina sequence data for error-correction. For the annotation, the transcriptome will be quality filtered and annotated, and incorporated into the pipeline as evidence in its improved state, and MAKER2 (genome curation and annotation pipeline) will be used in addition to GenSAS. Following the aforementioned improvements, an investigation into gene structure (introns and exons, CDS), telomeric and repetitive sequences will expand our knowledge into PW95's genome complexity and using *in silico* methods to predict subcellular localization of proteins and transcription factors in the genome will give support and expand our current knowledge of potential function. There are many genes

that are responsible for conserved biological functions that are tightly regulated and can give hints of the gene make up of PW95.

Lastly, the RNA sequencing reads on average mapped back to PW95's genome with an unexpectedly low percentage (5%), which could be attributed to the current state of the genome assembly. Annotation and mapping will potentially improve with an optimized version of the genome. An alternative strategy to improve read mapping could be performing the differential expression analysis with the assembled transcriptome (quality filtered and annotated) instead of the genome while an improved version of it is available.

Does PW95 show potential for biofuel production?

The experimental evidence gathered through this work points towards the implementation of PW95 as a production strain for outdoor cultivation in CBM water impoundments. PW95 is a robust strain in terms of adaptation to oligotrophic and fluctuating environmental conditions with minimal nutrient requirements (low concentration of nitrate, 6 times less than Bold's media). The perfect algal cultivation system is not yet available, even with the increasing knowledge in genome editing perfection is still a faraway reality; but genomic and functional studies are a way to make informed decision about cultivation strategies based on genomic capabilities and the gene expression of those functionalities. PW95 cultivated in unfiltered CBM water might have established diversity-productivity interactions with the accompanying microbial consortia, which had a positive impact on growth and is very important aspect of the outdoor cultivation. Compared to highly productive strains like *Botryococccus braunii*

(up to 60% FAME content) (Zhu, Li and Hiltunen 2016), PW95 accumulates an average percentage of FAMEs in CBM water (20–27%) but PW95 can be a potential source of C18/C16 feedstock, since it produces predominantly the FAMEs species desirable for biofuel applications.

The current thesis is part of an ongoing effort to gain an understanding of the genetic and genomic basis that defines and controls the physiological behavior of this unique alga with a view to harnessing its capabilities for use in water management and the production of biofuels and other value-added products.

REFERENCES CITED

Abo BO, Odey EA, Bakayoko M *et al.* Microalgae to biofuels production: A review on cultivation, application and renewable energy. *Rev Environ Health* 2019;**34**:91–9.

Adachi M, Kanno T, Okamoto R *et al.* Population structure of Alexandrium (Dinophyceae) cyst formation-promoting bacteria in Hiroshima Bay, Japan. *Appl Environ Microbiol* 2003;**69**:6560–8.

Adeniyi OM, Azimov U, Burluka A. Algae biofuel: Current status and future applications. *Renew Sustain Energy Rev* 2018;**90**:316–35.

Aleya L, Dauta A, Reynolds CS. Endogenous regulation of the growth-rate responses of a spring-dwelling strain of the freshwater alga, Chlorella minutissima, to light and temperature. *Eur J Protistol* 2011;**47**:239–44.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Anders S, Pyl PT, Huber W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**:166–9.

Andersen RA (Robert A. *Algal Culturing Techniques*. Elsevier/Academic Press, 2005.

Apprill A, Mcnally S, Parsons R *et al.* Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol* 2015;**75**:129–37.

Aravinthan V, Harrington D. Coal seam gas water as a medium to grow Dunalliella tertiolecta microalgae for lipid extraction. *Desalin Water Treat* 2014;**52**:947–58.

Avidan O, Brandis A, Rogachev I *et al.* Enhanced acetyl-CoA production is associated with increased triglyceride accumulation in the green alga Chlorella desiccata. *J Exp Bot* 2015;**66**:3725–35.

Bankevich A, Nurk S, Antipov D *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;**19**:455–77.

Barnhart EP, Weeks EP, Jones EJP *et al.* Hydrogeochemistry and coal-associated bacterial populations from a methanogenic coal bed. *Int J Coal Geol* 2016;**162**:14–26.

Beardall J, Giordano M. Ecological implications of microalgal and cyanobacterial CO2 concentrating mechanisms, and their regulation. *Funct Plant Biol* 2002;**29**:335–47.

Beardall J, Raven JA. Cyanobacteria vs green algae: Which group has the edge? *J Exp Bot* 2017;**68**:3697–9.

Bell TAS, Doig L, Peyton BM *et al.* Contributions of the microbial community to algal biomass and biofuel productivity in a wastewater treatment lagoon system. *Algal Res* 2019;**39**:101461.

Belzile N, Chen YW. Thallium in the environment: A critical review focused on natural waters, soils, sediments and airborne particles. *Appl Geochemistry* 2017;**84**:218–43.

Beyter D, Tang P-Z, Becker S *et al.* Diversity, Productivity, and Stability of an Industrial Microbial Ecosystem. *Appl Environ Microbiol* 2016;**82**:2494–505.

Biscotti MA, Olmo E, Heslop-Harrison JS. Repetitive DNA in eukaryotic genomes. *Chromosom Res* 2015;**23**:415–20.

Blaby IK, Blaby-Haas CE, Tourasse N *et al.* The Chlamydomonas genome project: a decade on. *Trends Plant Sci* 2014;**19**:672–80.

Blaby-Haas CE, Merchant SS. Comparative and functional algal genomics. *Annu Rev ofPlant Biol Comp* 2019;**70**:605–38.

Blaby-Haas CE, Merchant SS. The ins and outs of algal metal transport. *Biochim Biophys Acta - Mol Cell Res* 2012;**1823**:1531–52.

Bokulich NA, Kaehler BD, Rideout JR *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;**6**:1–17.

Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–7.

Bonnot C, Proust H, Pinson B *et al.* Functional PTB phosphate transporters are present in streptophyte algae and early diverging land plants. *New Phytol* 2017;**214**:1158–71.

Boyle NR, Sengupta N, Morgan JA. Metabolic flux analysis of heterotrophic growth in Chlamydomonas reinhardtii. *PLoS One* 2017;**12**:e0177292.

Brejovńa B, Vinař T, Chen Y *et al.* Finding genes in Schistosoma japonicum: Annotating novel genomes with help of extrinsic evidence. *Nucleic Acids Res* 2009;**37**:e52.

Breuer G, Evers WAC, de Vree JH *et al.* Analysis of fatty acid content and composition in microalgae. *J Vis Exp* 2013, DOI: 10.3791/50628.

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;**268**:78–94.

Callahan BJ, McMurdie PJ, Rosen MJ *et al.* Suppl. data DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinforma Appl NOTE* 2009;**25**:1972–3.

Chapman RL. Algae: The world's most important "plants"-an introduction. *Mitig Adapt Strateg Glob Chang* 2013;**18**:5–12.

Chapman RL. Algae: The world's most important "plants"-an introduction. *Mitig Adapt Strateg Glob Chang* 2013;**18**:5–12.

Cheloni G, Slaveykova V. Photo-Oxidative Stress in Green Algae and Cyanobacteria. *React Oxyg Species* 2018;**5**:126–33.

Chen W, Zhang C, Song L *et al.* A high throughput Nile red method for quantitative measurement of neutral lipids in microalgae. *J Microbiol Methods* 2009;**77**:41–7.

Chen X, Qiu CE, Shao JZ. Evidence for K+-dependent HCO3- utilization in the marine diatom Phaeodactylum tricornutum. *Plant Physiol* 2006;**141**:731–6.

Chen Z, Wang L, Qiu S *et al.* Determination of Microalgal Lipid Content and Fatty Acid for Biofuel Production. *Biomed Res Int* 2018;**2018**:1503126.

Chisti Y, Yan J. Energy from algae: Current status and future trends. Algal biofuels - A status report. *Appl Energy* 2011;**88**:3277–9.

Cho SH, Ji S-C, Hur SB *et al.* Optimum temperature and salinity conditions for growth of green algae Chlorella ellipsoidea and Nannochloris oculata. *Fish Sci* 2007;**73**:1050–1056.

Chokshi K, Pancha I, Ghosh A *et al.* Nitrogen starvation-induced cellular crosstalk of ROS-scavenging antioxidants and phytohormone enhanced the biofuel potential of green microalga Acutodesmus dimorphus. *Biotechnol Biofuels* 2017;**10**:60.

Clark ME, Edelmann RE, Duley ML *et al.* Biofilm formation in Desulfovibrio vulgaris Hildenborough is dependent upon protein filaments. *Environ Microbiol* 2007;**9**:2844–54.

Conesa A, Madrigal P, Tarazona S *et al.* A survey of best practices for RNA-seq data analysis. *Genome Biol* 2016;**17**:13.

Cooksey KE, Guckert JB, Williams SA *et al.* Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red. *J Microbiol Methods* 1987;**6**:333–45.

Davidson AL, Chen J. ATP-Binding Cassette Transporters in Bacteria. *Annu Rev Biochem* 2004;**73**:241–68.

Dobin A, Davis CA, Schlesinger F *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21.

Domozych DS, Ciancia M, Fangel JU *et al.* The Cell Walls of Green Algae: A Journey through Evolution and Diversity. *Front Plant Sci* 2012;**3**:82.

Doronina N V., Kaparullina EN, Trotsenko YA *et al.* Chelativorans multitrophicus gen. nov., sp. nov. and Chelativorans oligotrophicus sp. nov., aerobic EDTA-degrading bacteria. *Int J Syst Evol Microbiol* 2010;**60**:1044–51.

Dos Reis Falcão V, Pedroso Tonon A, Cabral Oliveira M *et al.* RNA isolation method for polysaccharide rich algae: Agar producing Gracilaria tenuistipitata (Rhodophyta). *J Appl Phycol* 2008;**20**:9–12.

Earl D, Bradnam K, St John J *et al.* Assemblathon 1: a competitive assessment of de novo short read assembly methods. *Genome Res* 2011;**21**:2224–41.

Eckardt NA. A new chlorophyll degradation pathway. *Plant Cell* 2009;**21**:700.

El-Gebali S, Mistry J, Bateman A *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res* 2019;**47**:D427–32.

Ellegaard M, Ribeiro S. The long-term persistence of phytoplankton resting stages in aquatic “seed banks.” *Biol Rev* 2018;**93**:166–83.

Eustance E, Gardner RD, Moll KM *et al.* Growth, nitrogen utilization and biodiesel potential for two chlorophytes grown on ammonium, nitrate or urea. *J Appl Phycol* 2013;**25**:1663–77.

Fakhry EM, El Maghraby DM. Lipid accumulation in response to nitrogen limitation and variation of temperature in Nannochloropsis salina. *Bot Stud* 2015;**56**:6.

Fields MW, Hise A, Lohman EJ *et al.* Sources and resources: Importance of nutrients, resource allocation, and ecology in microalgal cultivation for lipid accumulation. *Appl Microbiol Biotechnol* 2014;**98**:4805–16.

Floyd GL, Watanabe S, Deason TR. Comparative Ultrastructure of the Zoospores of Eight species of Characium (Chlorophyta). *Arch fur Protistenkd* 1993;**143**:63–73.

Fulbright SP, Robbins-Pianka A, Berg-Lyons D *et al.* Bacterial community changes in an industrial algae production system. *Algal Res* 2018;**31**:147–56.

Gardner RD, Cooksey KE, Mus F *et al.* Use of sodium bicarbonate to stimulate triacylglycerol accumulation in the chlorophyte Scenedesmus sp. and the diatom Phaeodactylum tricornutum. *J Appl Phycol* 2012;**24**:1311–20.

Gardner RD, Lohman E, Gerlach R *et al.* Comparison of CO2 and bicarbonate as inorganic carbon sources for triacylglycerol and starch accumulation in Chlamydomonas reinhardtii. *Biotechnol Bioeng* 2013;**110**:87–96.

Gardner RD, Lohman EJ, Cooksey KE *et al.* Cellular cycling, carbon utilization, and photosynthetic oxygen production during bicarbonate-induced triacylglycerol accumulation in a Scenedesmus sp. *Energies* 2013b;**6**:6060–607.

Georgianna DR, Mayfield SP. Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature* 2012;**488**:329–35.

Gerotto C, Alboresi A, Giacometti GM *et al.* Coexistence of plant and algal energy dissipation mechanisms in the moss Physcomitrella patens. *New Phytol* 2012;**196**:763–73.

Grossman AR. Paths toward Algal Genomics. *Plant Physiol* 2005;**137**:410–27.

Grubert E, Sanders KT. Water Use in the United States Energy System: A National Assessment and Unit Process Inventory of Water Consumption and Withdrawals. *Environ Sci Technol* 2018;**52**:39.

Guarnieri MT, Levering J, Henard CA *et al.* Genome Sequence of the Oleaginous Green Alga, Chlorella vulgaris UTEX 395. *Front Bioeng Biotechnol* 2018;**6**:37.

Haas BJ, Salzberg SL, Zhu W *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 2008;**9**:R7.

Han F, Wang W, Li Y *et al.* Changes of biomass, lipid content and fatty acids composition under a light–dark cyclic culture of Chlorella pyrenoidosa in response to different temperature. *Bioresour Technol* 2013;**132**:182–9.

Han Y, Gao S, Muegge K *et al.* Advanced Applications of RNA Sequencing and Challenges. *Bioinform Biol Insights* 2015;**9**:29–46.

Harrington D. CSG WATER AS A MEDIUM TO GROW MARINE MICROALGAE FOR BIOFUEL PRODUCTION. 2011.

Higgins BT, Gennity I, Fitzgerald PS *et al.* Algal–bacterial synergy in treatment of winery wastewater. *npj Clean Water* 2018;**1**, DOI: 10.1038/s41545-018-0005-y.

Hirooka S, Hirose Y, Kanesaki Y *et al.* Acidophilic green algal genome provides insights into adaptation to an acidic environment. *Proc Natl Acad Sci U S A* 2017;**114**:E8304–13.

Hodgskiss LH, Nagy J, Barnhart EP *et al.* Cultivation of a native alga for biomass and

biofuel accumulation in coal bed methane production water. *Algal Res* 2016;**19**:63–8.

Hodgskiss LH. EVALUATION OF A GREEN ALGA ISOLATE FOR GROWTH AND LIPID ACCUMULATION IN COAL BED METHANE WATER FROM THE POWDER RIVER BASIN. 2015.

Hoekman SK, Broch A, Robbins C *et al.* Review of biodiesel composition, properties, and specifications. *Renew Sustain Energy Rev* 2012;**16**:143–69.

Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 2011;**12**:491.

Hopkinson BM, Meile C, Shen C. Quantification of extracellular carbonic anhydrase activity in two marine diatoms and investigation of its role. *Plant Physiol* 2013;**162**:1142–52.

Hovde BT, Hanschen ER, Steadman Tyler CR *et al.* Genomic characterization reveals significant divergence within Chlorella sorokiniana (Chlorellales, Trebouxiophyceae). *Algal Res* 2018;**35**:449–61.

Hsia SY, Yang SK. Enhancing Algal Growth by Stimulation with LED Lighting and Ultrasound. *J Nanomater* 2015;**2015**, DOI: 10.1155/2015/531352.

Humann JL, Lee T, Ficklin S *et al.* Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. Humana, New York, NY, 2019, 29–51.

Idenyi J, Ebenyi L, Ogah O *et al.* Effect of Different Growth Media on the Cell Densities of Freshwater Microalgae Isolates. *IOSR J Pharm Biol Sci Ver* 2016;**11**:2319–7676.

Ihnken S, Beardall J, Kromkamp JC *et al.* Light acclimation and pH perturbations affect photosynthetic performance in chlorella mass culture. *Aquat Biol* 2014;**22**:95–110.

Imam SH, Buchanan MJ, Shin HC *et al.* The Chlamydomonas cell wall: Characterization of the wall framework. *J Cell Biol* 1985;**101**:1599–607.

Ivanova N, Rohde C, Munk C *et al.* Complete genome sequence of Truepera radiovictrix type strain (RQ-24). *Stand Genomic Sci* 2011;**4**:91–9.

Jayakumar V, Sakakibara Y. Comprehensive evaluation of non-hybrid genome assembly tools for third-generation PacBio long-read sequence data. *Brief Bioinform* 2019;**20**:866–76.

Jia H, Yuan Q. Removal of nitrogen from wastewater using microalgae and microalgae bacteria consortia. *Cogent Environ Sci* 2016;**2**:1–15.

Jiang H, Gao K. EFFECTS OF LOWERING TEMPERATURE DURING CULTURE ON THE PRODUCTION OF POLYUNSATURATED FATTY ACIDS IN THE MARINE DIATOM PHAEODACTYLUM TRICORNUTUM (BACILLARIOPHYCEAE)1. *J Phycol* 2004;**40**:651–4.

Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci* 2019, DOI: 10.1002/pro.3711.

Katoh K, Misawa K, Kuma K-I *et al.* MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* 2002;**30**:3059–66.

Khotimchenko S V., Vas'kovsky VE. An Inositol-Containing Sphingolipid from the Red Alga Gracilaria verrucosa. *Russ J Bioorganic Chem* 2004;**30**:168–71.

Kim BH, Ramanan R, Cho DH *et al.* Role of Rhizobium, a plant growth promoting bacterium, in enhancing algal biomass through mutualistic interaction. *Biomass and Bioenergy* 2014, DOI: 10.1016/j.biombioe.2014.07.015.

Kim KM, Park JH, Bhattacharya D *et al.* Applications of next-generation sequencing to unravelling the evolutionary history of algae. *Int J Syst Evol Microbiol* 2014;**64**:333–45.

Kim MS, Baik KS, Park SC *et al.* Roseomonas frigidaquae sp. nov., isolated from a water-cooling system. *Int J Syst Evol Microbiol* 2009;**59**:1630–4.

Koren S, Walenz BP, Berlin K *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 2017;**27**:722–36.

Korlach J, Bjornson KP, Chaudhuri BP *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Methods Enzymol* 2010;**472**:431–55.

Kothari R, Prasad R, Kumar V *et al.* Production of biodiesel from microalgae Chlamydomonas polypyrenoideum grown on dairy industry wastewater. *Bioresour Technol* 2013;**144**:499–503.

Kozich JJ, Westcott SL, Baxter NT *et al.* Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol* 2013;**79**:5112–20.

Krohn-Molt I, Wemheuer B, Alawi M *et al.* Metagenome Survey of a Multispecies and Alga-Associated Biofilm Revealed Key Elements of Bacterial-Algal Interactions in Photobioreactors. *Appl Environ Microbiol* 2013;**79**:6196–206.

Kurpan Nogueira DP, Silva AF, Araújo OQF *et al.* Impact of temperature and light intensity on triacylglycerol accumulation in marine microalgae. *Biomass and*

Bioenergy 2015;**72**:280–7.

Lakaniemi A-M, Intihar VM, Tuovinen OH *et al.* Growth of Chlorella vulgaris and associated bacteria in photobioreactors. *Microb Biotechnol* 2012;**5**:69–78.

Laurens LML, Chen-Glasser M, McMillan JD. A perspective on renewable bioenergy from photosynthetic algae as feedstock for biofuels and bioproducts. *Algal Res* 2017;**24**:261–4.

Lautenberger JA, Chen ZQ. A method for the preparation of high molecular weight yeast DNA. *Gene Anal Tech* 1987;**4**:87–8.

Lee T, Peace C, Jung S *et al.* GenSAS - An online integrated genome sequence annotation pipeline. *Proc - 2011 4th Int Conf Biomed Eng Informatics, BMEI 2011* 2011;**4**:1967–73.

Lepš J, Šmilauer P. *Multivariate Analysis of Ecological Data Using CANOCO*. Cambridge University Press, 2003.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* 2019, DOI: 10.1093/nar/gkz239.

Lever MA, Torti A, Eickenbusch P *et al.* A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. *Front Microbiol* 2015;**6**:476.

Li P-C, Yu S-W, Li K *et al.* The Mutation of Glu at Amino Acid 3838 of AtMDN1 Provokes Pleiotropic Developmental Phenotypes in Arabidopsis. *Sci Rep* 2016;**6**:36446.

Li S, Ma L, Li H *et al.* Snap: An integrated SNP annotation platform. *Nucleic Acids Res* 2007;**35**:707–10.

Li W, Gao K, Beardall J. Interactive Effects of Ocean Acidification and Nitrogen-Limitation on the Diatom Phaeodactylum tricornutum. *PLoS One* 2012;**7**:1–8.

Li Y, Horsman M, Wang B *et al.* Effects of nitrogen sources on cell growth and lipid accumulation of green alga Neochloris oleoabundans. *Appl Microbiol Biotechnol* 2008;**81**:629–36.

Lian J, Wijffels RH, Smidt H *et al.* The effect of the algal microbiome on industrial production of microalgae. *Microb Biotechnol* 2018;**11**:806–18.

Liu J, Lu Y, Hua W *et al.* A New Light on Photosystem II Maintenance in Oxygenic Photosynthesis. *Front Plant Sci* 2019;**10**, DOI: 10.3389/fpls.2019.00975.

Lohman EJ, Gardner RD, Halverson L *et al.* An efficient and scalable extraction and quantification method for algal derived biofuel. *J Microbiol Methods* 2013b;**94**:235–44.

Lohman EJ, Gardner RD, Halverson L *et al.* Carbon Partitioning in Lipids synthesized by Chlamydomonas reinhardtii when Cultured Under Three Unique Inorganic Carbon Regimes. *Bioresour Technol* 2013;**5**:171–80.

Lohman EJ, Gardner RD, Pedersen T *et al.* Optimized inorganic carbon regime for enhanced growth and lipid accumulation in Chlorella vulgaris. *Biotechnol Biofuels* 2015;**8**:82.

Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res* 2014;**42**:1–8.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:1–21.

Luttgeharm KD, Kimberlin AN, Cahoon EB. Plant Sphingolipid Metabolism and Function. Springer, Cham, 2016, 249–86.

Mackinder LCM. The Chlamydomonas CO2-concentrating mechanism and its potential for engineering photosynthesis in plants. *New Phytol* 2018;**217**:54–61.

Maltsev YI, Konovalenko T V. New finding of green algae with potential for algal biotechnology, Chlorococcum oleofaciens and its molecular investigation. *Regul Mech Biosyst* 2018;**8**:532–9.

Marín I, Arahal DR. The Family Beijerinckiaceae. *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, 115–33.

McDonald D, Clemente JC, Kuczynski J *et al.* The Biological Observation Matrix (BIOM) format or: how I learned to stop worrying and love the ome-ome. *Gigascience* 2012;**1**:7.

Merchant SS, Prochnik SE, Vallon O *et al.* The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 2007;**318**:245–50.

Merhi Z, Polotsky AJ, Bradford AP *et al.* Adiposity Alters Genes Important in Inflammation and Cell Cycle Division in Human Cumulus Granulosa Cell. *Reprod Sci* 2015;**22**:1220–8.

Miller SM. Volvox, Chlamydomonas, Evolution of Multicellularity. *Nat Educ* 2010:3(9):65.

Moll KM, Gardner RD, Eustance EO *et al.* Combining multiple nutrient stresses and bicarbonate addition to promote lipid accumulation in the diatom RGd-1. *Algal Res* 2014;**5**:7–15.

Moriya Y, Itoh M, Okuda S *et al.* KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res* 2007;**35**:W182-5.

Mus F, Toussaint J-P, Cooksey KE *et al.* Physiological and molecular analysis of carbon source supplementation and pH stress-induced lipid accumulation in the marine diatom Phaeodactylum tricornutum. *Appl Microbiol Biotechnol* 2013;**97**:3625–42.

Nagarajan N, Pop M. Sequence assembly demystified. *Nat Rev Genet* 2013;**14**, DOI: 10.1038/nrg3367.

Nawrocki EP. STRUCTURAL RNA HOMOLOGY SEARCH AND ALIGNMENT USING COVARIANCE MODELS. 2009.

Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74.

Pancha I, Chokshi K, Ghosh T *et al.* Bicarbonate supplementation enhanced biofuel production potential as well as nutritional stress mitigation in the microalgae Scenedesmus sp. CCNM 1077. *Bioresour Technol* 2015;**193**:315–23.

Parada AE, Needham DM, Fuhrman JA. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016;**18**:1403–14.

Paredes-Sabja D, Sarker N, Setlow B *et al.* Roles of DacB and spm proteins in clostridium perfringens spore resistance to moist heat, chemicals, and UV radiation. *Appl Environ Microbiol* 2008;**74**:3730–8.

Park JJ, Wang H, Gargouri M *et al.* The response of Chlamydomonas reinhardtii to nitrogen deprivation: A systems biology analysis. *Plant J* 2015;**81**:611–24.

Park S, Nguyen THT, Jin E. Improving lipid production by strain development in microalgae: Strategies, challenges and perspectives. *Bioresour Technol* 2019;**292**:121953.

Pedersen TC, Gardner RD, Gerlach R *et al.* Assessment of Nannochloropsis gaditana growth and lipid accumulation with increased inorganic carbon delivery. *J Appl Phycol* 2018;**30**:2155–66.

Pertea M, Salzberg SL. Using GlimmerM to Find Genes in Eukaryotic Genomes. *Curr Protoc Bioinforma* 2003;**00**:4.4.1-4.4.20.

Phillippy AM, Schatz MC, Pop M. Genome assembly forensics: Finding the elusive mis-assembly. *Genome Biol* 2008;**9**, DOI: 10.1186/gb-2008-9-3-r55.

Pollier J, Vancaester E, Kuzhiumparambil U *et al.* A widespread alternative squalene epoxidase participates in eukaryote steroid biosynthesis. *Nat Microbiol* 2019;**4**:226–33.

Pörtner HO. Ecosystem effects of ocean acidification in times of ocean warming: A physiologist's view. *Mar Ecol Prog Ser* 2008;**373**:203–17.

Prochnik SE, Umen J, Nedelcu AM *et al.* Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. *Science* 2010;**329**:223–6.

Ramamoorthy V, Govindaraj L, Dhanasekaran M *et al.* Combination of driselase and lysing enzyme in one molar potassium chloride is effective for the production of protoplasts from germinated conidia of Fusarium verticillioides. *J Microbiol Methods* 2015;**111**:127–34.

Ramanan R, Kang Z, Kim B-H *et al.* Phycosphere bacterial diversity in green algae reveals an apparent similarity across habitats. *Algal Res* 2015;**8**:140–4.

Ramanan R, Kim B-H, Cho D-H *et al.* Algae–bacteria interactions: Evolution, ecology and emerging applications. *Biotechnol Adv* 2016;**34**:14–29.

Ramaraj R, Tsai DD-W, Chen PH. Chlorophyll is not accurate measurement for algal biomass. *Chiang Mai J Sci* 2013;**40**:1–9.

Reinfelder JR. Carbon concentrating mechanisms in eukaryotic marine phytoplankton. *Ann Rev Mar Sci* 2011;**3**:291–315.

Ritchie RJ. Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth Res* 2006;**89**:27–41.

Roach T, Sedoud A, Krieger-Liszkay A. Acetate in mixotrophic growth medium affects photosystem II in Chlamydomonas reinhardtii and protects against photoinhibition. *Biochim Biophys Acta - Bioenerg* 2013;**1827**:1183–90.

Ronquist F, Teslenko M, Van Der Mark P *et al.* Software for Systematics and Evolution MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Syst Biol* 2012;**61**:539–42.

Roth MS, Cokus SJ, Gallaher SD *et al.* Chromosome-level genome assembly and transcriptome of the green alga Chromochloris zofingiensis illuminates astaxanthin production. *Proc Natl Acad Sci U S A* 2017;**114**:E4296–305.

Sager R, Granick S. Nutricional control of sexuality in Chlamydomonas reinhardtii. *J*

Gen Physiol 1954;**37**:729–42.

Saier MH, Beatty JT, Goffeau A *et al.* The major facilitator superfamily. *J Mol Microbiol Biotechnol* 1999;**1**:257–79.

Sambles C, Moore K, Lux TM *et al.* Metagenomic analysis of the complex microbial consortium associated with cultures of the oil-rich alga Botryococcus braunii. *Microbiologyopen* 2017;**6**, DOI: 10.1002/mbo3.482.

Sasso S, Stibor H, Mittag M *et al.* The natural history of model organisms from molecular manipulation of domesticated chlamydomonas reinhardtii to survival in nature. *Elife* 2018;**7**:1–14.

Sathasivam R, Ki J-S. Heat shock protein genes in the green alga Tetraselmis suecica and their role against redox and non-redox active metals. *Eur J Protistol* 2019;**69**:37–51.

Schaum E, Rost B, Millar AJ *et al.* Variation in plastic responses of a globally distributed picoplankton species to ocean acidification. *Nat Clim Chang* 2013;**3**:298–302.

Schnurr PJ, Espie GS, Allen DG. Algae biofilm growth and the potential to stimulate lipid accumulation through nutrient starvation. *Bioresour Technol* 2013;**136**:337–44.

Segata N, Izard J, Waldron L *et al.* Metagenomic biomarker discovery and explanation. *Genome Biol* 2011;**12**:R60.

Sharma M, Garg HS, Chandra K. Erythro-sphino-4,8-dienine-N-palmitate: an antiviral agent from the green alga Ulva fasciata. *Bot Mar* 1996;**39**:213–5.

Sharma PK, Saharia M, Srivstava R *et al.* Tailoring microalgae for efficient biofuel production. *Front Mar Sci* 2018;**5**:1–19.

Shetty P, Boboescu IZ, Pap B *et al.* Exploitation of Algal-Bacterial Consortia in Combined Biohydrogen Generation and Wastewater Treatment. *Front Energy Res* 2019;**7**:52.

Shubert LE. NONMOTILE COCCOID AND COLONIAL GREEN ALGAE. *Freshwater Algae of North America*. Academic Press, 2003, 253–309.

Shurin JB, Burkart MD, Mayfield SP *et al.* Recent progress and future challenges in algal biofuel production. *F1000Research* 2016;**5**, DOI: 10.12688/f1000research.9217.1.

Simão FA, Waterhouse RM, Ioannidis P *et al.* BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015;**31**:3210–2.

Singh NK, Upadhyay AK, Rai UN. Algal Technologies for Wastewater Treatment and

Biofuels Production: An Integrated Approach for Environmental Management. *Algal Biofuels*. Cham: Springer International Publishing, 2017, 97–107.

Singh SP, Singh P. Effect of temperature and light on the growth of algae species: A review. *Renew Sustain Energy Rev* 2015;**50**:431–44.

Stanke M, Diekhans M, Baertsch R *et al.* Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 2008;**24**:637–44.

Stockenreiter M, Graber A-K, Haupt F *et al.* The effect of species diversity on lipid production by micro-algal communities. *J Appl Phycol* 2012;**24**:45–54.

Talebi AF, Mohtashami SK, Tabatabaei M *et al.* Fatty acids profiling: A selective criterion for screening microalgae strains for biodiesel production. *Algal Res* 2013;**2**:258–67.

Tamas I, Smirnova A V, He Z *et al.* The (d)evolution of methanotrophy in the Beijerinckiaceae--a comparative genomics analysis. *ISME J* 2014;**8**:369–82.

Tanaka K, Shimizu N, Sugano S *et al.* Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. *Nature* 2004;**428**:653–7.

Tasić MB, Pinto LFR, Klein BC *et al.* Botryococcus braunii for biodiesel production. *Renew Sustain Energy Rev* 2016;**64**:260–70.

Taziki M, Ahmadzadeh H, A. Murry M. Growth of Chlorella vulgaris in High Concentrations of Nitrate and Nitrite for Wastewater Treatment. *Curr Biotechnol* 2016;**4**:441–7.

Tejada-Jiménez M, Galván A, Fernández E. Algae and humans share a molybdate transporter. *Proc Natl Acad Sci U S A* 2011;**108**:6420–5.

Tirichine L, Bowler C. Decoding algal genomes: Tracing back the history of photosynthetic life on Earth. *Plant J* 2011;**66**:45–57.

Urakami T, Oyanag H, Araki H *et al.* Recharacterization and Emended Description of the Genus Mycoplana and Description of Two New Species, Mycoplana ramosa and Mycoplana segnis. *Int J Syst Bacteriol* 1990;**40**:434–42.

Valenzuela J, Mazurie A, Carlson RP *et al.* Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. *Biotechnol Biofuels* 2012;**5**:40.

Valenzuela J, Mazurie A, Carlson RP *et al.* Potential role of multiple carbon fixation pathways during lipid accumulation in Phaeodactylum tricornutum. *Biotechnol Biofuels* 2012;**5**:40.

Varshney P, Mikulic P, Vonshak A *et al.* Extremophilic micro-algae and their potential contribution in biotechnology. *Bioresour Technol* 2015;**184**:363–72.

White DA, Pagarette A, Rooks P *et al.* The effect of sodium bicarbonate supplementation on growth and biochemical composition of marine microalgae cultures. *J Appl Phycol* 2013;**25**:153–65.

White TJ, Bruns T, Lee S *et al.* AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL RIBOSOMAL RNA GENES FOR PHYLOGENETICS. *PCR Protoc* 1990:315–22.

Wibberg D, Blom J, Jaenicke S *et al.* Complete genome sequencing of Agrobacterium sp. H13-3, the former Rhizobium lupini H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J Biotechnol* 2011;**155**:50–62.

Wickham H. *Ggplot2 : Elegant Graphics for Data Analysis*. Springer, 2016.

Wu Y-W. ezTree: an automated pipeline for identifying phylogenetic marker genes and inferring evolutionary relationships among uncultivated prokaryotic draft genomes. *BMC Genomics* 2018;**19**:921.

Xia L, Yang H, He Q *et al.* Physiological responses of freshwater oleaginous microalgae Desmodesmus sp. NMX451 under nitrogen deficiency and alkaline pH-induced lipid accumulation. *J Appl Phycol* 2014;**27**:649–59.

Xiao X, Pei M, Liu X *et al.* Planktonic algal bloom significantly alters sediment bacterial community structure. *J Soils Sediments* 2017;**17**:2547–56.

Yurkov V, Jappe J, Vermeglio A. Tellurite resistance and reduction by obligately aerobic photosynthetic bacteria. *Appl Environ Microbiol* 1996;**62**:4195–8.

Zdobnov EM, Apweiler R. InterProScan - An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 2001;**17**:847–8.

Zhang YM, Chen H, He CL *et al.* Nitrogen Starvation Induced Oxidative Stress in an Oil-Producing Green Alga Chlorella sorokiniana C3. *PLoS One* 2013;**8**:1–12.

Zhao L-S, Li K, Wang Q-M *et al.* Nitrogen Starvation Impacts the Photosynthetic Performance of Porphyridium cruentum as Revealed by Chlorophyll a Fluorescence. *Sci Rep* 2017;**7**:8542.

Zhao S, Zhang B. A comprehensive evaluation of ensembl, RefSeq, and UCSC annotations in the context of RNA-seq read mapping and gene quantification. *BMC Genomics* 2015;**16**:97.

Zhou J, Richlen ML, Sehein TR *et al.* Microbial Community Structure and Associations During a Marine Dinoflagellate Bloom. *Front Microbiol* 2018;**9**:1201.

Zhou W, Li Y, Min M *et al.* Local bioprospecting for high-lipid producing microalgal strains to be grown on concentrated municipal wastewater for biofuel production. *Bioresour Technol* 2011;**102**:6909–19.

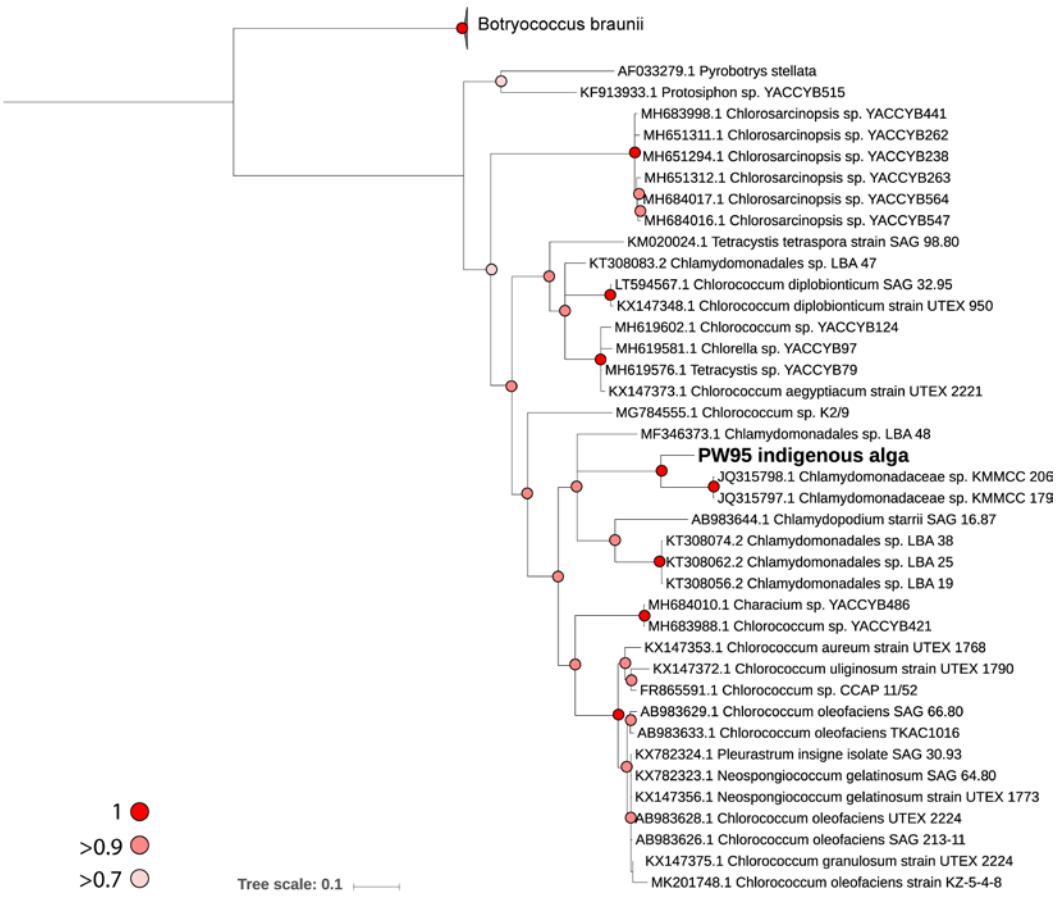
Zhu J, Chen W, Chen H *et al.* Improved Productivity of Neutral Lipids in Chlorella sp. A2 by Minimal Nitrogen Supply. *Front Microbiol* 2016;**7**, DOI: 10.3389/fmicb.2016.00557.

Zhu LD, Li ZH, Hiltunen E. Strategies for Lipid Production Improvement in Microalgae as a Biodiesel Feedstock. *Biomed Res Int* 2016;**2016**:1–8.

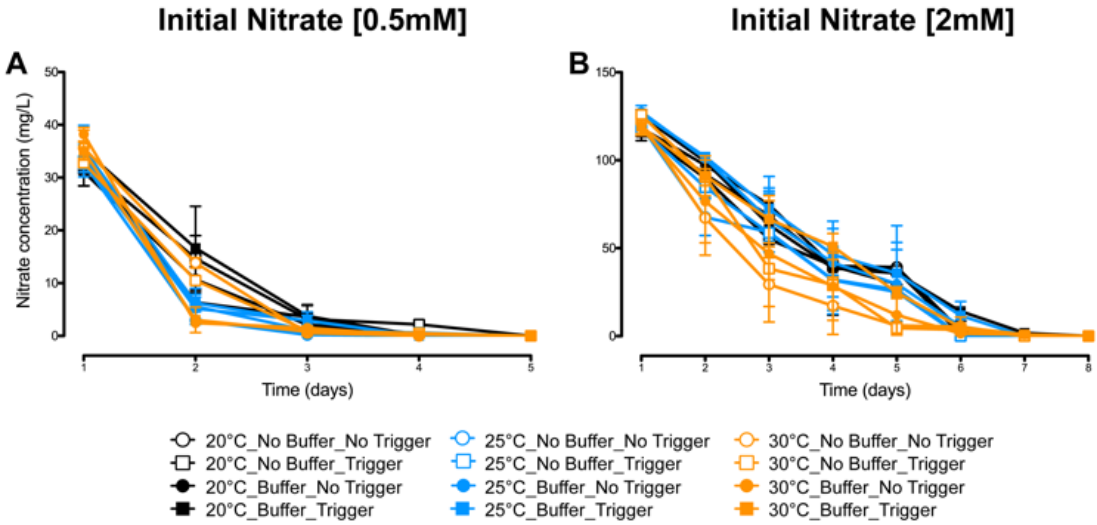
APPENDICES

APPENDIX A

SUPPLEMENTARY MATERIALS FOR CHAPTER TWO



Supplementary Figure A1. Phylogenetic relations of PW95 native isolate (indicated in bold) based on the Internal Transcribed Spacer-ITS consensus sequence. 46 sequences were used for the analysis and the phylogenetic tree was constructed using Bayesian Inference with the following parameters: General Time Reversible (GTR) model with 4 gamma categories and 2 parallel runs, 8 chains, 2 million iterations and 0.25 burn-in fraction. All Bayesian posterior probabilities are shown in colored dots on the branches. The global average standard deviation of split sequences was 0.003. *Boytriococcus braunii* was chosen as outgroup (sequences shown as a collapsed clade).



Supplementary Figure A2. Comparison of nitrate depletion between the two levels of initial nitrate (0.5 mM and 2 mM) for all the combinations of treatments evaluated in this study. Error bars represent the standard error of the mean of two independent replicates. The window of nitrate depletion for treatments as 0.5 mM of initial nitrate was between days 3-4 and for treatments at 2 mM was days 6-7. Temperatures are differentiated by color: 20°C (black), 25°C (blue) and 30°C (orange). Filled symbols represent treatments with Buffer and clear symbols represent No Buffer. Treatments with Trigger are represented by squares and No Trigger with circles. Error bars represent the standard error of the mean of two independent replicates.

APPENDIX B

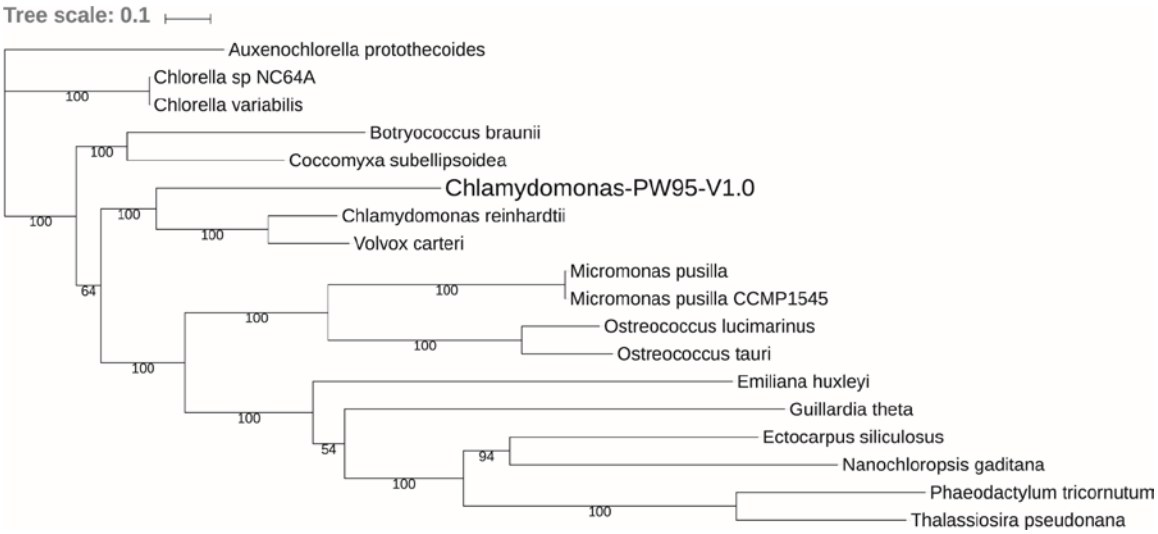
SUPPLEMENTARY MATERIALS FOR CHAPTER THREE

Supplementary Table B1. Conserved single copy marker protein set. Single copy marker proteins shared by all genomes including PW95 for phylogenetic tree construction.

ID	PROTEIN NAME
PF01092	Ribosomal protein S6e
PF01157	Ribosomal protein L21e
PF01198	Ribosomal protein L31e
PF01280	Ribosomal protein L19e
PF01282	Ribosomal protein S24e
PF01294	Ribosomal protein L13e
PF01491	Frataxin-like domain
PF01655	Ribosomal protein L32
PF01777	Ribosomal L27e protein family
PF02167	Cytochrome C1 family
PF02271	Ubiquinol-cytochrome C reductase complex 14kD subunit
PF02542	YgbB family
PF02548	Ketopantoate hydroxymethyltransferase
PF02966	Mitosis protein DIM1
PF03568	Peptidase family C50
PF03690	Uncharacterised protein family (UPF0160)
PF03849	Transcription factor Tfb2
PF05019	Coenzyme Q (ubiquinone) biosynthesis protein Coq4
PF05091	Eukaryotic translation initiation factor 3 subunit 7 (eIF-3)
PF05997	Nucleolar protein, Nop52
PF06047	NF-kappa-B-activating protein C-terminal domain
PF06246	Isy1-like splicing family
PF06799	Conserved in the green lineage and diatoms 27
PF07109	Magnesium-protoporphyrin IX methyltransferase C-terminus
PF08149	BING4CT (NUC141) domain
PF08153	NGP1NT (NUC091) domain
PF16399	Intron-binding protein aquarius N-terminus
PF17135	Ribosomal protein 60S L18 and 50S L18e
PF17144	Ribosomal large subunit proteins 60S L5, and 50S L18
PF17833	UPF0113 Pre-PUA domain

Supplementary Table B2. List of selected organisms for functional annotation in KEGG Automatic Annotation (KAAS). Functional annotation was performed using single-directional best hit (SBH) by BLAST against the manually curated KEGG GENES database against 40 algal, plant and bacterial species.

Group	Organism	Common name	KAAS ID
Algae	*Chlamydomonas reinhardtii*	Green algae	cre
	Ostreococcus lucimarinus		olu
	Ostreococcus tauri		ota
	Micromonas commoda		mis
	Emiliania huxleyi	Cryptophyte	ehx
	Guillardia theta		gtt
	Phaeodactylum tricornutum	Diatom	pti
	Thalassiosira pseudonana		tps
Plants	*Arabidopsis thaliana*	Thale cress	ath
	Arabidopsis lyrata	lyrate rockcress	aly
	Glycine max	Soybean	gmx
	Oryza sativa japonica	Japanese rice	osa
	Zea mays	Maize	zma
	Olea europaea var. sylvestris	Wild olive	oeu
Cyanobacteria	*Synechocystis sp. PCC 6803*	Cyanobacteria	syn
	Synechococcus sp. WH8102		syw
	Synechococcus elongatus PCC6301		syc
	Synechococcus elongatus PCC7942		syf
	Synechococcus sp. CC9605		syd
	Synechococcus sp. CC9902		sye
	Synechococcus sp. CC9311		syg
	Synechococcus sp. RCC307		syr
	Synechococcus sp. WH7803		syx
	Synechococcus sp. PCC7002		syp
	Synechococcus sp. JA-3-3Ab		cya
	Synechococcus sp. JA-2-3B'a		cyb
Unicellular amoeba	*Dictyostelium discoideum*	Cellular slime mold	ddi



Supplementary Figure B1. Maximum likelihood tree of PW95 phylogenetic relationships. IQ-TREE was used to generate a ML tree (100 replicates in ultrafast bootstrap approximation mode). model parameters to be optimal: LG substitution model with empirical residue frequencies, gamma-shaped rate variation (4 categories) with a proportion of invariable sites (LG+F+I+G4).

APPENDIX C

SUPPLEMENTARY MATERIALS FOR CHAPTER FOUR

Supplementary Table C1. Laboratory analytical report of CBM water collected from well FG-09. Reported values for the CBM water used in this study using standardized media components for algal growth as a reference.

Analyses	Result	Units
NON-METALS		
Inorganic Carbon, Dissolved (DIC)	126	mg/L
INORGANICS		
Chloride	64	mg/L
Sulfate	728	mg/L
AGGREGATE ORGANICS		
Organic Carbon, Dissolved (DOC)	3	mg/L
NUTRIENTS		
Nitrogen, Ammonia as N	0.67	mg/L
Nitrogen, Nitrite as N	ND	mg/L
Nitrogen, Nitrate as N	ND	mg/L
Nitrogen, Nitrate+Nitrite as N	ND	mg/L
Nitrogen, Kjeldahl, Total as N	0.8	mg/L
Nitrogen, Total	0.8	mg/L
Phosphorus, Total as P	0.029	mg/L
METALS, TOTAL		
Calcium	3	mg/L
Iron	0.04	mg/L
Magnesium	2	mg/L
Manganese	0.007	mg/L
Potassium	3	mg/L
Sodium	611	mg/L

Supplementary Table C2. List of selected organisms for functional annotation in KEGG Automatic Annotation (KAAS). Functional annotation was performed using single-directional best hit (SBH) by BLAST against the manually curated KEGG GENES database against 40 algal, plant and bacterial species.

Group	Organism	Common name	KAAS ID
Algae	*Chlamydomonas reinhardtii*	Green algae	cre
	Ostreococcus lucimarinus		olu
	Ostreococcus tauri		ota
	Micromonas commoda		mis
	Emiliania huxleyi		ehx
	Guillardia theta	Cryptophyte	gtt
	Phaeodactylum tricornutum	Diatom	pti
	Thalassiosira pseudonana		tps
Plants	*Arabidopsis thaliana*	Thale cress	ath
	Arabidopsis lyrata	lyrate rockcress	aly
	Glycine max	Soybean	gmx
	Oryza sativa japonica	Japanese rice	osa
	Zea mays	Maize	zma
	Olea europaea var. sylvestris	Wild olive	oeu
Cyanobacteria	*Synechocystis sp. PCC 6803*	Cyanobacteria	syn
	Synechococcus sp. WH8102		syw
	Synechococcus elongatus PCC6301		syc
	Synechococcus elongatus PCC7942		syf
	Synechococcus sp. CC9605		syd
	Synechococcus sp. CC9902		sye
	Synechococcus sp. CC9311		syg
	Synechococcus sp. RCC307		syr
	Synechococcus sp. WH7803		syx
	Synechococcus sp. PCC7002		syp
	Synechococcus sp. JA-3-3Ab		cya
	Synechococcus sp. JA-2-3B'a		cyb
Unicellular amoeba	*Dictyostelium discoideum*	Cellular slime mold	ddi

Supplementary Table C3. Significant DEGs for late exponential phase of growth, filtered vs. unfiltered water (E-F vs. E-UF). The table shows the PW95 DEG with the corresponding functional annotation and log2 fold change. Up-regulated DEGs have positive fold change and down regulated have negative fold change. NA stands for non-assigned DEGs.

PW95 Gene	Fold Change	Assigned Gene/ Protein	Name	Function
PW-WITH-RNA-SEQ.00g079640.m01-v1.0.a1	2.43	SOD2	superoxide dismutase,Fe-Mn family [EC:1.15.1.1]	ROS defense
PW-WITH-RNA-SEQ.00g044180.m01-v1.0.a1	2.24	SLC25A20_29	solute carrier family 25 (mitochondrial carnitine/acylcarnitine transporter)	Thermogenesis/ Transporters
PW-WITH-RNA-SEQ.00g072680.m01-v1.0.a1	1.98	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms/ / Exosome
PW-WITH-RNA-SEQ.00g145440.m01-v1.0.a1	1.92	NA	NA	NA
PW-WITH-RNA-SEQ.00g086330.m01-v1.0.a1	1.88	chlM	magnesium-protoporphyrin O-methyltransferase [EC:2.1.1.11]	Porphyrin and chlorophyll metabolism
PW-WITH-RNA-SEQ.00g204680.m01-v1.0.a1	1.86	NA	NA	NA
PW-WITH-RNA-SEQ.00g090780.m01-v1.0.a1	1.70	chlE	magnesium-protoporphyrin IX monomethyl ester (oxidative) cyclase [EC:1.14.13.81]	Porphyrin and chlorophyll metabolism
PW-WITH-RNA-SEQ.00g065060.m01-v1.0.a1	1.67	GAPA	glyceraldehyde-3-phosphate dehydrogenase	Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g093730.m01-v1.0.a1	-1.31	NA	NA	NA
PW-WITH-RNA-SEQ.00g113750.m01-v1.0.a1	-1.48	E3.4.23.40	phytepsin [EC:3.4.23.40]	Peptidases
PW-WITH-RNA-SEQ.00g200660.m01-v1.0.a1	-1.70	NA	NA	NA
PW-WITH-RNA-SEQ.00g022340.m01-v1.0.a1	-1.76	SELENBP1	methanethiol oxidase [EC:1.8.3.4]	Sulfur metabolism/ Exosome

PW-WITH-RNA-SEQ.00g043960.m01-v1.0.a1	-1.95	SLC10A7	solute carrier family 10 (sodium/bile acid cotransporter)	Transporters/ Exosome
PW-WITH-RNA-SEQ.00g072160.m01-v1.0.a1	-2.07	NA	NA	NA
PW-WITH-RNA-SEQ.00g201190.m01-v1.0.a1	-2.22	GST	glutathione S-transferase [EC:2.5.1.18]	Glutathione metabolism/ Metabolism of xenobiotics by cytochrome P450/ Transporters
PW-WITH-RNA-SEQ.00g023820.m01-v1.0.a1	-2.59	NA	NA	NA
PW-WITH-RNA-SEQ.00g054620.m01-v1.0.a1	-2.79	CDC25	Cdc25 family phosphatase [EC:3.1.3.48 1.20.4.1]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g161440.m01-v1.0.a1	-3.25	PCCB	propionyl-CoA carboxylase beta chain [EC:6.4.1.3 2.1.3.15]	Glyoxylate and dicarboxylate metabolism/ Propanoate metabolism/ Valine, leucine and isoleucine degradation
PW-WITH-RNA-SEQ.00g184920.m01-v1.0.a1	-4.16	PCCB	propionyl-CoA carboxylase beta chain [EC:6.4.1.3 2.1.3.15]	Glyoxylate and dicarboxylate metabolism/ Propanoate metabolism/ Valine, leucine and isoleucine degradation
PW-WITH-RNA-SEQ.00g163440.m01-v1.0.a1	-4.42	NA	NA	NA
PW-WITH-RNA-SEQ.00g022150.m01-v1.0.a1	-5.09	NA	NA	NA
PW-WITH-RNA-SEQ.00g201260.m01-v1.0.a1	-5.40	NA	NA	NA
PW-WITH-RNA-SEQ.00g023780.m01-v1.0.a1	-6.27	NA	NA	NA
PW-WITH-RNA-SEQ.00g201250.m01-v1.0.a1	-7.17	NA	NA	NA
PW-WITH-RNA-SEQ.00g149320.m01-v1.0.a1	-16.86	nrtA	nitrate/nitrite transport system substrate-binding protein	Nitrogen metabolism/ ABC transporters/ Transporters

Supplementary Table C4. Significant DEGs for late stationary phase of growth, filtered vs. unfiltered water (S-F vs. S-UF). The table shows the PW95 DEG with the corresponding functional annotation and log2 fold change. Up-regulated DEGs have positive fold change and down regulated have negative fold change. NA stands for non-assigned DEGs.

PW95 Gene	Fold Change	Assigned Gene/ Protein	Name	Function
PW-WITH-RNA-SEQ.00g092220.m01-v1.0.a1	2.55	SLHCA1	light-harvesting complex I chlorophyll a/b binding protein 1	Photosynthesis - antenna proteins
PW-WITH-RNA-SEQ.00g099790.m01-v1.0.a1	2.45	LHCA4	light-harvesting complex I chlorophyll a/b binding protein 4	Photosynthesis - antenna proteins
PW-WITH-RNA-SEQ.00g002950.m01-v1.0.a1	2.42	NDUFS8	NADH dehydrogenase (ubiquinone) Fe-S protein 8 [EC:7.1.1.2 1.6.99.3]	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g152080.m01-v1.0.a1	2.26	metK	S-adenosylmethionine synthetase [EC:2.5.1.6]	Cysteine and methionine metabolism
PW-WITH-RNA-SEQ.00g038750.m01-v1.0.a1	2.17	RP-L40e	large subunit ribosomal protein L40e	Ribosome
PW-WITH-RNA-SEQ.00g005090.m01-v1.0.a1	2.15	ftsH	cell division protease FtsH [EC:3.4.24.-]	Peptidases - Chaperones and folding catalysts
PW-WITH-RNA-SEQ.00g079640.m01-v1.0.a1	2.07	SOD2	superoxide dismutase	
PW-WITH-RNA-SEQ.00g177230.m01-v1.0.a1	2.05	psbQ	photosystem II oxygen-evolving enhancer protein 3	Photosynthesis
PW-WITH-RNA-SEQ.00g185990.m01-v1.0.a1	1.89	IDH1	isocitrate dehydrogenase [EC:1.1.1.42]	Citrate cycle (TCA cycle)

PW-WITH-RNA-SEQ.00g123990.m01-v1.0.a1	1.87	RP-S6e	small subunit ribosomal protein S6e	Ribosome
PW-WITH-RNA-SEQ.00g146300.m01-v1.0.a1	1.86	NA	NA	NA
PW-WITH-RNA-SEQ.00g092240.m01-v1.0.a1	1.85	RP-L18Ae	large subunit ribosomal protein L18Ae	Ribosome
PW-WITH-RNA-SEQ.00g072680.m01-v1.0.a1	1.85	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g113120.m01-v1.0.a1	1.85	RACK1	guanine nucleotide-binding protein subunit beta-2-like 1 protein	Messenger RNA biogenesis
PW-WITH-RNA-SEQ.00g112550.m01-v1.0.a1	1.84	H2B	histone H2B	Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g097000.m01-v1.0.a1	1.82	RP-S11e	small subunit ribosomal protein S11e	Ribosome
PW-WITH-RNA-SEQ.00g195400.m01-v1.0.a1	1.80	trpB	tryptophan synthase beta chain [EC:4.2.1.20]	Glycine, serine and threonine metabolism/ Phenylalanine, tyrosine and tryptophan biosynthesis
PW-WITH-RNA-SEQ.00g165000.m01-v1.0.a1	1.79	RP-S27Ae	small subunit ribosomal protein S27Ae	Ribosome
PW-WITH-RNA-SEQ.00g045570.m01-v1.0.a1	1.77	RP-S3e	small subunit ribosomal protein S3e	Ribosome
PW-WITH-RNA-SEQ.00g145470.m01-v1.0.a1	1.76	RP-L12e	large subunit ribosomal protein L12e	Ribosome
PW-WITH-RNA-SEQ.00g027920.m01-v1.0.a1	1.73	RP-S14e	small subunit ribosomal protein S14e	Ribosome

PW-WITH-RNA-SEQ.00g133960.m01-v1.0.a1	1.73	RP-L11e	large subunit ribosomal protein L11e	Ribosome
PW-WITH-RNA-SEQ.00g026690.m01-v1.0.a1	1.73	RP-L18e	large subunit ribosomal protein L18e	Ribosome
PW-WITH-RNA-SEQ.00g175810.m01-v1.0.a1	1.72	RP-L44e	large subunit ribosomal protein L44e	Ribosome
PW-WITH-RNA-SEQ.00g071600.m01-v1.0.a1	1.72	RP-S9e	small subunit ribosomal protein S9e	Ribosome
PW-WITH-RNA-SEQ.00g034650.m01-v1.0.a1	1.71	E4.1.3.1	isocitrate lyase [EC:4.1.3.1]	Glyoxylate and dicarboxylate metabolism
PW-WITH-RNA-SEQ.00g175110.m01-v1.0.a1	1.70	RP-S8e	small subunit ribosomal protein S8e	Ribosome
PW-WITH-RNA-SEQ.00g119750.m02-v1.0.a1	1.70	RP-L9e	large subunit ribosomal protein L9e	Ribosome
PW-WITH-RNA-SEQ.00g141570.m01-v1.0.a1	1.70	RP-L31e	large subunit ribosomal protein L31e	Ribosome
PW-WITH-RNA-SEQ.00g188450.m01-v1.0.a1	1.70	RP-S20e	small subunit ribosomal protein S20e	Ribosome
PW-WITH-RNA-SEQ.00g095260.m01-v1.0.a1	1.69	RP-L10e	large subunit ribosomal protein L10e	Ribosome
PW-WITH-RNA-SEQ.00g048860.m01-v1.0.a1	1.69	RP-L15e	large subunit ribosomal protein L15e	Ribosome
PW-WITH-RNA-SEQ.00g001950.m01-v1.0.a1	1.68	trpB	tryptophan synthase beta chain [EC:4.2.1.20]	Glycine, serine and threonine metabolism
PW-WITH-RNA-SEQ.00g096000.m01-v1.0.a1	1.68	RP-L30e	large subunit ribosomal protein L30e	Ribosome

PW-WITH-RNA-SEQ.00g134590.m01-v1.0.a1	1.66	RP-S15e	small subunit ribosomal protein S15e	Ribosome
PW-WITH-RNA-SEQ.00g117120.m01-v1.0.a1	1.65	RP-L14e	large subunit ribosomal protein L14e	Ribosome
PW-WITH-RNA-SEQ.00g131590.m01-v1.0.a1	1.65	RP-S23e	small subunit ribosomal protein S23e	Ribosome
PW-WITH-RNA-SEQ.00g150450.m01-v1.0.a1	1.64	RP-S7e	small subunit ribosomal protein S7e	Ribosome
PW-WITH-RNA-SEQ.00g103150.m01-v1.0.a1	1.62	RP-L4e	large subunit ribosomal protein L4e	Ribosome
PW-WITH-RNA-SEQ.00g175130.m01-v1.0.a1	1.61	RP-S18e	small subunit ribosomal protein S18e	Ribosome
PW-WITH-RNA-SEQ.00g115510.m01-v1.0.a1	1.61	RP-S15Ae	small subunit ribosomal protein S15Ae	Ribosome
PW-WITH-RNA-SEQ.00g074070.m01-v1.0.a1	1.61	RP-S13e	small subunit ribosomal protein S13e	Ribosome
PW-WITH-RNA-SEQ.00g170330.m01-v1.0.a1	1.60	RP-SAe	small subunit ribosomal protein SAe	Ribosome
PW-WITH-RNA-SEQ.00g049440.m01-v1.0.a1	1.60	RP-L3e	large subunit ribosomal protein L3e	Ribosome
PW-WITH-RNA-SEQ.00g074880.m01-v1.0.a1	1.59	RP-L6e	large subunit ribosomal protein L6e	Ribosome
PW-WITH-RNA-SEQ.00g030690.m01-v1.0.a1	1.59	RP-S10e	small subunit ribosomal protein S10e	Ribosome
PW-WITH-RNA-SEQ.00g105450.m01-v1.0.a1	1.59	RP-S2e	small subunit ribosomal protein S2e	Ribosome

PW-WITH-RNA-SEQ.00g174550.m01-v1.0.a1	1.58	RP-L8e	large subunit ribosomal protein L8e	Ribosome
PW-WITH-RNA-SEQ.00g030350.m01-v1.0.a1	1.57	RP-L39e	large subunit ribosomal protein L39e	Ribosome
PW-WITH-RNA-SEQ.00g165370.m01-v1.0.a1	1.56	RP-L19e	large subunit ribosomal protein L19e	Ribosome
PW-WITH-RNA-SEQ.00g114390.m01-v1.0.a1	1.54	ALDO	fructose-bisphosphate aldolase	Glycolysis / Gluconeogenesis / Pentose phosphate pathway / Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g142130.m01-v1.0.a1	1.53	fabD, MCT1	[acyl-carrier-protein] S-malonyltransferase	Fatty acid biosynthesis
PW-WITH-RNA-SEQ.00g122590.m01-v1.0.a1	1.53	FARSA	phenylalanyl-tRNA synthetase alpha chain [EC:6.1.1.20]	Aminoacyl-tRNA biosynthesis
PW-WITH-RNA-SEQ.00g097080.m01-v1.0.a1	1.51	RP-S12e	small subunit ribosomal protein S12e	Ribosome
PW-WITH-RNA-SEQ.00g024300.m01-v1.0.a1	1.51	sat	sulfate adenylyltransferase [EC:2.7.7.4]	Sulfur metabolism / Purine metabolism
PW-WITH-RNA-SEQ.00g117190.m01-v1.0.a1	1.50	RP-L5e	large subunit ribosomal protein L5e	Ribosome
PW-WITH-RNA-SEQ.00g085290.m01-v1.0.a1	1.50	RP-L37e	large subunit ribosomal protein L37e	Ribosome
PW-WITH-RNA-SEQ.00g065060.m01-v1.0.a1	1.50	GAPA	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating) [EC:1.2.1.13]	Carbon fixation in photosynthetic organisms

PW-WITH-RNA-SEQ.00g159440.m01-v1.0.a1	1.50	RP-L7e	large subunit ribosomal protein L7e	Ribosome
PW-WITH-RNA-SEQ.00g107830.m01-v1.0.a1	1.48	RP-L7Ae	large subunit ribosomal protein L7Ae	Ribosome biogenesis
PW-WITH-RNA-SEQ.00g199720.m01-v1.0.a1	1.47	ppa	inorganic pyrophosphatase [EC:3.6.1.1]	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g118630.m01-v1.0.a1	1.46	rpiA	ribose 5-phosphate isomerase A [EC:5.3.1.6]	Pentose phosphate pathway / Carbon fixation in photosynthetic organisms
__ambiguous	1.45	NA	NA	NA
PW-WITH-RNA-SEQ.00g024420.m01-v1.0.a1	1.45	E3.1.3.37	sedoheptulose-bisphosphatase [EC:3.1.3.37]	Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g075540.m01-v1.0.a1	1.44	acpP	acyl carrier protein	Fatty acid biosynthesis
PW-WITH-RNA-SEQ.00g187650.m01-v1.0.a1	1.43	ATPF1D	F-type H+-transporting ATPase subunit delta	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g071100.m01-v1.0.a1	1.42	NSA2	ribosome biogenesis protein NSA2	Ribosome biogenesis
PW-WITH-RNA-SEQ.00g070290.m01-v1.0.a1	1.41	RP-LP0	large subunit ribosomal protein LP0	Ribosome
PW-WITH-RNA-SEQ.00g138770.m01-v1.0.a1	1.40	RP-S3Ae	small subunit ribosomal protein S3Ae	Ribosome
PW-WITH-RNA-SEQ.00g103380.m01-v1.0.a1	1.39	RP-LP1	large subunit ribosomal protein LP1	Ribosome
PW-WITH-RNA-SEQ.00g085760.m01-v1.0.a1	1.39	RP-L13e	large subunit ribosomal protein L13e	Ribosome
PW-WITH-RNA-SEQ.00g165610.m01-v1.0.a1	1.38	RP-L32e	large subunit ribosomal protein L32e	Ribosome

PW-WITH-RNA-SEQ.00g193110.m01-v1.0.a1	1.38	PDHB	pyruvate dehydrogenase E1 component beta subunit [EC:1.2.4.1]	Glycolysis / Gluconeogenesis / TCA cycle
PW-WITH-RNA-SEQ.00g121580.m01-v1.0.a1	1.37	E2.6.1.83	LL-diaminopimelate aminotransferase [EC:2.6.1.83]	Lysine biosynthesis
PW-WITH-RNA-SEQ.00g055670.m01-v1.0.a1	1.35	PDHA	pyruvate dehydrogenase E1 component alpha subunit [EC:1.2.4.1]	Glycolysis / Gluconeogenesis
PW-WITH-RNA-SEQ.00g101380.m01-v1.0.a1	1.33	fabF	3-oxoacyl-[acyl-carrier-protein] synthase II	Fatty acid biosynthesis
PW-WITH-RNA-SEQ.00g138980.m01-v1.0.a1	1.33	RP-L35Ae	large subunit ribosomal protein L35Ae	Ribosome
PW-WITH-RNA-SEQ.00g163770.m01-v1.0.a1	1.33	DLD	dihydrolipoamide dehydrogenase	TCA cycle/ Pyruvate metabolism/ Glyoxylate and dicarboxylate metabolism/ Propanoate metabolism/ Glycine, serine and threonine metabolism/ Valine, leucine and isoleucine degradation/ Lysine degradation/ Tryptophan metabolism
PW-WITH-RNA-SEQ.00g191140.m01-v1.0.a1	1.33	argC	N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]	Arginine biosynthesis
PW-WITH-RNA-SEQ.00g196000.m01-v1.0.a1	1.33	accA	acetyl-CoA carboxylase carboxyl transferase subunit alpha [EC:6.4.1.2 2.1.3.15]	Pyruvate metabolism

PW-WITH-RNA-SEQ.00g113180.m01-v1.0.a1	1.31	H2A	histone H2B	Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g112940.m01-v1.0.a1	1.30	RP-L27e	large subunit ribosomal protein L27e	Ribosome
PW-WITH-RNA-SEQ.00g135750.m01-v1.0.a1	1.30	ATPF1G	F-type H+-transporting ATPase subunit gamma	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g186000.m01-v1.0.a1	1.30	IDH1	isocitrate dehydrogenase	Citrate cycle (TCA cycle)/ Photosynthesis
PW-WITH-RNA-SEQ.00g089850.m01-v1.0.a1	1.26	EEF1A	elongation factor 1-alpha	RNA transport
PW-WITH-RNA-SEQ.00g047540.m01-v1.0.a1	1.25	EF3	elongation factor 3	Translation factors
PW-WITH-RNA-SEQ.00g131410.m01-v1.0.a1	1.22	DYNLL	dynein light chain LC8-type	Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g104920.m01-v1.0.a1	1.20	ATPeV0C	V-type H+-transporting ATPase 16kDa proteolipid subunit	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g042120.m01-v1.0.a1	1.20	RP-L17e	large subunit ribosomal protein L17e	Ribosome
PW-WITH-RNA-SEQ.00g201950.m01-v1.0.a1	1.20	RP-S16e	small subunit ribosomal protein S16e	Ribosome
PW-WITH-RNA-SEQ.00g048380.m01-v1.0.a1	1.20	NA	NA	NA
PW-WITH-RNA-SEQ.00g027930.m01-v1.0.a1	1.19	RP-L35e	large subunit ribosomal protein L35e	Ribosome
PW-WITH-RNA-SEQ.00g126850.m01-v1.0.a1	1.18	RP-L36e	large subunit ribosomal protein L36e	Ribosome
PW-WITH-RNA-SEQ.00g127400.m01-v1.0.a1	1.18	TXNL4A	U5 snRNP protein	Spliceosome
PW-WITH-RNA-SEQ.00g165270.m01-v1.0.a1	1.14	PSMD14	26S proteasome regulatory subunit N11	Proteasome
PW-WITH-RNA-SEQ.00g083770.m01-v1.0.a1	1.12	E5.2.1.8	peptidylprolyl isomerase [EC:5.2.1.8]	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g200140.m01-v1.0.a1	1.12	NA	NA	NA

PW-WITH-RNA-SEQ.00g069560.m01-v1.0.a1	1.08	SF3A2	splicing factor 3A subunit 2	Spliceosome
PW-WITH-RNA-SEQ.00g146150.m01-v1.0.a1	1.07	RP-S5e	small subunit ribosomal protein S5e	Ribosome
PW-WITH-RNA-SEQ.00g042130.m01-v1.0.a1	1.04	ISCU	iron-sulfur cluster assembly enzyme ISCU	Mitochondrial biogenesis
PW-WITH-RNA-SEQ.00g068500.m01-v1.0.a1	1.03	PABPC	polyadenylate-binding protein	RNA transport/ Messenger RNA biogenesis
PW-WITH-RNA-SEQ.00g162390.m01-v1.0.a1	1.02	NA	NA	NA
PW-WITH-RNA-SEQ.00g157420.m01-v1.0.a1	1.01	RP-L10Ae	large subunit ribosomal protein L10Ae	Ribosome
PW-WITH-RNA-SEQ.00g200660.m01-v1.0.a1	-1.22	NA	NA	NA
PW-WITH-RNA-SEQ.00g146770.m01-v1.0.a1	-1.29	NA	NA	NA
PW-WITH-RNA-SEQ.00g093850.m01-v1.0.a1	-1.38	NA	NA	NA
PW-WITH-RNA-SEQ.00g022340.m01-v1.0.a1	-1.56	SELENBP1	methanethiol oxidase [EC:1.8.3.4]	Sulfur metabolism
PW-WITH-RNA-SEQ.00g107780.m01-v1.0.a1	-1.59	NA	NA	NA
PW-WITH-RNA-SEQ.00g106900.m01-v1.0.a1	-1.62	SACS		Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g099180.m01-v1.0.a1	-1.74	NA	NA	NA
PW-WITH-RNA-SEQ.00g075400.m01-v1.0.a1	-1.84	NA	NA	NA
PW-WITH-RNA-SEQ.00g142530.m01-v1.0.a1	-2.08	RRP8	ribosomal RNA-processing protein 8 [EC:2.1.1.287]	Ribosome biogenesis
PW-WITH-RNA-SEQ.00g116500.m01-v1.0.a1	-2.15	NA	NA	NA
PW-WITH-RNA-SEQ.00g163440.m01-v1.0.a1	-2.27	NA	NA	NA
PW-WITH-RNA-SEQ.00g183380.m01-v1.0.a1	-2.27	NA	NA	NA
PW-WITH-RNA-SEQ.00g161710.m01-v1.0.a1	-2.47	NA	NA	NA
PW-WITH-RNA-SEQ.00g102080.m01-v1.0.a1	-2.50	E2.4.1.34	3-beta-glucan synthase [EC:2.4.1.34]	Starch and sucrose metabolism
PW-WITH-RNA-SEQ.00g204890.m01-v1.0.a1	-2.61	LUT5	beta-ring hydroxylase	Carotenoid biosynthesis
PW-WITH-RNA-SEQ.00g088930.m01-v1.0.a1	-2.95	NA	NA	NA
PW-WITH-RNA-SEQ.00g023830.m01-v1.0.a1	-3.00	NA	NA	NA

PW-WITH-RNA-SEQ.00g023780.m01-v1.0.a1	-3.11	NA	NA	NA
PW-WITH-RNA-SEQ.00g023820.m01-v1.0.a1	-3.79	NA	NA	NA
PW-WITH-RNA-SEQ.00g017630.m01-v1.0.a1	-4.11	ABCG2	ATP-binding cassette	ABC transporters
PW-WITH-RNA-SEQ.00g043600.m01-v1.0.a1	-6.00	NA	NA	NA
PW-WITH-RNA-SEQ.00g022150.m01-v1.0.a1	-6.29	NA	NA	NA
PW-WITH-RNA-SEQ.00g201250.m01-v1.0.a1	-6.80	NA	NA	NA
PW-WITH-RNA-SEQ.00g201240.m01-v1.0.a1	-6.87	NA	NA	NA
PW-WITH-RNA-SEQ.00g058850.m01-v1.0.a1	-7.46	NA	NA	NA
PW-WITH-RNA-SEQ.00g201260.m01-v1.0.a1	-7.77	NA	NA	NA

Supplementary Table C5. Significant DEGs for filtered water, exponential vs. stationary phase (E-F vs. S-F). The table shows the PW95 DEG with the corresponding functional annotation and log2 fold change. Up-regulated DEGs have positive fold change and down regulated have negative fold change. NA stands for non-assigned DEGs.

PW95 Gene	Fold Change	Assigned Gene/ Protein	Name	Function
PW-WITH-RNA-SEQ.00g075440.m01-v1.0.a1	6.75	NA	NA	NA
PW-WITH-RNA-SEQ.00g099410.m01-v1.0.a1	6.39	CCNB	G2/mitotic-specific cyclin-B	DNA replication proteins/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g057740.m01-v1.0.a1	6.38	NA	NA	NA
PW-WITH-RNA-SEQ.00g203800.m01-v1.0.a1	6.22	NA	NA	NA
PW-WITH-RNA-SEQ.00g136430.m01-v1.0.a1	5.84	NA	NA	NA
PW-WITH-RNA-SEQ.00g159680.m01-v1.0.a1	5.62	TOP2	DNA topoisomerase II [EC:5.6.2.2]	Platinum drug resistance/ DNA replication proteins/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g036620.m01-v1.0.a1	5.29	KIF11	kinesin family member 11	Chromosome and associated proteins/ Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g080810.m01-v1.0.a1	4.71	NA	NA	NA
PW-WITH-RNA-SEQ.00g036630.m01-v1.0.a1	4.66	KIF11	kinesin family member 11	Chromosome and associated proteins/ Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g133180.m01-v1.0.a1	3.98	NA	NA	NA
PW-WITH-RNA-SEQ.00g059630.m01-v1.0.a1	3.65	NA	NA	NA
PW-WITH-RNA-SEQ.00g086140.m01-v1.0.a1	3.62	NA	NA	NA
PW-WITH-RNA-SEQ.00g159930.m01-v1.0.a1	3.22	NA	NA	NA

PW-WITH-RNA-SEQ.00g165350.m01-v1.0.a1	3.15	rpaA	two-component system	Two-component system/ Two-component system
PW-WITH-RNA-SEQ.00g196790.m01-v1.0.a1	2.99	NA	NA	NA
PW-WITH-RNA-SEQ.00g052960.m01-v1.0.a1	2.91	NA	NA	NA
PW-WITH-RNA-SEQ.00g062260.m01-v1.0.a1	2.84	RRM1	ribonucleoside-diphosphate reductase subunit M1 [EC:1.17.4.1]	Purine metabolism/ Pyrimidine metabolism/ DNA repair and recombination proteins
PW-WITH-RNA-SEQ.00g132950.m01-v1.0.a1	2.84	SMC4	structural maintenance of chromosome 4	Cell cycle - yeast/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g199370.m01-v1.0.a1	2.78	NA	NA	NA
PW-WITH-RNA-SEQ.00g075400.m01-v1.0.a1	2.72	NA	NA	NA
PW-WITH-RNA-SEQ.00g065950.m01-v1.0.a1	2.71	NA	NA	NA
PW-WITH-RNA-SEQ.00g124090.m01-v1.0.a1	2.68	NA	NA	NA
PW-WITH-RNA-SEQ.00g053410.m01-v1.0.a1	2.62	TUBA	tubulin alpha	Messenger RNA biogenesis/ Chromosome and associated proteins/ Cytoskeleton proteins/ Exosome
PW-WITH-RNA-SEQ.00g056000.m01-v1.0.a1	2.58	NA	NA	NA
PW-WITH-RNA-SEQ.00g203360.m01-v1.0.a1	2.57	NA	NA	NA
PW-WITH-RNA-SEQ.00g131350.m01-v1.0.a1	2.55	TUBB	tubulin beta	Chromosome and associated proteins/ Cytoskeleton proteins/ Exosome
PW-WITH-RNA-SEQ.00g180490.m01-v1.0.a1	2.48	uaZ	urate oxidase [EC:1.7.3.3]	Purine metabolism
PW-WITH-RNA-SEQ.00g028530.m01-v1.0.a1	2.43	NA	NA	NA
PW-WITH-RNA-SEQ.00g200120.m01-v1.0.a1	2.39	GME	GDP-D-mannose 3'	Ascorbate and aldarate metabolism/ Amino sugar and nucleotide sugar metabolism
PW-WITH-RNA-SEQ.00g141760.m01-v1.0.a1	2.37	NA	NA	NA
PW-WITH-RNA-SEQ.00g110500.m01-v1.0.a1	2.34	MBD4	methyl-CpG-binding domain protein 4 [EC:3.2.2.-]	Base excision repair/ Chromosome and associated proteins/ DNA repair and recombination proteins
PW-WITH-RNA-SEQ.00g109770.m01-v1.0.a1	2.31	NA	NA	NA
PW-WITH-RNA-SEQ.00g183380.m01-v1.0.a1	2.31	NA	NA	NA
PW-WITH-RNA-SEQ.00g073010.m01-v1.0.a1	2.22	MDN1	midasin	Ribosome biogenesis in eukaryotes/ Ribosome biogenesis
PW-WITH-RNA-SEQ.00g036670.m01-v1.0.a1	2.21	NA	NA	NA

PW-WITH-RNA-SEQ.00g142530.m01-v1.0.a1	2.14	RRP8	ribosomal RNA-processing protein 8 [EC:2.1.1.287]	Ribosome biogenesis
PW-WITH-RNA-SEQ.00g159740.m01-v1.0.a1	2.13	AURKX	aurora kinase	Protein kinases/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g178050.m01-v1.0.a1	2.06	HSFF	heat shock transcription factor	Transcription factors
PW-WITH-RNA-SEQ.00g053420.m01-v1.0.a1	2.06	NA	NA	NA
PW-WITH-RNA-SEQ.00g099200.m01-v1.0.a1	2.03	POR	NADPH-ferrihemoprotein reductase [EC:1.6.2.4]	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g053030.m01-v1.0.a1	1.96	TC.SULP	sulfate permease	Transporters
PW-WITH-RNA-SEQ.00g039840.m01-v1.0.a1	1.90	QSOX	thiol oxidase [EC:1.8.3.2]	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g073870.m01-v1.0.a1	1.87	NA	NA	NA
PW-WITH-RNA-SEQ.00g108900.m01-v1.0.a1	1.86	DNAH	dynein heavy chain	Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g099180.m01-v1.0.a1	1.85	NA	NA	NA
PW-WITH-RNA-SEQ.00g187080.m01-v1.0.a1	1.84	CETN1	centrin-1	Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g093730.m01-v1.0.a1	1.79	NA	NA	NA
PW-WITH-RNA-SEQ.00g107810.m01-v1.0.a1	1.79	lepA	GTP-binding protein LepA	
PW-WITH-RNA-SEQ.00g129190.m01-v1.0.a1	1.79	P4HA	prolyl 4-hydroxylase [EC:1.14.11.2]	Arginine and proline metabolism
PW-WITH-RNA-SEQ.00g027140.m01-v1.0.a1	1.73	NA	NA	NA
PW-WITH-RNA-SEQ.00g055940.m01-v1.0.a1	1.73	NA	NA	NA
PW-WITH-RNA-SEQ.00g150620.m01-v1.0.a1	1.71	SMARCAD1	matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1 [EC:3.6.4.12]	Signaling pathways regulating pluripotency of stem cells/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g138010.m01-v1.0.a1	1.69	NA	NA	NA
PW-WITH-RNA-SEQ.00g039970.m01-v1.0.a1	1.69	NA	NA	NA
PW-WITH-RNA-SEQ.00g128010.m01-v1.0.a1	1.68	NA	NA	NA
PW-WITH-RNA-SEQ.00g040920.m01-v1.0.a1	1.67	NA	NA	NA
PW-WITH-RNA-SEQ.00g109260.m01-v1.0.a1	1.66	NA	NA	NA

PW-WITH-RNA-SEQ.00g065520.m01-v1.0.a1	1.64	HNRNPF_H	heterogeneous nuclear ribonucleoprotein F/H	Spliceosome
PW-WITH-RNA-SEQ.00g051060.m01-v1.0.a1	1.63	topA	DNA topoisomerase I [EC:5.6.2.1]	DNA replication proteins/ DNA repair and recombination proteins
PW-WITH-RNA-SEQ.00g114450.m01-v1.0.a1	1.59	CYP51	sterol 14alpha-demethylase [EC:1.14.14.154 1.14.15.36]	Steroid biosynthesis/ Cytochrome P450
PW-WITH-RNA-SEQ.00g066160.m01-v1.0.a1	1.59	NA	NA	NA
PW-WITH-RNA-SEQ.00g109780.m01-v1.0.a1	1.56	NA	NA	NA
PW-WITH-RNA-SEQ.00g205260.m01-v1.0.a1	1.55	NA	NA	NA
PW-WITH-RNA-SEQ.00g142850.m01-v1.0.a1	1.55	SEC61A	protein transport protein SEC61 subunit alpha	Protein export/ Protein processing in endoplasmic reticulum/ Secretion system
PW-WITH-RNA-SEQ.00g200110.m01-v1.0.a1	1.53	P4HA	prolyl 4-hydroxylase [EC:1.14.11.2]	Arginine and proline metabolism
PW-WITH-RNA-SEQ.00g095630.m01-v1.0.a1	1.53	NA	NA	NA
PW-WITH-RNA-SEQ.00g148380.m01-v1.0.a1	1.52	udk	uridine kinase [EC:2.7.1.48]	Pyrimidine metabolism/ Drug metabolism - other enzymes
PW-WITH-RNA-SEQ.00g079530.m01-v1.0.a1	1.52	GTF3C5	general transcription factor 3C polypeptide 5 (transcription factor C subunit 1)	Transcription machinery
PW-WITH-RNA-SEQ.00g062150.m01-v1.0.a1	1.50	NA	NA	NA
PW-WITH-RNA-SEQ.00g055660.m01-v1.0.a1	1.50	VCP	transitional endoplasmic reticulum ATPase	Protein processing in endoplasmic reticulum/ Messenger RNA biogenesis/ Membrane trafficking/ Exosome
PW-WITH-RNA-SEQ.00g023110.m01-v1.0.a1	1.49	mmsB	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31]	Valine, leucine and isoleucine degradation
PW-WITH-RNA-SEQ.00g053940.m01-v1.0.a1	1.49	NA	NA	NA
PW-WITH-RNA-SEQ.00g088960.m01-v1.0.a1	1.48	AURKX	aurora kinase	Protein kinases/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g039580.m01-v1.0.a1	1.47	NA	NA	NA

PW-WITH-RNA-SEQ.00g086310.m01-v1.0.a1	1.47	NA	NA	NA
PW-WITH-RNA-SEQ.00g107560.m01-v1.0.a1	1.46	NA	NA	NA
PW-WITH-RNA-SEQ.00g017600.m01-v1.0.a1	1.44	H2A	histone H2A	Chromosome and associated proteins/ Exosome
PW-WITH-RNA-SEQ.00g125870.m01-v1.0.a1	1.44	NA	NA	NA
PW-WITH-RNA-SEQ.00g145620.m01-v1.0.a1	1.41	SQLE	squalene monooxygenase [EC:1.14.14.17]	Steroid biosynthesis/ Sesquiterpenoid and triterpenoid biosynthesis
PW-WITH-RNA-SEQ.00g113020.m01-v1.0.a1	1.40	NA	NA	NA
PW-WITH-RNA-SEQ.00g119050.m01-v1.0.a1	1.39	NA	NA	NA
PW-WITH-RNA-SEQ.00g109290.m01-v1.0.a1	1.38	DGAT2	diacylglycerol O-acyltransferase 2	Glycerolipid metabolism
PW-WITH-RNA-SEQ.00g196690.m01-v1.0.a1	1.37	NA	NA	NA
PW-WITH-RNA-SEQ.00g066740.m01-v1.0.a1	1.33	pcm	protein-L-isoaspartate(D-aspartate) O-methyltransferase [EC:2.1.1.77]	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g126630.m01-v1.0.a1	1.33	mnmE	tRNA modification GTPase [EC:3.6.-.-]	Transfer RNA biogenesis
PW-WITH-RNA-SEQ.00g145200.m01-v1.0.a1	1.33	NA	NA	NA
PW-WITH-RNA-SEQ.00g065420.m01-v1.0.a1	1.33	E3.1.3.25	myo-inositol-1(or 4)-monophosphatase [EC:3.1.3.25]	Inositol phosphate metabolism/ Streptomycin biosynthesis/ Phosphatidylinositol signaling system
PW-WITH-RNA-SEQ.00g039960.m01-v1.0.a1	1.30	P4HA	prolyl 4-hydroxylase [EC:1.14.11.2]	Arginine and proline metabolism
PW-WITH-RNA-SEQ.00g068010.m01-v1.0.a1	1.30	PRPS	ribose-phosphate pyrophosphokinase [EC:2.7.6.1]	Pentose phosphate pathway/ Purine metabolism
PW-WITH-RNA-SEQ.00g140770.m01-v1.0.a1	1.30	NA	NA	NA
PW-WITH-RNA-SEQ.00g096530.m01-v1.0.a1	1.28	RRM2	ribonucleoside-diphosphate reductase subunit M2 [EC:1.17.4.1]	Purine metabolism/ Pyrimidine metabolism/ p53 signaling pathway/ DNA repair and recombination proteins
PW-WITH-RNA-SEQ.00g123670.m01-v1.0.a1	1.27	NA	NA	NA
PW-WITH-RNA-SEQ.00g045830.m01-v1.0.a1	1.25	NNT	H+-translocating NAD(P) transhydrogenase [EC:1.6.1.2 7.1.1.1]	Nicotinate and nicotinamide metabolism

PW-WITH-RNA-SEQ.00g136500.m01-v1.0.a1	1.24	NA	NA	NA
PW-WITH-RNA-SEQ.00g032770.m02-v1.0.a1	1.24	K07034	uncharacterized protein	Function unknown
PW-WITH-RNA-SEQ.00g142650.m01-v1.0.a1	1.23	RPS6KB	ribosomal protein S6 kinase beta [EC:2.7.11.1]	Axon regeneration/ Protein kinases
PW-WITH-RNA-SEQ.00g085880.m01-v1.0.a1	1.22	NA	NA	NA
PW-WITH-RNA-SEQ.00g062470.m01-v1.0.a1	1.22	ABCC1	ATP-binding cassette	ABC transporters/ Sphingolipid signaling pathway/ Vitamin digestion and absorption/
PW-WITH-RNA-SEQ.00g095250.m01-v1.0.a1	1.16	MARS	methionyl-tRNA synthetase [EC:6.1.1.10]	Aminoacyl-tRNA biosynthesis/ Amino acid related enzymes/ Transfer RNA biogenesis
PW-WITH-RNA-SEQ.00g116110.m01-v1.0.a1	1.16	NA	NA	NA
PW-WITH-RNA-SEQ.00g055010.m01-v1.0.a1	1.15	NA	NA	NA
PW-WITH-RNA-SEQ.00g133950.m01-v1.0.a1	1.14	NA	NA	NA
PW-WITH-RNA-SEQ.00g029790.m01-v1.0.a1	1.12	NA	NA	NA
PW-WITH-RNA-SEQ.00g127450.m01-v1.0.a1	1.11	NA	NA	NA
PW-WITH-RNA-SEQ.00g038250.m01-v1.0.a1	1.10	NA	NA	NA
PW-WITH-RNA-SEQ.00g035560.m01-v1.0.a1	1.10	EP300	CREBBP	Cell cycle
PW-WITH-RNA-SEQ.00g103650.m01-v1.0.a1	1.09	R1	alpha-glucan	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g127390.m01-v1.0.a1	1.09	PPP2R2	serine/threonine-protein phosphatase 2A regulatory subunit B	mRNA surveillance pathway/ Sphingolipid signaling pathway/ Cell cycle - yeast/ Protein phosphatases and associated proteins/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g132940.m01-v1.0.a1	1.07	ABCB9	TAPL	ABC transporters/ Lysosome/ Transporters
PW-WITH-RNA-SEQ.00g018550.m01-v1.0.a1	1.04	XEG113	arabinosyltransferase [EC:2.4.2.-]	Other types of O-glycan biosynthesis/ Glycosyltransferases
PW-WITH-RNA-SEQ.00g043560.m01-v1.0.a1	1.04	NA	NA	NA
PW-WITH-RNA-SEQ.00g131430.m01-v1.0.a1	1.03	PDIA1	protein disulfide-isomerase A1 [EC:5.3.4.1]	Protein processing in endoplasmic reticulum/ Chaperones and folding catalysts/ Membrane trafficking/ Exosome
PW-WITH-RNA-SEQ.00g099350.m01-v1.0.a1	1.03	E4.1.1.17	speC	Arginine and proline metabolism/ Glutathione metabolism

PW-WITH-RNA-SEQ.00g160080.m01-v1.0.a1	-1.01	RP-L13Ae	large subunit ribosomal protein L13Ae	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g135790.m01-v1.0.a1	-1.01	NA	NA	NA
PW-WITH-RNA-SEQ.00g183790.m01-v1.0.a1	-1.01	NA	NA	NA
PW-WITH-RNA-SEQ.00g153430.m01-v1.0.a1	-1.07	ATPeF0C	F-type H+-transporting ATPase subunit c	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g191250.m01-v1.0.a1	-1.07	NA	NA	NA
PW-WITH-RNA-SEQ.00g030350.m01-v1.0.a1	-1.10	RP-L39e	large subunit ribosomal protein L39e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g087290.m01-v1.0.a1	-1.11	NA	NA	NA
PW-WITH-RNA-SEQ.00g112940.m01-v1.0.a1	-1.13	RP-L27e	large subunit ribosomal protein L27e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g074790.m01-v1.0.a1	-1.13	NA	NA	NA
PW-WITH-RNA-SEQ.00g063840.m01-v1.0.a1	-1.13	NA	NA	NA
PW-WITH-RNA-SEQ.00g104660.m01-v1.0.a1	-1.14	RP-S24e	small subunit ribosomal protein S24e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g135200.m01-v1.0.a1	-1.15	RP-L23Ae	large subunit ribosomal protein L23Ae	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g024140.m01-v1.0.a1	-1.17	NA	NA	NA
PW-WITH-RNA-SEQ.00g084680.m01-v1.0.a1	-1.19	ATPeF1G	F-type H+-transporting ATPase subunit gamma	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g100900.m01-v1.0.a1	-1.19	wrbA	NAD(P)H dehydrogenase (quinone) [EC:1.6.5.2]	Ubiquinone and other terpenoid-quinone biosynthesis
PW-WITH-RNA-SEQ.00g109170.m01-v1.0.a1	-1.20	groEL	chaperonin GroEL	RNA degradation/ Messenger RNA biogenesis/ Chaperones and folding catalysts/ Mitochondrial biogenesis/ Exosome
PW-WITH-RNA-SEQ.00g041760.m01-v1.0.a1	-1.20	NRT	MFS transporter	Nitrogen metabolism/ Transporters
PW-WITH-RNA-SEQ.00g074360.m01-v1.0.a1	-1.20	PSAP	saposin	Lysosome/ Exosome

PW-WITH-RNA-SEQ.00g146150.m01-v1.0.a1	-1.20	RP-S5e	small subunit ribosomal protein S5e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g126600.m01-v1.0.a1	-1.20	RP-L28e	large subunit ribosomal protein L28e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g069560.m01-v1.0.a1	-1.22	SF3A2	splicing factor 3A subunit 2	Spliceosome/ Spliceosome
PW-WITH-RNA-SEQ.00g030690.m01-v1.0.a1	-1.22	RP-S10e	small subunit ribosomal protein S10e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g103290.m01-v1.0.a1	-1.23	NA	NA	NA
PW-WITH-RNA-SEQ.00g122090.m01-v1.0.a1	-1.25	H4	histone H4	Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g063360.m01-v1.0.a1	-1.25	TOM40	mitochondrial import receptor subunit TOM40	Mitochondrial biogenesis/ Transporters
PW-WITH-RNA-SEQ.00g134230.m01-v1.0.a1	-1.25	NA	NA	NA
PW-WITH-RNA-SEQ.00g103380.m01-v1.0.a1	-1.26	RP-LP1	large subunit ribosomal protein LP1	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g042120.m01-v1.0.a1	-1.26	RP-L17e	large subunit ribosomal protein L17e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g147940.m01-v1.0.a1	-1.26	NA	NA	NA
PW-WITH-RNA-SEQ.00g201960.m01-v1.0.a1	-1.28	NA	NA	NA
PW-WITH-RNA-SEQ.00g074070.m01-v1.0.a1	-1.29	RP-S13e	small subunit ribosomal protein S13e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g096000.m01-v1.0.a1	-1.31	RP-L30e	large subunit ribosomal protein L30e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g145470.m01-v1.0.a1	-1.31	RP-L12e	large subunit ribosomal protein L12e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g053800.m01-v1.0.a1	-1.33	leuD	3-isopropylmalate/(R)-2-methylmalate dehydratase small subunit [EC:4.2.1.33 4.2.1.35]	C5-Branched dibasic acid metabolism/ Valine, leucine and isoleucine biosynthesis

PW-WITH-RNA-SEQ.00g041020.m01-v1.0.a1	-1.34	VTC4	inositol-phosphate phosphatase / L-galactose 1-phosphate phosphatase [EC:3.1.3.25 3.1.3.93]	Ascorbate and aldarate metabolism/ Inositol phosphate metabolism/ Phosphatidylinositol signaling system
PW-WITH-RNA-SEQ.00g091680.m01-v1.0.a1	-1.35	XBAT32_33	E3 ubiquitin-protein ligase XBAT32/33 [EC:2.3.2.27]	Ubiquitin system
PW-WITH-RNA-SEQ.00g087170.m01-v1.0.a1	-1.36	RP-L35	large subunit ribosomal protein L35	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g201950.m01-v1.0.a1	-1.38	RP-S16e	small subunit ribosomal protein S16e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g204540.m01-v1.0.a1	-1.38	EXD1	exonuclease 3'-5' domain-containing protein 1	Messenger RNA biogenesis
PW-WITH-RNA-SEQ.00g141570.m01-v1.0.a1	-1.38	RP-L31e	large subunit ribosomal protein L31e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g202750.m01-v1.0.a1	-1.39	PSAP	saposin	Lysosome/ Exosome
PW-WITH-RNA-SEQ.00g087310.m01-v1.0.a1	-1.39	SLC25A32	solute carrier family 25 (mitochondrial folate transporter)	Transporters
PW-WITH-RNA-SEQ.00g126590.m01-v1.0.a1	-1.42	NA	NA	NA
PW-WITH-RNA-SEQ.00g053970.m01-v1.0.a1	-1.42	MDH1	malate dehydrogenase [EC:1.1.1.37]	Citrate cycle (TCA cycle)/ Pyruvate metabolism/ Glyoxylate and dicarboxylate metabolism/ Carbon fixation in photosynthetic organisms/ Cysteine and methionine metabolism
PW-WITH-RNA-SEQ.00g203720.m01-v1.0.a1	-1.42	glgC	glucose-1-phosphate adenylyltransferase [EC:2.7.7.27]	Starch and sucrose metabolism/ Amino sugar and nucleotide sugar metabolism/ Biofilm formation - Escherichia coli

PW-WITH-RNA-SEQ.00g131370.m01-v1.0.a1	-1.42	PFN	profilin	Regulation of actin cytoskeleton/Membrane trafficking/ Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g187800.m01-v1.0.a1	-1.42	FKBP4_5	FK506-binding protein 4/5 [EC:5.2.1.8]	Chaperones and folding catalysts
PW-WITH-RNA-SEQ.00g138980.m01-v1.0.a1	-1.45	RP-L35Ae	large subunit ribosomal protein L35Ae	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g104470.m01-v1.0.a1	-1.46	trxA	thioredoxin 1	Chaperones and folding catalysts
PW-WITH-RNA-SEQ.00g140680.m01-v1.0.a1	-1.46	NA	NA	NA
PW-WITH-RNA-SEQ.00g027930.m01-v1.0.a1	-1.47	RP-L35e	large subunit ribosomal protein L35e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g085190.m01-v1.0.a1	-1.48	NA	NA	NA
PW-WITH-RNA-SEQ.00g116340.m01-v1.0.a1	-1.48	NA	NA	NA
PW-WITH-RNA-SEQ.00g055840.m01-v1.0.a1	-1.49	NA	NA	NA
PW-WITH-RNA-SEQ.00g152920.m01-v1.0.a1	-1.49	NA	NA	NA
PW-WITH-RNA-SEQ.00g072690.m01-v1.0.a1	-1.50	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms/ HIF-1 signaling pathway/ Exosome
PW-WITH-RNA-SEQ.00g064720.m01-v1.0.a1	-1.50	nadB	L-aspartate oxidase [EC:1.4.3.16]	Alanine, aspartate and glutamate metabolism/ Nicotinate and nicotinamide metabolism
PW-WITH-RNA-SEQ.00g087280.m01-v1.0.a1	-1.51	NA	NA	NA
PW-WITH-RNA-SEQ.00g038750.m01-v1.0.a1	-1.52	RP-L40e	large subunit ribosomal protein L40e	Ribosome/ Ribosome
PW-WITH-RNA-SEQ.00g074780.m01-v1.0.a1	-1.58	NRT	MFS transporter	Nitrogen metabolism/ Transporters
PW-WITH-RNA-SEQ.00g064520.m01-v1.0.a1	-1.65	SKP1	S-phase kinase-associated protein 1	Protein processing in endoplasmic reticulum/ Ubiquitin mediated proteolysis
PW-WITH-RNA-SEQ.00g176990.m01-v1.0.a1	-1.65	SFRS2	arginine/serine-rich 2	Spliceosome/ Herpes simplex virus 1 infection/ Spliceosome
PW-WITH-RNA-SEQ.00g178780.m01-v1.0.a1	-1.66	moaA	GTP 3' 8-cyclase [EC:4.1.99.22]	Folate biosynthesis/ Sulfur relay system
PW-WITH-RNA-SEQ.00g140020.m01-v1.0.a1	-1.66	psbW	photosystem II PsbW protein	Photosynthesis/ Photosynthesis proteins

PW-WITH-RNA-SEQ.00g095020.m01-v1.0.a1	-1.67	RGP	reversibly glycosylated polypeptide / UDP-arabinopyranose mutase [EC:2.4.1.- 5.4.99.30]	Amino sugar and nucleotide sugar metabolism/ Glycosyltransferases
PW-WITH-RNA-SEQ.00g079560.m01-v1.0.a1	-1.70	NA	NA	NA
PW-WITH-RNA-SEQ.00g200080.m01-v1.0.a1	-1.70	NA	NA	NA
PW-WITH-RNA-SEQ.00g038470.m01-v1.0.a1	-1.72	NA	NA	NA
PW-WITH-RNA-SEQ.00g032820.m01-v1.0.a1	-1.72	CPK	calcium-dependent protein kinase [EC:2.7.11.1]	Plant-pathogen interaction
PW-WITH-RNA-SEQ.00g142130.m01-v1.0.a1	-1.76	fabD	[acyl-carrier-protein] S-malonyltransferase [EC:2.3.1.39]	Fatty acid biosynthesis/ Lipid biosynthesis proteins
PW-WITH-RNA-SEQ.00g095150.m01-v1.0.a1	-1.78	PDK2_3_4	pyruvate dehydrogenase kinase 2/3/4 [EC:2.7.11.2]	Protein kinases
PW-WITH-RNA-SEQ.00g087300.m01-v1.0.a1	-1.78	cobA	uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	Porphyrin and chlorophyll metabolism
PW-WITH-RNA-SEQ.00g023520.m01-v1.0.a1	-1.80	VDAC2	voltage-dependent anion channel protein 2	Calcium signaling pathway/ cGMP-PKG signaling pathway/ Ferroptosis/ Necroptosis/ Cellular senescence/ NOD-like receptor signaling pathway/ Cholesterol metabolism/ Mitochondrial biogenesis/ Ion channels
PW-WITH-RNA-SEQ.00g117890.m01-v1.0.a1	-1.82	NA	NA	NA
PW-WITH-RNA-SEQ.00g123520.m01-v1.0.a1	-1.85	grxC	glutaredoxin 3	Chaperones and folding catalysts
PW-WITH-RNA-SEQ.00g158520.m01-v1.0.a1	-1.90	NA	NA	NA
PW-WITH-RNA-SEQ.00g079540.m01-v1.0.a1	-1.92	NA	NA	NA
PW-WITH-RNA-SEQ.00g113750.m01-v1.0.a1	-1.96	E3.4.23.40	phytepsin [EC:3.4.23.40]	Peptidases
PW-WITH-RNA-SEQ.00g076140.m01-v1.0.a1	-1.98	NA	NA	NA
PW-WITH-RNA-SEQ.00g045010.m01-v1.0.a1	-2.06	NA	NA	NA

PW-WITH-RNA-SEQ.00g191140.m01-v1.0.a1	-2.07	argC	N-acetyl-gamma-glutamyl-phosphate reductase [EC:1.2.1.38]	Arginine biosynthesis
PW-WITH-RNA-SEQ.00g072680.m01-v1.0.a1	-2.07	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g074480.m01-v1.0.a1	-2.08	NA	NA	NA
PW-WITH-RNA-SEQ.00g199720.m01-v1.0.a1	-2.12	ppa	inorganic pyrophosphatase [EC:3.6.1.1]	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g054620.m01-v1.0.a1	-2.15	CDC25	Cdc25 family phosphatase [EC:3.1.3.48 1.20.4.1]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g067350.m01-v1.0.a1	-2.15	NA	NA	NA
PW-WITH-RNA-SEQ.00g074610.m01-v1.0.a1	-2.16	NA	NA	NA
PW-WITH-RNA-SEQ.00g201330.m01-v1.0.a1	-2.19	NA	NA	NA
PW-WITH-RNA-SEQ.00g158590.m01-v1.0.a1	-2.20	NA	NA	NA
PW-WITH-RNA-SEQ.00g122410.m01-v1.0.a1	-2.29	NA	NA	NA
PW-WITH-RNA-SEQ.00g158740.m01-v1.0.a1	-2.43	NA	NA	NA
PW-WITH-RNA-SEQ.00g066080.m01-v1.0.a1	-2.44	PRDX2_4	peroxiredoxin (alkyl hydroperoxide reductase subunit C) [EC:1.11.1.15]	Apoptosis - fly/ Exosome
PW-WITH-RNA-SEQ.00g163440.m01-v1.0.a1	-2.46	NA	NA	NA
PW-WITH-RNA-SEQ.00g043950.m01-v1.0.a1	-2.48	NEDD8	ubiquitin-like protein Nedd8	Ubiquitin system
PW-WITH-RNA-SEQ.00g071460.m01-v1.0.a1	-2.50	NA	NA	NA
PW-WITH-RNA-SEQ.00g066700.m01-v1.0.a1	-2.50	PPM1L	protein phosphatase 1L [EC:3.1.3.16]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g041770.m01-v1.0.a1	-2.50	NRT	nitrate/nitrite transport system substrate-binding protein	Nitrogen metabolism/ Transporters
PW-WITH-RNA-SEQ.00g170270.m01-v1.0.a1	-2.51	UCP2_3	solute carrier family 25 (mitochondrial uncoupling protein)	Transporters

PW-WITH-RNA-SEQ.00g125930.m01-v1.0.a1	-2.56	DUR3	NA	NA
PW-WITH-RNA-SEQ.00g135260.m01-v1.0.a1	-2.57	H4	NA	NA
PW-WITH-RNA-SEQ.00g026130.m01-v1.0.a1	-2.65	yneE	ion channel-forming bestrophin family protein	Transporters
PW-WITH-RNA-SEQ.00g082470.m01-v1.0.a1	-2.70	NA	NA	NA
PW-WITH-RNA-SEQ.00g203790.m01-v1.0.a1	-2.85	PPM1L	protein phosphatase 1L [EC:3.1.3.16]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g196840.m01-v1.0.a1	-2.95	NA	NA	NA
PW-WITH-RNA-SEQ.00g063340.m01-v1.0.a1	-2.96	SOD2	superoxide dismutase	
PW-WITH-RNA-SEQ.00g069170.m01-v1.0.a1	-3.04	NA	NA	NA
PW-WITH-RNA-SEQ.00g197370.m01-v1.0.a1	-3.05	NA	NA	NA
PW-WITH-RNA-SEQ.00g172990.m01-v1.0.a1	-3.17	NA	NA	NA
PW-WITH-RNA-SEQ.00g134620.m01-v1.0.a1	-3.25	NA	NA	NA
PW-WITH-RNA-SEQ.00g020500.m01-v1.0.a1	-4.19	NA	NA	NA
PW-WITH-RNA-SEQ.00g201250.m01-v1.0.a1	-4.24	NA	NA	NA
PW-WITH-RNA-SEQ.00g067150.m01-v1.0.a1	-5.18	H4	histone H4	Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g149280.m01-v1.0.a1	-8.14	NA	NA	NA
PW-WITH-RNA-SEQ.00g149320.m01-v1.0.a1	-19.90	nrtA	nitrate/nitrite transport system substrate-binding protein	Nitrogen metabolism/ ABC transporters/ Transporters

Supplementary Table C6: Significant DEGs for unfiltered water, exponential vs. stationary phase (E-UF vs. S-UF). The table shows the PW95 DEG with the corresponding functional annotation and log2 fold change. Up-regulated DEGs have positive fold change and down regulated have negative fold change. NA stands for non-assigned DEGs.

PW95 Gene	Fold Change	Assigned Gene/ Protein	Name	Function
PW-WITH-RNA-SEQ.00g057740.m01-v1.0.a1	8.77	NA	NA	NA
PW-WITH-RNA-SEQ.00g075440.m01-v1.0.a1	7.55	NA	NA	NA
PW-WITH-RNA-SEQ.00g133180.m01-v1.0.a1	7.47	NA	NA	NA
PW-WITH-RNA-SEQ.00g036620.m01-v1.0.a1	7.41	KIF11	kinesin family member 11	Chromosome and associated proteins/ Cytoskeleton proteins

PW-WITH-RNA-SEQ.00g036630.m01-v1.0.a1	7.12	KIF11	kinesin family member 11	Chromosome and associated proteins/ Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g080810.m01-v1.0.a1	6.00	NA	NA	NA
PW-WITH-RNA-SEQ.00g086140.m01-v1.0.a1	5.66	NA	NA	NA
PW-WITH-RNA-SEQ.00g082330.m01-v1.0.a1	5.58	NA	NA	NA
PW-WITH-RNA-SEQ.00g122830.m01-v1.0.a1	5.00	NA	NA	NA
PW-WITH-RNA-SEQ.00g159680.m01-v1.0.a1	4.97	TOP2	DNA topoisomerase II [EC:5.6.2.2]	DNA replication proteins/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g099410.m01-v1.0.a1	4.94	CCNB	G2/mitotic-specific cyclin-B	DNA replication proteins/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g118550.m01-v1.0.a1	4.69	NA	NA	NA
PW-WITH-RNA-SEQ.00g180490.m01-v1.0.a1	3.27	uaZ	urate oxidase [EC:1.7.3.3]	Purine metabolism
PW-WITH-RNA-SEQ.00g165350.m01-v1.0.a1	3.14	rpaA	two-component system	Two-component system
PW-WITH-RNA-SEQ.00g062260.m01-v1.0.a1	2.61	RRM1	ribonucleoside-diphosphate reductase subunit M1 [EC:1.17.4.1]	Purine Pyrimidine metabolism/ DNA repair and recombination proteins
PW-WITH-RNA-SEQ.00g038540.m01-v1.0.a1	2.57	cah	carbonic anhydrase [EC:4.2.1.1]	Nitrogen metabolism
PW-WITH-RNA-SEQ.00g010040.m01-v1.0.a1	2.55	MCM4	DNA replication licensing factor MCM4 [EC:3.6.4.12]	DNA replication/ Cell cycle
PW-WITH-RNA-SEQ.00g093730.m01-v1.0.a1	2.50	NA	NA	NA
PW-WITH-RNA-SEQ.00g146300.m01-v1.0.a1	2.45	NA	NA	NA
PW-WITH-RNA-SEQ.00g068440.m01-v1.0.a1	2.37	NA	NA	NA
PW-WITH-RNA-SEQ.00g203810.m01-v1.0.a1	2.32	MMAB	cob(I)alamin adenosyltransferase [EC:2.5.1.17]	Porphyrin and chlorophyll metabolism
PW-WITH-RNA-SEQ.00g196790.m01-v1.0.a1	2.17	NA	NA	NA
PW-WITH-RNA-SEQ.00g160880.m01-v1.0.a1	2.12	NA	NA	NA
PW-WITH-RNA-SEQ.00g131350.m01-v1.0.a1	2.07	TUBB	tubulin beta	Chromosome and associated proteins/ Cytoskeleton proteins
PW-WITH-RNA-SEQ.00g187080.m01-v1.0.a1	2.07	CETN1	centrin-1	Chromosome and associated proteins

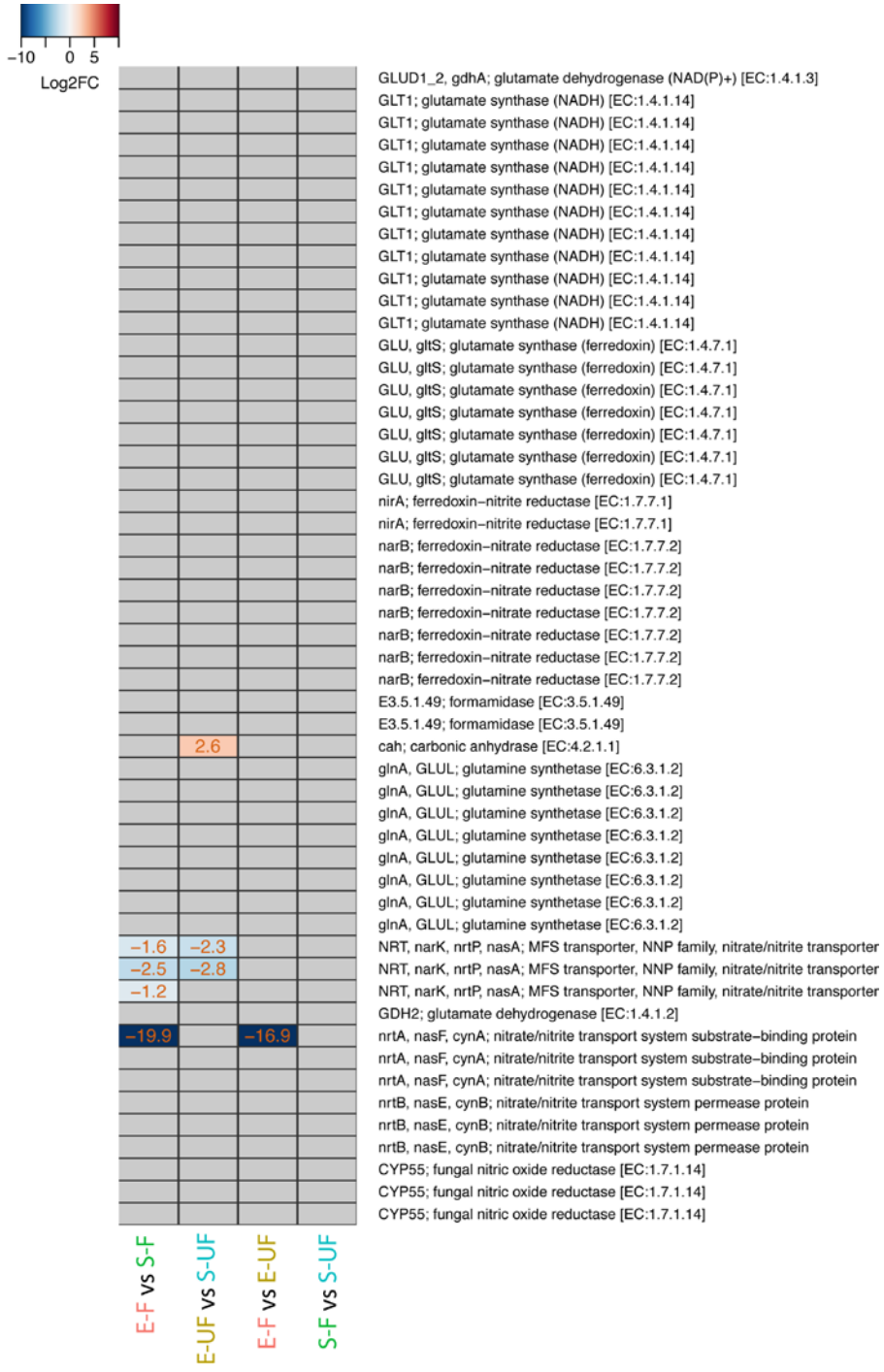
PW-WITH-RNA-SEQ.00g028530.m01-v1.0.a1	2.01	NA	NA	NA
PW-WITH-RNA-SEQ.00g072160.m01-v1.0.a1	1.97	NA	NA	NA
PW-WITH-RNA-SEQ.00g092140.m01-v1.0.a1	1.91	NA	NA	NA
PW-WITH-RNA-SEQ.00g056000.m01-v1.0.a1	1.90	NA	NA	NA
PW-WITH-RNA-SEQ.00g200120.m01-v1.0.a1	1.88	GME	GDP-D-mannose 3'	Ascorbate and aldarate metabolism/ Amino sugar and nucleotide sugar metabolism
PW-WITH-RNA-SEQ.00g088960.m01-v1.0.a1	1.86	AURKX	aurora kinase	Protein kinases
PW-WITH-RNA-SEQ.00g141760.m01-v1.0.a1	1.81	NA	NA	NA
PW-WITH-RNA-SEQ.00g103130.m01-v1.0.a1	1.80	NA	NA	NA
PW-WITH-RNA-SEQ.00g099200.m01-v1.0.a1	1.79	POR	NADPH-ferrihemoprotein reductase [EC:1.6.2.4]	Enzymes with EC numbers
PW-WITH-RNA-SEQ.00g200110.m01-v1.0.a1	1.74	P4HA	prolyl 4-hydroxylase [EC:1.14.11.2]	Arginine and proline metabolism
PW-WITH-RNA-SEQ.00g116530.m01-v1.0.a1	1.73	NA	NA	NA
PW-WITH-RNA-SEQ.00g165340.m01-v1.0.a1	1.72	HSP20	HSP20 family protein	Protein processing in endoplasmic reticulum/ Chaperones and folding catalysts
PW-WITH-RNA-SEQ.00g142850.m01-v1.0.a1	1.70	SEC61A	protein transport protein SEC61 subunit alpha	Protein export
PW-WITH-RNA-SEQ.00g129190.m01-v1.0.a1	1.69	P4HA	prolyl 4-hydroxylase [EC:1.14.11.2]	Arginine and proline metabolism
PW-WITH-RNA-SEQ.00g068050.m01-v1.0.a1	1.64	NA	NA	NA
PW-WITH-RNA-SEQ.00g050660.m01-v1.0.a1	1.63	PPCDC	phosphopantothenoylcysteine decarboxylase [EC:4.1.1.36]	Pantothenate and CoA biosynthesis
PW-WITH-RNA-SEQ.00g205260.m01-v1.0.a1	1.62	NA	NA	NA
PW-WITH-RNA-SEQ.00g119040.m01-v1.0.a1	1.58	NA	NA	NA
PW-WITH-RNA-SEQ.00g138010.m01-v1.0.a1	1.57	NA	NA	NA
PW-WITH-RNA-SEQ.00g065520.m01-v1.0.a1	1.47	HNRNPF_H	heterogeneous nuclear ribonucleoprotein F/H	Spliceosome
PW-WITH-RNA-SEQ.00g128010.m01-v1.0.a1	1.45	NA	NA	NA
PW-WITH-RNA-SEQ.00g065420.m01-v1.0.a1	1.42	IMPA	myo-inositol-1(or 4)-monophosphatase [EC:3.1.3.25]	Inositol phosphate metabolism/ Phosphatidylinositol signaling system

PW-WITH-RNA-SEQ.00g039900.m01-v1.0.a1	1.37	CMT1	3-mannosyltransferase [EC:2.4.1.-]	Glycosyltransferases
PW-WITH-RNA-SEQ.00g145620.m01-v1.0.a1	1.36	SQLE	squalene monooxygenase [EC:1.14.14.17]	Steroid biosynthesis/Sesquiterpenoid and triterpenoid biosynthesis
PW-WITH-RNA-SEQ.00g078380.m01-v1.0.a1	1.35	CPK	calcium-dependent protein kinase [EC:2.7.11.1]	Plant-pathogen interaction
PW-WITH-RNA-SEQ.00g109500.m01-v1.0.a1	1.35	NA	NA	NA
PW-WITH-RNA-SEQ.00g109780.m01-v1.0.a1	1.35	NA	NA	NA
PW-WITH-RNA-SEQ.00g159740.m01-v1.0.a1	1.34	AURKX	aurora kinase	Protein kinases/ Chromosome and associated proteins
PW-WITH-RNA-SEQ.00g053030.m01-v1.0.a1	1.27	TC.SULP	sulfate permease	Transporters
PW-WITH-RNA-SEQ.00g181310.m01-v1.0.a1	1.20	NA	NA	NA
PW-WITH-RNA-SEQ.00g032770.m02-v1.0.a1	1.18	K07034	uncharacterized protein	Function unknown
PW-WITH-RNA-SEQ.00g021560.m01-v1.0.a1	1.13	GDE1	glycerophosphodiester phosphodiesterase [EC:3.1.4.46]	Glycerophospholipid metabolism
PW-WITH-RNA-SEQ.00g023110.m01-v1.0.a1	1.05	mmsB	3-hydroxyisobutyrate dehydrogenase [EC:1.1.1.31]	Valine, leucine and isoleucine degradation
PW-WITH-RNA-SEQ.00g142650.m01-v1.0.a1	1.02	RPS6KB	ribosomal protein S6 kinase beta [EC:2.7.11.1]	Axon regeneration
PW-WITH-RNA-SEQ.00g068950.m01-v1.0.a1	1.01	NA	NA	NA
PW-WITH-RNA-SEQ.00g110480.m01-v1.0.a1	1.01	NA	NA	NA
PW-WITH-RNA-SEQ.00g126590.m01-v1.0.a1	-1.02	NA	NA	NA
PW-WITH-RNA-SEQ.00g113480.m01-v1.0.a1	-1.07	NA	NA	NA
PW-WITH-RNA-SEQ.00g048460.m01-v1.0.a1	-1.08	NA	NA	NA
PW-WITH-RNA-SEQ.00g074790.m01-v1.0.a1	-1.13	NA	NA	NA
PW-WITH-RNA-SEQ.00g047940.m01-v1.0.a1	-1.14	NA	NA	NA
PW-WITH-RNA-SEQ.00g116640.m01-v1.0.a1	-1.16	glpX-SEBP	fructose-1	Glycolysis/ Gluconeogenesis/ Pentose phosphate pathway/ Fructose and mannose metabolism/ Carbon fixation in photosynthetic organisms

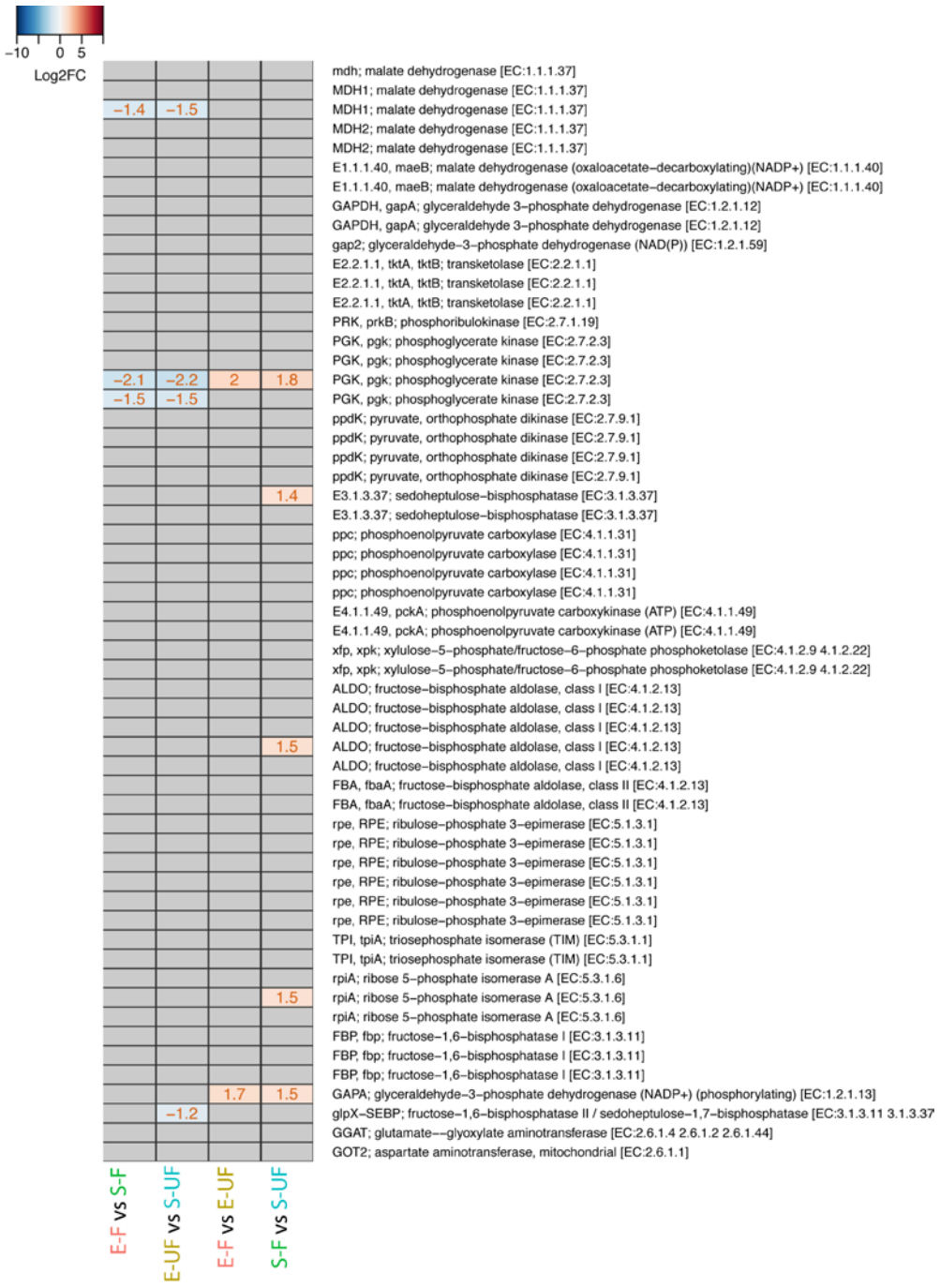
PW-WITH-RNA-SEQ.00g153730.m01-v1.0.a1	-1.21	SLC25A23S	solute carrier family 25 (mitochondrial phosphate transporter)	Transporters
PW-WITH-RNA-SEQ.00g080560.m01-v1.0.a1	-1.25	DDX56	ATP-dependent RNA helicase DDX56/DBP9 [EC:3.6.4.13]	Ribosome biogenesis
PW-WITH-RNA-SEQ.00g156030.m01-v1.0.a1	-1.28	NA	NA	NA
PW-WITH-RNA-SEQ.00g067340.m01-v1.0.a1	-1.31	ZFP36L	butyrate response factor	Messenger RNA biogenesis/ Messenger RNA biogenesis
PW-WITH-RNA-SEQ.00g103290.m01-v1.0.a1	-1.36	NA	NA	NA
PW-WITH-RNA-SEQ.00g109170.m01-v1.0.a1	-1.37	groEL	chaperonin GroEL	RNA degradation/ Mitochondrial biogenesis
PW-WITH-RNA-SEQ.00g064720.m01-v1.0.a1	-1.39	nadB	L-aspartate oxidase [EC:1.4.3.16]	Alanine, aspartate and glutamate metabolism
PW-WITH-RNA-SEQ.00g176990.m01-v1.0.a1	-1.44	SFRS2	splicing factor	Spliceosome
PW-WITH-RNA-SEQ.00g053970.m01-v1.0.a1	-1.48	MDH1	malate dehydrogenase [EC:1.1.1.37]	TCA cycle/ Pyruvate metabolism/ Glyoxylate and dicarboxylate metabolism/ Carbon fixation in photosynthetic organisms/Cysteine and methionine metabolism
PW-WITH-RNA-SEQ.00g110080.m01-v1.0.a1	-1.49	NA	NA	NA
PW-WITH-RNA-SEQ.00g116340.m01-v1.0.a1	-1.49	NA	NA	NA
PW-WITH-RNA-SEQ.00g072690.m01-v1.0.a1	-1.52	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms
PW-WITH-RNA-SEQ.00g201960.m01-v1.0.a1	-1.52	NA	NA	NA
PW-WITH-RNA-SEQ.00g130980.m01-v1.0.a1	-1.57	NA		NA
PW-WITH-RNA-SEQ.00g063340.m01-v1.0.a1	-1.64	SOD2	superoxide dismutase	
PW-WITH-RNA-SEQ.00g125930.m01-v1.0.a1	-1.65	DUR3	urea-proton symporter	Transporters
PW-WITH-RNA-SEQ.00g119870.m01-v1.0.a1	-1.66	MYBP	transcription factor MYB	Transcription factors
PW-WITH-RNA-SEQ.00g203720.m01-v1.0.a1	-1.66	glgC	glucose-1-phosphate adenylyltransferase [EC:2.7.7.27]	Starch and sucrose metabolism

PW-WITH-RNA-SEQ.00g091680.m01-v1.0.a1	-1.69	XBAT32_33	E3 ubiquitin-protein ligase XBAT32/33 [EC:2.3.2.27]	Ubiquitin system
PW-WITH-RNA-SEQ.00g204540.m01-v1.0.a1	-1.70	EXD1	exonuclease 3'-5' domain-containing protein 1	Messenger RNA biogenesis
PW-WITH-RNA-SEQ.00g147940.m01-v1.0.a1	-1.71	NA	NA	NA
PW-WITH-RNA-SEQ.00g199720.m01-v1.0.a1	-1.79	ppa	inorganic pyrophosphatase [EC:3.6.1.1]	Oxidative phosphorylation
PW-WITH-RNA-SEQ.00g032820.m01-v1.0.a1	-1.82	CPK	calcium-dependent protein kinase [EC:2.7.11.1]	Plant-pathogen interaction
PW-WITH-RNA-SEQ.00g128420.m01-v1.0.a1	-1.82	NA	NA	NA
PW-WITH-RNA-SEQ.00g041780.m01-v1.0.a1	-1.89	NA	NA	NA
PW-WITH-RNA-SEQ.00g122310.m01-v1.0.a1	-1.95	NA	NA	NA
PW-WITH-RNA-SEQ.00g178780.m01-v1.0.a1	-1.98	moaA	GTP 3'	Folate biosynthesis
PW-WITH-RNA-SEQ.00g066700.m01-v1.0.a1	-2.05	PPM1L	protein phosphatase 1L [EC:3.1.3.16]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g072710.m01-v1.0.a1	-2.08	GCH1	GTP cyclohydrolase IA [EC:3.5.4.16]	Folate biosynthesis
PW-WITH-RNA-SEQ.00g075170.m01-v1.0.a1	-2.12	NA	NA	NA
PW-WITH-RNA-SEQ.00g087300.m01-v1.0.a1	-2.16	cobA	uroporphyrin-III C-methyltransferase [EC:2.1.1.107]	Porphyrin and chlorophyll metabolism
PW-WITH-RNA-SEQ.00g072680.m01-v1.0.a1	-2.21	PGK	phosphoglycerate kinase [EC:2.7.2.3]	Glycolysis / Gluconeogenesis
PW-WITH-RNA-SEQ.00g117890.m01-v1.0.a1	-2.23	NA	NA	NA
PW-WITH-RNA-SEQ.00g074780.m01-v1.0.a1	-2.25	NRT	MFS transporter	Nitrogen metabolism
PW-WITH-RNA-SEQ.00g057450.m01-v1.0.a1	-2.27	PABPC	polyadenylate-binding protein	RNA transport
PW-WITH-RNA-SEQ.00g079560.m01-v1.0.a1	-2.33	NA	NA	NA
PW-WITH-RNA-SEQ.00g046800.m01-v1.0.a1	-2.34	NA	NA	NA
PW-WITH-RNA-SEQ.00g129720.m01-v1.0.a1	-2.36	NA	NA	NA
PW-WITH-RNA-SEQ.00g026130.m01-v1.0.a1	-2.39	yneE	ion channel-forming bestrophin family protein	Transporters
PW-WITH-RNA-SEQ.00g201330.m01-v1.0.a1	-2.43	NA	NA	NA
PW-WITH-RNA-SEQ.00g069170.m01-v1.0.a1	-2.69	NA	NA	NA
PW-WITH-RNA-SEQ.00g074610.m01-v1.0.a1	-2.70	NA	NA	NA
PW-WITH-RNA-SEQ.00g079540.m01-v1.0.a1	-2.70	NA	NA	NA

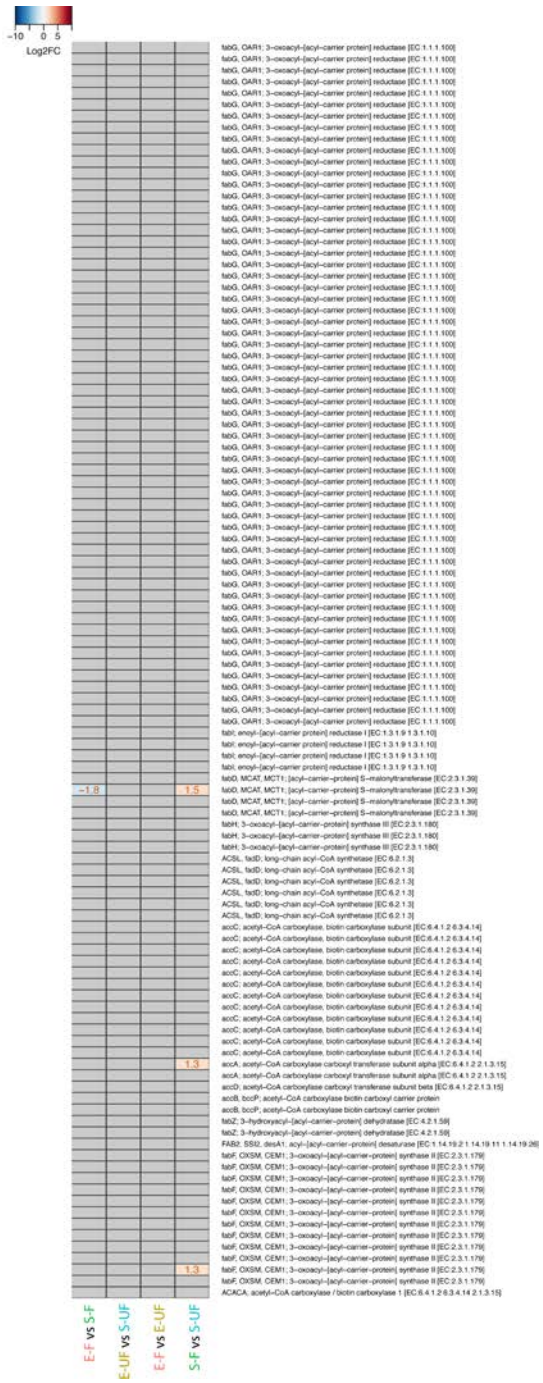
PW-WITH-RNA-SEQ.00g020500.m01-v1.0.a1	-2.71	NA	NA	BLAST HIT: Maltose export-like protein, maltose excess protein 1 chloroplastic Chlamy and Chlorella
PW-WITH-RNA-SEQ.00g203790.m01-v1.0.a1	-2.73	PPM1L	protein phosphatase 1L [EC:3.1.3.16]	Protein phosphatases and associated proteins
PW-WITH-RNA-SEQ.00g067350.m01-v1.0.a1	-2.79	NA	NA	NA
PW-WITH-RNA-SEQ.00g041770.m01-v1.0.a1	-2.81	NRT	MFS transporter NNP family nitrate/nitrite transporter	Nitrogen metabolism
PW-WITH-RNA-SEQ.00g022150.m01-v1.0.a1	-2.92	NA	NA	NA
PW-WITH-RNA-SEQ.00g172990.m01-v1.0.a1	-3.08	NA	NA	NA
PW-WITH-RNA-SEQ.00g082470.m01-v1.0.a1	-3.13	NA	NA	NA
PW-WITH-RNA-SEQ.00g074480.m01-v1.0.a1	-3.20	NA	NA	NA
PW-WITH-RNA-SEQ.00g086010.m01-v1.0.a1	-3.32	NA	NA	NA
PW-WITH-RNA-SEQ.00g196840.m01-v1.0.a1	-3.34	NA	NA	NA
PW-WITH-RNA-SEQ.00g182110.m01-v1.0.a1	-3.37	NA	NA	NA
PW-WITH-RNA-SEQ.00g043950.m01-v1.0.a1	-3.50	NEDD8	ubiquitin-like protein Nedd8	Ubiquitin system
PW-WITH-RNA-SEQ.00g095650.m01-v1.0.a1	-3.75	NA	NA	NA
PW-WITH-RNA-SEQ.00g201250.m01-v1.0.a1	-3.87	NA	NA	NA
PW-WITH-RNA-SEQ.00g043600.m01-v1.0.a1	-4.37	NA	NA	NA
PW-WITH-RNA-SEQ.00g058850.m01-v1.0.a1	-4.51	NA	NA	NA
PW-WITH-RNA-SEQ.00g201260.m01-v1.0.a1	-4.60	NA	NA	NA
PW-WITH-RNA-SEQ.00g201240.m01-v1.0.a1	-5.20	NA	NA	NA
PW-WITH-RNA-SEQ.00g043610.m01-v1.0.a1	-5.76	AUX1	auxin influx carrier (AUX1 LAX family)	Plant hormone signal transduction



Supplementary Figure C1. Preliminary assessment of metabolic pathway reconstruction for nitrogen assimilation in PW95 with the significant DEGs. Heat map shows the level of up- (red) and down-expression (blue) of DEGs in a relevant metabolic for this study. Pathway analysis was not possible given the low number of DEGs for the pathway.



Supplementary Figure C2. Preliminary assessment of metabolic pathway reconstruction for carbon fixation in PW95 with the significant DEGs. Heat map shows the level of up- (red) and down-regulation (blue) of DEGs in a relevant metabolic for this study. Pathway analysis was not possible given the low number of DEGs for the pathway.



Supplementary Figure C3. Preliminary assessment of metabolic pathway reconstruction for fatty acid biosynthesis in PW95 with the significant DEGs. Heat map shows the level of up- (red) and down-expression (blue) of DEGs in a relevant metabolic for this study. Pathway analysis was not possible given the low number of DEGs for the pathway.

APPENDIX D

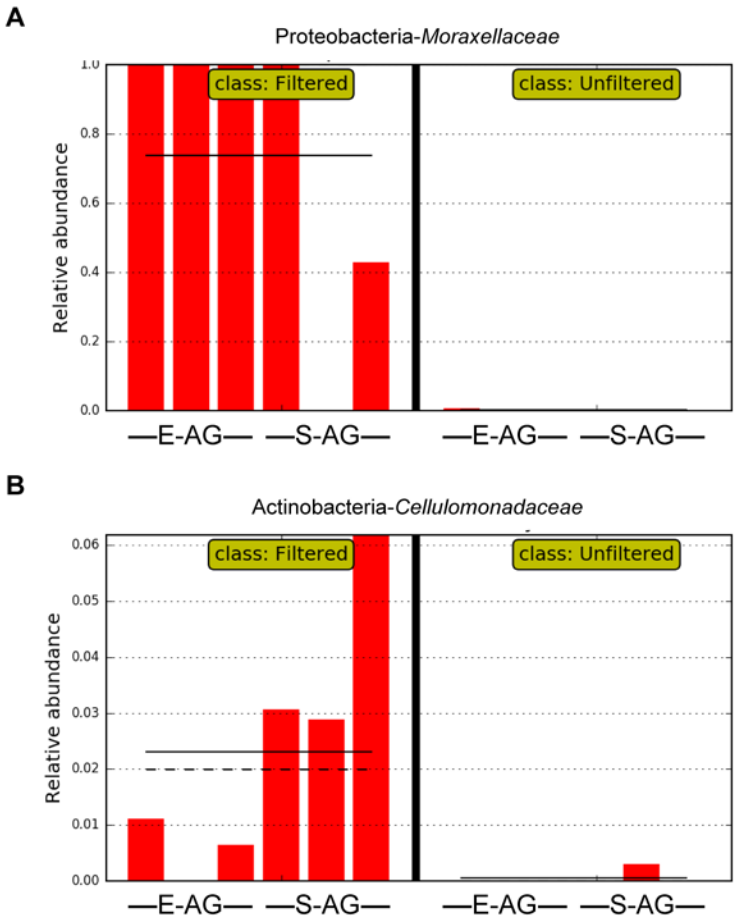
SUPPLEMENTARY MATERIALS FOR CHAPTER FIVE

Supplementary Table D1. Laboratory analytical report of CBM water collected from well FG-09. Reported values for the CBM water used in this study using standardized media components for algal growth as a reference.

Analyses	Result	Units
NON-METALS		
Inorganic Carbon, Dissolved (DIC)	126	mg/L
INORGANICS		
Chloride	64	mg/L
Sulfate	728	mg/L
AGGREGATE ORGANICS		
Organic Carbon, Dissolved (DOC)	3	mg/L
NUTRIENTS		
Nitrogen, Ammonia as N	0.67	mg/L
Nitrogen, Nitrite as N	ND	mg/L
Nitrogen, Nitrate as N	ND	mg/L
Nitrogen, Nitrate+Nitrite as N	ND	mg/L
Nitrogen, Kjeldahl, Total as N	0.8	mg/L
Nitrogen, Total	0.8	mg/L
Phosphorus, Total as P	0.029	mg/L
METALS, TOTAL		
Calcium	3	mg/L
Iron	0.04	mg/L
Magnesium	2	mg/L
Manganese	0.007	mg/L
Potassium	3	mg/L
Sodium	611	mg/L

Supplementary Table D2. Dominant species of FAMEs produced by PW95 relative to the total content of FAMEs per condition. Replicate samples are displayed as the average percentage of total FAMEs.

Dominant FAME species	E-F	Proportion in sample (%)	S-F	Proportion in sample (%)	E-UF	Proportion in sample (%)	S-UF	Proportion in sample (%)
C16:0	0.30	1.94	1.42	6.16	1.56	9.76	0.97	7.26
C16:1 (z-9)	2.93	19.10	3.28	14.18	1.98	12.36	1.62	12.10
C18:0	0.86	5.57	1.17	5.06	0.43	2.66	0.32	2.42
C18:1 (e-9)	0.18	1.18	0.01	0.03	0.01	0.05	0.01	0.08
C18:1 (z-9)	0.00	0.00	3.25	14.09	3.07	19.21	0.00	0.00
C18:2 (e-9,12)	6.17	40.15	6.24	27.01	3.22	20.14	5.52	41.13
C18:2 (z-9,12)	1.71	11.13	2.91	12.59	1.75	10.93	1.54	11.46
C18:3 (z-9,12,15)	1.57	10.23	2.19	9.49	1.40	8.73	1.53	11.42
C20:0	0.30	1.93	0.39	1.68	0.15	0.94	0.16	1.20
C20:1 (z-11)	0.20	1.29	0.33	1.42	0.21	1.32	0.19	1.40
C22:0	0.07	0.44	0.07	0.32	0.55	3.44	0.05	0.36
Average Total FAME (%)	**15.37**		**23.10**		**15.99**		**13.42**	



Supplementary Figure D1. Comparison between algal aggregate samples versus the type of water (filtered and unfiltered) using LEfSe. High dimensional biomarker discovery algorithm (LEfSe) analysis output used to identify the abundant bacterial taxa characterizing the differences between the algal aggregate samples in filtered water versus unfiltered water. The figure illustrates examples of the LEfSe output determining the differences between filtered and unfiltered water per ASV analyzed. In this case showing the results from two ASVs that were considered exclusive to the aggregates in filtered water.