UNIVERSITY OF CALIFORNIA SAN DIEGO

Expanding an algae production platform: Industrially relevant advancement of *Chlamydomonas reinhardtii*

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Joseph Thomas Ostrand

Committee in charge:

Stephen Mayfield, Chair
Eric Allen
Steven Briggs
Theresa Gaasterland
Mark Hildebrand
Jose Pruneda-Paz

2018

The dissertation of Joseph Thomas Ostrand is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____
Chair

University of California San Diego

2018

DEDICATION

This dissertation is dedicated to my family, both new and old,

who gave me the opportunity to imagine and realize this lifelong goal.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ARG7 – gene for arginosuccinate lyase

ARS2 – gene for arylsulfatase

bHLH – basic helix-loop-helix

CRISPR – clustered regularly interspaced palindromic repeats

DREME – discriminative regular expression motif elicitation

GFP – green fluorescent protein

HR – homologous recombination

HSM – high salt media

HSP70A – heat shock protein 70A

NHEJ – non-homologous end-joining

OFP – orange fluorescent protein

POWRS – position-sensitive word set

RBCS2 – rubisco small subunit

SAP – synthetic algal promoter

SLICE – seamless ligation cloning extract

TAP – tris acetate phosphate

TF – transcription factor

TSS – transcriptional start site

UTR – untranslated region

Y1H – yeast one-hybrid

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

As part of a project in the 7th grade, our teacher asked us what we wanted to achieve in our academic careers. As a 12 year old, I wrote that I wanted to earn a PhD degree from UCSD. This body of work has been made possible by the tremendous opportunities that I have been afforded by my parents. They have supported my pursuit for success in academics from a very young age, and without their hard work, this lifelong goal would never have been considered.

I also have to recognize the unwavering love and support from my wife Jilli. She was a smiling face when the science wasn't cooperating, and would always find a way to keep my life balanced and fulfilling. I couldn't ask for a better partner with which to tackle all of life's challenges.

Chapter 1, in full, is a reprint of the material as it appears in *The Plant Journal* 2015. Scranton, Melissa; Ostrand, Joseph; Fields, Francis; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this chapter.

Chapter 2, in full, is a reprint of material in preparation for submission for publication. Ostrand, Joseph; Mayfield, Stephen. The dissertation author is the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in *Algal Research 2018*. Fields, Francis; Ostrand, Joseph; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this paper.

Chapter 4, in full, is a reprint of material in preparation for submission for publication. Ostrand, Joseph; Fields, Francis; Tran, Miller; Mayfield, Stephen. The dissertation author is the primary investigator and author of this paper.

Chapter 5, in full, is a reprint of the material as it appears in *Algal Research* 2016. Scranton, Melissa; Ostrand, Joseph; Georgianna, David; Lofgren, Shane; Li, Daphne; Ellis, Rosalie;

Carruthers, David; Dräger, Andreas; Masica, David; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this paper.

Chapter 7, in full, is a reprint of material as it appears in *Algal Research* 2018. Nour-Eldin, Hussam; Specht, Elizabeth; Ostrand, Joseph; Karunanithi, Prema; Hoang, Kevin; Mayfield, Stephen. The dissertation author is the tertiary investigator and author of this paper.

VITA

2011	Bachelor of Science, Department of Chemistry and Biochemistry, University of California Santa Barbara, CA
2011	Bachelor of Arts, Department of Environmental Studies, University of California Santa Barbara, CA
2012-2018	Graduate researcher, Division of Biological Sciences, University of California San Diego, La Jolla, CA
2018	Doctor of Philosophy, Division of Biological Sciences, University of California San Diego, La Jolla, CA

PUBLICATIONS

Ostrand JT, Fields FJ, Tran M, Mayfield SP. Genome shuffling and selection for improved recombinant protein accumulation in the chloroplast of *Chlamydomonas reinhardtii*. **2018**. (In preparation.)

Ostrand JT, Mayfield SP. Antibiotic-free selection in the nuclear genome of *Chlamydomonas reinhardtii* by multicolor cell sorting. **2018**. (In preparation.)

Fields FJ, **Ostrand JT**, Mayfield SP. Fed-batch mixotrophic cultivation of *Chlamydomonas reinhardtii* for high-density cultures. Algal Research, 2018.

Nour-Eldin HH, Specht EA, **Ostrand JT**, Karunanithi PS, Hoang KTD, Mayfield SP. High throughput system for detecting improvements in homologous recombination efficiency in *Chlamydomonas reinhardtii*. Algal Research, 2018.

Scranton M, **Ostrand JT**, Georgianna DR, Lofgren S, Li D, Ellis R, Carruthers D, Dräger A, Masica D, and Mayfield, SP. Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. Algal Research, 2016.

Scranton M, **Ostrand JT**, Fields FJ, and Mayfield SP. *Chlamydomonas* as a model for biofuels and bio-products production. The Plant Journal, 2015.

FIELDS OF STUDY

Major field: Biology

Studies in Molecular Biology
Professors Stephen Mayfield, Eric Allen, and Jose Pruneda-Paz

Studies in Microbiology
Professor Mark Hildebrand

Studies in Computational Biology and Bioinformatics
Professors Steven Briggs and Theresa Gaasterland

ABSTRACT OF THE DISSERTATION

Expanding an algae production platform:
Industrially relevant advancement of *Chlamydomonas reinhardtii*

by

Joseph Thomas Ostrand

Doctor of Philosophy in Biology

University of California San Diego, 2018

Professor Stephen P. Mayfield, Chair

The initial surge of industrial interest in green algae for the production of renewable fuel has given way to a blossoming industry with the potential to contribute to the commercial production of food, materials, and modern medicine. Algae are biologically diverse, edible,

genetically tractable and have high lipid and protein content. Their unicellular nature facilitates the efficient conversion of energy into biological material without the need to produce structural materials like that of higher plants. The diversity of the fields in which algae have become a viable production platform has necessitated development of a broad range of cultivation strategies and tools for improving their industrial relevance.

Molecular research on green algae has yielded a growing encyclopedia of genetic tools for manipulation of the chloroplast and nuclear genomes. However, as the landscape of desired algal products has changed, some of these tools need to be refined for new cultivation systems and advanced product engineering. Although the future is bright for algae as a bio-manufacturing platform, many aspects of their production lag behind their established counterparts like yeast, *E. coli* and mammalian cell culture. The research outlined in this dissertation presents significant gains in the advancement of green algae as an industrial organism by refining cultivation strategies and genetic tools to foster success in commercially relevant systems.

Production of algae in both open and closed systems is addressed. For open pond cultivation, environmental concerns render certain genetic markers unusable. A system for selection of transgenic algae without the need for antibiotic resistance cassettes has been developed. In general, closed systems offer more tight regulation of abiotic growth conditions and optimization of growth and product accumulation in these systems is imperative to the future of industrial algae. In closed systems, high density cultures are required for maximum biomass yield with limited infrastructure. Research presented here shows the first look at *C. reinhardtii* grown in fermentors for high-density cultivation and provides a metabolic comparison with more standard growth conditions.

An analysis of transgene expression in high-density cultures expressed a need for new genetic tools that are functional under these conditions. The second half of the dissertation focuses

on the development of advanced genetic tools and synthetic systems for optimization of transgene expression. Synthetic tools have the power to revolutionize recombinant protein production in green algae. Ultimately, this dissertation provides an extensive body of work which identifies gaps in the commercial viability of green algae, and provides transformational solutions to many of the problems hampering industrial relevance.

INTRODUCTION

Throughout the course of human history, we have taken advantage of biological systems to satisfy our needs and make our world a more pleasant place to live. Technological advances in the biological space allowed humans to shift from a hunting and gathering lifestyle into settlements. Reliable crop cultivation and the domestication of animals gave early humans the surplus of resources that facilitated the specialization of labor and development of new materials and ideas.

As our understanding of the world around us grew, agricultural techniques became more refined, producing greater crop yields to support more dense populations. The population density we experience today can be attributed to the industrial revolution in the late 1700s. Steam powered tractors and processing equipment were made accessible due to cheap energy in the form of fossil fuels. The discovery and implementation of fossil fuels coupled agricultural yield with the availability of cheap, transportable energy and dramatically increased the amount of land that a single farmer could manage. As more land was cultivated for agriculture, the focus shifted from geographic expansion to harvest yield per hectare. A renaissance of farming strategies including chemical synthesis of fertilizers and pesticides expanded food security and further optimized human control over the biological world.

In the late 20th century, research into the building blocks of life opened the door to a brand new realm for generating deliberate, targeted changes in biological systems worldwide. A complex understanding of the genetic material inherent to all life on the planet allowed humans to engineer organisms with new attributes. Transgenic organisms can be made more resistant to pests and disease, reducing the risk of catastrophic crop loss. Clever exploitation of this knowledge has led to a worldwide spread in genetically engineered crops. The power of deliberate genetic manipulation has expanded well beyond crop engineering, and has been monumental to progress in human longevity.

However, as humans have expanded globally, the impact of our actions to the health of the planet has expanded as well. A renewed interest on the sustainability of our farming and industrial practices has restricted the window of acceptable technologies to our future development. As we look to the future, progressive technological advances will be required to supply clean energy, sufficient food, sustainable materials, and affordable medicine to a rapidly increasing global population. The solution to these problems may lie in our ability to cultivate and manipulate a new commercial organism: microalgae.

Fuel

Algae-derived biodiesel has been an obvious target for renewable fuels because algae was one of the major contributors of the organic material that comprises the petroleum we use today. Algae and other microorganisms which populated the earth's oceans found their way into the earth's crust where they were subjected to high temperature and pressure to create the energy-dense reserves we refine to power the world. The same process can be imitated by factories in a process called hydrothermal liquefaction to turn algae grown today into the exact same hydrocarbons which power machines around the world. Algae farms can renewably provide a fungible fuel source at a global scale.

Food

The livestock industry currently uses one-third of the total cropland worldwide to grow feed. A global drop in meat consumption or an alternative source of animal feed will be required as the availability of arable land worldwide continues to dwindle. Algae's rapid generation times and ability to grow on non-arable land make it an intriguing source for animal feed. Unlike higher plants which dedicate much of their energy to developing carbohydrates for structure, microalgae are rich in lipids and protein which are critically important to feed composition. Many species of

algae naturally produce omega-3 fatty acids, namely EPA and DHA, are known to reduce heart disease and inflammation in humans, making them also an intriguing resource for human supplements. Growing particular algal strains which are high in these unsaturated fatty acids will also reduce the need to farm and extract these oils from fish.

Materials

Today, nearly all carbon-based materials require petroleum-derived products for their manufacture. The chemical process of synthesizing carbon networks for plastics and polyurethanes renders these materials largely inaccessible to the enzymatic degradation, meaning they cannot be naturally broken down in the environment. A social push for biodegradable materials has led researchers to photosynthetically derived oils as a source of renewable, buildable carbon. Algae are a promising host for petroleum-alternative materials because petroleum was derived from algae to begin with. Naturally occurring algal diversity provides a source for pathways and products which are difficult to produce in other systems. The lack of tissue differentiation minimizes the energy conversion to undesired products. Lastly, algae does not interfere with the food supply chain.

Medicine

Recombinant protein-based therapeutics have brought a new wave of ideas into modern medicine. However, current production systems are wrought with issues. Mammalian tissue cultures provide a quality platform for high-quality therapeutics, but they are expensive to cultivate, and can be infected by viruses which can crash the culture and affect regulatory control of the purified therapeutic. Prokaryotic hosts like *E. coli* lack the advanced folding machinery to properly assemble many complex protein therapeutics. Algal hosts are cheap to cultivate in photosynthetic or fermentation systems and can assemble complex proteins. Genetic fusions facilitate production of chimeric proteins such as antibody-toxin fusions which are traditionally synthesized chemically,

a process which needs to be optimized for each antibody and toxin combination. In addition, many algal species have been designated by the FDA as "generally regarded as safe", meaning that they can be safely consumed. Oral consumption of whole-cell algae which contain the therapeutic alleviates the need for costly protein purification and storage. An algae-expressed gut-active malaria vaccine is being investigated as a suitable host for bringing immunogenicity to developing countries.

Altogether, the future of green algae to provide a sustainable, affordable and renewable resource for fuel, food, materials, and therapeutics exists because of our capacity to understand these organisms at a molecular level. Genetic tools can be used to make algae more robust to biotic and abiotic challenges, and can create whole new markets for the platform. The relatively short history of relevance to the commercial market has limited the breadth of molecular knowledge in algae. Not surprisingly, the diversity of industrially interesting species of algae is vast, and detailed characterization of every species is far from complete. In an effort to consolidate research efforts and improve the depth of algal knowledge, many researchers to date have focused on a singular green alga, *Chlamydomonas reinhardtii*.

Chapter 1 provides an overview of why *C. reinhardtii* is a suitable model organism for green algae and reviews the critical aspects of development of *C. reinhardtii* pertaining to entering the industrial space. Although fuel production is not economically feasible for microalgae with the current infrastructure, the addition of high-value coproducts may help cover the costs for oil cultivation. A succinct description of the genetic toolset used to engineer the chloroplast is presented, focusing on the success of endogenous regulatory elements to engineer transgenic strains and the diversity of recombinant proteins which can be produced. This analysis continues by

outlining the success of synthetic regulatory elements albeit with some caveats in the genetic toolset currently available for chloroplast engineering. Proteins expressed in the chloroplast cannot be transported throughout the cell, which significantly limits opportunities for metabolic engineering. Next, the advantages and roadblocks of nuclear genome engineering in *C. reinhardtii* are explored. Although genetic engineering in the nuclear genome has yet to show substantial accumulation of recombinant proteins, peptide transit sequences allow protein secretion as well as transport to many intracellular organelles, which will be required for advanced metabolic engineering. A major limitation of nuclear genome editing is the lack of sophisticated gene targeting technology, which is addressed more extensively in Chapter 7. Chapter 1 then describes aspects of *C. reinhardtii* as it fits into the industrial space. Basic descriptions are laid out for outdoor cultivation strategies, including both open pond systems and closed systems. An analysis of biomass and lipid accumulation of green algae in both open and closed cultivation identifies *C. reinhardtii* as an effective closed-system producer with relatively high lipid content.

In accordance with current regulations, transgenic microalgae need to be grown in closed systems to prevent the spread of heterologous genes to native species. This is particularly important with respect to antibiotic resistance cassettes which can lead to the unintended development of antibiotic-resistant microbes. A pilot experiment conducted at UCSD showed that a relatively benign fluorescent reporter GFP gene did not show significant spread to native species. This marker lends itself to unique selection strategies for development of transgenic strains. Chapter 2 presents a FACS-based strategy for developing transgenic *C. reinhardtii* utilizing expression of a fluorescent protein in lieu of an antibiotic resistance cassette. Cells were transformed with a GFP expression cassette containing a transcriptionally fused gene for resistance to zeocin. *C. reinhardtii* was electroporated to produce nuclear genome lesions and subsequent integrations of our expression vector. Instead of selecting by antibiotic resistance, the transformed cells were run

through a sorting flow cytometer, and selected for recovery based on their GFP fluorescence. Of the 39 individuals which survived the sorting process, six were resistant to zeocin on selection plates, and of those six transformants, four showed statistically significant GFP accumulation by plate reader. This enrichment of successful transformants which are expressing the gene of interest is comparable to selection strategies utilizing antibiotic resistance.

Even with clever selection strategies that eliminate the need for antibiotic resistance, closed systems will be required for the industrial relevance of *C. reinhardtii* and green algae as a whole. Many strains will require the precise control of their environment that can only be achieved in closed systems. In addition, the high value coproducts produced in transgenic strains will likely require closed growth chambers to facilitate reliable purification strategies and minimize product loss. Chapter 3 provides the first characterization of chloroplast-expressed transgene expression from *C. reinhardtii* in a closed fermentor system. The growth rate and accumulation of a chloroplast-expressed GFP marker in *C. reinhardtii* was compared in a closed fermentation system under batch growth and fed-batch growth conditions. The batch growth strategy resulted in a density of 2.33g/L AFDW. In the fed-batch approach, a predominantly acetic-acid based feed was administered under the control of a pH-stat, which maintained a constant pH of 7.0, and pushed the final maximum density to 23.7g/L AFDW. Although the biomass yield was 10-fold higher in the fed batch system, the relative GFP expression level per cell was significantly diminished. This could be due to expression being regulated by a light-driven promoter, and as culture density increases, light penetration is diminished and gene activation slows. Overall, the fed-batch system was able to generate higher levels of recombinant protein per liter of culture. Cultures from both growth strategies were analyzed for metabolite composition to determine potential stresses caused by the density of the culture and the feed. High-density *C. reinhardtii* culture maintenance and feed

development look to be fruitful targets to providing major improvements to recombinant protein production strategies.

As mentioned previously, the most productive recombinant protein production strategies in *C. reinhardtii* hijack endogenous regulatory regions which ordinarily regulate photosystem proteins to drive expression of exogenous genes. With the future of recombinant protein production likely to occur in high-density cultures in which light availability is limited, or in stainless steel fermenters where light is completely excluded, strains and tools must be optimized to express transgenes in the dark. The endogenous regulatory elements in the chloroplast used in current expression vectors have shown minimal dark activity, and testing a catalog of new endogenous elements is not a guaranteed solution.

Chapter 4 investigates a nonspecific approach in which a chloroplast GFP gene was optimized for dark expression through rounds of mutagenesis and breeding and selection by FACS. Genome shuffling through mutagenesis and breeding can lead to chromatin restructuring and changes in gene regulation which can have dramatic effects on the phenotype of progeny. This technique has been used previously in *C. reinhardtii* in which breeding induced variability led to increased cell viability in higher salt concentrations. This project intended to exploit the phenotypic variance induced by breeding to rapidly generate a novel strain which would have improved accumulation of recombinant protein in the dark.

A strain which expressed chloroplast GFP was either subjected to UV mutagenesis or bred with one of 10 genetically distinct wild-type strains to encourage changes in gene regulation via genome shuffling. Resulting progeny were investigated by flow cytometry, and the individuals which showed increases in GFP fluorescence above the parent strain were isolated by FACS. The best expressors from a single round of mating or mutagenesis and selection had increased GFP

accumulation in the dark by approximately 10-fold. Additional rounds of mating and mutagenesis showed further increases in GFP accumulation, although relative gains were substantially lower in the second round. Although many species of algae are not known to undergo sexual reproduction, the success of the UV mutagenesis and sorting shows that this strategy can be amended to most species. Ultimately, the success of this program shows that novel phenotypes which are difficult to engineer deliberately can be created as long as there is an effective selection strategy. However, it can be difficult to use this strategy to investigate the basic biology underlying the developed phenotypes. A thorough understanding of the biological pieces which dictate a phenotype help lay the foundation for building next-generation tools.

Chapter 5 presents a much more investigative approach to try to improve exogenous gene expression in the nuclear genome. Nuclear gene expression in *C. reinhardtii* lags well behind gene expression in the chloroplast, and is orders of magnitude away from the industry leading platforms for recombinant protein accumulation like *E. coli*, yeast and CHO cells. Due to low natural rates of homologous recombination in the nuclear genome, exogenous expression vectors are integrated randomly. Positional effects due to the presence of local regulatory elements or chromatin structure can dramatically alter gene expression from one transformant to the next. In addition, high GC content can substantially restrict codon usage in exogenous genes, and nuclear gene silencing has been shown to stifle recombinant protein accumulation.

Although transgenic protein accumulation is low, genes expressed in the nuclear genome undergo post-translational modifications which can make them much more desirable for many applications. Metabolic engineering routinely requires expression of proteins within specific organelles, and proteins translated by nuclear-encoded genes can be tagged with localization signals that direct the products to precise locations. Other tags send translated proteins out of the cell

entirely, providing a circumstance by which cells do not need to be lysed in order to purify the desired product. Finally, proteins which are encoded in the nuclear genome can be glycosylated in the ER and Golgi complex. Glycosylation of proteins has an array of functions, including assistance in folding, protein maintenance, and cell recognition. From a recombinant protein standpoint, glycosylation is often required for bioactivity, and the pattern of glycosylation varies from one production system to the next. Yeast, for example, is known to hyperglycosylate its protein products. Proteins with substantially different glycosylation patterns can be recognized as foreign or misfolded and subjected to degradation. These criteria make improved gene expression in the nuclear genome a major objective.

The report in Chapter 5 takes a synthetic biology approach to improve nuclear transgene expression. The study investigates the building blocks of promoter regions in *C. reinhardtii* by determining the ability of synthetic promoters to drive expression of a fluorescent protein. Synthetic promoters were designed by first determining the conservation of sequence motifs within the regulatory regions of the most highly expressed genes in the genome by use of a program called POWRS. Though the exact biological role of these individual motifs is unknown, their presence in multiple highly expressed genes suggests they may have an effect on transcriptional regulation. These motifs were assembled into a promoter backbone based on their positional dependence in relation to the transcriptional start site. By this process, 25 unique synthetic algal promoters were assembled and cloned in front of a fluorescent protein. The strength of each promoter was validated by determining the fluorescence of at least 5,000 independent transformation events to account for positional effects. In this first trial of synthetic promoter construction, we were able to create functional promoters, some of which were stronger than the best hybrid promoter described in the literature. Targeted deletions of synthetic promoter sequences were performed to determine which

motifs were critical to the phenotype shown. One motif was identified as being necessary but not sufficient for transcriptional activation.

The process of narrowing down effective motifs is elaborate, and involves transformation of many uniquely built expression constructs. In an effort to more rapidly characterize the positional dependence and overall effect of individual motifs, we developed two promoter libraries, each with 1000 unique partial-promoter sequences. This generation of synthetic promoters contained fewer motifs, and could therefore be used to more accurately assign activity to specific motifs. The promoter libraries were cloned in front of a fluorescent reporter and FACS was used to isolate the most fluorescent individuals. By comparing the sequences of highly fluorescent individuals with the sequences of low and non-expressors, the goal was to categorize sequence motifs by their activity and positional dependence. However, road blocks associated with reliable cell sorting and recovery of stable phenotypes complicated downstream analysis. Chapter 6 provides an overview of this work as well as a look ahead at how we may utilize our understanding of cis- and trans-activation in *C. reinhardtii* to develop a second generation of synthetic promoters and genetic tools for many algal species. For example, we show that constitutive expression of a native transcription factor can impact recombinant protein accumulation driven by a hybrid promoter. By using our understanding of trans-activating factors in other systems, motifs which correspond to transcription factor binding sites can be added to promoter sequences to increase transcription. Similar to the orthogonal systems developed in S. cerevisiae, synthetic regulatory circuits could be designed in *C. reinhardtii* to drive exceptional nuclear gene expression.

As mentioned previously, a major limitation to recombinant protein platform development in the nuclear genome of *C. reinhardtii* is the lack of reliable gene editing tools and targeted vector introduction. Many labs are researching the potential of systems like CRISPR/Cas9 and TALENs, but a major target to solve this problem is the low rate of native homologous recombination

mediated repair in the nuclear genome. The final chapter of this dissertation presents a tool which reliably measures the rate of homologous recombination. There are many theories as to how the rate of HR could be increased, but reliable analysis of the rate of HR has been limited. Previous systems for HR measurement do not capture instances in which imperfect HR occurs. This report utilizes a large intron as the homology region, meaning that sequence alterations due to imperfect HR mediated repair are not detrimental to the selection mechanism. By using this tool, future studies which seek to increase HR rates for nuclear gene editing can verify the efficacy of their strategies.

Ultimately, the studies presented here show concrete advances to the field of algal biology, and provide important steps to the commercial relevance of algae as a production platform for a variety of products. The production systems that are used today have reached their scale because of the dedication to research on their fundamental biology, and the clever generation of molecular tools by scientists to investigate and alter their being. As the tools for the algae world approach the sophistication of these model organisms, the full potential of green algae to be an industrial powerhouse will be realized.

CHAPTER 1:

CHLAMYDOMONAS AS A MODEL FOR BIOFUELS AND BIO-PRODUCTS PRODUCTION

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2015) **82,** 523–531 doi: 10.1111/tpj.12780

SI CHLAMYDOMONAS

Chlamydomonas as a model for biofuels and bio-products production

Melissa A. Scranton, Joseph T. Ostrand, Francis J. Fields and Stephen P. Mayfield*

California Center for Algae Biotechnology, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Dr, La Jolla, CA 92093-0368, USA

Received 9 December 2014; revised 8 January 2015; accepted 12 January 2015; published online 1 February 2015.
*For correspondence (e-mail smayfield@ucsd.edu).

SUMMARY

Developing renewable energy sources is critical to maintaining the economic growth of the planet while protecting the environment. First generation biofuels focused on food crops like corn and sugarcane for ethanol production, and soybean and palm for biodiesel production. Second generation biofuels based on cellulosic ethanol produced from terrestrial plants, has received extensive funding and recently pilot facilities have been commissioned, but to date output of fuels from these sources has fallen well short of what is needed. Recent research and pilot demonstrations have highlighted the potential of algae as one of the most promising sources of sustainable liquid transportation fuels. Algae have also been established as unique biofactories for industrial, therapeutic, and nutraceutical co-products. *Chlamydomonas reinhardtii*'s long established role in the field of basic research in green algae has paved the way for understanding algal metabolism and developing genetic engineering protocols. These tools are now being utilized in *C. reinhardtii* and in other algal species for the development of strains to maximize biofuels and bio-products yields from the lab to the field.

Keywords: *Chlamydomonas reinhardtii*, recombinant proteins, biofuels, bio-products, molecular engineering.

INTRODUCTION

Access to affordable and environmentally sustainable fuels and energy sources may be the greatest challenge of this century. With demand continuing to increase and new supplies costing ever more to extract, the availability of fossil fuels will inevitably shrink, resulting in rising energy prices worldwide. With the rising cost of energy also comes the rising cost of food, as food and fuel prices are closely linked. Recently, algae-based biofuels have been highlighted as one of the best current alternative source of renewable energy (Merchant *et al.*, 2011; Georgianna and Mayfield, 2012; Leite *et al.*, 2013; Oncel, 2013). Algae do not compete for arable land, have fast generation times, can grow in salt and waste water, and have the potential to produce more oil per acre than land plants (Dismukes *et al.*, 2008; Demirbas and Demirbas, 2011). While most focus has been on the production of biodiesel, algae can also be a source of other fuels, such as bioethanol, biohydrogen, and biogas (Jones and Mayfield, 2011; Oncel, 2013).

While algal biofuels hold significant promise to meet future energy demands, improvements are needed at all levels of production in order to realize this potential. Algal biofuels are not economically viable on their own at current production levels (Brownbridge *et al.*, 2013; Nagarajan *et al.*, 2013). One near-term solution is to couple biofuel production with high-value co-products to increase the commercial value of the entire algal biomass. Current high-value bio-products produced in algae include industrial and therapeutic proteins as well as nutraceuticals and other high-value small molecules (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Rasala and Mayfield, 2014).

As the most characterized algal species with the largest set of genetic tools and techniques, *Chlamydomonas reinhardtii* is an excellent model organism to understand and improve biofuels and bio-products production in algae. *Chlamydomonas reinhardtii* has led the field in the development of molecular tools for strain selection and engineering for green alga. By far, more recombinant proteins have been expressed in *C. reinhardtii* than all other algal species combined (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Rasala and Mayfield, 2014). Studies in *C. reinhardtii* have also helped elucidate the molecular mechanisms behind algal lipid and hydrogen metabolism

(Merchant *et al.*, 2011; Torzillo and Seibert, 2013). *C. reinhardtii* is also among the first of the engineered algal species to be studied in commercial settings, which allows academic researchers to begin to understand the challenges of bringing transgenic algae to commercial-scale production (Scoma *et al.*, 2012; Gimpel *et al.*, 2014; Schoepp *et al.*, 2014).

BIOENGINEERING ALGAE

Chloroplast engineering

Over the last 70 years, *C. reinhardtii* has become the flagship alga for laboratory studies and genetic manipulation. The eukaryotic green alga has three modifiable genomes and is capable of producing a wide variety of protein products (Rosales-Mendoza *et al.*, 2012; Barrera and Mayfield, 2013; Rasala *et al.*, 2013b). The efficient manufacture of these products at commercial viability will require a myriad of genetic tools to enhance protein accumulation and bioactivity (Figure 1). To date, the chloroplast genome has been the primary target for engineering protein production, predominantly because it readily performs homologous recombination and is easily transformed (Boynton *et al.*, 1988). Typically, recombinant protein expressed from the chloroplast accounts for 1–10% of total protein (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Rasala and Mayfield, 2014).

One current limitation of chloroplast engineering in *C. reinhardtii* is the use of native promoters to achieve high levels of protein accumulation. The best exogenous gene expression achieved utilizes the endogenous *psbA* promoter (Manuell *et al.*, 2007). However, the effective use of the *psbA* promoter requires that the endogenous *psbA* regulatory region be removed due to a phenomenon known as auto-attenuation (Minai *et al.*, 2006; Rasala and Mayfield, 2011). This renders the organism non-photosynthetic, eliminating one of the unique attributes of algae. One way to alleviate this issue is to reintroduce the *psbA* gene under the control of a different endogenous promoter (Gimpel *et al.*, 2014). This method results in photosynthetic *C. reinhardtii* that can produce recombinant proteins, albeit at reduced levels compared with the non-photosynthetic versions. Exogenous transcriptional regulation machinery have been used successfully in *C. reinhardtii* (Kato *et al.*, 2007), however exogenous promoters remain an underutilized resource for chloroplast engineering. Synthetic regulatory elements, such as UTRs, that modify gene expression have also been used in the *C. reinhardtii* chloroplast (Specht and Mayfield, 2012), although the successful use of fully synthetic promoters has yet to be described. A full understanding of synthetic promoter synthesis would allow for fine-tuning of gene expression, which is particularly useful for metabolic engineering.

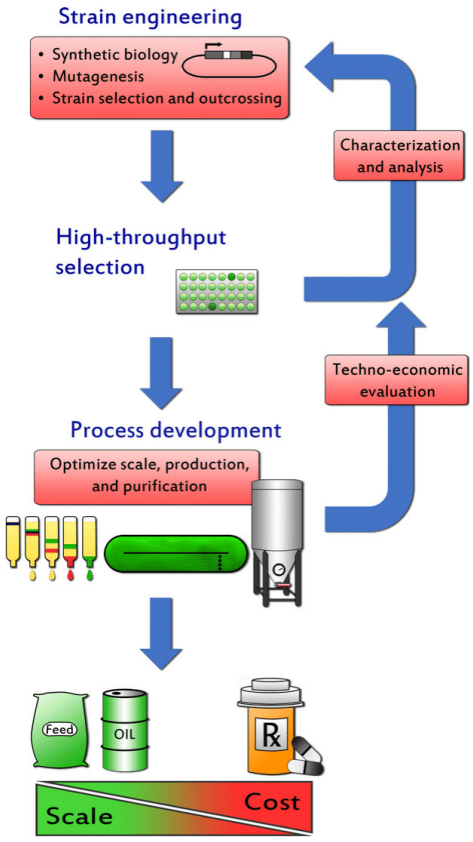

Figure 1. Steps to engineering a production strain of green algae. Genetic manipulation using advanced tools will result in a desired phenotype. The isolated strain will be tested for scale-up and reanalyzed for characteristics such as growth rate, population density, recombinant protein accumulation, and lipid profile. If these traits do not overcome economic constraints of production, additional genetic manipulation in the laboratory is required. Once an ideal strain is identified, fuels and co-products can be extracted from scaled-up cultures.

In addition to engineering strains to produce maximal amounts of recombinant protein, chloroplast engineering is necessary for developing strains for robust growth, particularly in outdoor settings. Light intensity and availability is a major concern for maximizing growth rates and cell density. Excessive solar irradiation leads to photoinhibition, while low light flux due to weather or cell culture density can lead to reduced photosynthetic activity and slower growth. Chloroplast engineering for improved function under varying light conditions would allow maximum production of algae throughout the year. Early data have shown that the light antenna of *C. reinhardtii* can be modified to alter chlorophyll content and subsequently adjust

its absorbance spectrum, altering photosynthetic productivity (Kirst *et al.*, 2012).

An alternative method of chloroplast engineering to alter photosynthetic activity has focused on the protein subunits of the photosystem itself. Photosystem II is a highly conserved protein complex, which performs the initial rate limiting step of photosynthesis. It is known that PSII is in part responsible for enhanced fitness in various light conditions (Mulo *et al.*, 2012). Recently, the D1 protein of *C. reinhardtii* was replaced with the homologous D1 proteins from *Synechococcus sp. 7942* (Vinyard *et al.*, 2013, 2014). These studies showed that complementation of the native D1 protein with natural variants from other species could increase photosynthetic efficiencies under varying light conditions, and also that this increased efficiency led to improved overall growth. In addition, due to *C. reinhardtii*'s unique low fluorescence background, these studies were the first to elucidate the mechanisms behind the optimized photosynthetic activity. Thus, not only were improved strains generated, but *C. reinhardtii* proved to be an ideal model for understanding PSII function.

Nuclear engineering

Although the chloroplast can effectively accumulate high levels of recombinant protein, nuclear transformation technology is required to fully engineer algae. Nuclear expression allows for organelle targeting or secretion of proteins, and enables more complex post-translational modifications of proteins such as glycosylation. In particular, proper protein localization is required for metabolic engineering, which relies on compartmentalization for some steps in syntheses. The nuclear genome of *C. reinhardtii* has been much more difficult to develop for recombinant protein production than the chloroplast. Low protein accumulation levels due in part to transgene silencing and insertion positional effects have limited the use of the nuclear genome from a protein manufacturing standpoint (Cerutti *et al.*, 1997; De Wilde *et al.*, 2000; Wu-Scharf *et al.*, 2000). However, recently it was shown that fusing recombinant proteins to the Sh Ble antibiotic resistance gene, which requires high protein accumulation, could substantially increase recombinant protein accumulation (Rasala *et al.*, 2012). In addition, *C. reinhardtii* lacks strong constitutive promoters comparable to the viral promoters used in plants. While hybrid promoters have had some success in increasing transgene expression in *C. reinhardtii* (Schroda *et al.*, 2000), truly synthetic promoters may hold great promise for tight regulation of robust recombinant protein expression (Venter, 2007).

Homologous recombination

Homologous recombination (HR) allows for targeted gene knock-outs as well as targeted insertion for reducing positional effects for recombinant gene expression. This mechanism has enabled sophisticated engineering in the *C. reinhardtii* chloroplast; however at present, HR does not occur at sufficient levels in the nuclear genome to be of practical use. Knockdowns have been created by artificial microRNAs, which is also a valuable tool for genetic engineering (Molnar *et al.*, 2009; Zhao *et al.*, 2009; Moellering and Benning, 2010; Schmollinger *et al.*, 2010). Alternatively, a knockout strain library is currently being developed by insertional mutagenesis. Because of the scale of this project, a high-throughput screen known as ChlaMmeSeq has been established to identify genes disrupted by insertional mutagenesis. In a pilot screen, a pool of insertional mutants was shown to cover 39% of known protein coding genes (Zhang *et al.*, 2014). Although each of these methods have been valuable for progress of the field, a rapid method for developing novel knock-outs and targeted gene introduction is essential for future development of *C. reinhardtii*.

To date, methods of identifying HR in *C. reinhardtii* have not been reliable, nor have there been ways to significantly increase the low basal rate of HR in Chlamydomonas (Sodeinde and Kindle, 1993; Gumpel *et al.*, 1994; Nelson and Lefebvre, 1995). However, there have been many advances in targeted gene delivery across other platforms in the past decade. Transcription-activator-like effectors (TALEs) allow one to design a unique DNA-binding target sequence and the addition of a nuclease can achieve targeted restriction digestion and subsequent homology directed repair (Li *et al.*, 2011). Successful gene targeting by a TALE has been shown by a subsequent increase in endogenous ARS expression (Gao *et al.*, 2014). However, to this point, TALENs have not successfully been utilized for targeted gene disruption in *C. reinhardtii*. The CRISPR/Cas9 system is an alternative method for inducing targeted gene delivery. Here, a guide RNA will direct restriction by the Cas9 nuclease to a specific sequence of DNA. The CRISPR/Cas9 system has been successful in mammals, fish, fungus, and plants (tobacco, Arabidopsis, sorghum and rice; Jiang *et al.*, 2013). Cas9 has recently been transiently expressed in *C. reinhardtii,* although stable cell lines have not been established (Jiang *et al.*, 2014a). Successful transient expression shows that Cas9 is capable of modifying the *C. reinhardtii* genome.

The *C. reinhardtii* nuclear genome generally performs repairs of double-stranded breaks almost exclusively by non-homologous end joining (NHEJ) and errors are often introduced during the process. Unfortunately, because *Chlamydomonas* relies so heavily on NHEJ as a repair mechanism, even if TALENs or the CRISPR/Cas9 systems are developed in *C. reinhardtii,* it is unlikely that a large percentage of gene delivery will occur by targeted homology repair. NHEJ machinery is comprised of three main subunits: Ku70, Ku80, and DNA ligase IV. In yeast, Ku70 and

Ku80 form a heterodimer that binds to double stranded breaks and recruits DNA ligase IV to mediate NHEJ (Kanaar *et al.*, 1998). The *C. reinhardtii* genome contains homologues to each of these genes. Based on a study in *Pichia pastoris,* one strategy may be to stifle expression of these repair proteins in *Chlamydomonas* in order to increase the level of HR activity (Näätsaari *et al.*, 2012).

High-throughput screening

High-throughput screening allows detection of rare phenotypes including those following gene insertion or mutagenesis. One of the fastest quantitative methods of screening mutants is fluorescence-activated cell sorting (FACS). This technique has been utilized for detection of high lipid accumulating mutants by a process utilizing Nile Red, a lipid dye, to analyze lipid levels in a set of mutants generated using EMS (Xie *et al.*, 2014). This method successfully isolated *C. reinhardtii* strains that accumulated 23–58% higher fatty acid content by dry mass when compared to their parent strain. Alternatively, Terashima *et al.* (2014) developed a high lipid-accumulating mutant pool by insertional mutagenesis, a strategy termed *Chlamydomonas* High Lipid Sorting (CHiLiS). This method generates a detectable insertion that can be used to rapidly identify the disrupted gene(s). Characterization of the strains isolated by this method showed a significant increase in triacylglycerol accumulation. These papers demonstrate that rapid screening can identify mutants with altered metabolism for the production of biofuels, which can easily be translated beyond *C. reinhardtii.* In addition to fluorescent stains, fluorescent protein reporters in *C. reinhardtii* have been shown to effectively sort using FACs (Rasala *et al.*, 2013a). By tagging recombinant proteins with fluorescent protein tags, production strains can be rapidly screened and genes that affect recombinant protein accumulation can be identified. Fluorescent reporters fused to antibodies have also been successfully used to sort for wild isolates with specific cell wall components (Jiang *et al.*, 2014b). Development of techniques using these reporters will be necessary to rapidly isolate and engineer production strains.

Ultimately, high-throughput screening technologies will be an essential tool for generating strains for outdoor production. High-throughput screening allows one to select exclusively based on desirable traits, rather than validating transformation with linkage to an antibiotic marker. Horizontal gene transfer of antibiotic resistance cassettes to environmental microbes poses a theoretical threat to human health and safety. Therefore, high-throughput screening mechanisms that avoid generation of transgenic algae with drug resistant markers, and that can be scaled up to outdoor ponds for economically feasible biofuels production, are an area where more research is required.

ALGAL BIOFUELS

Biodiesel

Chlamydomonas reinhardtii has been a key model in understanding complex algae lipid metabolism. Current efforts have focused on the use of triacylglycerols (TAGs) as a first generation biodiesel. TAGs from algae are also of interest as alternatives to plant-based edible oils (Klok *et al.*, 2014). Traditionally, *C. reinhardtii* has not been seen as an oleaginous species (James *et al.*, 2011). However, studies have shown that natural levels of lipids can vary widely in *C. reinhardtii* strains (Siaut *et al.*, 2011). In addition, wild-type and mutant *C. reinhardtii* can accumulate significant amounts of TAGs in response to nitrogen or salt stress (Wang *et al.*, 2009; Li *et al.*, 2010; Work *et al.*, 2010; Goodson *et al.*, 2011; James *et al.*, 2011; Siaut *et al.*, 2011). In fact, *C. reinhardtii* studies have helped elucidate the mechanisms behind lipid accumulation in response to nitrogen deprivation (Boyle *et al.*, 2012; Msanne *et al.*, 2012; Goodenough *et al.*, 2014). *Chlamydomonas reinhardtii* has also been an excellent model to identify algal genes involved in TAG metabolism (Merchant *et al.*, 2011; Klok *et al.*, 2014).

Targeted overexpression of putative TAG metabolic genes in *C. reinhardtii* has been met with mixed success. Overexpression of type 2 diacylglycerol acyltransferases (DGATs) DGAT2-1 and DGAT2-5 led to increased lipid content, while DGAT2-a,b,c overexpression had no effect (Deng *et al.*, 2012; La Russa *et al.*, 2012). Overexpression of acyl-ACP (acyl carrier protein) esterase (AAE) led to an altered lipid profile but not an increase in lipid content (Blatti *et al.*, 2011). It is clear that even in *C. reinhardtii,* a better understanding of lipid metabolism is required in order to fully utilize algae's potential as a source of TAGs for biodiesel or edible oil production.

Biohydrogen

As an alternative to liquid fuels, *C. reinhardtii* has also been extensively studied as a model for the photoproduction of biohydrogen. Renewed interest in *C. reinhardtii* biohydrogen production began when significant and sustained H_2 production was demonstrated in sulfur starved strains (Melis *et al.*, 2000). The theoretical maximum light conversion efficiency for H_2 production is 13.4%, which is higher than the 11.2% limit for biodiesel (Torzillo and Seibert, 2013). In addition, purities of up to 98% can be achieved without the need of extraction (Torzillo *et al.*, 2009). However, several factors significantly limit the actual production of H_2 in algae. The most significant limitation is the sensitivity of hydrogenase to O_2, which inhibits hydrogenase activity by affecting transcription and protein maturation (Ghirardi *et al.*, 2007). Additional constraints include competition for electrons from alternative pathways, inefficient light conversion of large light-har-

vesting antennae, and inherent limitations of anaerobic growth (Torzillo and Seibert, 2013).

Studies in *C. reinhardtii* to address these limitations have included the generation of improved strains through either mutagenesis or targeted genetic engineering (Melis *et al.*, 2007; Esquível *et al.*, 2011; Torzillo and Seibert, 2013; Baltz *et al.*, 2014; Xu *et al.*, 2014). Studies have also shown significantly improved H_2 production through the optimization of growth conditions including periodic sulfur starvation (Melis *et al.*, 2000) and optimized light, media and photobioreactor conditions (Torzillo and Seibert, 2013; Oncel *et al.*, 2014). However, despite these optimizations, the best light conversion efficiency reached has been approximately 3%. Even if these optimal conditions could be perfectly scaled to outdoor growth, the cost of algal H_2 would be well over US$8 per gallon gasoline equivalent (gge) (James *et al.*, 2009). However, continued research holds promise to optimize algae strains and growth conditions to make biohydrogen a competitive fuel in the future.

HIGH-VALUE BIO-PRODUCTS

Production of recombinant proteins in *C. reinhardtii* has been a fruitful area of research in recent years. Several recent articles have thoroughly reviewed the plethora of protein products that have been successfully expressed within *C. reinhardtii* (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Rasala and Mayfield, 2014). To date, 34 different protein co-products have been produced in algae, 29 of which were produced in *C. reinhardtii.*

The vast majority of targeted proteins have been therapeutic proteins. With a current market value of US$140 billion, therapeutic proteins provide an attractive co-product target to increase the economic value of algal biomass (Walsh, 2014). Therapeutic proteins produced include the UV-protectant metallothionein (Zhang *et al.*, 2006), antibody mimics 14FN3 and SAA-10FN3 (Rasala *et al.*, 2010), and anti-cancer proteins TNF-Related Apoptosis Inducing Ligand (TRAIL; Yang *et al.*, 2006), and allophycocyanin (Su *et al.*, 2005). The glycoprotein hormone erythropoietin has also been produced from the *C. reinhardtii* nucleus, demonstrating the ability to generate recombinant proteins with glycosylation, the most common post-translational modification found on protein therapeutics (Eichler-Stahlberg *et al.*, 2009; Walsh, 2014).

The largest class of therapeutic proteins that have been produced in *C. reinhardtii* are subunit vaccines (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Rasala and Mayfield, 2014). Vaccines are a particularly attractive target because algae are generally regarded as safe (GRAS) by the FDA. Therefore, if orally available vaccines can be produced, these vaccines may be stored and administered as lyophilized algal pellets. Edible algal vaccines would reduce the cost of vaccines by orders of magnitude and vastly increase their availability since algal vaccine pellets do not need to purified, do not require needle injection, and can be stored for up to 20 months at RT (Dreesen *et al.*, 2010; Gregory *et al.*, 2013).

Chlamydomonas reinhardtii has also been shown to be able to produce a novel type of antibody therapeutic. Monoclonal antibodies (mAbs) are the largest growing segment of protein therapeutics with almost 27% of the new approvals and six of the top 10 selling biotherapeutics products in 2013 (Walsh, 2014). In particular, antibody-drug conjugates (ADCs) are gaining in popularity as more effective cancer treatment. The most significant drawback of ADCs is the requirement for chemical linkage of the toxic molecule, which leads to increased production costs and potential off-target toxicity of free toxins. In contrast, the unique environment of the *C. reinhardtii* chloroplast is able to produce immunotoxins in which toxic proteins are genetically linked to the targeting antibody (Tran *et al.*, 2013a,b). This reduces the cost of production since the immunotoxin is produced as a single protein and this production method can also help eliminate off-target toxicity.

Protein co-products expressed in *C. reinhardtii* also include a range of nutraceuticals for use as feed additives and/or human nutritional supplements. Feed additives include phytase, xylanase, and flounder growth hormone (Kim *et al.*, 2002; Yoon *et al.*, 2011; Georgianna *et al.*, 2013). Mammary-associated serum amyloid A (MAA) is one of the best expressed bio-products in *C. reinhardtii,* accumulating to over 5% of total soluble protein (Manuell *et al.*, 2007). MAA can be used as a prophylaxis against enteric bacterial infections in humans as well as replacement for antibiotics in animal feed (Larson *et al.*, 2003; Mack *et al.*, 2003). Recombinant proteins have also been engineered to increase high-value nutritional compounds in *C. reinhardtii* such as organic selenium (Hou *et al.*, 2013), carotenoids (Baldo *et al.*, 2011; Couso *et al.*, 2011), and triacylglycerols (La Russa *et al.*, 2012). Carotenoids and triacylglycerols have the added benefit of being used in biofuels formulations (Merchant *et al.*, 2011; Peralta-Yahya *et al.*, 2011).

OUTDOOR CULTIVATION SYSTEMS

Large-scale microalgae cultivation can occur in either open or closed systems (Ugwu *et al.*, 2008; Singh *et al.*, 2011). Open systems range from small artificial ponds to large open bodies of water; however the most productive open systems currently implemented are shallow raceway ponds (Singh *et al.*, 2011). Although these open systems require large areas of land, they can be constructed in arid environments and avoid competition with preexisting agriculture (Georgianna and Mayfield, 2012). Raceways can be constructed relatively cheaply and require little maintenance (Singh *et al.*, 2011; Schoepp *et al.*, 2014). Open

systems are inherently susceptible to contamination and thus are predominantly used to culture only robust photoautotrophic species (Ugwu *et al.*, 2008; Rasala and Mayfield, 2014). Heterotrophic growth in open systems is not a viable option due to the heightened susceptibility of contamination. Greenhouses can be used to cover smaller open systems in order to deter contamination and potentially increase productivity. Additionally, greenhouses allow for the growth of genetically engineered strains in contained 'open' systems, but these structures will add to the initial cost of a facility.

Closed systems are highly variable in design but overall provide a more productive environment while protecting the culture from contamination (Ugwu *et al.*, 2008; Singh *et al.*, 2011). More strains of algae can be cultivated in these closed systems, including mixotrophic and hydrogen-producing species of microalgae (Scoma *et al.*, 2012; Gimpel *et al.*, 2014). The productivity and sterility benefits of a closed system are countered by their heightened costs as they may cost up to 10 times that of open systems due to the additional infrastructure and operational costs (Del Campo *et al.*, 2007).

Recently, pilot facilities containing a variety of these systems have been built to test strains in a setting more comparable to conditions encountered at commercial facilities (Scoma *et al.*, 2012; Gimpel *et al.*, 2014; Schoepp *et al.*, 2014). Experiments with *C. reinhardtii* have been conducted in these outdoor systems to test the viability of the strain as a producer of biomass, biohydrogen and even high-value bio-products. These studies are invaluable for understanding and addressing the challenges of bringing laboratory strains into commercial production.

Table 1 Biomass production of various algal species in an outdoor setting compared to average reported lipid content

Species	Closed system production (g L^{-1} day^{-1})	Open system production (g L^{-1} day^{-1})	Lipid content (nutrient-replete, % dry weight)
Scenedesmus dimorphus	0.095	0.090	26
Chlorella vulgaris	0.047	0.035	25
Chlamydomonas reinhardtii	0.078	0.015	21
Arthrospira platensis	0.040	0.018	13
Anabaena sp.	0.055	–	5
Porphyridium purpureum	0.074	0.036	11
Nannochloropsis salina	0.043	0.028	27
Dunaliella tertiolecta	0.039	0.031	15
Phaeodactylum tricornutum	0.021	–	21

Biomass production data from Schoepp *et al.* (2014) and lipid content from Griffiths and Harrison (2009).

OUTDOOR CULTIVATION OF *C. REINHARDTII*

Researchers at the University of California, San Diego have developed one of the first systems to measure research-scale production of algal strains in closed bags and open ponds. Biomass production (g L^{-1} day^{-1}) was reported for several algae and cyanobacteria species (Table 1; Schoepp *et al.*, 2014). *Chlamydomonas reinhardtii* (CC-1690) out produced eight other species in a 100-L closed system, but production was reduced by 80% when grown in an open 800-L pond. Lipid content was not measured in this study; however, reported average lipid content in *C. reinhardtii* is comparable to other species (Table 1; Griffiths and Harrison, 2009). This suggests that *C. reinhardtii* is competitive with other algal species in closed systems. While closed systems may be economical for high-value products, the cost and size limitations of closed systems may hinder sufficient production of lipid-accumulating strains for biofuel production (Figure 1).

Chlamydomonas reinhardtii was also tested in an outdoor system for its ability to produce biohydrogen gas. Scoma *et al.* cultivated *C. reinhardtii* heterotrophically in a 50-L tubular closed system. By starving the culture of sulfur, photosystem II was down-regulated to the point where oxygen consumption was greater than oxygen evolution, creating the necessary anaerobic conditions for biohydrogen production. However, the combination of uncontrolled solar intensities and nutrient deprivation can lead to photo-damage and the production of reactive oxygen species in microalgae, thus hindering production. To address this photoinhibition, Scoma *et al.* acclimated their cultures to natural light for 1 week before experimentation to induce physiological changes such as reduced chlorophyll, higher levels of xanthophylls, and increased photosynthetic and respiratory rates. When cells were allowed to acclimate to solar light, a more robust photoprotection system was observed, but biohydrogen production was still only equal to 20% of the laboratory results.

Recombinant protein production in *C. reinhardtii* has been demonstrated in a pilot commercial setting. Bovine Milk Amyloid A (MAA), a protein with anti-microbial properties, has previously been produced to high levels in *C. reinhardtii* in the laboratory (Manuell *et al.*, 2007). High accumulation of MAA in this strain required removal of the *psbA* gene, thus making the strain non-photosynthetic and unfit for outdoor growth. Recently, *psbA* was rescued using the alternative regulatory regions to eliminate issues of auto-attenuation (Gimpel *et al.*, 2014). The best rescued strain was then grown in 100-L hanging bags and accumulated MAA at approximately 3 mg L^{-1} of culture for a maximum yield of 11.8 g MAA kg^{-1} dry weight biomass. Even

with lower photosynthetic efficiency and lower total soluble protein production versus the knockout laboratory strain, the production of MAA in *C. reinhardtii* has the potential to be highly profitable since it can be administered as a solid algal pellet at a cost 60 times lower than traditional purified bovine colostrum MAA.

CONCLUSION

Chlamydomonas reinhardtii with its extensive research history and genetic tool infrastructure has been an excellent system to begin to understand algal metabolism and strain development. *Chlamydomonas reinhardtii*'s ease of transformation, large toolset, and natural variety has led to strains with significantly increased biofuels and bio-product yields. The biggest successes with *C. reinhardtii* have been the production of high-value products at laboratory scale in closed systems (Almaraz-Delgado *et al.*, 2013; Barrera and Mayfield, 2013; Gimpel *et al.*, 2014; Rasala and Mayfield, 2014). However, *C. reinhardtii* remains limited by its naturally low oil content, sensitivity to high solar irradiation and its weak growth in outdoor ponds, thus making it a relatively poor choice for large-scale biofuels production (James *et al.*, 2011; Scoma *et al.*, 2012; Schoepp *et al.*, 2014). While genetic engineering may help to overcome some of *C. reinhardtii*'s limitations, it will be essential to also expand the knowledge and tools developed in *C. reinhardtii* to other algal species.

Algae are the most diverse organisms on the planet, able to grow in almost any environment, from marine systems to hot springs to desert soil crusts and even sewer drains (Norton *et al.*, 1996; Dufresne *et al.*, 2008; Parker *et al.*, 2008; Tirichine and Bowler, 2011; Blunt *et al.*, 2012). Through these adaptations, strains have naturally developed favorable traits for specific manufacturing applications. Many wild-type species of algae will be economically viable as production platforms for bio-products with minimal strain optimization. For instance, cold-tolerant microorganisms have to maintain higher levels of polyunsaturated fatty acids to maintain membrane fluidity at colder temperatures. These extremophiles make obvious choices for sources of omega-3 fatty acids. To date, several natural species have already demonstrated robust growth in commercial-scale systems, mostly for nutraceuticals (Oncel, 2013). In addition, over the past 20 years, we have expanded the use of genetic tools to a wide variety of algal species, albeit with varying success. These tools will be needed to engineer an arsenal of pest-resistant strains with optimal growth rates and production capabilities as well as to utilize algae as a protein biomanufacturing platform. The combination of improved strain selection and engineering in more diverse species is our best chance of developing algal strains for large-scale renewable energy and sustainable recombinant protein production in the future.

ACKNOWLEDGEMENTS

This work is supported by the US Department of Energy (DE-EE0003373) and the California Energy Commission Initiative for Large Molecule Sustainable Fuels (CILMSF 500-10-039).

REFERENCES

Almaraz-Delgado, A.L., Flores-Uribe, J., Pérez-España, V.H., Salgado-Manjarrez, E. and Badillo-Corona, J.A. (2013) Production of therapeutic proteins in the chloroplast of *Chlamydomonas reinhardtii*. *AMB Express*, **4**, 57.

Baldo, F.C., Inmaculada, C., Rosa, L., Herminia, R. and Vargas, M.Á. (2011) Enhancement of carotenoids biosynthesis in *Chlamydomonas reinhardtii* by nuclear transformation using a phytoene synthase gene isolated from *Chlorella zofingiensis*. *Appl. Microbiol. Biotechnol.* **91**, 341–351.

Baltz, A., Dang, K.V., Beyly, A., Auroy, P., Richaud, P., Cournac, L. and Peltier, G. (2014) Plastidial expression of type II NAD(P)H dehydrogenase increases the reducing state of plastoquinones and hydrogen photoproduction rate by the indirect pathway in *Chlamydomonas reinhardtii*. *Plant Physiol.* **165**, 1344–1352.

Barrera, D.J. and Mayfield, S.P. (2013) High-value recombinant protein production in microalgae. In *Handbook of Microalgal Culture: Applied Phycology and Biotechnology* (Richmond, A. and Hu, Q. eds). Oxford, UK: John Wiley & Sons, Ltd, pp. 532–544.

Blatti, J.L., Beld, J., Behnke, C.A., Mendez, M., Mayfield, S.P. and Burkart, M.D. (2011) Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. *PLoS One*, **7**, e42949.

Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H. and Prinsep, M.R. (2012) Marine natural products. *Nat. Prod. Rep.* **29**, 144–222.

Boyle, N., Page, M., Liu, B. *et al.* (2012) Three acyltransferases and nitrogen-responsive regulator are implicated in nitrogen starvation-induced triacylglycerol accumulation in *Chlamydomonas*. *J. Biol. Chem.* **287**, 15811–15825.

Boynton, J.E., Gillham, N.W., Harris, E.H. *et al.* (1988) Chloroplast transformation in *Chlamydomonas* with high-velocity microprojectiles. *Science*, **240**, 1534–1538.

Brownbridge, G., Azadi, P., Smallbone, A., Bhave, A., Taylor, B. and Kraft, M. (2013) The future viability of algae-derived biodiesel under economic and technical uncertainties. *Bioresour. Technol.* **151**, 166–173.

Cerutti, H., Johnson, A.M., Gillham, N.W. and Boynton, J.E. (1997) Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell*, **9**, 925–945.

Couso, I., Vila, M., Rodriguez, H. and Vargas, M.A. (2011) Overexpression of an exogenous phytoene synthase gene in the unicellular alga *Chlamydomonas reinhardtii* leads to an increase in the content of carotenoids. *Biotechnol. Prog.* **27**, 54–60.

De Wilde, C., Van Houdt, H., De Buck, S., Angenon, G., De Jaeger, G. and Depicker, A. (2000) Plants as bioreactors for protein production: avoiding the problem of transgene silencing. *Plant Mol. Biol.* **43**, 347–359.

Del Campo, J.A., Garcia-Gonzalez, M. and Guerrero, M.G. (2007) Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. *Appl. Microbiol. Biotechnol.* **74**, 1163–1174.

Demirbas, A. and Demirbas, F.M. (2011) Importance of algae oil as a source of biodiesel. *Energy Convers. Manage.* **52**, 163–170.

Deng, X.D., Gu, B., Li, Y.J., Hu, X.W. and Guo, J.C. (2012) The roles of acyl-CoA: diacylglycerol acyltransferase 2 genes in the biosynthesis of triacylglycerols by the green algae *Chlamydomonas reinhardtii*. *Mol. Plant*, **5**, 945–947.

Dismukes, C.G., Carrieri, D., Bennette, N., Ananyev, G.M. and Posewitz, M.C. (2008) Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. *Curr. Opin. Biotechnol.* **19**, 235–240.

Dreesen, I.A., Charpin-El Hamri, G. and Fussenegger, M. (2010) Heat-stable oral alga-based vaccine protects mice from *Staphylococcus aureus* infection. *J. Biotechnol.* **145**, 273–280.

Dufresne, A., Ostrowski, M., Scanlan, D.J. *et al.* (2008) Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* **9**, R90.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O. and Heitzer, M. (2009) Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta*, **229**, 873–883.

Esquível, M.G., Amaro, H.M., Pinto, T.S. and Fevereiro, P.S. (2011) Efficient H_2 production via *Chlamydomonas reinhardtii*. *Trends Biotechnol.* **29**, 595–600.

Gao, H., Wright, D.A., Li, T., Wang, Y.J., Horken, K., Weeks, D.P., Yang, B. and Spalding, M.H. (2014) TALE activation of endogenous genes in *Chlamydomonas reinhardtii*. *Algal Res.* **5**, 52–60.

Georgianna, D.R. and Mayfield, S.P. (2012) Exploiting diversity and synthetic biology for the production of algal biofuels. *Nature,* **488**, 329–335.

Georgianna, D.R., Michael, J.H., Marina, M., Shuiqin, W., Kyle, B., Alex, J.L., James, H., Michael, M. and Stephen, P.M. (2013) Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*. *Algal Res.* **2**, 2–9.

Ghirardi, M.L., Posewitz, M.C. and Maness, P.C. (2007) Hydrogenases and hydrogen photoproduction in oxygenic photosynthetic organisms*. *Annu. Rev. Plant Biol.* **58**, 71–91.

Gimpel, J.A., Hyun, J.S. and Schoepp, N.G. (2014) Production of recombinant proteins in microalgae at pilot greenhouse scale. *Biotechnol. Bioeng.* **112**, 339–345.

Goodenough, U., Blaby, I., Casero, D. *et al.* (2014) The path to triacylglyceride obesity in the sta6 strain of *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **13**, 591–613.

Goodson, C., Roth, R., Wang, Z.T. and Goodenough, U. (2011) Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost. *Eukaryot. Cell*, **10**, 1592–1606.

Gregory, J.A., Topol, A.B., Doerner, D.Z. and Mayfield, S.P. (2013) Alga-produced cholera toxin-Pfs25 fusion proteins as oral vaccines. *Appl. Environ. Microbiol.* **79**, 3917–3925.

Griffiths, M.J. and Harrison, S.T.L. (2009) Lipid productivity as a key characteristic for choosing algal species for biodiesel production. *J. Appl. Phycol.* **21**, 493–507.

Gumpel, N.J., Rochaix, J.D. and Purton, S. (1994) Studies on homologous recombination in the green alga *Chlamydomonas reinhardtii*. *Curr. Genet.* **26**, 438–442.

Hou, Q.T., Qiu, S., Liu, Q., Tian, J., Hu, Z.L. and Ni, J.Z. (2013) Selenoprotein-transgenic *Chlamydomonas reinhardtii*. *Nutrients*, **5**, 624–636.

James, B.D., Baum, G.N., Perez, J. and Baum, K.N. (2009) Technoeconomic boundary analysis of biological pathways to hydrogen production. US Department of Energy, Office of Energy Efficiency & Renewable Energy ed. Subcontract report NREL/SR-560-46674. p. 207, U.S. Dept. of Energy, Washington, DC.

James, G.O., Hocart, C.H., Hillier, W., Chen, H., Kordbacheh, F., Price, D.G. and Djordjevic, M.A. (2011) Fatty acid profiling of *Chlamydomonas reinhardtii* under nitrogen deprivation. *Bioresour. Technol.* **102**, 3343–3351.

Jiang, W., Zhou, H., Bi, H., Fromm, M., Yang, B. and Weeks, D.P. (2013) Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in *Arabidopsis,* tobacco, sorghum and rice. *Nucleic Acids Res.* **41**, e188.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M. and Weeks, D.P. (2014a) Successful transient expression of Cas9 and gingle guide RNA genes in *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **13**, 1465–1469.

Jiang, W., Cossey, S., Rosenberg, J.N., Oyler, G.A. and Weeks, D. (2014b) A rapid live-cell ELISA for characterizing antibodies against cell surface antigens of *Chlamydomonas reinhardtii* and its use in isolating algae from natural environments with related cell wall components. *BMC Plant Biol.* **14**, 244.

Jones, C.S. and Mayfield, S.P. (2011) Algae biofuels: versatility for the future of bioenergy. *Curr. Opin. Biotechnol.* **23**, 346–351.

Kanaar, R., Hoeijmakers, J.H.J. and van Gent, D.C. (1998) Molecular mechanisms of DNA double-strand break repair. *Trends Cell Biol.* **8**, 483–489.

Kato, K., Marui, T., Kasai, S. and Shinmyo, A. (2007) Artificial control of transgene expression in *Chlamydomonas reinhardtii* chloroplast using the *lac* regulation system from *Escherichia coli*. *J. Biosci. Bioeng.* **104**, 207–213.

Kim, D.H., Kim, Y.T., Cho, J.J., Bae, J.H., Hur, S.B., Hwang, I. and Choi, T.J. (2002) Stable integration and functional expression of flounder growth hormone gene in transformed microalga, *Chlorella ellipsoidea*. *Mar. Biotechnol.* **4**, 63–73.

Kirst, H., Garcia-Cerdan, J.G., Zurbriggen, A. and Melis, A. (2012) Assembly of the light-harvesting chlorophyll antenna in the green alga *Chlamydomonas reinhardtii* requires expression of the *TLA2-CpFTSY* gene. *Plant Physiol.* **158**, 930–945.

Klok, A.J., Lamers, P.P., Martens, D.E., Draaisma, R.B. and Wijffels, R.H. (2014) Edible oils from microalgae: insights in TAG accumulation. *Trends Biotechnol.* **32**, 521–528.

La Russa, M., Bogen, C., Uhmeyer, A., Doebbe, A., Filippone, E., Kruse, O. and Mussgnug, J.H. (2012) Functional analysis of three type-2 DGAT homologue genes for triacylglycerol production in the green microalga *Chlamydomonas reinhardtii*. *J. Biotechnol.* **162**, 13–20.

Larson, M.A., Wei, S.H., Weber, A., Mack, D.R. and McDonald, T.L. (2003) Human serum amyloid A3 peptide enhances intestinal MUC3 expression and inhibits EPEC adherence. *Biochem. Biophys. Res. Commun.* **300**, 531–540.

Leite, G.B., Abdelaziz, A.E.M. and Hallenbeck, P.C. (2013) Algal biofuels: challenges and opportunities. *Bioresour. Technol.* **145**, 134–141.

Li, Y., Han, D., Hu, G., Dauvillee, D., Sommerfeld, M., Ball, S. and Hu, Q. (2010) *Chlamydomonas* starchless mutant defective in ADP-glucose pyrophosphorylase hyper-accumulates triacylglycerol. *Metab. Eng.* **12**, 387–391.

Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res.* **39**, 6315–6325.

Mack, D.R., McDonald, T.L., Larson, M.A., Wei, S. and Weber, A. (2003) The conserved TFLK motif of mammary-associated serum amyloid A3 is responsible for up-regulation of intestinal MUC3 mucin expression *in vitro*. *Pediatr. Res.* **53**, 137–142.

Manuell, A.L., Beligni, M.V., Elder, J.H., Siefker, D.T., Tran, M., Weber, A., McDonald, T.L. and Mayfield, S.P. (2007) Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast. *Plant Biotechnol. J.* **5**, 402–412.

Melis, A., Zhang, L., Forestier, M. and Ghirardi, M.L. (2000) Sustained photobiological hydrogen gas production upon reversible inactivation of oxygen evolution in the green alga *Chlamydomonas reinhardtii*. *Plant Physiol.* **122**, 127–136.

Melis, A., Seibert, M. and Ghirardi, M.L. (2007) Hydrogen fuel production by transgenic microalgae. In *Transgenic Microalgae as Green Cell Factories* (León, R., Galván Cejudo, A. and Fernández, E. eds). New York: Springer New York, pp. 110–121.

Merchant, S.S., Kropat, J., Liu, B., Shaw, J. and Warakanont, J. (2011) TAG, You're it! *Chlamydomonas* as a reference organism for understanding algal triacylglycerol accumulation. *Curr. Opin. Biotechnol.* **23**, 352–363.

Minai, L., Wostrikoff, K. and Wollman, F. (2006) Chloroplast biogenesis of photosystem II cores involves a series of assembly-controlled steps that regulate translation. *Plant Cell*, **18**, 159–175.

Moellering, E.R. and Benning, C. (2010) RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **9**, 97–106.

Molnar, A., Bassett, A., Thuenemann, E., Schwach, F., Karkare, S., Ossowski, S., Weigel, D. and Baulcombe, D. (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. *Plant J.* **58**, 165–174.

Msanne, J., Xu, D., Konda, A.R. and Casas-Mollano, J.A. (2012) Metabolic and gene expression changes triggered by nitrogen deprivation in the photoautotrophically grown microalgae *Chlamydomonas reinhardtii* and *Coccomyxa sp. C-169*. *Phytochemistry*, **75**, 50–59.

Mulo, P., Sakurai, I. and Aro, E.-M. (2012) Strategies for *psbA* gene expression in cyanobacteria, green algae and higher plants: from transcription to PSII repair. *Biochim. Biophys. Acta*, **1817**, 247–257.

Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F.S. and Glieder, A. (2012) Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. *PLoS One*, **7**, e39720.

Nagarajan, S., Chou, S.K., Cao, S., Wu, C. and Zhou, Z. (2013) An updated comprehensive techno-economic analysis of algae biodiesel. *Bioresour. Technol.* **145**, 150–156.

Nelson, J.A. and Lefebvre, P.A. (1995) Targeted disruption of the *NIT8* gene in *Chlamydomonas reinhardtii*. *Mol. Cell. Biol.* **15**, 5762–5769.

Norton, T.A., Melkonian, M. and Andersen, R.A. (1996) Algal biodiversity. *Phycologia*, **35**, 308–326.

Oncel, S.S. (2013) Microalgae for a macroenergy world. *Renew. Sustain. Energy Rev.* **26**, 241–264.

Oncel, S.S., Kose, A., Faraloni, C., Imamoglu, E., Elibol, M., Torzillo, G. and Sukan, F.V. (2014) Biohydrogen production using mutant strains of *Chlamydomonas reinhardtii:* the effects of light intensity and illumination patterns. *Biochem. Eng. J.* **92**, 47–52.

Parker, M.S., Mock, T. and Armbrust, E.V. (2008) Genomic insights into marine microalgae. *Annu. Rev. Genet.* **42**, 619–645.

Peralta-Yahya, P.P., Ouellet, M. and Chan, R. (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**, 483.

Rasala, B.A. and Mayfield, S.P. (2011) The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. *Bioeng. Bugs*, **2**, 50–54.

Rasala, B.A. and Mayfield, S.P. (2014) Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. *Photosynth. Res.* **123**, 227–239.

Rasala, B.A., Muto, M., Lee, P.A., Jager, M., Cardoso, R.M.F., Behnke, C.A., Kirk, P., Hokanson, C.A., Crea, R., Mendez, M. and **others** (2010) Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*. *Plant Biotechnol. J.* **8**, 719–733.

Rasala, B.A., Lee, P.A., Shen, Z.X., Briggs, S.P., Mendez, M. and Mayfield, S.P. (2012) Robust expression and secretion of Xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A Peptide. *PLoS One*, **7**, e43349.

Rasala, B., Barrera, D., Ng, J., Plucinak, T., Rosenberg, J., Weeks, D., Oyler, G., Peterson, T., Haerizadeh, F. and Mayfield, S. (2013a) Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*. *Plant J.* **74**, 545–556.

Rasala, B.A., Gimpel, J.A., Tran, M. and Hannon, M.J. (2013b) Genetic engineering to improve algal biofuels production. In *Algae for Biofuels and Energy* (Borowitzka, M.A. and Moheimani, N.R. eds). Dordrecht, Netherlands: Springer Netherlands, pp. 99–113.

Rosales-Mendoza, S., Paz-Maldonado, L.M. and Soria-Guerra, R.E. (2012) *Chlamydomonas reinhardtii* as a viable platform for the production of recombinant proteins: current status and perspectives. *Plant Cell Rep.* **31**, 479–494.

Schmollinger, S., Strenkert, D. and Schroda, M. (2010) An inducible artificial microRNA system for *Chlamydomonas reinhardtii* confirms a key role for heat shock factor 1 in regulating thermotolerance. *Curr. Genet.* **56**, 383–389.

Schoepp, N.G., Stewart, R.L., Sun, V. and Quigley, A.J. (2014) System and method for research-scale outdoor production of microalgae and cyanobacteria. *Bioresour. Technol.* **166**, 273–281.

Schroda, M., Blocker, D. and Beck, C.F. (2000) The *HSP70A* promoter as a tool for the improved expression of transgenes in *Chlamydomonas*. *Plant J.* **21**, 121–131.

Scoma, A., Giannelli, L., Faraloni, C. and Torzillo, G. (2012) Outdoor H_2 production in a 50-L tubular photobioreactor by means of a sulfur-deprived culture of the microalga *Chlamydomonas reinhardtii*. *J. Biotechnol.* **157**, 620–627.

Siaut, M., Cuiné, S., Cagnon, C. and Fessler, B. (2011) Oil accumulation in the model green alga *Chlamydomonas reinhardtii:* characterization, variability between common laboratory strains and relationship with starch reserves. *BMC Biotechnol.* **11**, 7.

Singh, A., Nigam, P.S. and Murphy, J.D. (2011) Mechanism and challenges in commercialisation of algal biofuels. *Bioresour. Technol.* **102**, 26–34.

Sodeinde, O.A. and Kindle, K.L. (1993) Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*. *Proc. Natl Acad. Sci. USA*, **90**, 9199–9203.

Specht, E. and Mayfield, S.P. (2012) Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs. *ACS Synth. Biol.* **2**, 34–46.

Su, Z.L., Qian, K.X., Tan, C.P., Meng, C.X. and Qin, S. (2005) Recombination and heterologous expression of allophycocyanin gene in the chloroplast of *Chlamydomonas reinhardtii*. *Acta Biochim. Biophys. Sin.* **37**, 709–712.

Terashima, M., Freeman, E.S., Jinkerson, R.E. and Jonikas, M.C. (2014) A fluorescence-activated cell sorting-based strategy for rapid isolation of high-lipid *Chlamydomonas* mutants. *Plant J.* **81**, 147–159.

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

Torzillo, G. and Seibert, M. (2013) Hydrogen production by *Chlamydomonas reinhardtii*. In *Handbook of Microalgal Culture: Applied Phycology and Biotechnology*, 2nd edn (Richmond, A. and Hu, Q. eds). Oxford, UK: John Wiley & Sons, Ltd., pp. 417–432.

Torzillo, G., Scoma, A., Faraloni, C. and Ena, A. (2009) Increased hydrogen photoproduction by means of a sulfur-deprived *Chlamydomonas reinhardtii* D1 protein mutant. *Int. J. Hydrogen Energy*, **34**, 4529–4536.

Tran, M., Henry, R.E., Siefker, D., Van, C., Newkirk, G., Kim, J., Bui, J. and Mayfield, S.P. (2013a) Production of anti-cancer immunotoxins in algae: ribosome inactivating proteins as fusion partners. *Biotechnol. Bioeng.* **110**, 2826–2835.

Tran, M., Van, C., Barrera, D.J., Pettersson, P.L., Peinado, C.D., Bui, J. and Mayfield, S.P. (2013b) Production of unique immunotoxin cancer therapeutics in algal chloroplasts. *Proc. Natl Acad. Sci. USA*, **110**, E15–E22.

Ugwu, C.U., Aoyagi, H. and Uchiyama, H. (2008) Photobioreactors for mass cultivation of algae. *Bioresour. Technol.* **99**, 4021–4028.

Venter, M. (2007) Synthetic promoters: genetic control through cis engineering. *Trends Plant Sci.* **12**, 118–124.

Vinyard, D.J., Gimpel, J., Ananyev, G.M., Cornejo, M.A., Golden, S.S., Mayfield, S.P. and Dismukes, G.C. (2013) Natural variants of photosystem II subunit D1 tune photochemical fitness to solar intensity. *J. Biol. Chem.* **288**, 5451–5462.

Vinyard, D.J., Gimpel, J. and Ananyev, G.M. (2014) Engineered photosystem II reaction centers optimize photochemistry versus photoprotection at different solar intensities. *J. Am. Chem. Soc.* **136**, 4048–4055.

Walsh, G. (2014) Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **32**, 992–1000.

Wang, Z.T., Ullrich, N., Joo, S. and Waffenschmidt, S. (2009) Algal lipid bodies: stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. *Eukaryot. Cell*, **8**, 1856–1868.

Work, V.H., Radakovits, R., Jinkerson, R.E. and Meuser, J.E. (2010) Increased lipid accumulation in the *Chlamydomonas reinhardtii sta7-10* starchless isoamylase mutant and increased carbohydrate synthesis in complemented strains. *Eukaryot. Cell*, **9**, 1251–1261.

Wu-Scharf, D., Jeong, B., Zhang, C. and Cerutti, H. (2000) Transgene and transposon silencing in *Chlamydomonas reinhardtii* by a DEAH-box RNA helicase. *Science*, **290**, 1159–1162.

Xie, B., Stessman, D., Hart, J.H., Dong, H., Wang, Y., Wright, D.A., Nikolau, B.J., Spalding, M.H. and Halverson, L.J. (2014) High throughput fluorescence activated cell sorting for lipid hyperaccumulating *Chlamydomonas reinhardtii* mutants. *Plant Biotechnol. J.* **12**, 872–882.

Xu, L., Wang, Q., Wu, S. and Li, D. (2014) Improvement of hydrogen yield of *Iba*-transgenic *Chlamydomonas reinhardtii* caused by increasing respiration and impairing photosynthesis. *Int. J. Hydrogen Energy*, **39**, 13347–13352.

Yang, Z.Q., Li, Y.N., Chen, F., Li, D., Zhang, Z.F., Liu, Y.X., Zheng, D.X., Wang, Y. and Shen, G.F. (2006) Expression of human soluble TRAIL in *Chlamydomonas reinhardtii* chloroplast. *Chin. Sci. Bull.* **51**, 1703–1709.

Yoon, S.-M.M., Kim, S.Y., Li, K.F., Yoon, B.H., Choe, S. and Kuo, M.M. (2011) Transgenic microalgae expressing *Escherichia coli* AppA phytase as feed additive to reduce phytate excretion in the manure of young broiler chicks. *Appl. Microbiol. Biotechnol.* **91**, 553–563.

Zhang, Y.K., Shen, G.F. and Ru, B.G. (2006) Survival of human metallothionein-2 transplastomic *Chlamydomonas reinhardtii* to ultraviolet B exposure. *Acta Biochim. Biophys. Sin. (Shanghai)*, **38**, 187–193.

Zhang, R., Patena, W., Armbruster, U., Gang, S.S., Blum, S.R. and Jonikas, M.C. (2014) High-throughput genotyping of green algal mutants reveals random distribution of mutagenic insertion sites and endonucleolytic cleavage of transforming DNA. *Plant Cell*, **26**, 1398–1409.

Zhao, T., Wang, W., Bai, X. and Qi, Y. (2009) Gene silencing by artificial microRNAs in *Chlamydomonas*. *Plant J.* **58**, 157–164.

Acknowledgements

Chapter 1, in full, is a reprint of the material as it appears in *The Plant Journal* 2015. Scranton, Melissa; Ostrand, Joseph; Fields, Francis; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this chapter.

CHAPTER 2:

ANTIBIOTIC-FREE SELECTION IN THE NUCLEAR GENOME OF *CHLAMYDOMONAS REINHARDTII* BY MULTICOLOR CELL SORTING

Antibiotic-free selection in the nuclear genome of *C. reinhardtii* by multicolor cell sorting

Ostrand, J., and Mayfield, S.

Introduction

Chlamydomonas reinhardtii presents a unique photosynthetic platform for a wide variety of potential products. Advances in genetic manipulation in the model green alga have opened the door for metabolic engineering and heterologous protein accumulation. Genetically engineered strains will undoubtedly be required for sophisticated strain optimization and biological manufacturing of high-value products(1). Two major advantages of *C. reinhardtii* as an industrial organism are photosynthetic outdoor cultivation and Generally Regarded As Safe (GRAS) approval for human and animal consumption(2). The fear of horizontal gene transfer and unintended development of antibiotic-resistant microbes will likely prevent outdoor cultivation and consumption of organisms with antibiotic resistance cassettes for the foreseeable future(3,4). Therefore, strategies which either remove the resistance marker post-selection or avoid these markers altogether must be devised.

In the chloroplast genome, high rates of native homologous recombination facilitate removal of antibiotic resistance cassettes through direct-repeat-mediated excision(5). In the nuclear genome, a few strategies have been developed for marker removal, but they predominantly rely on expression of a non-native excision protein as in the Cre/lox system or outcrossing to wild-type strains(6). Outcrossing, though useful as a diagnostic tool, presents major limitations in the development of commercial *C. reinhardtii* strains. As industrial growth strategies become more refined, precise genetic backgrounds will be required for phenotype stability and reliable cultivation. Additionally, development of outdoor production strains may look to sterile backgrounds to prevent spontaneous breeding with naturally occurring *C. reinhardtii* strains(7).

Gene selection strategies which do not require breeding as a means of marker removal are therefore more desired for commercial strain development.

Current antibiotic-free selection strategies in *C. reinhardtii* rely on restoration of native metabolism in mutant strains. There are quite a few mutant strains that have been used for this strategy, including those with impaired ARG7, NIT1, NIC7, and THI10(8,9). Though this is a reliable means of gene introduction, it requires that the background strain contains a mutation for one of these genes, which limits the scope of potential production strains and prevents multiple rounds of gene introduction in a single strain.

The EPA recently approved a pilot study for outdoor cultivation of genetically engineered *C. reinhardtii* expressing GFP(10). Researchers determined that cultivation of the GFP-containing strain did not affect the growth of native algae strains. Presented here is a FACS-based selection strategy utilizing a benign GFP isoform which allows for isolation of genetic transformants without the need for antibiotics or specific background strains.

Results

The mEmerald gene was codon optimized for nuclear expression in *C. reinhardtii*, and cloned into the pBR9 Ble-2a expression vector developed by Rasala et al. in 2011(11). A schematic of the expression vector used is shown in Figure 1. The cassette is driven by the AR1 promoter, a well described regulatory element comprised of the RBCS2 intron and the HSP70a promoter(12).

Figure 2.1. Nuclear expression vector diagram. The nuclear expression of mEmerald is driven by the hybrid ar1 promoter. The vector contains a Ble gene, which provides resistance to zeocin, that is transcriptionally fused to mEmerald by the 2A FMDV peptide.

The Ble gene provides resistance to zeocin, which was used for downstream verification of gene expression and to estimate transformation efficiency(13). The mEmerald gene is an improved form of GFP which has been reported to have increased fluorescence. By using the 2a peptide, the

resistance gene and the fluorescent protein are transcriptionally fused, and upon translation they separate into two unique protein products(14).

This vector was transformed into the cc1690 *C. reinhardtii* background strain via electroporation. Cc1690 which was grown to $1x10^8$ cells per mL were incubated with 1µg of linearized pBR9, and electroporated using Thermo Fisher MAX Efficiency buffer. Cells were

Table 2.1. Transformation and gating efficiency. Cells transformed with the mEmerald expression vector and those which were electroporated with no vector were measured for their transformation efficiency by counting zeocin resistant colonies from an aliquot of the electroporated cells. The rate of survival corresponds to the transformation rate. The majority of the electroporated cells were analyzed by flow cytometer. The total number of cells investigated by flow cytometry for the mEmerald transformation and the negative control wild-type strain is shown, as well as the selectivity of the gating strategy for mEmerald recognition. The gating strategy yields only 4 false positives in the negative control from 20,000,000 investigated events.

	Transformation rate based on zeocin plates	**# of cells investigated by flow cytometry**	**# individuals in mEmerald gate**
No DNA control	0	$2.0x10^7$	4
mEmerald	$3x10^{-6}$	$2.3x10^7$	164

allowed to recover in 10mL of TAP media for 16 hours, after which $1x10^7$ of the shocked cells were plated onto a TAP-Agar plate containing 15µg/mL zeocin. The remaining $9x10^7$ cells were diluted 1:10 into TAP media and maintained on a shaker flask for 24 hours. The pool of unselected transformants was analyzed by flow cytometry and, using the gating hierarchy shown in Figure 2, GFP-positive individuals were identified. Cells which pass through the provided gating strategy are considered to be GFP positive. To prove the selectivity of our gating strategy, cc1690 was electroporated without addition of the pBR9 vector. When compared, the pool transformed with pBR9-Emerald showed an enrichment of individuals in the final GFP-positive gate of approximately 40-fold (Table 1), and only 4 false-positives were identified in the negative control out of 20,000,000 events investigated. The transformation efficiency as determined by zeocin resistance is approximately $3x10^{-6}$. According to GFP data, the transformation efficiency was

approximately $7x10^{-6}$. This discrepancy can be attributed to both the fact that low expression levels

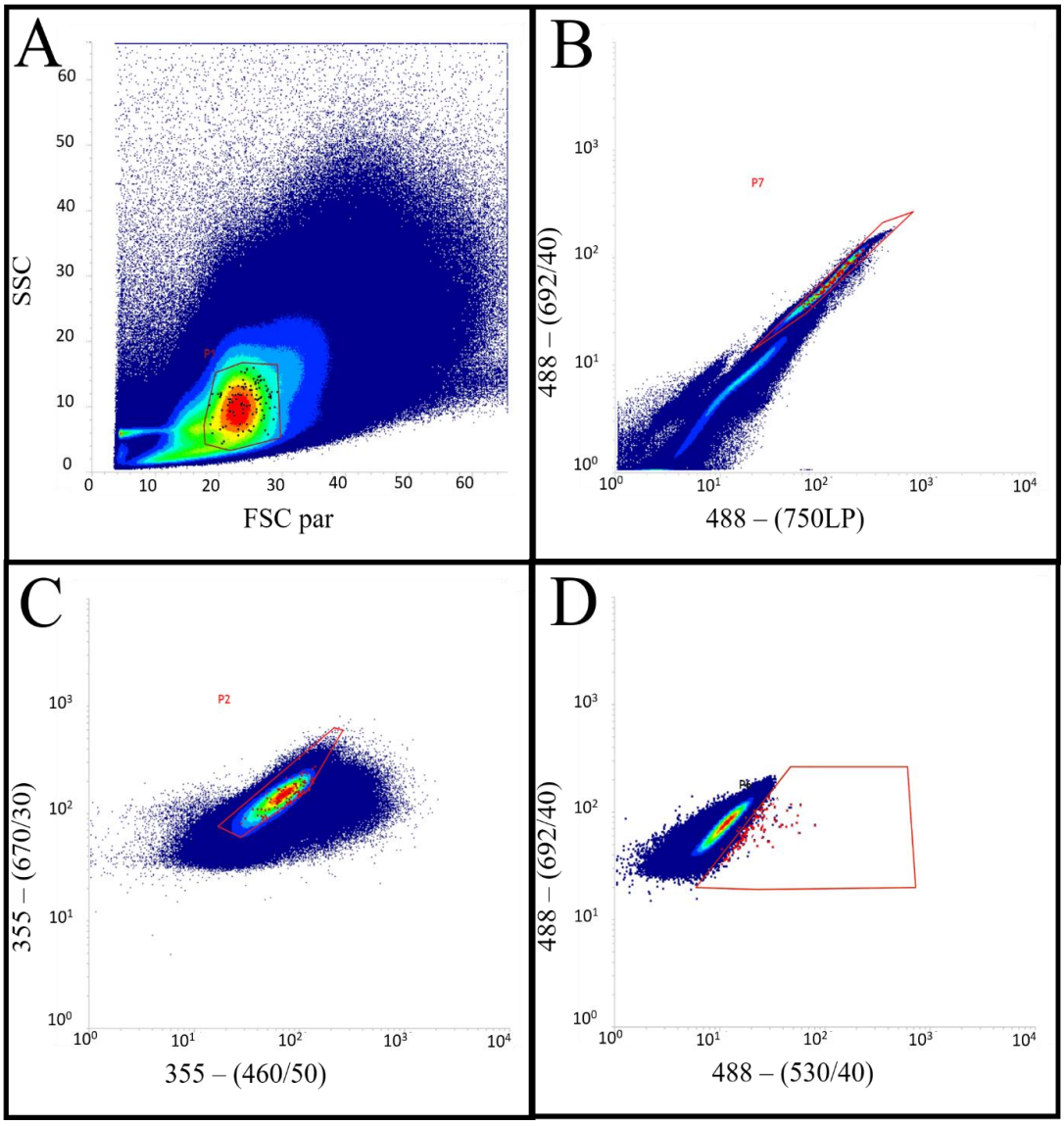

Figure 2.2. Gating strategy for GFP identification. Each pane depicts an inclusion gate to narrow down the electroporated culture to select for *C. reinhardtii* cells which contain mEmerald within which the larger dots (black in A, red in B, C, and D) correspond to the individual 164 cells which were sorted for the experiment. As the gating strategy progresses from A to D, cells excluded at each step are removed from the subsequent analysis. Cells outside of the gate in A are not plotted in B. Cells outside of the gate in B are not plotted in C. etc. The gate in D selects for the mEmerald positive population which was sorted.

of the Ble gene do not result in resistance to zeocin, which artificially lowers the perceived transformation rate by antibiotic resistance, as well as the imperfect gating strategy which results in the false characterization of some negative individuals as mEmerald positive.

Cells which fell within the Emerald-positive gate from the unselected pBR9-Emerald transformed pool were sorted individually into 96-well plates containing TAP media. Using a technique known as Index Sorting, fluorescence properties of the sorted individuals have been plotted as larger red or black dots on the flow cytometry density plots shown in Figure 2. Due to the rarity of the phenotype, only 164 individuals were sorted. The sorted individuals were left to

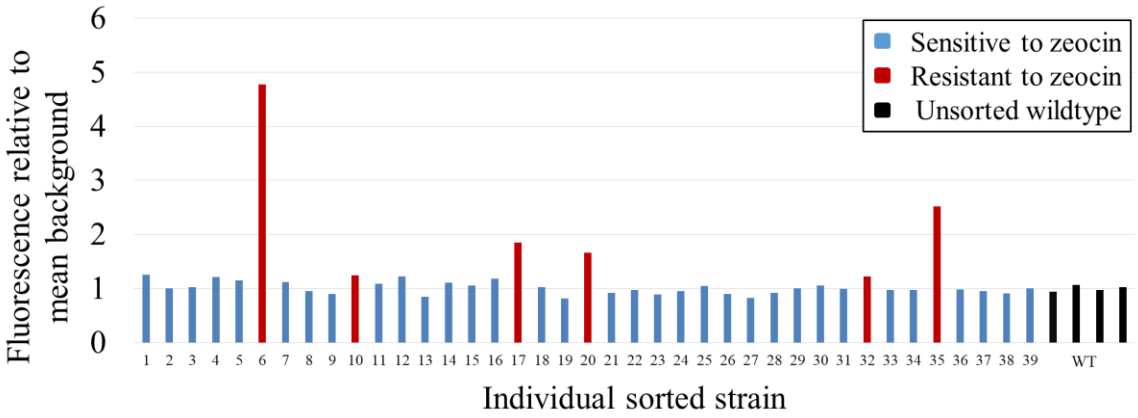

Figure 2.3. **Sorted individuals investigated by fluorescent plate reader and antibiotic.** Each strain which survived sorting was grown in a 500µL TAP culture and measured for mEmerald fluorescence by plate reader. The fluorescence of each strain is reported as a relative expression level when compared to the fluorescence of the unsorted parent strain. The average of four biological replicates of the unsorted parent is set as 1. In addition, each strain was plated onto TAP plates containing 15µg/mL zeocin. Strains which were resistant to zeocin are shown in red, while those that were sensitive to zeocin are shown in blue.

recover in the dark for 48 hours, and were then shifted to 100µE light to grow for one week. Of the 164 individuals sorted, 39 strains survived the sorting process. These 39 strains were grown to saturation in 2mL cultures and analyzed by fluorescent plate reader to determine if they were legitimate hits (Figure 3).The 39 strains were also plated onto 15 µg/mL zeocin plates. The strains which were found to be resistant to zeocin are shown in red on Figure 3. From these data, it is clear that we successfully sorted individuals which have integrated the transformation vector.

The most exceptional GFP expressor as determined by plate reader was confirmed to have GFP expression by fluorescence microscopy, as shown in Figure 4. The phenotype seen is indicative of nuclear-expressed cytoplasmic fluorescent protein, as it is predominantly excluded from the chloroplast.

Discussion

Here we demonstrate that expression of a fluorescent protein can be effectively used as a selection marker to identify successful nuclear transformation. By sorting for GFP accumulation, we have effectively enriched the population of transformed individuals from approximately $1x10^{-6}$ to $1x10^{-1}$, a much more manageable concentration for downstream analysis. The presence

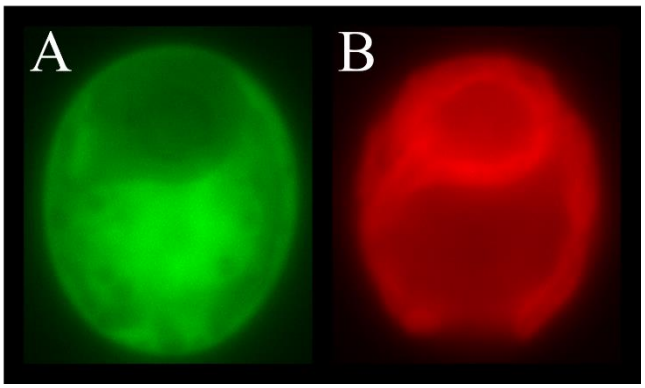

Figure 2.4. Fluorescence microscopy image of sorted strain. The most highly mEmerald fluorescent strain as determined by plate reader assay was visualized by fluorescent microscope. (A) shows the cell when analyzed by a GFP filter set, and exhibits cytoplasmic accumulation of mEmerald. (B) visualizes chlorophyll in the cell and provides a clear outline of the chloroplast.

of the GFP marker makes it simple to screen sorted individuals for expression of the gene of interest, in this case the Ble gene, and relative fluorescence levels can be used for rapid phenotype verification. Of the four individuals which showed statistically significant increases in GFP fluorescence, all four showed resistance to zeocin. In this instance, the Ble gene provides a simple phenotype to detect expression of our gene of interest, but enrichment of the transformant pool by

10^5-fold by GFP sorting makes it realistic to identify expressing strains by more complicated strategies, such as immunoblotting, which would otherwise be too labor-intensive.

Based on the evidence presented in Rasala et al. 2012(11), flow cytometery is capable of differentiating many different fluorescing strains (mCherry, mCerulean, GFP and tdTomato) from each other and their wild-type background. Presumably a variety of fluorophores could be used to identify successful transformations which would facilitate gene stacking in a single organism. Based on the rarity of successful transformants identified in this study, gene stacking would likely need to occur in successive rounds of transformation and sorting, as opposed to cotransformation and simultaneous gating for dual-fluorescing individuals.

Ultimately, this tool will be incredibly useful for specific applications in the commercial development of algal strains, and provides a strategy for creating nuclear transformants in an unbiased background strain with no requirement for antibiotic selection genes or exogenous gene-excision proteins.

Methods and Materials

Plasmid Construction

The vector used in this study is based on the pBR9 ble-2a vector described in Rasala et al. 2013. The mEmerald gene was codon optimized using the *C. reinhardtii* codon usage table provided by the Kazusa DNA Research Institute. The mEmerald gene was cloned into the pBR9 backbone by simple XhoI/BamHI restriction-based ligation.

Transformation protocol

The nuclear transformation protocol was adapted from the protocol provided with the GeneArt MAX Efficiency Transformation Reagent (Thermo #A24229). *C. reinhardtii* strain cc-1690 was grown on a shaker in 250mL of TAP media under 100µE to $5x10^6$ cells mL^{-1}. This cell count is indicative of log-phase growth. The cells were pelleted by centrifugation and the growth

media was decanted. The cells were resuspended in GeneArt MAX Efficiency Transformation Reagent to a concentration of $3x10^8$ cells mL^{-1} to wash any remaining growth media. The cells were pelleted and the transformation reagent was decanted. Again, the culture was resuspended in GeneArt MAX Efficiency Transformation Reagent to a cell concentration of $3x10^8$ cells mL^{-1}. 300µL of the resulting solution was transferred to a 1.5mL microcentrifuge tube along with 1µg of linearized plasmid and placed on ice for 5 minutes. The mixture was transferred to a chilled electroporation cuvette (Thermo #P46050). The cuvette was placed in the cuvette holder of a BioRad Gene Pulser Xcell and pulsed with the following settings: Voltage 500V, Capacity 50µF and Resistance 800Ω. The cuvette was left at room temperature for 10 minutes, at which point the 300µL solution was transferred to 10mL of TAP media in a 50mL centrifuge tube. After 16 hours, 10^7 cells were plated onto TAP agar plates with 15µg mL^{-1} zeocin to determine transformation efficiency. The remainder of the 10mL culture was added to 50mL of TAP solution in a small flask and placed on a shaker in 100µE of light to grow for an additional day prior to flow cytometer analysis.

Flow cytometer setup

Cells were analyzed and sorted by a 5-laser Becton Dickinson Influx flow cytometer running BD FACS Sortware. The instrument was equipped with a 70µm nozzle and operated on BD FACSFlow sheath fluid at a constant sheath pressure of 33.0psi. Cells were sorted into a 96-well clear polystyrene microplate. Each well contained 100uL of autoclaved TAP media, and the reservoir to prevent evaporation was filled with autoclaved TAP media. The post-sort plates were incubated at room temperature in the dark for 24 hours and were then moved under 75uE of light for ten days.

References

1. Scranton MA, Ostrand JT, Fields FJ, Mayfield SP. Chlamydomonas as a model for biofuels and bio-products production. Plant J Cell Mol Biol. 2015 May;82(3):523–31.

2. Press Release - GRAS [Internet]. Triton. [cited 2018 May 22]. Available from: https://www.tritonai.com/press-releases-1/

3. Bennett PM. An assessment of the risks associated with the use of antibiotic resistance genes in genetically modified plants: report of the Working Party of the British Society for Antimicrobial Chemotherapy. J Antimicrob Chemother. 2004 Feb 12;53(3):418–31.

4. Toomey N, Monaghan Á, Fanning S, Bolton D. Transfer of Antibiotic Resistance Marker Genes between Lactic Acid Bacteria in Model Rumen and Plant Environments. Appl Environ Microbiol. 2009 May 15;75(10):3146–52.

5. Day A, Goldschmidt-Clermont M. The chloroplast transformation toolbox: selectable markers and marker removal. Plant Biotechnol J. 2011 Jun 1;9(5):540–53.

6. Kasai Y, Harayama S. Construction of Marker-Free Transgenic Strains of Chlamydomonas reinhardtii Using a Cre/loxP-Mediated Recombinase System. PLoS ONE [Internet]. 2016 Aug 26 [cited 2018 May 22];11(8). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5001723/

7. Daniell H. Molecular strategies for gene containment in transgenic crops. Nat Biotechnol. 2002 Jun;20(6):581–6.

8. Grossman AR, Harris EE, Hauser C, Lefebvre PA, Martinez D, Rokhsar D. Chlamydomonas reinhardtii at the Crossroads of Genomics. Eukaryot Cell. 2003 Dec 1;2(6):1137–50.

9. Meslet-Cladière L, Vallon O. Novel Shuttle Markers for Nuclear Transformation of the Green Alga Chlamydomonas reinhardtii ▿ . Eukaryot Cell. 2011 Dec;10(12):1670–8.

10. Szyjka SJ, Mandal S, Schoepp NG, Tyler BM, Yohn CB, Poon YS. Evaluation of phenotype stability and ecological risk of a genetically engineered alga in open pond production. Algal Res. 2017 Jun;24:378–86.

11. Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP. Expanding the spectral palette of fluorescent proteins for the green microalga Chlamydomonas reinhardtii. Plant J Cell Mol Biol. 2013 May;74(4):545–56.

12. Schroda M, Blöcker D, Beck CF. The HSP70A promoter as a tool for the improved expression of transgenes in Chlamydomonas. Plant J Cell Mol Biol. 2000 Jan;21(2):121–31.

13. Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP. Robust Expression and Secretion of Xylanase1 in Chlamydomonas reinhardtii by Fusion to a Selection Gene and Processing with the FMDV 2A Peptide. Stiller JW, editor. PLoS ONE. 2012 Aug 24;7(8):e43349.

14. Ryan MD, King AMQ, Thomas GP. Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J Gen Virol. 1991 Nov 1;72(11):2727–32.

Acknowledgements

Chapter 2, in full, is a reprint of material in preparation for submission for publication. Ostrand, Joseph; Mayfield, Stephen. The dissertation author is the primary investigator and author of this paper.

CHAPTER 3:

FED-BATCH MIXOTROPHIC CULTIVATION OF *CHLAMYDOMONAS REINHARDTII* FOR HIGH-DENSITY CULTURES

Algal Research 33 (2018) 109–117

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Fed-batch mixotrophic cultivation of *Chlamydomonas reinhardtii* for high-density cultures

Francis J. Fields*, Joseph T. Ostrand, Stephen P. Mayfield

The California Center for Algae Biotechnology, Department of Biological Sciences, University of California San Diego, 9500 Gilman Dr., La Jolla, California 92093, United States

ARTICLE INFO

Keywords:
Chlamydomonas reinhardtii
Heterotrophic
Fed-batch
High density
Recombinant protein production
Primary metabolites

ABSTRACT

The green algae *Chlamydomonas reinhardtii* is a model organism that has been proposed as a potential production platform for high-value recombinant products, but to date, there has been little development of a commercially viable cultivation process. Here we demonstrate a novel mixotrophic fed-batch cultivation strategy in which average biomass density, productivity, and total amount of recombinant GFP significantly increases. Systematic feeding of acetic acid and nutrients into a bioreactor resulted in culture densities increasing from an initial 0.45 ± 0.03 to 23.69 ± 0.5 g L^{-1} AFDW after 168 h at an average productivity of 181.01 ± 8.8 mg L^{-1} h^{-1}, a 10-fold increase in comparison to traditional batch cultures. GFP expression was low under both conditions, but fed-batch cultivation resulted in a 2.5-fold increase in total GFP upon culture termination. The effect of fed-batch cultivation on lipid composition and primary metabolites was investigated and elevated levels of the osmoregulatory molecules proline and glycerol were found, suggesting that salt accumulation may have increased over time in fed-batch cultures and eventually limited growth.

1. Introduction

Chlamydomonas reinhardtii is a eukaryotic single-celled soil-dwelling green microalgae (Chlorophyta). For over fifty years it has been a premier model organism used to understand the biology, physiology, and genetics of algae, as well as a model organism to understand photosynthesis in both plants and algae [1,2]. One fundamental reason this species became a model organism is its ability to heterotrophically metabolize acetate, allowing for growth in complete darkness and the isolation of non-photosynthetic mutants. The cumulative research into this organism has resulted in a significant understanding of the algal cell cycle, metabolism, genomics, and genetic circuits [3]; making *C. reinhardtii* potentially the best studied and well characterized species of algae. Due to this extensive research, *C. reinhardtii* has been targeted for use in biotechnology, most frequently proposed as a potential production platform for high-value recombinant products for medical, nutritional, and industrial use [4,5]. Additionally, a recently published toxicological evaluation demonstrated that the consumption of *C. reinhardtii* biomass presents no health concerns, thus opening the door for *C. reinhardtii* biomass to be used as a nutritional additive in feeds and foods [6].

The molecular tools developed for *C. reinhardtii* have allowed for facile transformation of both the chloroplast and nuclear genomes [5], and to date, over one hundred high-value recombinant proteins have been expressed in *C. reinhardtii*, including multiple therapeutic proteins, such as antibodies and immunotoxins, vaccines, industrial enzymes, animal feeds, and nutritional supplements [4,7–11]. While the techniques for nuclear transformation are well established in *C. reinhardtii*, the lack of directed homologous recombination for transgene integration results recombinant protein expression being strongly influenced by positional effects in the genome, leading to low and inconsistent protein accumulation levels from nuclear transformants [5,10]. Unlike gene integration in the nuclear genome, chloroplast transformation proceeds by homologous recombination. Coupled with the sophisticated protein-folding machinery found in the chloroplast, the plastid genome is a more practical expression platform for transgene integration and recombinant protein expression. Even with these advantages in the chloroplast, expression remains low and available genetic constructs are light-regulated, resulting in variable expression as light penetration into the culture changes [12]. The accumulation of recombinant protein expressed in the *C. reinhardtii* chloroplast rarely exceeds 5% total protein and is more commonly < 1% [13]. In comparison, the expression of intracellular recombinant proteins in bacteria, yeasts, and mammalian cells can be much higher, reaching above 10% of total protein content in some cases. However, these platforms can face difficulties producing many sought-after high-value proteins, often due to

* Corresponding author.
E-mail address: ffields@ucsd.edu (F.J. Fields).

https://doi.org/10.1016/j.algal.2018.05.006
Received 20 March 2017; Received in revised form 8 May 2018; Accepted 10 May 2018

Table 1
Maximum observed biomass density and productivity of *C. reinhardtii* strains grown mixotrophically in TAP media.

CC-strain	Biomass density ($DW\,g\,L^{-1}$)	Cell density ($cells \times 10^7\,mL^{-1}$)	Productivity ($DW\,g\,L^{-1}\,d^{-1}$)
125	1.76	1.9	0.25
1009	1.74	1.75	0.24
1373	1.8	2.58	0.25
1690	1.76	2.21	0.25
2290	1.78	1.07	0.27
2343	1.82	1.88	0.27
2344	1.9	3.44	0.24
2931	1.64	1.62	0.27
2935	1.84	4.62	0.28
2937	2.08	4.5	0.29
2938	1.64	3.24	0.26
4147	0.33	0.34	0.1

challenges involving protein folding, post-translational modification, or production costs [14].

A practical consequence of low expression in *C. reinhardtii* is an inability to produce the minimum quantities of heterologous products needed for characterization of post-translational modifications, bioactivity assays, or in vivo experiments (e.g., animal trials). In order to generate meaningful quantities of recombinant products and make *C. reinhardtii* a realistic production platform, there are two main developments that must occur: (1) development of robust and controllable gene expression systems that are comparable to alternative production platforms and (2) development of cultivation processes that allow for rapid growth and maximize biomass concentration.

Research to improve transgene integration and expression in *C. reinhardtii* is ongoing and new technologies, like synthetic promoters and CRISPR, show great promise in this area [15,16]. Overlapping these technologies with classic host improvement strategies, like new strain discovery, breeding, and mutagenesis, will potentially lead to the necessary advancements in gene expression and ultimately product production.

However, improving gene expression and protein accumulation is not sufficient in itself to allow *C. reinhardtii* to become a commercially viable production platform. To be cost competitive with other microbial systems, greatly improved biomass production processes must be developed to enable the rapid generation of large amounts of algal biomass. Today, large volume cultures of algae are commonly grown under phototrophic conditions in ponds and specialized photobioreactors [17,18]. However, low biomass productivity of *C. reinhardtii* in open ponds [19], coupled with regulatory obstacles faced with producing genetically modified strains outdoors [20], has directed the cultivation of *C. reinhardtii* into closed systems operating under either mixotrophic or heterotrophic conditions. These systems (i.e., bioreactors and fermenters) allow for highly controlled commercial-scale alternatives and have been successfully implemented in the cultivation of numerous microbial hosts [14,21,22]. However, these systems come at a much higher capital equipment cost, and can require extensive optimization.

Most commonly, *C. reinhardtii* is cultivated in batches of media containing acetate as a carbon source [23], resulting in dry cell mass concentrations of $1–2\,g\,L^{-1}$. In the most notable attempt to achieve high-density *C. reinhardtii* cultures, Chen & Johns [24] conducted comparative studies using sodium acetate as a carbon source in batch, fed-batch, chemostat, and perfusion culture. Using a combination of fed-batch and hollow-fiber cell-recycle systems, biomass densities of $9\,g\,L^{-1}$ dry weight were achieved. They theorized the culture growth became limited over time due to buildup of sodium ions as a consequence of using sodium acetate as the primary carbon source in the feed [25].

In the present study, we present an alternative method for generating high-density cultures of *C. reinhardtii* by using fed-batch mixotrophic growth and examine the effects it has on growth, biomass composition, and accumulation of a reporter recombinant protein. A fed-batch cultivation strategy was chosen in order to maximize final density of biomass and recombinant protein so as to simplify downstream processing. A strain of *C. reinhardtii* was engineered to produce a recombinant green fluorescent protein (GFP) in the chloroplast, used as a general model for recombinant protein production, and cultured under both batch and fed-batch conditions, in which the latter was operated as a pH-stat with a concentrated feed of acetic acid and nutrients pumped into the bioreactor. The primary goals of this study were to provide a cultivation strategy that would allow the generation of increased quantities of algal biomass and recombinant protein and to advance the development of *C. reinhardtii* towards cultivation in scalable industrial heterotrophic bioreactors.

2. Materials and methods

2.1. Strain selection

A preliminary study was conducted to characterize maximum biomass density, cell density, and productivity of our laboratory's *C. reinhardtii* collection of strains obtained from www.chlamycollection.org (Table 1). Strains were grown in Erlenmeyer flasks containing 50 mL of TAP media on a shaker under constant light ($75\,\mu Mol\,m^{-2}\,s^{-1}$) and sampled every day for one week. Among the strains included in this analysis, the strain CC-2937, a wild isolate from Quebec, Canada [26], had the highest maximum biomass density and productivity and was therefore chosen for further experimentation.

2.2. Vector construction and transformation

A construct (Fig. 1) was devised for chloroplast expression of a codon-optimized GFP [27] using the *C. reinhardtii psbD* promoter and 3′ and 5′ UTR regulatory elements [28]. This cassette was cloned into the 3HB vector [13], which contains 5′ and 3′ homologous chloroplast genomic sequence that confers transformation into the 3HB silent site. The *aphA-6* gene with the *atpA* promoter and 5′UTR and *rbcL* 3′UTR was subcloned into the upstream *BamHI* site. CC-2937 cells were cultured as described in Rasala et al. [29] and the vector was transformed via particle bombardment as described in Rasala et al. [28]. Transformants were selected by plating on TAP-agar plates containing $150\,mg\,L^{-1}$ kanamycin. Strains that showed resistance to kanamycin were put through rounds of selection on kanamycin, and PCR screened for homoplasmy as described in Rasala et al. [29] using the forward primer CGTCCACTAAAATTTATTTACCCGAAGGGG and the reverse primer GTTAAGGCTAGCTGCTAAGTCTTCTTTTCGC which bind in the homology sequence of *psbH*. A homoplasmic colony was eventually identified after several weeks of growth on kanamycin-containing TAP-agar plates. This transformed strain was maintained and used in subsequent experiments.

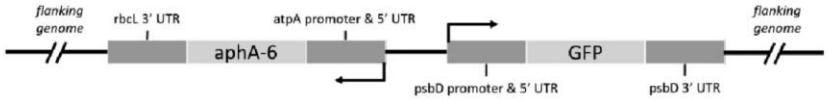

Fig. 1. Construct design for expression of kanamycin resistance (*aphA-6*) and of green fluorescent protein (*GFP*) in the *C. reinhardtii* chloroplast.

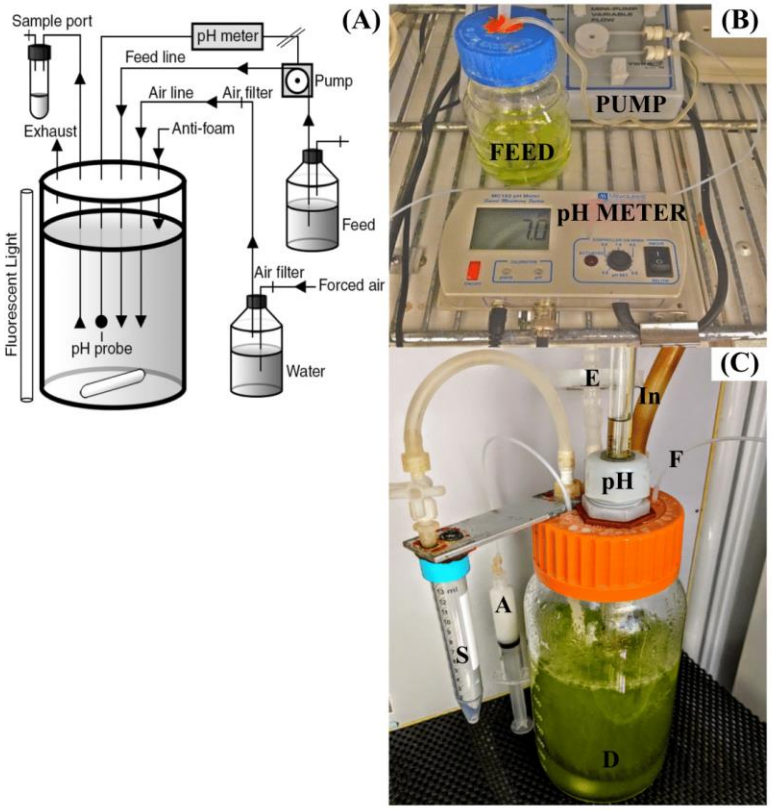

Fig. 2. Bioreactor design depicting general setup and the pH-driven feeding mechanism used for fed-batch cultivation (A). Images on the right show the final setup of one of the reactors including the feeding system (B) and labeled bioreactor components (C): In = forced air; *D* = air diffusion; *F* = feed line; *pH* = pH probe; *A* = antifoam; *S* = sample port; *E* = exhaust. Both batch and fed-batch cultures were cultivated in these bioreactors, but feeding was disabled during batch cultivation.

2.3. Growth media and feed formulation

A modified TAP-media was used for growing inoculum and as the starting media for all experiments. This media was prepared by dissolving per liter: 400 mg NH_4Cl, 143.4 mg K_2HPO_4, 72.6 mg KH_2PO_4, 100 mg $MgSO_4 \cdot 7H_2O$, 50 mg $CaCl_2 \cdot 2H_2O$, 2.42 g Tris, 1 mL Kropat's trace metals [30], and 1.25 mL glacial acetic acid into Milli-Q filtered water; the resulting pH of the media was 6.8. Media was autoclaved at 121 °C for 60 min before use.

A concentrated feed mimicking the basal media was prepared by mixing 290 mL Milli-Q filtered water with 333 mL glacial acetic acid and 250 mL trace metal solution into which 28.87 g NH_4Cl, 10.35 g K_2HPO_4, 5.24 g KH_2PO_4, 7.22 g $MgSO_4 \cdot 7H_2O$, 3.61 g $CaCl_2 \cdot 2H_2O$, and 174.70 g Tris were dissolved; the resulting pH was 4.0. The feed was filter sterilized with a 0.22-μm vacuum cup filter and stored at room temperature. Prior to each fed-batch experiment, 200 mL of feed was transferred into small sterile glass bottles that connected to the bioreactors.

2.4. Bioreactor design

A low-cost bioreactor was designed and built for axenic heterotrophic cultivation of *C. reinhardtii* with the capability of operating under fed-batch conditions (Fig. 2). Two 1-L glass bottles (Product #1397-1L; Corning) were each custom-fitted with a forced air diffuser, stir bar, feed line, sample port, antifoam injector, exhaust port, and pH probe (EasyFerm Plus 325; Hamilton Company). Each bioreactor was placed on a stir plate (StableTemp Top Stirrer; Cole-Parmer) set at 700-rpm inside an incubator (Model I-35LL; Percival Scientific) maintained at 25 °C under constant fluorescent light (150 $\mu Mol\ m^{-2}\ s^{-1}$). Each pH probe was connected to a pH meter (MC 122 pH Controller; Milwaukee Instruments) that controlled a peristaltic pump (Mini Pump Variable Flow; VWR) to regulate the flow of feed into the culture for fed-batch experiments (Fig. 2B). The pH meter was set to turn the pump "ON" when the pH rose above 7.0 and turn the pump "OFF" once the pH fell below 7.0. Batch experiments had no feed additions. Small amounts of antifoam (SE-15; Sigma-Aldrich) were injected manually with a syringe as needed.

2.5. Inoculation, sampling, and biomass quantification

For each experiment, a bioreactor was filled with 550 mL of filtered growth media, covered in foil, and autoclaved at 121 °C for 60 min. Bioreactors were allowed to cool in a UV-sterilized biosafety cabinet before inoculum was added. To prepare inoculum, a flask containing 50 mL of growth media was inoculated with the transformed strain of CC-2937 from a TAP-agar plate and grown to saturation on a shaker under constant light (150 μMol $m^{-2} s^{-1}$). The culture was then diluted 1:10 and allowed to grow for 3 days under the same conditions; 250 mL were then added to each bioreactor as inoculum, bringing the initial volume to 800 mL. Both batch and fed-batch experiments were conducted in triplicate and allowed to grow for 8 days.

A 10-mL liquid sample was drawn daily from sampling ports and used to directly quantitate cell and biomass density. A hemocytometer was loaded with a 10-μL sample and cells were counted under 40× magnification on a light microscope. Ash free dry weight (AFDW) biomass was determined according to methods described by Zhu & Lee [31], however, only 2 mL of culture were filtered and samples washed with 100 mL of Milli-Q filtered water. A one-tailed unpaired *t*-test was used to determine any significant increases in average biomass density as a result of fed-batch cultivation.

2.6. Analysis of lipid composition and primary metabolites

A 50-mL sample was taken at 72 h and 168 h from the last replica of the batch and fed-batch experiments. The sample was centrifuged, supernatant decanted, and the biomass pellet was flash frozen then lyophilized for 48 h to remove all water content. The samples were sent on dry ice to the NIH: West Coast Metabolomics Center at the University of California, Davis (http://metabolomics.ucdavis.edu) for lipidomics and primary metabolite analysis. The samples were fractionated and processed separately for each of these analyses. Lipidomics were performed using a charged surface hybrid column electrospray (CSH-ESI) to identify both positively and negatively charged molecules by quadruple time of flight tandem mass spectrometry (QTOF-MS/MS) [32] (Table S1). Primary metabolites were identified using an automated liner exchange-cold injection system gas chromatography time of flight mass spectrometer (ALEX-CIS GCTOF MS) [33] (Table S2).

2.7. GFP quantification

GFP was quantified at 48, 96, 144, and 168 h for each replica. A 1-mL sample aliquot was centrifuged and the supernatant replaced with an equal volume of TBST buffer. The aliquot was then sonicated at 20% amplitude for 15 s while on ice to lyse cells and solubilize proteins. A Lowry assay was used to determine the concentration of soluble protein for each sample. 5 μg of soluble protein and three purified GFP standards (0.15 ng, 0.3 ng, and 0.6 ng) were loaded and separated via SDS-PAGE gel. GFP was visualized via Western blot using an anti-GFP antibody. Densitometry was performed on the Western blots with ImageJ 1.50i software (https://imagej.nih.gov/ij/index.html) following methods described by Miller [34]. A regression of the GFP standards versus their ImageJ density was plotted ($R^2 > 0.97$; Table S3) and the resulting trend line was used to calculate the GFP quantity in samples taken from the bioreactors. A one-tailed unpaired *t*-test was used to determine any significant increase in average GFP content at each time point.

3. Results and discussion

3.1. Fed-batch growth results

During fed-batch cultivation, the transgenic strain of *C. reinhardtii* CC-2937 grew from 0.45 ± 0.03 to 23.69 ± 0.5 g L^{-1} AFDW after 168 h at an average productivity of 181.01 ± 8.8 mg $L^{-1} h^{-1}$ (Fig. 3A). Viable cell concentrations increased from $8.23 \times 10^5 \pm 2.14 \times 10^5$ to $1.31 \times 10^8 \pm 1.45 \times 10^7$ cells mL^{-1} during this same time period (Fig. 3B). Maximum biomass density, cell concentration, and average productivity were all significantly higher under fed-batch conditions compared to batch growth ($P < 0.001$). The fed-batch cultures received an average of 133.33 ± 9.02 mL of feed after 168 h (Fig. 4A), after which the cultures began to crash rapidly, resulting in 20% of cells dying by 192 h with corresponding reduction in feeding rates. The pH of all fed-

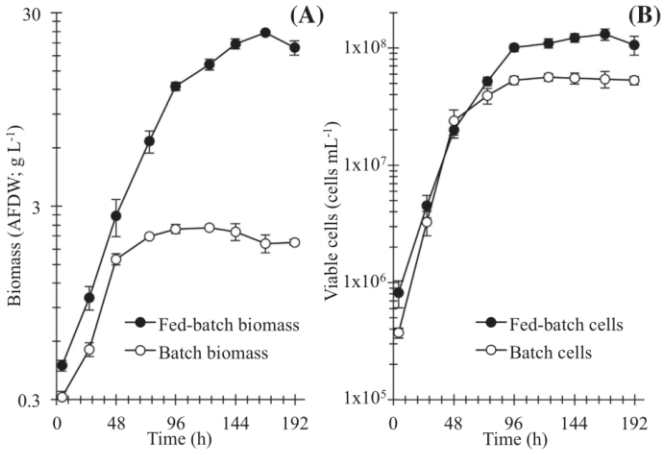

Fig. 3. Average biomass density (A) and concentration of viable cells (B) of batch (open points) and fed-batch (filled points) cultures. Error bars indicate standard error of the mean ($n = 3$).

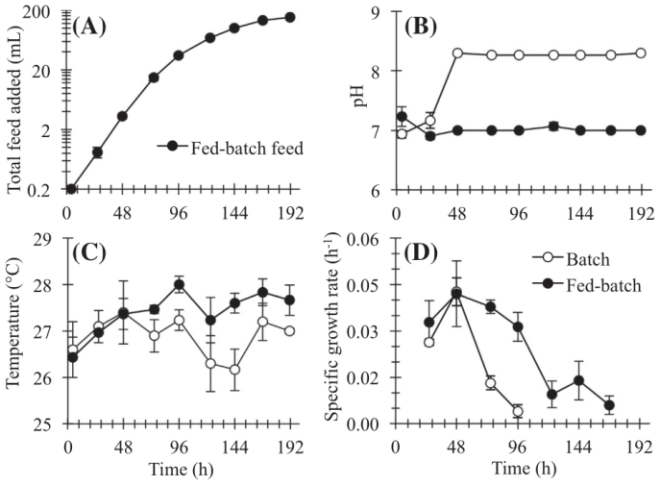

Fig. 4. Averages of the total feed added to the fed-batch cultures (A), pH (B), temperature (C), and specific growth rate (D) of batch (open points) and fed-batch (filled points) cultures. Error bars indicate standard error of the mean (n = 3).

batch cultures remained between 6.9 and 7.1 (Fig. 4B) as a result of the automated pH-controlled feeding mechanism (Fig. 2B). Average culture temperature trended upwards throughout the experiments, ranging from a low of 26.43 ± 0.43 to a high of 28.00 ± 0.17 °C during the experiments (Fig. 4C).

3.2. Batch growth results

During batch cultivation, cultures grew from 0.31 ± 0.02 to 2.29 ± 0.12 g L^{-1} AFDW after 96 h at an average productivity of 22.62 ± 3.3 mg L^{-1} h^{-1} (Fig. 3A). Viable cell concentrations increased from $3.73 \times 10^5 \pm 3.71 \times 10^4$ to $5.31 \times 10^7 \pm 4.37 \times 10^6$ cells mL^{-1} during this same time period (Fig. 3B). These cultures became saturated after 123 h; reaching a peak density of 2.33 ± 0.09 g L^{-1} AFDW biomass and $5.63 \times 10^7 \pm 4.03 \times 10^6$ cells mL^{-1} and then remained in a stationary phase for the rest of the experiment. In each replica, the pH of the cultures rose rapidly from 7.0 to 8.3 within 48 h, where it remained for the duration of the experiments (Fig. 4B). Culture temperatures ranged from a low of 26.17 ± 0.45 to a high of 27.40 ± 0.68 °C during the experiments (Fig. 4C).

3.3. Comparisons of specific growth rates

The average specific growth rates during the first 96 h were 0.02 and 0.04 h^{-1} in batch and fed-batch cultures, respectively (Fig. 4D). The maximum specific growth rates (μ_{max}) observed were 0.049 and 0.063 h^{-1} in batch and fed-batch, respectively, both occurring between 24 and 48 h. This is similar to other examinations of *C. reinhardtii*, in which μ_{max} has been reported as 0.035 h^{-1} under heterotrophic growth and 0.066 h^{-1} under mixotrophic growth [35]. In comparison, other species of microalgae grown heterotrophically at commercial scale, such as *Chlorella* and *Schizochytrium*, range from μ_{max} of 0.2 to 0.07 h^{-1}, respectively [21], suggesting that, with further optimization, fed-batch cultivation of *C. reinhardtii* has potential to contest certain production platforms.

3.4. Cell morphology and the relationship between biomass and cell concentration

Cell morphology in batch cultures was characteristic of *C. reinhardtii* cultures; all cells were approximately 10 µm in length and the flagella, pyrenoids, and eyespots were easily viewed with a light microscope (Fig. 5A). There is an unchanging linear trend during batch growth when comparing cell concentration to biomass (Fig. 5C).

In comparison, the cell morphology of fed-batch cultures looked identical to batch cultures for the first 96 h of growth. However, after this point, the cells began to undergo a morphological change; the cells became spherical and doubled in diameter, which equates to an approximate eight-fold increase in cell volume. This shift was accompanied by eventual arrest of cell division and loss of both the flagella and eyespots; pyrenoids were visible in few cells (Fig. 5B). This change can be further visualized in the comparison between cell concentration and biomass, which requires two separate trends in order to accurately explain the relationship (Fig. 5D). This transformation in cell size explains the greater percent increase observed between 96 and 168 h in biomass (89% increase) compared to cell concentration (30% increase) in fed-batch cultures (Fig. 3).

3.5. Lipid composition and primary metabolites

Lipid composition relative to total extracted lipids is displayed in Table 2. Betaine lipids comprised the largest fraction, specifically diacylglyceryl trimethylhomoserine (DGTS), which is the primary membrane lipid in *C. reinhardtii* [36]. There is a correlating decrease in DGTS and increase in triacylglycerols (TG) between 72 and 168 h in batch cultures. The shift in metabolism from membrane lipids to storage lipids, like TG, is commonly observed in algal cultures that are depleted of elements critical for growth, (e.g., nitrogen) [37]. The lack of change in lipid composition in fed-batch cultures between 72 and 168 h suggests that the nutrients are consistently replete within the media. The elevated amounts of nonpolar lipids (NPL) in fed-batch

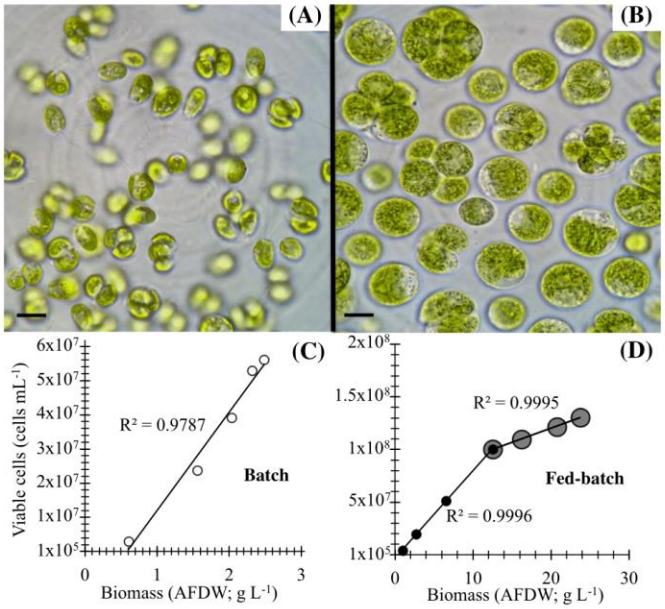

Fig. 5. Morphology of *C. reinhardtii* cultures observed during the first 96 h of growth in both batch and fed-batch culture (A) and the observed change in morphology in fed-batch cultures after 96 h (B); scale bars 10 μm. Comparison of trends between average viable cell concentration and biomass density in batch (C) and fed-batch cultures (D); two correlating trends are observed in the latter.

Table 2
Relative composition of lipids (% of total lipids).

	Batch		Fed-batch	
Lipid	72 h	168 h	72 h	168 h
Free fatty acids (FA)	14.86%	15.37%	13.01%	13.90%
Phospholipids (PL)				
Phosphatidylcholine (PC)	0.48%	0.43%	0.63%	0.63%
Lyso-PC (LPC)	0.03%	0.05%	0.02%	0.04%
Phosphatidyl ethanolamines (PE)	0.28%	0.16%	0.72%	1.56%
Lyso-PE (LPE)	0.02%	0.01%	0.07%	0.04%
Phosphatidylinositols (PI)	0.81%	0.87%	1.62%	1.02%
Phosphatidylglycerols (PG)	3.72%	1.85%	5.87%	3.90%
Total PL	5.34%	3.36%	8.94%	7.19%
Glycolipids (GL)				
Sulfoquinovosyl diacylglycerols (SQDG)	1.91%	1.00%	3.99%	4.08%
Betaine lipids (BL)				
Diacylglyceryl trimethylhomoserine (DGTS)	60.00%	39.15%	58.54%	52.99%
Lyso-DGTS	14.98%	13.76%	7.81%	13.50%
Total BL	74.98%	52.91%	66.35%	66.49%
Nonpolar lipids (NPL)				
Diacylglycerols (DG)	1.10%	1.02%	1.84%	1.52%
Triacylglycerols (TG)	1.80%	26.34%	5.85%	6.82%
Total NPL	2.89%	27.35%	7.69%	8.34%

Table 3
Relative fatty acid composition (% of total free fatty acids).

	Batch		Fed-batch	
Fatty acid	72 h	168 h	72 h	168 h
Saturated (SFA)				
Caprylic (8:0)	0.01%	0.01%	0.03%	0.02%
Myristic (14:0)	0.35%	0.40%	0.19%	0.27%
Margaric (17:0)	0.12%	0.17%	0.09%	0.13%
Arachidic (20:0)	0.11%	0.13%	0.16%	0.45%
Cerotic (26:0)	0.07%	0.08%	0.11%	0.19%
Total SFA	0.67%	0.80%	0.58%	1.06%
Monounsaturated (MUFA)				
Palmitoleic (16:1)	0.57%	0.72%	0.56%	0.97%
Heptadecenoic (17:1)	0.07%	0.08%	0.02%	0.03%
Oleic (18:1)	8.21%	5.51%	22.15%	18.56%
Eicosenoic (20:1)	0.09%	0.05%	0.59%	0.47%
Total MUFA	8.95%	6.36%	23.32%	20.03%
Polyunsaturated (PUFA)				
Linoleic (18:2)	10.94%	23.23%	21.56%	29.69%
Linolenic (18:3)	75.68%	67.01%	52.17%	46.33%
Stearidonic (18:4)	3.69%	2.55%	2.16%	2.69%
Eicosadienoic (20:2)	0.05%	0.03%	0.17%	0.17%
Dihomo-γ-linolenic (20:3)	0.03%	0.02%	0.04%	0.03%
Total PUFA	90.38%	92.84%	76.10%	78.92%

cultures would also suggest that a sufficient amount carbon is being provided to the culture.

Relative fatty acid composition as percent of total free fatty acids is provided in Table 3. Relative amounts of fatty acids (FA) were slightly higher in batch cultures than fed-batch (Table 2). Linolenic acid (18:3) was in the highest abundance in all samples and decreased slightly between time points. Oleic acid (18:1) was elevated in fed-batch and linoleic acid (18:2) more than doubled in batch cultures at 168 h. Saturated fatty acids were extremely low in both conditions.

A summary of primary metabolite composition as relative percent of the total soluble fraction within biomass is presented in Table 4; metabolites < 0.25% were omitted and can be found in supplemental material. In general, relative concentrations of metabolites within the

Table 4
Relative composition of primary metabolites (% of total soluble fraction).

		Batch		Fed-batch	
Metabolite	PubChem	72 h	168 h	72 h	168 h
α-Aminobutyric acid	6657	0.22%	0.27%	0.12%	0.11%
Alanine	5950	11.76%	6.67%	14.62%	12.31%
Aspartic acid	5960	0.77%	0.40%	0.11%	0.48%
β-Alanine	239	1.20%	1.16%	0.38%	0.31%
Ethanolamine	700	10.01%	6.97%	3.01%	4.26%
Glutamic acid	33032	1.00%	0.76%	0.66%	1.76%
Glycine	750	0.56%	0.28%	0.25%	0.27%
Isoleucine	6306	2.62%	2.03%	0.95%	0.67%
Leucine	6106	5.65%	4.32%	1.87%	1.42%
Oxoproline	7405	1.78%	1.28%	0.61%	0.73%
Phenylalanine	6140	0.84%	0.79%	0.24%	0.17%
Proline	145742	4.45%	4.55%	36.06%	33.40%
Putrescine	1045	0.72%	0.19%	1.24%	1.08%
Serine	5951	1.34%	0.92%	0.64%	0.72%
Threonine	6288	0.80%	0.63%	0.42%	0.38%
Tyrosine	6057	0.43%	0.36%	0.22%	0.18%
Valine	6287	4.04%	2.47%	1.45%	1.07%
Guanosine	6802	1.41%	1.03%	0.47%	0.55%
Uracil	1174	0.27%	0.17%	0.05%	0.05%
Lactic acid	612	1.43%	1.02%	1.69%	0.75%
Succinic acid	1110	0.62%	0.43%	0.39%	0.19%
Ergosterol	444679	0.55%	0.63%	0.12%	0.06%
Linoleic acid	5280450	0.46%	0.89%	0.09%	0.13%
Linolenic acid	5280934	4.29%	4.02%	0.45%	0.44%
Palmitic acid	985	1.48%	1.77%	0.25%	0.28%
Pelargonic acid	8158	0.41%	0.41%	0.11%	0.12%
Stearic acid	5281	4.72%	3.41%	0.71%	1.05%
Galactinol	11727586	0.23%	4.67%	0.03%	0.26%
Galactose	439357	0.48%	0.31%	0.48%	0.43%
Glucose	64689	0.14%	0.08%	0.13%	0.13%
Glycerol	753	8.66%	14.57%	25.34%	25.95%
Trehalose	7427	0.03%	0.06%	0.02%	1.28%
4-Hydroxymandelic	328	0.60%	0.82%	0.14%	0.09%
β-Glycerolphosphate	2526	0.41%	0.02%	0.01%	0.01%
Glycerol-3-galactoside	16048618	1.81%	11.58%	0.46%	1.81%
Glycerol-3-phosphate	754	1.61%	0.71%	0.50%	0.29%
Methylgalactoside	94214	3.88%	4.35%	1.00%	0.17%
Methylhexose	560150	0.27%	0.14%	0.07%	0.01%
Phytol	5280435	11.58%	7.84%	1.19%	3.48%
α-Tocopherol	638015	0.17%	0.38%	0.04%	0.02%

batch cultures were evenly distributed; alanine and glycerol were the highest concentrated metabolites of 11.76% at 72 h and 14.57% at 168 h, respectively. Contrary to this, in fed-batch cultures, proline and glycerol comprised approximately 35% and 25%, respectively, relative to other metabolites. Glycerol and proline have both been described as osmoregulators in *C. reinhardtii* and accumulation of these metabolites are correlated with high-salt stress [38,39]. The accumulation of these osmoregulators therefore indicates that the feed added to the fed-batch cultures is resulting in a toxic accumulation of salt ions and the feed composition requires optimization in order to increase productivity and culture stability.

3.6. GFP content

GFP content as a fraction of total soluble protein was lower than anticipated in CC-2937, occurring at < 0.01% of total soluble protein. During cultivation, GFP content increased over time in both batch and fed batch cultures (Fig. 6A). However, GFP content was significantly higher by an average of 2 times in batch cultures compared to fed-batch cultures from 144 to 168 h of cultivation ($P < 0.05$). Potential factors limiting this expression under fed-batch cultivation are: light limitation due to significantly higher culture density, therefore depressing the light-regulated *psbD* promoter driving GFP production; elevated levels

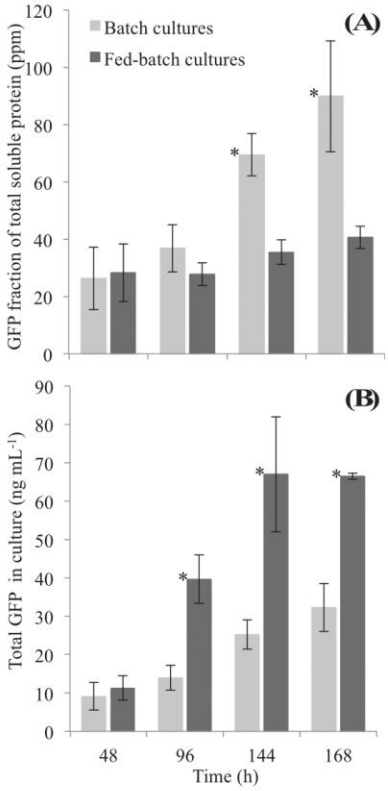

Fig. 6. Average GFP content of batch (light bars) and fed-batch (dark bars) cultures as: (A) a fraction of soluble protein and (B) total GFP in the culture. Error bars indicate standard error of the mean ($n = 3$); *indicates a significant increase in GFP content ($P < 0.05$).

of stress on the cells and shifts in metabolism brought about by the accumulation of salt in the media; or unforeseen biological restraints present in the host strain, CC-2937, that limit accumulation of recombinant protein under any condition.

Nevertheless, due to the dramatic increases in biomass as a result of fed-batch cultivation, total GFP content in the fed-batch cultures (Fig. 6B) was significantly higher by an average of 2.5 times compared to batch cultures from 96 to 168 h of cultivation ($P < 0.05$). This supports the concept that fed-batch cultivation is a viable strategy to increase production of recombinant products, but for optimal expression of heterologous products in the chloroplast, the invention of light-independent genetic constructs will be necessary for this technology to be implemented at commercial scale.

4. Conclusions

Here we demonstrate cultivation of a transgenic strain of *Chlamydomonas reinhardtii* under mixotrophic fed-batch conditions, where a concentrated feed of acetic acid and nutrients were

systematically added to the culture. This cultivation strategy resulted in significant increases in biomass density, cell concentration, productivity, and recombinant protein accumulation in comparison to traditional batch cultivation. We conclude, based on the increased production of osmoregulatory metabolites and shift in cell morphology, that the growth of fed-batch cultures ultimately became limited by the accumulation of salt, not the depletion of carbon or nutrients. These findings indicate that with the continued development of new of molecular tools and optimization of growth media, *C. reinhardtii* has great potential to produce meaningful quantities of both biomass and high-value recombinant products in industrial bioreactors.

Author contributions

FJF conceived, designed, and executed experiments, analyzed data, and drafted the manuscript. JTO assisted in execution of experiments, data interpretation, figure generation, and critical review of the manuscript. SPM supervised this project and critically reviewed the manuscript. FJF takes full responsibility for the integrity of this work, from inception to finished article.

Conflict of interest and informed consent

The authors declare that they have no financial or competing interests that influence the outcomes of this study. No conflicts, informed consent, human or animal rights applicable.

Declaration of authors' agreements to authorship and submission

The authors confirm that the work is original and agree to authorship and submission to *Algal Research* for peer review.

Acknowledgements

We thank the UC Davis West Coast Metabolomics Center for processing lipidomics and metabolomics samples, Anthony Berndt and Svein Horn for discussion of results and reading of the manuscript, and partial funding support from the U.S. Department of Energy (DOE-EE0007689).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2018.05.006.

References

[1] E.H. Harris, Chlamydomonas as a model organism, Annu. Rev. Plant Biol. 52 (1) (2001) 363–406, http://dx.doi.org/10.1146/annurev.arplant.52.1.363.
[2] S.S. Merchant, S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, ... W.F. Marshall, The *Chlamydomonas* genome reveals the evolution of key animal and plant functions, Science 318 (5848) (2007) 245–250, http://dx.doi.org/10.1126/science.1143609.
[3] E.H. Harris, The *Chlamydomonas* Sourcebook, Second Edition, Academic Press, Oxford, United Kingdom, 2009.
[4] B.A. Rasala, S.P. Mayfield, Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses, Photosynth. Res. 123 (3) (2015) 227–239, http://dx.doi.org/10.1007/s11120-014-9994-7.
[5] M.A. Scranton, J.T. Ostrand, F.J. Fields, S.P. Mayfield, *Chlamydomonas* as a model for biofuels and bio-products production, Plant J. 82 (3) (2015) 523–531, http://dx.doi.org/10.1111/tpj.12780.
[6] T.S. Murbach, R. Glavits, J.R. Endres, G. Hirka, A. Vertesi, E. Beres, I.P. Szakonyine, A toxicological evaluation of *Chlamydomonas reinhardtii*, a green algae, Int. J. Toxicol. 37 (1) (2018) 53–62, http://dx.doi.org/10.1177/1091581817746109.
[7] D.J. Barrera, S.P. Mayfield, High-value recombinant protein production in microalgae, Handbook of Microalgal Culture: Applied Phycology and Biotechnology, 2 2013, pp. 532–544.
[8] C. Griesbeck, I. Kobl, M. Heitzer, Chlamydomonas reinhardtii, Mol. Biotechnol. 34 (2) (2006) 213–223, http://dx.doi.org/10.1385/MB:34:2:213.
[9] M.A. Scaife, G.T. Nguyen, J. Rico, D. Lambert, K.E. Helliwell, A.G. Smith, Establishing *Chlamydomonas reinhardtii* as an industrial biotechnology host, Plant J. 82 (3) (2015) 532–546, http://dx.doi.org/10.1111/tpj.12781.
[10] E. Specht, S. Miyake-Stoner, S. Mayfield, Micro-algae come of age as a platform for recombinant protein production, Biotechnol. Lett. 32 (10) (2010) 1373–1383, http://dx.doi.org/10.1007/s10529-010-0326-5.
[11] M. Tran, C. Van, D.J. Barrera, P.L. Pettersson, C.D. Peinado, J. Bui, S.P. Mayfield, Production of immunotoxins in algal chloroplasts, Proc. Natl. Acad. Sci. 110 (1) (2013) E15–E22, http://dx.doi.org/10.1073/pnas.1214638110.
[12] S.P. Mayfield, A.L. Manuell, S. Chen, J. Wu, M. Tran, D. Siefker, M. Muto, J. Marin-Navarro, *Chlamydomonas reinhardtii* chloroplasts as protein factories, Curr. Opin. Biotechnol. 18 (2) (2007) 126–133, http://dx.doi.org/10.1016/j.copbio.2007.02.001.
[13] A.L. Manuell, M.V. Beligni, J.H. Elder, D.T. Siefker, M. Tran, A. Weber, T.L. McDonald, S.P. Mayfield, Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast, Plant Biotechnol. J. 5 (3) (2007) 402–412, http://dx.doi.org/10.1111/j.1467-7652.2007.00249.x.
[14] A.L. Demain, P. Vaishnav, Production of recombinant proteins by microbes and higher organisms, Biotechnol. Adv. 27 (2009) 297–306, http://dx.doi.org/10.1016/j.biotechadv.2009.01.008.
[15] S.E. Shin, J.M. Lim, H.G. Koh, E.K. Kim, N.K. Kang, S. Jeon, S. Kwon, W.S. Shin, B. Lee, K. Hwangbo, J. Kim, S.H. Ye, J.Y. Yun, H. Seo, H.M. Oh, K.J. Kim, J.S. Kim, W.J. Jeong, Y.K. Chang, B.R. Jeong, CRISPR/Cas9-induced knockout and knock-in mutations in *Chlamydomonas reinhardtii*, Sci. Rep. 6 (27810) (2016), http://dx.doi.org/10.1038/srep27810.
[16] M. Venter, Synthetic promoters: genetic control through *cis* engineering, Trends Plant Sci. 12 (3) (2007) 118–124, http://dx.doi.org/10.1016/j.tplants.2007.01.002.
[17] M.A. Borowitzka, N.R. Moheimani, Algae for Biofuels and Energy, Springer, Dordrecht, Netherlands, 2013.
[18] C.Y. Chen, K.L. Yeh, R. Aisyah, D.J. Lee, J.S. Chang, Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: a critical review, Bioresour. Technol. 102 (1) (2011) 71–81, http://dx.doi.org/10.1016/j.biortech.2010.06.159.
[19] N.G. Schoepp, R.L. Stewart, V. Sun, A.J. Quigley, D. Mendola, S.P. Mayfield, M.D. Burkart, System and method for research-scale outdoor production of microalgae and cyanobacteria, Bioresour. Technol. 166 (2014) 273–281, http://dx.doi.org/10.1016/j.biortech.2014.05.046.
[20] S.J. Szyjka, S. Mandal, N.G. Schoepp, B.M. Tyler, C.B. Yohn, Y.S. Poon, S. Villareal, M.D. Burkart, J.B. Shurin, S.P. Mayfield, Evaluation of phenotype stability and ecological risk of a genetically engineered alga in open pond production, Algal Res. 24 (2017) 378–386, http://dx.doi.org/10.1016/j.algal.2017.04.006.
[21] F. Bumbak, S. Cook, V. Zachleder, S. Hauser, K. Kovar, Best practices in heterotrophic high-cell-density microalgal processes: achievements, potential and possible limitations, Appl. Microbiol. Biotechnol. 91 (1) (2011) 31, http://dx.doi.org/10.1007/s00253-011-3311-6.
[22] O. Perez-Garcia, F.M. Escalante, L.E. de-Bashan, Y. Bashan, Heterotrophic cultures of microalgae: metabolism and potential products, Water Res. 45 (1) (2011) 11–36, http://dx.doi.org/10.1016/j.watres.2010.08.037.
[23] D.S. Gorman, R.P. Levine, Cytochrome f and plastocyanin: their sequence in the photosynthetic electron transport chain of *Chlamydomonas reinhardi*, Proc. Natl. Acad. Sci. 54 (6) (1965) 1665–1669.
[24] F. Chen, M.R. Johns, A strategy for high cell density culture of heterotrophic microalgae with inhibitory substrates, J. Appl. Phycol. 7 (1995) 43–46, http://dx.doi.org/10.1016/j.watres.2010.08.037.
[25] F. Chen, M.R. Johns, Heterotrophic growth of *Chlamydomonas reinhardtii* on acetate in chemostat culture, Process Biochem. 31 (6) (1996) 601–604, http://dx.doi.org/10.1016/S0032-9592(96)00006-4.
[26] L. Sack, C. Zeyl, G. Bell, T. Sharbel, X. Reboud, T. Bernhardt, H. Koelewyn, Isolation of four new strains of *Chlamydomonas reinhardtii* (Chlorophyta) from soil samples, J. Phycol. 30 (1994) 770–773, http://dx.doi.org/10.1111/j.0022-3646.1994.00770.x.
[27] D. Barnes, S. Franklin, J. Schultz, R. Henry, E. Brown, A. Coragliotti, S. Mayfield, Contribution of 5′- and 3′-untranslated regions of plastid mRNAs to the expression of *Chlamydomonas reinhardtii* chloroplast genes, Mol. Gen. Genomics. 274 (6) (2005) 625–636, http://dx.doi.org/10.1007/s00438-005-0055-y.
[28] B.A. Rasala, M. Muto, P.A. Lee, M. Jager, R.M.F. Cardoso, C.A. Behnke, P. Kirk, C.A. Hokanson, R. Crea, M. Mendez, S.P. Mayfield, Production of therapeutic proteins in algae, analysis of expression of seven human proteins in the chloroplast of *Chlamydomonas reinhardtii*, Plant Biotechnol. J. 8 (6) (2010) 719–733, http://dx.doi.org/10.1111/j.1467-7652.2010.00503.x.
[29] B.A. Rasala, M. Muto, J. Sullivan, S.P. Mayfield, Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5′ untranslated region optimization, Plant Biotechnol. J. 9 (6) (2011) 674–683, http://dx.doi.org/10.1111/j.1467-7652.2011.00620.x.
[30] J. Kropat, A. Hong-Hermesdorf, D. Casero, P. Ent, M. Castruita, M. Pellegrini, S.S. Mechant, D. Malasarn, A revised mineral nutrient supplement increases biomass and growth rate in *Chlamydomonas reinhardtii*, Plant J. 66 (5) (2011) 770–780, http://dx.doi.org/10.1111/j.1365-313X.2011.04537.x.
[31] C.J. Zhu, Y.K. Lee, Determination of biomass dry weight of marine microalgae, J. Appl. Phycol. 9 (2) (1997) 189–194, http://dx.doi.org/10.1023/

A:1007914806640.
[32] T. Kind, J.K. Meissen, D. Yang, F. Nocito, A. Vaniya, Y.S. Cheng, J.S. VanderGheynst, O. Fiehn, Qualitative analysis of algal secretions with multiple mass spectrometric platforms, J. Chromatogr. A 1244 (2012) 139–147, http://dx.doi.org/10.1016/j.chroma.2012.04.074.
[33] O. Fiehn, G. Wohlgemuth, M. Scholz, T. Kind, D.Y. Lee, Y. Lu, S. Moon, B. Nikolau, Quality control for plant metabolomics: reporting MSI-compliant studies, Plant J. 53 (4) (2008) 691–704, http://dx.doi.org/10.1111/j.1365-313X.2007.03387.x.
[34] L. Miller, Analyzing gels and western blots with ImageJ, http://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j/, (November 4 2010) , Accessed date: 29 August 2016.
[35] N.R. Boyle, J.A. Morgan, Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*, BMC Syst. Biol. 3 (2009) 4, http://dx.doi.org/10.1186/1752-0509-3-4.
[36] T.S. Moore, Z. Du, Z. Chen, Membrane lipid biosynthesis in *Chlamydomonas reinhardtii*. In vitro biosynthesis of diacylglyceryltrimethylhomoserine, Plant Physiol. 125 (1) (2001) 423–429.
[37] K.K. Sharma, H. Schuhmann, P.M. Schenk, High lipid induction in microalgae for biodiesel production, Energies 5 (2012) 1532–1553, http://dx.doi.org/10.3390/en5051532.
[38] H.D. Husic, N.E. Tolbert, Effect of osmotic stress on carbon metabolism in *Chlamydomonas reinhardtii*, Plant Physiol. 82 (1986) 594–596.
[39] G. Mastrobuoni, S. Irgang, M. Pietzke, H.E. Assmus, M. Wenzel, W.X. Schulze, S. Kempa, Proteome dynamics and early salt stress response of the photosynthetic organism *Chlamydomonas reinhardtii*, BMC Genomics 13 (2012) 215, http://dx.doi.org/10.1186/1471-2164-13-215.

Acknowledgements

Chapter 3, in full, is a reprint of the material as it appears in *Algal Research 2018*. Fields, Francis; Ostrand, Joseph; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this paper.

CHAPTER 4:

GENOME SHUFFLING AND SELECTION FOR IMPROVED RECOMBINANT PROTEIN ACCUMULATION IN THE CHLOROPLAST OF *CHLAMYDOMONAS REINHARDTII*

Genome shuffling and selection for improved recombinant protein accumulation in the chloroplast of *C. reinhardtii*

Joseph T. Ostrand, Francis J. Fields, Miller Tran, and Stephen P. Mayfield

Abstract

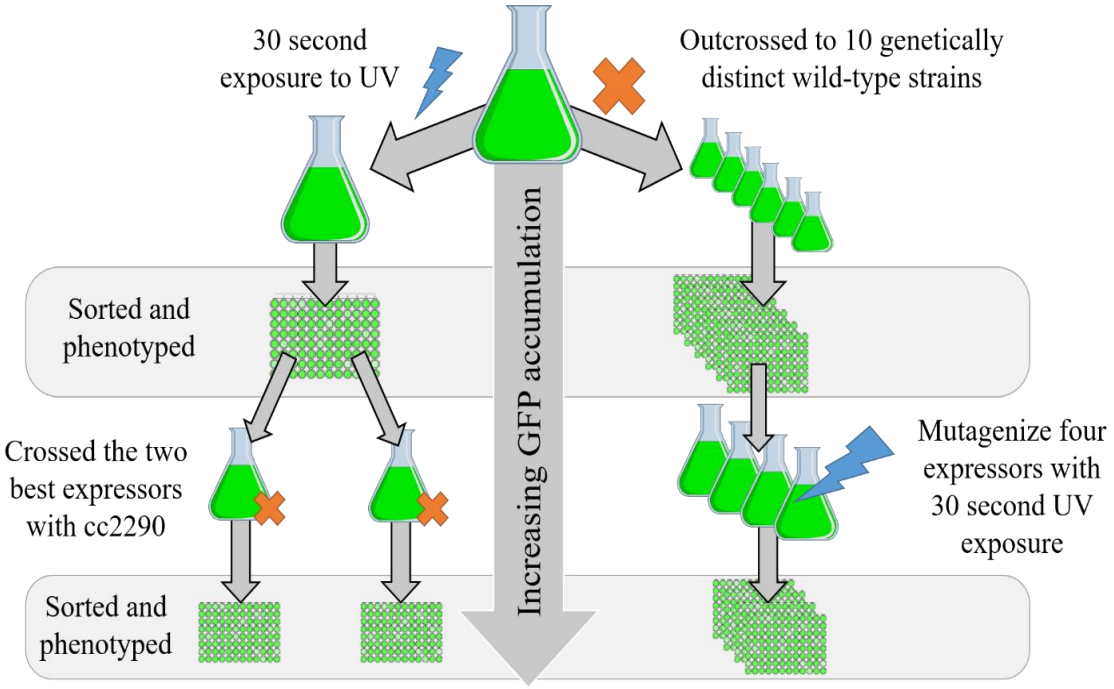

Figure 4.1. Visual abstract illustrating the breeding and mutagenesis program for increased GFP accumulation. Wild-type *C. reinhardtii* was transformed with a chloroplast GFP expression vector. This strain was either subjected to UV mutagenesis or was independently bred with one of 10 genetically distant wild-type strains. The resulting mutants and progeny were analyzed by flow cytometry and the most fluorescent individuals were sorted by FACS. 96 individual progeny were recovered from the sort and were analyzed by fluorescent plate reader. Two of the best expressors which were sorted from the mutagenized pool were bred with the wild-type strain cc2290. In addition, four individuals with varying GFP expression levels isolated from the breeding program were mutagenized. Both the first and second rounds of mutagenesis, breeding and sorting yielded individual strains with substantially higher GFP accumulation than the original strain.

In *C. reinhardtii*, endogenous light-driven regulatory elements have been used to drive recombinant protein accumulation to up to 10% of total soluble protein. However, as cultivation improvements facilitate higher density growth conditions, light penetration becomes limiting and light-activated promoters become less desirable. This report describes a high-throughput genome shuffling and selection based approach to optimize the *C. reinhardtii* background strain for dark expression of recombinant protein. This was accomplished by both outcrossing and mutagenesis with high-throughput selection by FACS for top-expression mutants. The result is a strain with >10-fold increase in reporter protein accumulation.

Introduction

As the price of oil has dropped and the demand for algal-based biofuels has backed off over the past decade, there has been an increased interest in high-value product development in algal hosts(1,2). Commercial development of recombinant protein production strains shifts the primary growth conditions away from open ponds into more tightly controlled bioreactors. Closed bioreactors are necessary for the precise optimization of temperature, pH and cell cycle that will produce the highest yields of product, and the consistency necessary to comply with regulatory standards(3,4). Smaller growth chambers facilitate high concentrations of biomass accumulation for optimal product yield per volume. As cell concentrations increase, light penetration into photobioreactors diminishes rapidly. This limits the capacity for phototrophic growth, as well as the efficacy of light-driven regulatory elements in production vectors. To date, the strategies which result in the highest yields of recombinant protein accumulation in *C. reinhardtii* utilize native light-regulated machinery to drive exogenous gene expression(5,6). Development of dark-optimized tools for increased recombinant protein accumulation in the chloroplast of *C. reinhardtii* will facilitate commercial relevance in the high-value recombinant protein market.

Whole-genome modification strategies for phenotype optimization can be incredibly powerful(7–9). Selective breeding in higher plants has been used for centuries to encourage complex phenotypes, and can be used effectively without an intricate understanding of the species. Breeding within a species causes recombination of the genome, which can result in certain genes being eliminated from the population, in addition to major alterations in chromatin structuring and gene regulation(10,11). This shuffling of the genome is well documented as a potential strategy for phenotype development across industrial organisms(7–9). When combined with robust selection, a phenotype which has been difficult to create de novo can be generated through the process of genome shuffling. Selective breeding in *C. reinhardtii* has already been shown to generate salt-tolerant progeny(12).

C. reinhardtii has two distinct mating types, designated mt(+), and mt(-). Upon nitrogen starvation, mt+ and mt- cells express complementary proteins on their flagella which help to bind the two individuals together(13). Upon flagellar binding, the membranes fuse to form a zygote and undergo meiosis, resulting in 4 daughter cells each with unique nuclear genomes. Each of the progeny will exclusively receive the chloroplast genome of the mt+ parent and the mt- chloroplast genome is selectively degraded during the mating process(14). This uniparental inheritance facilitates a breeding program in which all of the generated progeny will contain the heterologous chloroplast gene of interest as long as it was initially present in the mt+ parent.

In *C. reinhardtii*, proteins expressed by the nuclear genome can affect gene regulation in the chloroplast(15,16). Therefore, shuffling of the nuclear genome can result in modified expression levels of a heterologous gene in the chloroplast. GFP will be used as a means of quantifying chloroplast gene expression within progeny(17). By using a fluorescent marker, high-throughput analysis by flow cytometry can rapidly identify progeny which express increased levels of GFP and isolate them for individual analysis(18). By breeding and mutagenizing a *C. reinhardtii* strain expressing GFP, we were able to create and sort for exemplary strains for

chloroplast gene expression. The precise breeding and mutagenesis program is detailed in the first figure.

Results

Based on the phylogenetic tree generated by Flowers et al. 2015 (19), we gathered a collection of *C. reinhardtii* wildtype strains of varying backgrounds. Due to the uniparental inheritance of the chloroplast genome, the starting strain is required to be mt(+) in order to maintain the marker gene throughout the breeding program. A growth curve was generated for each of the wild-type strains and it was determined that cc4414 worked most optimally with our laboratory growth conditions (data not shown). cc4414 was transformed with a chloroplast GFP expression vector. One of the hybrid promoters developed as part of the Rasala et al. 2011 papers which showed some activity under dark conditions was selected to drive GFP expression in our chloroplast expression vector.

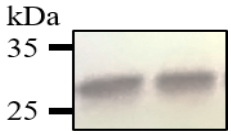

Figure 4.2. Western blot confirms GFP accumulates in the starting strain. Two biological replicates of the GFP-expressing strain 178 were grown in liquid TAP cultures in the dark. Harvested cells were lysed and 50ng of total soluble protein as measured by Bradford assay was loaded into each lane. Anti-GFP conjugated to AP was used to perform and develop the immunoblot.

The strain was shown to successfully accumulate GFP in its chloroplast when grown in the dark (Figure 2). This initial GFP-expressing strain will hereafter be referred to as strain "178".

Strain 178 was individually bred with a library of mt(-) strains to induce genome shuffling-mediated expression changes in GFP. The specific mt(-) strains were selected because their genetic ancestry had been previously characterized. Progeny from the mating were sonicated to help break apart the robust pellicle and to help remove parent populations which did not mate effectively. In addition, strain 178 was exposed to UV for varying times to induce mutagenesis. Each of the

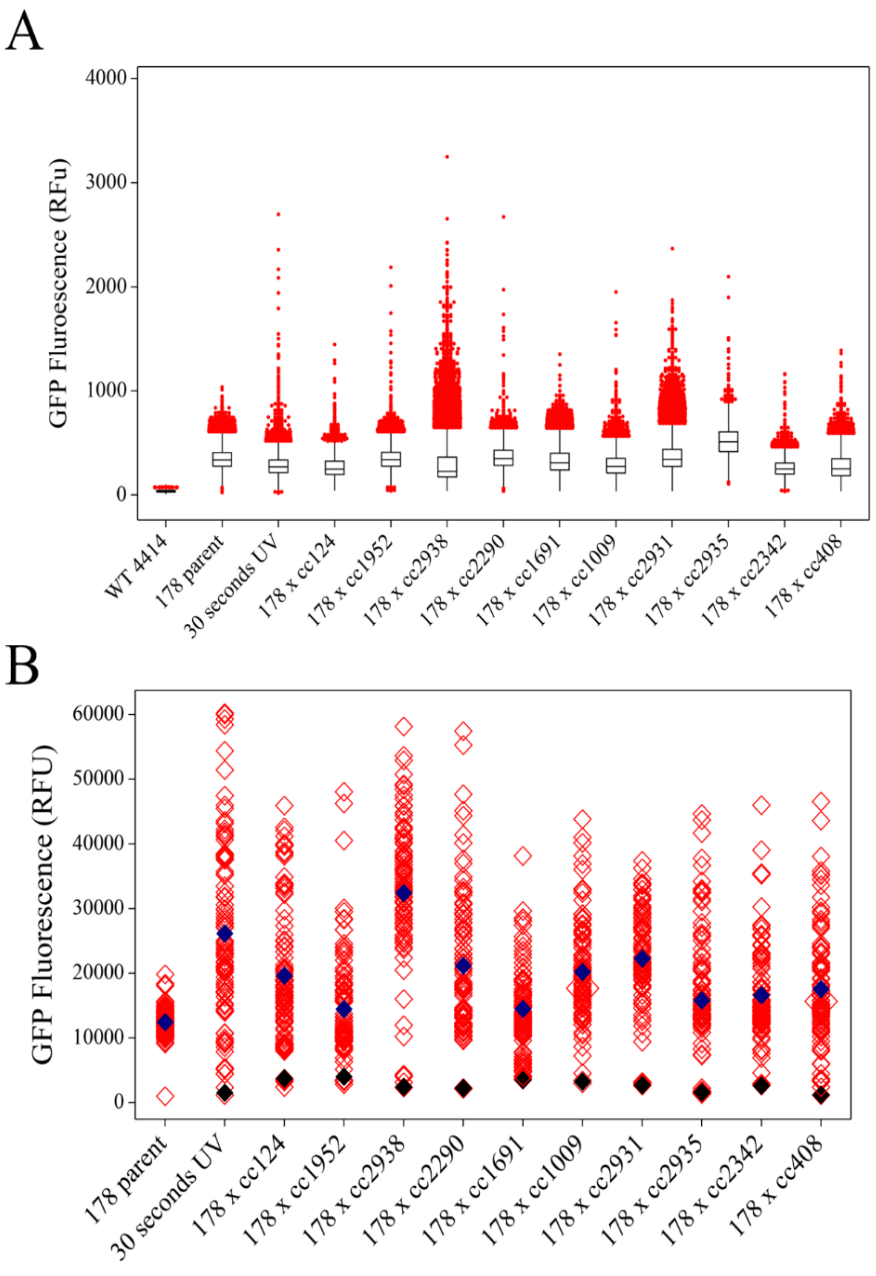

Figure 4.3. GFP fluorescence of progeny generated from mating and mutagenesis. (A)Mating and mutagenesis with the starting GFP-expressing strain 178 generated variability of GFP expression levels among dark-adapted progeny. Flow cytometry analysis of >3,500 cells is presented, and statistical outliers, shown in red, describe the likelihood that high-GFP progeny were created for each condition. (B) Highly fluorescent cells were isolated from each condition by FACS. 96 individuals from each condition were individually cultured and assayed for GFP by fluorescent plate reader. On this graph, the relative GFP fluorescence of each of the 96 isolated progeny is represented in red. The mean fluorescence of the group is shown in blue, and the background strain which was used for each circumstance is shown in black.

resulting pools was analyzed by flow cytometry to determine the variability in GFP expression due to breeding and mutagenesis (Figure 3A and S1). An increase in the number of outliers in progeny pools shows that breeding and mutagenesis have increased the variability of GFP expression within the population. The individuals with the greatest GFP fluorescence were sorted by FACS to determine the stability of the phenotype in an isogenic population and to more accurately quantify the change in GFP accumulation. 96 individuals which were sorted from mating with each breeding partner, as well as 96 individuals from the 30 second UV mutagenesis and 96 individuals from an unmodified parent strain 178 were grown in separate 500µL cultures and analyzed by fluorescent plate reader (Figure 3B). These results show that breeding with any strain and sorting for high expression of GFP resulted in individuals with substantial increases in GFP fluorescence. Breeding with cc2938 resulted in the largest increase in the mean fluorescence of the offspring and some of the most fluorescent individual strains. Mutagenesis by UV generated the three most fluorescent progeny strains. Although mating with cc2290 did not result in large population variability as measured by flow cytometry, two of the top ten most fluorescent individuals in the entire study were generated from this circumstance.

UV mutagenizing the best isolates from the breeding program

Four individuals from the breeding program were selected to undergo a round of UV mutagenesis and sorting to determine if the phenotype could be further optimized by utilizing both genome alteration techniques. Two of the four individuals, M22 and M26 were sorted from the cc2938 progeny, and showed a dramatic increase in GFP accumulation. M30 and M34 were selected from the cc2290 progeny, and were selected due to their more nominal increase in GFP accumulation. As shown in Figure 4A, the two strains from the breeding program which had already dramatically increased their GFP accumulation (M22, M26) did not yield additional increases in response to UV mutagenesis. By contrast, the two strains which had marginally increased their GFP accumulation from the breeding program (M30, M34) successfully generated additional high-GFP

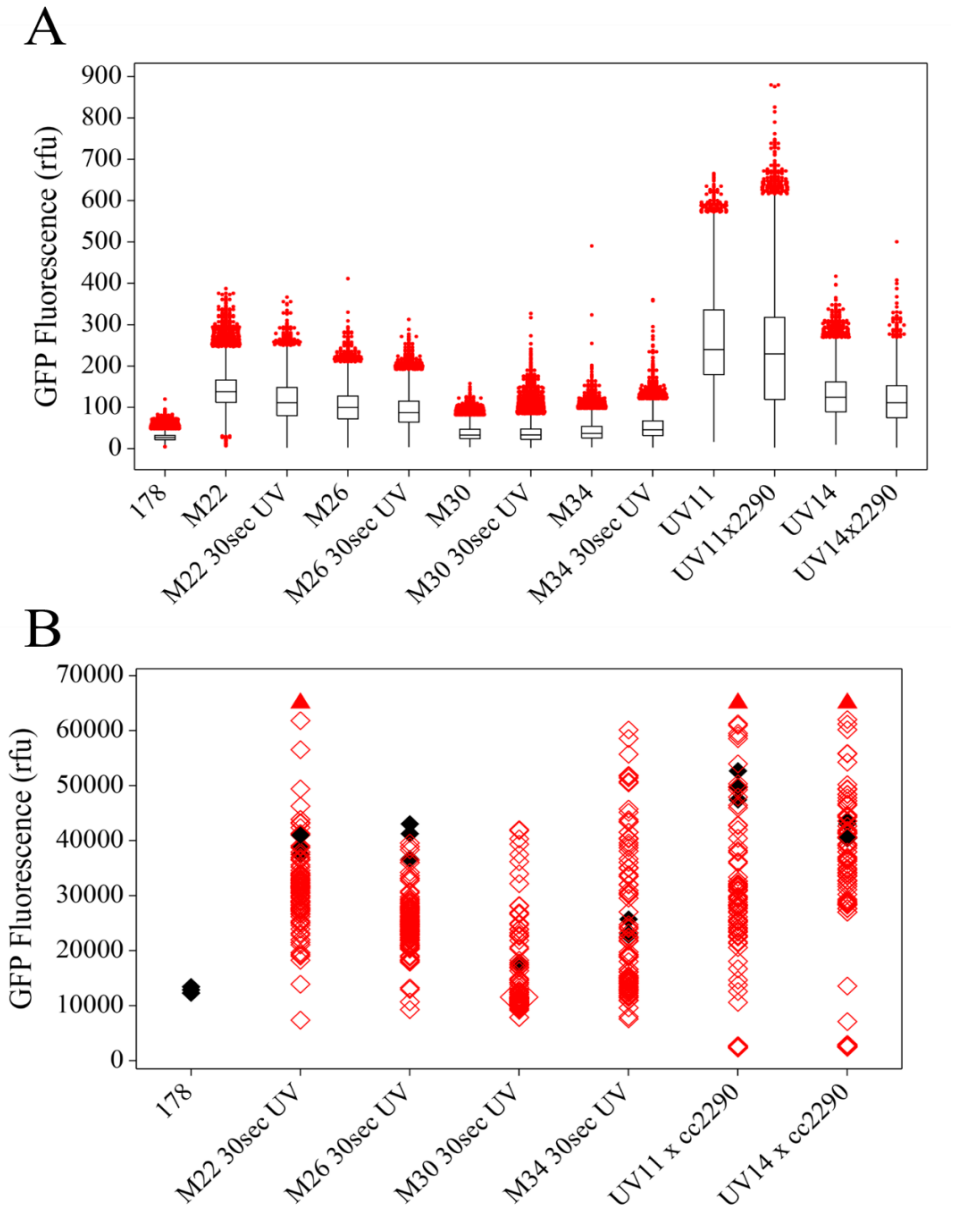

Figure 4.4. GFP fluorescence from a second round of mating or mutagenesis. (A) Four individuals were selected from the sorted offspring (M22, M26, M30 and M34), and two were selected from the sorted pool of UV mutants (UV11 and UV14). The four offspring strains were UV mutagenized and the UV mutants were bred with cc2290 to determine the possibility of further pushing the phenotype. The effect of a second round of breeding or mutagenesis on the GFP fluorescence of >15,000 individuals is compared with the starting strain by flow cytometry. (B) Optimal expressors from each circumstance were sorted and recovered. 96 individuals from each condition were grown independently in liquid culture and assayed by fluorescent plate reader for GFP accumulation. Red symbols correspond to sorted progeny from the second round of breeding and mutagenesis. The black symbols are the fluorescence readings for the parent strains prior to the second round of genome shuffling. Triangles represent an individual which reached the detection limit of the plate reader. No sample had more than one such event.

outliers in response to UV mutagenesis. A population of the best accumulators from each of these mutagenized pools were sorted, and 96 individuals were assayed by fluorescent plate reader (Figure 4B). With the exception of M26, each of the sorted populations contained individuals with greater GFP accumulation. The single best expressor from this program, isolated from the mutagenesis of M22, was renamed "I67" and kept for phenotyping by immunoblot and fluorescence microscopy (Figure 5 and Figure 6).

Breeding the best isolates from the pool of UV mutants

The two best expressors from the initial experiment were isolated from the pool of UV mutants. These two strains, UV11 and UV14, were bred with cc2290. The offspring were measured by flow cytometry and sorted for high expressors as described previously (Figure 4A). Although the accumulation of UV11 and UV14 was exceptionally high to begin with, the offspring generated from breeding them with cc2290 and sorting yielded additional increases to GFP accumulation. The best individual from the UV11xcc2290 cross was renamed "D52", and the highest expressor from the UV14xcc2938 cross renamed "C89".

The selected strains from each round of the study were run simultaneously on a single western blot as shown in Figure 5. The immunostaining verifies increases in GFP accumulation with each round of mutagenesis or mating and selection.

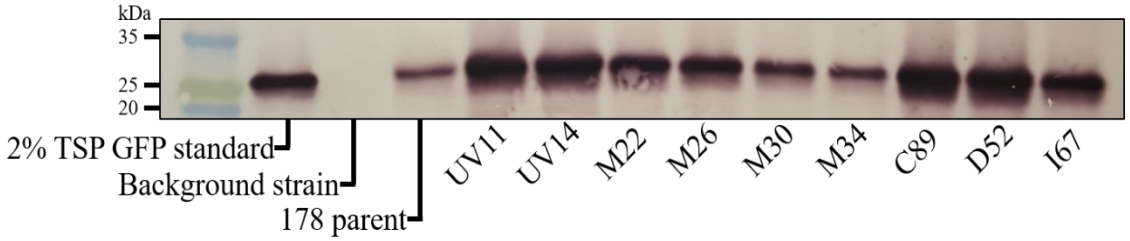

Figure 4.5. Anti-GFP western blot. Immunoblot confirms the difference in accumulation of GFP through successive rounds of genome shuffling and selection. 50ng of soluble protein as measured by Bradford assay was purified from dark-adapted liquid cultures and loaded in each lane. 1ng of purified GFP was loaded into the first lane as a marker to estimate total GFP accumulation.

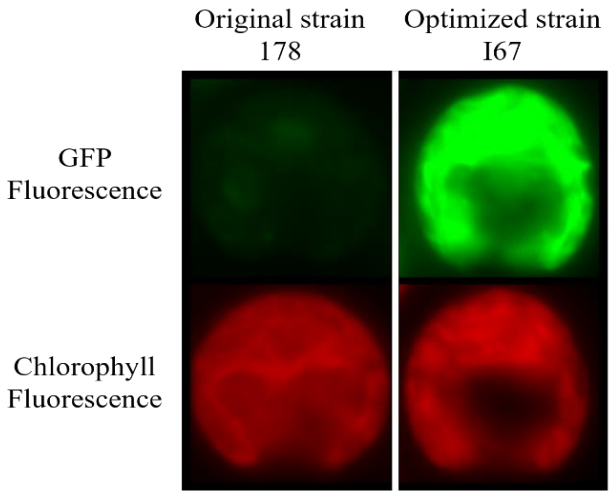

Figure 4.6. GFP accumulation in starting strain compared with the final optimized strain. Fluorescence microscopy image depicting the change in GFP fluorescence intensity from the starting strain (left) to the final optimized strain (right). Dark-adapted liquid cultures of 178 and I67 were confirmed to be homogenous for their phenotype throughout the culture. The cultures were subsequently mixed and analyzed simultaneously. Both images displayed were cropped from the same photograph after image processing.

Discussion

The data presented in this study show that mating and mutagenesis in combination with fluorescence-activated cell sorting are successful strategies for generating strains with increased heterologous protein accumulation. The breeding and mutagenesis process outlined here can be applied to any phenotype with a robust means of selection. In this report, a chloroplast expressed GFP marker was used to show changes in nuclear-encoded chloroplast gene regulation. In contrast to the nuclear genome which has a single copy, there are roughly 80 copies of the chloroplast genome in a single *C. reinhardtii* cell. The chloroplast has a high rate of homologous recombination which rapidly repairs changes to genetic sequence in the chloroplast genome. Because of this, changes to the chloroplast genome are often fixed and do not persist within the strain. Stable genetic transformation of the chloroplast requires multiple generations of constant selection, something not recapitulated by the FACS strategy described here. Therefore, changes to chloroplast gene expression which last multiple generations are likely due to nuclear-expressed factors. Due to the

generational persistence of the phenotype, we do not believe we have simply altered the GFP amino acid sequence in these strains to be a more fluorescent or stable version.

As strains proceed through successive rounds of mating and mutagenesis, it is clear that the rarity of the desired phenotype increases. Subsequent genome shuffling leads to predominantly reversions of GFP expression, and precise selection markers are required to identify the unique event in which genome shuffling led to the desired phenotype. This highlights the strength of high-throughput technologies such as FACS for this type of analysis. In higher plants, identification of exceptional offspring either requires full maturation of the plant to physically measure the phenotype, or approximations of phenotype must be made based on genotype. These limitations vastly restrict the number of offspring which can be screened. As shown above, the development of extreme phenotypes can be a very rare event, and effectively selecting for the preferred phenotype often requires many undesired offspring to be discarded.

Microalgae are a suitable model for large breeding programs due to their fast generation times and their unicellular nature facilitates ability to rapidly screen thousands of unique genotypes in high-throughput systems. We do not report any inherent correlation between phylogenetic distance and diversity created. However, cc2938, which created the most diversity in GFP accumulation in the progeny pool, had the most difficulty mating with cc4414. It did not form a robust pellicle, and repeated attempts to generate a mating response from cc2938 with UV mutants UV11 and UV14 were unsuccessful. In future experiments, it will be interesting to investigate additional Chlamydomonas species which are interfertile with *C. reinhardtii*, such as C. moewusii. However, not every strain of algae being investigated for commercial applications are capable of breeding. There will not be a one-size-fits-all species of algae for all production efforts. Translational techniques which are effective for diverse algal species will be more beneficial for the field than optimization strategies which are species-dependent. As shown above, any strains

which can be mutagenized by UV exposure will generate a diversity of phenotypes in their mutant pool.

The improved strains that were generated in this study have been shown to be great accumulators of GFP. However, GFP does not have standalone value, and production of GFP alone in algae is not useful from a commercial perspective. However, it will be intriguing to investigate if the strains that were generated in this study have made systemic changes which affect the accumulation of any exogenous gene driven by the regulatory elements present in our expression vector. Because there have likely been no changes to the chloroplast genome, the changes to GFP expression are not exclusive to the GFP protein. Homologous recombination in the chloroplast genome allows us to modify the high-expression strains C89, D52 and I67 to replace the GFP gene with other proteins of industrial relevance. If these strains have not generated a global increase in recombinant protein accumulation, in the future a GFP marker can be translationally fused to the protein of interest to track its expression through a breeding program. Altogether, the results from this study open the door to a variety of new industrial strategies for strain optimization in *C. reinhardtii*.

Materials and Methods

Chlamydomonas reinhardtii strains

C. reinhardtii strains were ordered from the Chlamydomonas Resource Center at the University of Minnesota and are listed here with their full names to facilitate simple ordering. The mt+ parent strain in which GFP was transformed is cc-4414 (DN2). The mt- strains that were used as breeding partners are cc408 (C9), cc124 (137c), cc2931 (North Carolina), cc2935 (Quebec), cc2938 (Quebec), cc2290 (S1 D2), cc1952 (S-1 C-5), cc1691 (Sager 6145), cc1009 (UTEX 89), and cc2342 (Jarvik #6, Pittsburgh, PA). *C. reinhardtii* wild-type strains were maintained on TAP plates for the length of this experiment.

Mating protocol

Wild-type strains were grown on TAP media in 75µE until approximately 0.5g of wet biomass was accumulated. The biomass is transferred to nitrogen-free TAP plates to induce gamete formation. After 48 hours of nitrogen starvation, the cells are scraped off the plate, resuspended in a small flask with 5mL of sterile water and placed on a shaker in 100µE light for 1 hour. At this point, the two desired mating partners (one mt+ and one mt-) are combined into one 10mL solution in a new flask. This mixture is shaken for 10 minutes under the same conditions. The mating flask is then left overnight in 75µE light without shaking to ensure proper flagellar adhesion. After 12 hours, strains which have successfully begun the mating process will have formed a pellicle on the bottom of the flask. The water is decanted from the mating flask, and the pellicle is transferred to a new TAP agar plate. After the plate is left to dry in a sterile hood, it is placed in the dark for 7 days to allow for zygote maturation. The pellicle is then returned to a flask, this time containing 50mL of TAP media. The flask is sonicated for 30 seconds at 25% amplitude to help break apart the pellicle and to kill cells which have not properly formed the zygospore. The flask is then left in the light for three days to facilitate release of the progeny from the zygospore.

UV Mutagenesis

UV mutagenesis was performed in a Bio-Rad Genelinker UV chamber. The culture to be mutagenized was grown to late-log phase and 25 mL of culture was transferred to a 10cm diameter petri dish. The petri dish was placed in the UV chamber for 0 to 90 seconds at approximately .09mJ of energy directed at the plate per second of exposure. The culture was transferred back into a flask and allowed to recover for two days shaking at 75µE.

Influx flow cytometer setup

The flow cytometry data collection and cell-sorting were performed by a Becton Dickinson Influx running BD FACS Sortware. The instrument was outfitted with a 70µm nozzle and was run with BD FACSFlow sheath fluid (BD #342003) at 33psi. Sample pressure varied from 33.1 to 34.0psi depending on sample concentration.

A series of inclusion gates were made to narrow the population in order to achieve a clean sample window for successful gating of desired GFP-positive progeny. This multicolor analysis required a 488nm laser with SSC and FSC detection as well as the following filter sets: 692/40 (FL4), 750LP (FL5) and 532/20RB (FL1), as well as a 355nm laser with filter sets for 458/20RB (FL10) and 670/30 (FL11).

Sorting protocol

Desired cells were sorted into 2mL of TAP media in Costar flat-bottom clear polystyrene 6 or 12 well plates (Corning #3513). The 2mL of media containing sorted individuals was left in the dark overnight and plated onto TAP-agar plates 16 hours later. The plates were left under 75µE light at room temperature for 1-2 weeks to allow colonies to form. Colonies were replated onto catalog TAP-agar plates and each individual colony was also used to start a 500uL TAP liquid culture in an Axygen 96 round deep-well storage microplate (Fisher #14-222-357). Each plate was sealed with a Research Products International regular strength Breathe-Easy sealing membrane (RPI #248738), and allowed to grow in 100µE light for one week until cultures reached saturation. Each well was diluted 1:10 into a new deep-well microplate and allowed to grow for a subsequent week under the same conditions to normalize for differences in starting material size. These cultures were used as experimental material for a plate reader assay.

Tecan plate reader protocol

Plate reader data was collected using a Tecan Infinite m200 Pro running Tecan i-control software. 100uL of culture was transferred from each well of the deep-well culture plates and transferred to a Costar flat bottom non-treated clear polystyrene assay plates (Corning #3370). Assay plates were shaken for 5 seconds with a 1mm amplitude prior to top fluorescence detection. Wells were excited by 25 flashes of 488nm wavelength light with a 9nm bandwidth. GFP fluorescence was monitored by emission at 522nm with a 20nm bandwidth and a gain setting of 129. Integration time was set at 20µs, lag time and settle time were both set at 0µs, and Z-position was manually set at 17877µm after determining it was the optimal read depth. Each plate was subsequently monitored for chlorophyll content by top fluorescence detection. Wells were excited by 25 flashes of 440nm wavelength with a 9nm bandwidth. Chlorophyll content was monitored by emission at 680nm with a 20nm bandwidth and a gain setting of 116. Integration time was set at 20µs, lag time and settle time were both set at 0µs, and Z-position was manually set at 21200µm.

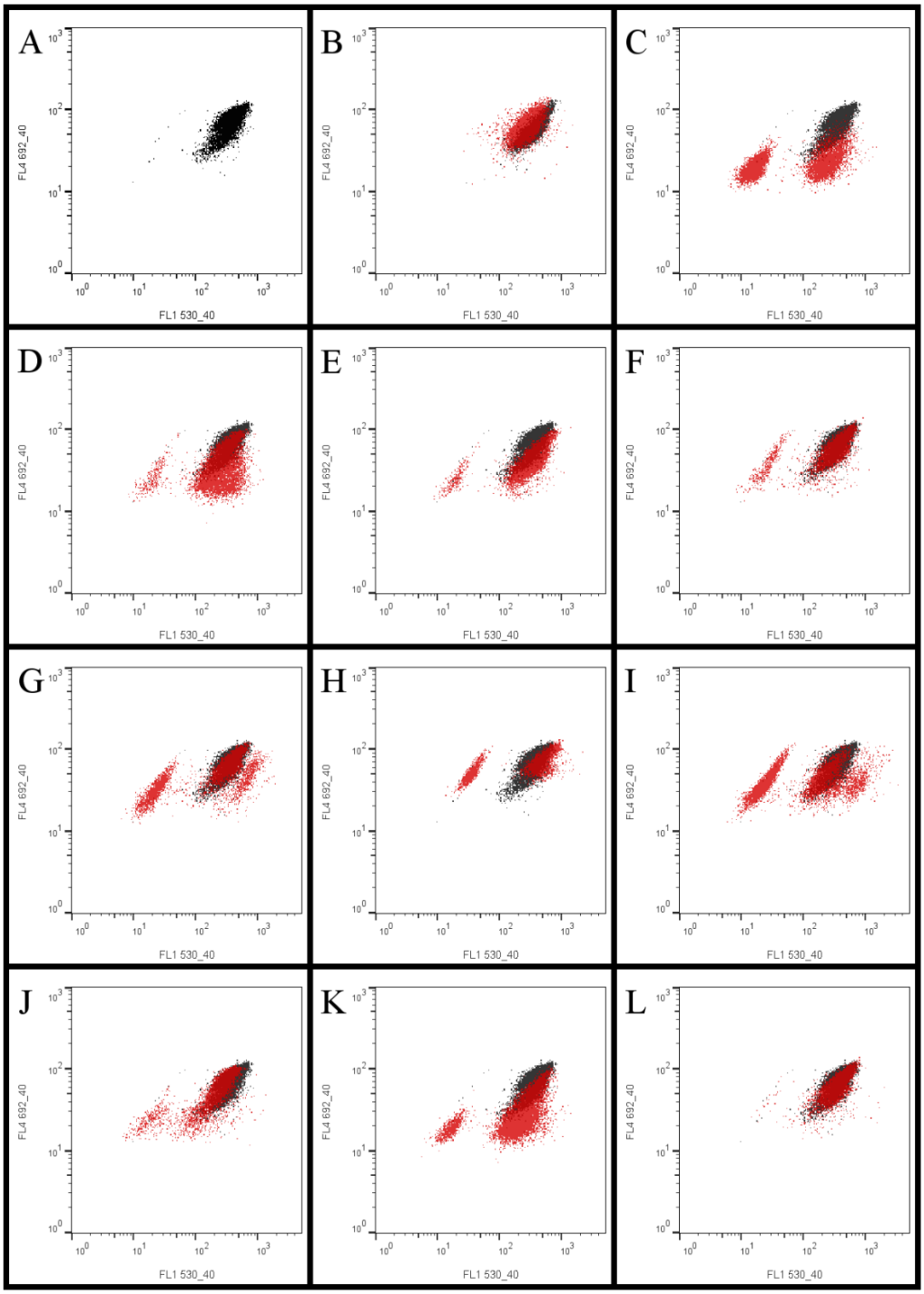

Figure 4.S1. Graphical representation of the flow cytometry data used to generate Figure 3A. In each pane, the black population is dark-adapted 178. The red population is either (B) UV mutagenized 178, or 178 after breeding with (C) cc124, (D) cc1009, (E) cc1691, (F) cc1952, (G) cc2931, (H) cc2935, (I) cc2938, (J) cc2342, (K) cc408 or (L) cc2290. Although the additional population which is present in the offspring samples does contain cells which have lost their ability to express GFP, it is predominantly contaminating parent cells which did not breed and were not killed by sonication.

1. Scranton MA, Ostrand JT, Fields FJ, Mayfield SP. Chlamydomonas as a model for biofuels and bio-products production. Plant J Cell Mol Biol. 2015 May;82(3):523–31.

2. Rasala BA, Mayfield SP. The microalga Chlamydomonas reinhardtii as a platform for the production of human protein therapeutics. Bioeng Bugs. 2011;2(1):50–4.

3. Fields FJ, Ostrand JT, Mayfield SP. Fed-batch mixotrophic cultivation of Chlamydomonas reinhardtii for high-density cultures. Algal Res. 2018 Jul;33:109–17.

4. Ramos-Martinez EM, Fimognari L, Sakuragi Y. High-yield secretion of recombinant proteins from the microalga Chlamydomonas reinhardtii. Plant Biotechnol J. 2017 Sep;15(9):1214–24.

5. Manuell AL, Beligni MV, Elder JH, Siefker DT, Tran M, Weber A. Robust expression of a bioactive mammalian protein in Chlamydomonas chloroplast. Plant Biotechnol J. 2007 May;5(3):402–12.

6. Rasala BA, Muto M, Sullivan J, Mayfield SP. Improved heterologous protein expression in the chloroplast of Chlamydomonas reinhardtii through promoter and 5′ untranslated region optimization: Improved heterologous protein expression in the chloroplast. Plant Biotechnol J. 2011 Aug;9(6):674–83.

7. Zhang Y-X, Perry K, Vinci VA, Powell K, Stemmer WPC, Cardayré SB del. Genome shuffling leads to rapid phenotypic improvement in bacteria. Nature. 2002 Feb;415(6872):644–6.

8. Gong J, Zheng H, Wu Z, Chen T, Zhao X. Genome shuffling: Progress and applications for phenotype improvement. Biotechnol Adv. 2009 Nov 1;27(6):996–1005.

9. Pinel D, D'Aoust F, Cardayre SB del, Bajwa PK, Lee H, Martin VJJ. Saccharomyces cerevisiae Genome Shuffling through Recursive Population Mating Leads to Improved Tolerance to Spent Sulfite Liquor. Appl Environ Microbiol. 2011 Jul 15;77(14):4736–43.

10. Lämke J, Bäurle I. Epigenetic and chromatin-based mechanisms in environmental stress adaptation and stress memory in plants. Genome Biol. 2017 Jun 27;18:124.

11. Wusirika R, Bohn M, Lai J, Kole C. Genetics, Genomics and Breeding of Maize. CRC Press; 2014. 307 p.

12. Perrineau M-M, Zelzion E, Gross J, Price DC, Boyd J, Bhattacharya D. Evolution of salt tolerance in a laboratory reared population of Chlamydomonas reinhardtii. Environ Microbiol. 2014 Jun;16(6):1755–66.

13. Ferris PJ, Goodenough UW. Mating Type in Chlamydomonas Is Specified by Mid, the minus-Dominance Gene. Genetics. 1997 Jul;146(3):859–69.

14. Vanwinkle-Swift KP. Uniparental inheritance is promoted by delayed division of the zygote in Chlamydomonas. Nature. 1978 Oct;275(5682):749–51.

15. Barkan A, Goldschmidt-Clermont M. Participation of nuclear genes in chloroplast gene expression. Biochimie. 2000 Jul;82(6–7):559–72.

16. Matsuo T, Onai K, Okamoto K, Minagawa J, Ishiura M. Real-Time Monitoring of Chloroplast Gene Expression by a Luciferase Reporter: Evidence for Nuclear Regulation of Chloroplast Circadian Period. Mol Cell Biol. 2006 Feb 1;26(3):863–70.

17. Franklin S, Ngo B, Efuet E, Mayfield* SP. Development of a GFP reporter gene for Chlamydomonas reinhardtii chloroplast. Plant J. 2002 Jun;30(6):733–44.

18. Rasala BA, Barrera DJ, Ng J, Plucinak TM, Rosenberg JN, Weeks DP. Expanding the spectral palette of fluorescent proteins for the green microalga Chlamydomonas reinhardtii. Plant J Cell Mol Biol. 2013 May;74(4):545–56.

19. Flowers JM, Hazzouri KM, Pham GM, Rosas U, Bahmani T, Khraiwesh B. Whole-Genome Resequencing Reveals Extensive Natural Variation in the Model Green Alga Chlamydomonas reinhardtii. Plant Cell. 2015 Sep;27(9):2353–69.

Acknowledgements

Chapter 4, in full, is a reprint of material in preparation for submission for publication. Ostrand, Joseph; Fields, Francis; Tran, Miller; Mayfield, Stephen. The dissertation author is the primary investigator and author of this paper.

CHAPTER 5:

SYNTHETIC PROMOTERS CAPABLE OF DRIVING ROBUST NUCLEAR GENE EXPRESSION IN THE GREEN ALGA *CHLAMYDOMONAS REINHARDTII*

Algal Research 15 (2016) 135–142

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*

Melissa A. Scranton [a], Joseph T. Ostrand [a], D. Ryan Georgianna [a,1], Shane M. Lofgren [b,2], Daphne Li [a], Rosalie C. Ellis [a], David N. Carruthers [a,3], Andreas Dräger [c,4], David L. Masica [d], Stephen P. Mayfield [a,*]

[a] California Center for Algae Biotechnology, Division of Biological Sciences, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0368, USA
[b] Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[c] Systems Biology Research Group, University of California, San Diego, 9500 Gilman Drive, La Jolla, 92093-0412, CA, USA
[d] Department of Biomedical Engineering, Institute for Computational Medicine, Johns Hopkins University, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 13 November 2015
Received in revised form 7 February 2016
Accepted 8 February 2016
Available online 27 February 2016

Keywords:
Algal engineering
C. reinhardtii
Synthetic biology
Nuclear promoters
Cis-motifs
Flow cytometry

ABSTRACT

Algae have enormous potential as bio-factories for the efficient production of a wide array of high-value products, and eventually as a source of renewable biofuels. However, tools for engineering the nuclear genomes of algae remain scarce and limited in functionality. In this study, synthetic algal promoters (*saps*) were generated as a tool for increasing nuclear gene expression and as a model for understanding promoter elements and structure in green algae. Promoters were generated to mimic native cis-motif elements, structure, and overall nucleotide composition of top expressing genes from *Chlamydomonas reinhardtii*. Twenty five *saps* were used to drive expression of a fluorescent reporter in transgenic algae. A majority of the promoters were functional *in vivo* and seven were identified to drive expression of the fluorescent reporter better than the current best endogenous promoter in *C. reinhardtii*, the chimeric *hsp70/rbs2* promoter. Further analysis of the best synthetic promoter, *sap11*, revealed a new DNA motif essential for promoter function that is widespread and highly conserved in *C. reinhardtii*. These data demonstrate the utility of synthetic promoters to drive gene expression in green algae, and lays the groundwork for the development of a suite of *saps* capable of driving the robust and complex gene expression that will be required for algae to reach their potential as an industrial platform for photosynthetic bio-manufacturing.

1. Introduction

Algae are among the most ancient and diverse organisms on the planet. Microalgae have evolved to adapt to a wide range of environments and consequently have proven to be a rich source of genetic and chemical diversity [7,23,41]. This diversity has been exploited as a unique source of bioactive compounds, including antioxidants, omega 3 fatty acids, and potentially novel therapeutic drugs [9]. In addition, microalgae have also proven to be cost-effective and safe hosts for expressing a wide array of recombinant proteins, including human and animal therapeutics, vaccines, and industrial enzymes [22,25,44,51].

Chlamydomonas reinhardtii is a long established model system for studying molecular and genetic systems of algae. The most successful advances in recombinant protein expression within *C. reinhardtii* have been within the chloroplast where exogenous protein levels have reached almost 10% of total soluble protein [37]. This progress has been aided by the fact that gene integration occurs exclusively by homologous recombination within the plastid [18]. The chloroplast also has strong, well-characterized promoters and regulatory untranslated regions (UTRs) to enable high levels of transgene expression [44,51]. The most successful regulatory elements are those from endogenous highly expressed photosynthetic proteins [24,44,51]. However, recent work in the Mayfield laboratory has shown that high-throughput analysis of synthetic 5′ UTRs can identify novel regulatory elements and lead to increased transgene expression within the plastid [50].

While advancements have been made in heterologous nuclear gene expression in *C. reinhardtii* over the last several years [42,43,47], these

* Corresponding author.
E-mail addresses: mscranton@ucsd.edu (M.A. Scranton), jostrand@ucsd.edu (J.P. Ostrand), ryan.georgianna@regi.com (D.R. Georgianna), slofgren@stanford.edu (S.M. Lofgren), dal029@ucsd.edu (D. Li), rellis@ucsd.edu (R.C. Ellis), dcarruth@umich.edu (D.N. Carruthers), andraeger@eng.ucsd.edu (A. Dräger), dmasica2@jhu.edu (D.L. Masica), smayfield@ucsd.edu (S.P. Mayfield).
[1] Current address: REG Life Sciences, 600 Gateway Blvd, South San Francisco, CA 94080, USA.
[2] Current Address: Institute for Immunity, Transplantation and Infection, and Department of Medicine, Stanford University School of Medicine, CA 94305, USA.
[3] Current Address: Chemical Engineering Department, University of Michigan, Ann Arbor, MI 48109, USA.
[4] Current Address: Center for Bioinformatics Tuebingen (ZBIT), University of Tuebingen, Sand 1, Tübingen 72076, Germany.

http://dx.doi.org/10.1016/j.algal.2016.02.011
2211-9264/

tools still lag significantly behind both plastid gene expression in algae, as well as heterologous gene expression in many other eukaryotic organisms. Controlled nuclear gene expression is an essential tool for synthetic biology in any industrial microorganisms. Recent advances also allow protein products to be targeted to any cellular location in *C. reinhardtii* [42]. Targeted expression is essential for metabolic engineering, since enzymes need to be localized to their functional site. Proper localization is also important for the production of high-value protein products. Specific organelles may be better suited for proper post-translational modification and folding of complex proteins. In particular, chloroplasts lack the enzymes involved in protein glycosylation, an essential modification for many therapeutic proteins [33]. Finally, nuclear expression allows for the secretion of recombinant proteins, which can lead to simpler and cheaper downstream processing [12].

One of the main reasons for poor heterologous gene expression from the nuclear genome of algae is the lack of strong promoters [44,51]. Studies have identified several endogenous promoters that promote exogenous gene expression, including those from the well-characterized and highly expressed genes such as those for the Rubisco small subunit (*RBCS2*), heat shock protein 70A (*HSP70A*), and photosystem I protein *psaD* [11,17,47]. In an attempt to increase expression above the modest levels achieved with these native promoters, chimeric promoters have been developed that contain the heat shock 70A promoter region fused upstream of the *RBCS2* promoter (*ar1*), which has led to increased transcription [46,47,54]. However, protein accumulation from exogenous genes expressed using this best chimeric promoter is still poor, with recombinant protein levels peaking around 0.25% of total soluble protein, which is well below the level of economic viability for almost any recombinant protein product. Finally, viral promoters that are favored in higher plant expression systems have been shown to be minimally successful in algal systems [14]. Therefore, novel regulatory elements must be identified or generated and combined into robust promoters capable of driving high rates of transcription in order to achieve the robust exogenous protein expression required to make algae a true industrial organisms.

Several recent reviews have highlighted the generation of synthetic promoters and promoter libraries as important biobricks for protein expression and, in particular, systems engineering [6,28,40,45]. Engineered promoters have demonstrated the ability to drive exogenous gene expression above levels achieved by the best native promoter systems. In addition, development of libraries of designer promoters is essential for systems engineering. The synthetic nature of these promoters reduces or eliminates the chance of homology dependent gene silencing and can potentially allow them to be utilized in multiple species or cell lines. In this study, publicly available mRNA expression data was utilized to identify cis-motifs found in promoters of highly expressed *C. reinhardtii* genes. These motifs were then used to generate a novel set of completely synthetic algal promoters (*saps*) that allowed for high constitutive gene expression within the *C. reinhardtii* nucleus. A combination of analysis of these native promoters and novel *saps* revealed previously uncharacterized *C. reinhardtii* promoter structures including a newly identified core DNA motif important for promoter function in highly transcribed genes.

2. Results

2.1. Native motif identification and saps generation

In order to generate *saps* capable of driving high heterologous gene expression, native *C. reinhardtii* genes were analyzed that showed the highest RNA accumulation in wild type (wt) cells grown under ambient conditions. The top 50 genes were identified based on previously published RNA-seq data [16] (Table S1). This data set was chosen because the growth conditions best match typical ambient small scale laboratory growth conditions for green algae. Promoter regions (−1000 to +50 nt from the transcription start site) from these genes were analyzed using the POWRS software [13]. POWRS identifies motifs based not only on enriched sequences but also on the position of these elements within the promoter region. POWRS clusters sequences together based on similarity to create motif clusters that can be characterized by position weight matrixes. POWRS identified 127 motif clusters containing 979 unique motifs within the top 50 native gene promoters (Table S2; Figs. S1 & S2). Upon inspection of the motifs, nine TC rich motifs were identified, some of which were localized around the transcription start site (TSS; Fig. S3). In *Arabidopsis thaliana*, a TC-like motif near the TSS may function similarly to the TATA box [4]. Therefore, these TC rich motifs were added to every synthetic promoter and enriched around the TSS.

Analysis of the top 50 native promoters also revealed that there is a decrease in the GC content within 500 bp around the transcription start site (Fig. 1a). This trend is in direct contrast with the promoters of higher plant species, which skew towards higher GC content near the TSS [8,20]. *C. reinhardtii* promoter GC content structure most resembles *Saccharomyces cerevisiae* and some prokaryotic species that increase AT-content towards the TSS. This trend in *C. reinhardtii* does not appear to be due simply to the higher overall GC content of its nuclear genome, since species like the red alga *Cyanidioschyzon merolae* also have high GC content but have an increase in GC towards the TSS [8]. In addition to a general AT-increase at the TSS, there also appeared to be smaller dips in GC content at approximately −280 and −140 bp upstream of the TSS. These AT-rich regions have a similar periodicity as that of nucleosome wrapped DNA, which is around 147 bp [34]. These AT-rich regions were incorporated in the synthetic promoters.

Synthetic promoters were generated to include nucleotide backbones that had a similar GC profile as the native promoters, including the aforementioned AT-bias towards the TSS and AT rich regions at −280 and −140 bp (Fig. 1a). Promoters were designed to be 500 bp in length for ease of synthesis and analysis. Since many motifs are localized across and downstream of the TSS, promoters were designed to mimic −450 bp upstream and 50 bp downstream of the TSS in order to not cutoff important motifs. This is a similar strategy to previous native hybrid promoter designs [47]. Motifs were overlaid onto nucleotide backbones constrained to a similar region to where they were found in the native sequences ([13]; Figs. S1 & S2, Fig. 1b).

2.2. Synthetic promoters drive transcription in vivo

Twenty five *saps* were studied for their ability to drive the expression of the mCherry fluorescent reporter protein. The *saps* were synthesized and cloned in front of an *mCherry* reporter gene, which also contained the 5′ and 3′ *RBCS2* UTRs as well as the first *RBCS2* intron (Fig. 1c). These elements have all been previously shown to improve mRNA accumulation and protein synthesis of heterologous genes in *C. reinhardtii* [36,42]. The vector construct also included a *hygromycin* resistance cassette, which was driven by the *beta tubulin* (*TUBB2*) promoter to select for transformed algae independent of synthetic promoter function [5]. This allowed large scale mCherry analysis of all promoters including weak or non-functioning promoters.

Transformation of the *C. reinhardtii* nucleus occurs almost exclusively through non-homologous end-joining [26,49]. This results in random insertion, multiple insertions, and highly variable exogenous gene expression. Typical promoter analysis involves measuring the expression of 10–50 individual transformants. However, measuring individual transformants is time and resource consuming, and the variability in expression is still high unless many individual are measured. Alternatively, if many transformants are pooled and protein or RNA levels are measured of the total population, noise from positional insertion effects can be reduced, but this does not allow measurement of the range of expression over the population pool. Therefore, for this study flow cytometry was used to measure promoter strength. Flow cytometry allows measurement of both a large number of transformants while also

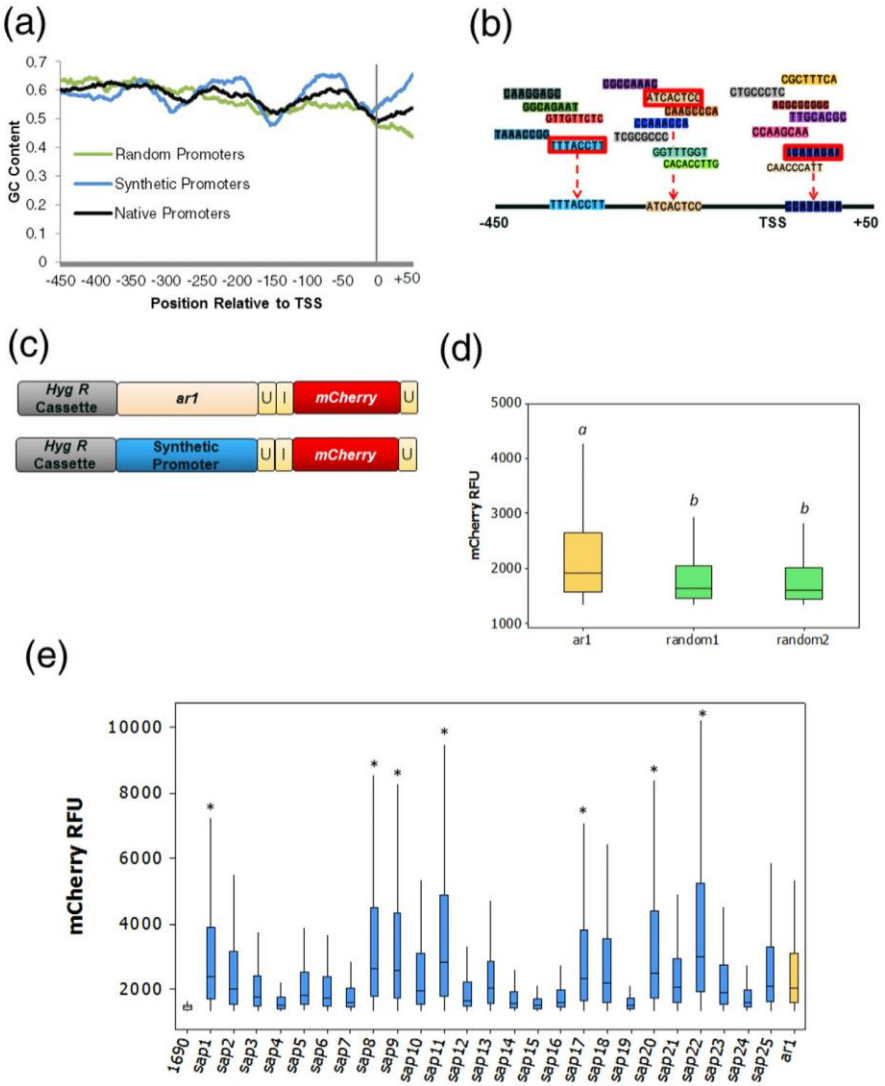

Fig. 1. Design of synthetic algal promoters and expression vector construction. (a) Relative GC content of the top 50 native promoters was analyzed (moving window 20 bp). Synthetic and random promoters were generated to mimic the AT-skew. (b) Motifs discovered in the top 50 native promoters were placed in a synthetic backbone in positions similar to their position in the native promoters. The overall promoter was designed to mimic −450 to +50 bp relative to TSS. (c) Synthetic algal promoters (*saps*) were placed upstream of mCherry expression cassette, which included the *RBCS2* 5′ and 3′ UTR (U) and first intron (I) in order to drive expression. A separate *hygromycin* expression cassette was place upstream of the mCherry cassette to allow for screening of transformants independent of synthetic promoter function. Synthetic promoters were compared to the *hsp70/rbcs2* hybrid promoter (*ar1*). (d) Randomly generated sequences are used to drive mCherry. The relative mCherry fluorescence of 5000 transformants is compared to 5000 transformants of the *ar1* construct by flow cytometry. Populations that are statistically different are indicated (a–b, Tukey's test, $p < 0.05$) Box and whisker plot indicates max (top of line), min (bottom of line), first quartile (bottom of box), second quartile (median; middle line), third quartile (top of box). (e) *sap* transformants were compared to *ar1* transformants by flow cytometry. Populations transformed with seven of the sap promoters have more mCherry fluorescence than *ar1* transformed cells (*, Tukey's test, $p < 0.05$).

recording the data for individual transgenic cells. This provides a highly confident average as well as the range of expression for our reporter gene for each promoter tested.

To determine if our synthetic promoters were functional based on our design principles, and not just coincidental, random promoters were also generated whose sequence had a similar GC content to both

native and our synthetic promoters, but with no periodical AT rich regions upstream or placement of motifs (Fig. 1a, Table S3). These promoters would also serve as a negative control for random positional effects since exogenous gene expression can occur simply due to insertion next to a native promoter [29].

Analysis of mCherry expression driven by the 25 *saps* revealed a wide range of functionality compared to *ar1*. As expected, there was low level of mCherry fluorescence above the WT background in our random promoter transformants (Fig. 1d). It is important to note that while five random promoters were generated, only two provided had enough mCherry positive transformants to perform proper statistical analysis and are shown in Fig. 1d. Multiple transformations and screenings were performed to increase the number of positive events for statistical analysis, but none could be successfully reproduced. Eight *saps* were found to be no better than these randomly generated promoters (Fig. 1e). However, 10 *saps* were not only better than our random controls, but were as good as *ar1*. Encouragingly, seven *saps* were actually better than *ar1* (Tukey HSD, $p < 0.05$) with both average and max mCherry fluorescent levels almost twice as high as *ar1*. These results were consistent over multiple transformations and screenings (Fig. S4a).

2.3. sap11 contains a positive cis-effector motif

In order to determine which motifs contribute to the promoter strength of the high-expressing *saps*, we chose *sap11* for further analysis, as it consistently produced the greatest amount of mCherry. First, a deletion series was performed in which nucleotides were deleted from the 5′ end so that −250, −150, or −50 bp upstream of the TSS remained (Fig. 2a). For this study, the expression vector was rearranged so that the *hygromycin* resistance cassette was downstream of the *mCherry* cassette. This rearrangement avoided any confounding data due to the relative shift of the position of the 3'UTR from the *hygromycin* cassette after promoter deletion. Rearrangement did not affect the promoter function of either *ar1* or *sap11* (Fig. S4b). The relative mCherry fluorescence from *sap11* in this rearranged vector was unchanged from the original design (Figs. 1e & 2b). Analysis of mCherry fluorescence in *sap11*Δ mutants revealed only a slight reduction in expression in *sap11*Δ*-250* and *sap11*Δ*-150* mutants (Fig. 2b). However, a significant drop in expression was observed in *sap11*Δ*-50* where there was no expression above those found for the random promoters. These results are consistent with the fact that core motifs are often found within 200 bp upstream of the TSS [3,38,55].

To further narrow down specific motifs essential for *sap11* function, motif deletion analysis was performed. Four regions contained POWRS identified motifs between −150 and −50 bp from the TSS (Fig. 2b). Eight A residues were used to replace the entire motif or the majority of the bases of the motif for those longer than 8 nucleotides. For motif 2, polyT residues were used to replace the motif since the region was highly A rich. Motif 5 comprised of a TC-rich motif that resided around the TSS. This motif was also deleted since it is homologous to the TC motifs found in Arabidopsis, and was therefore thought to be a functional element [4]. However, deletion of motif 5 (*sap11*Δ*m5*) did not result in significant reduction in mCherry production (Fig. 2d). Therefore, either this particular iteration of the motif was not utilized in *sap11* or

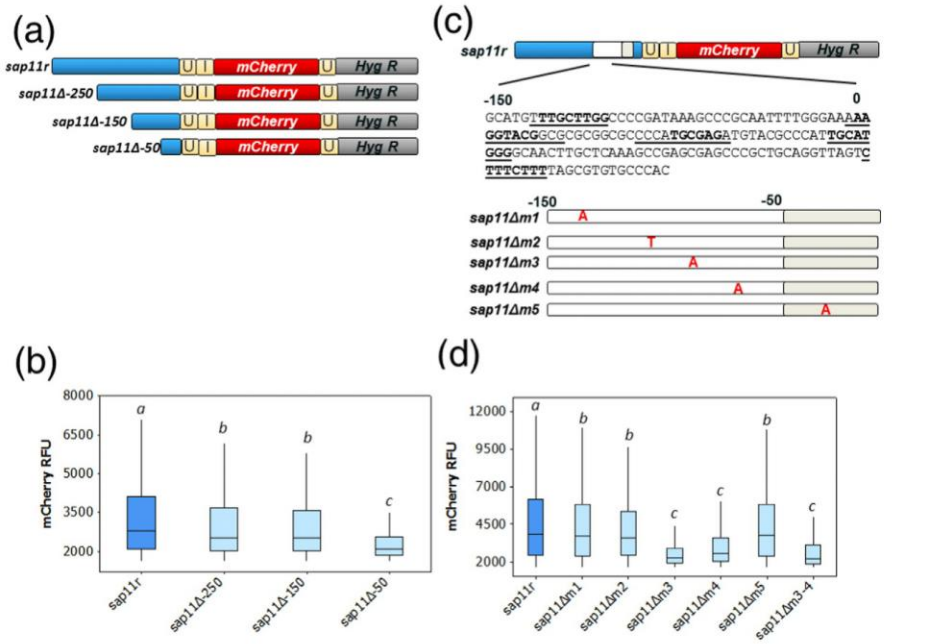

Fig. 2. Promoter and motif deletions of *sap11*. (a) The expression vector was rearranged to have the hygromycin resistance cassette downstream of the mCherry cassette. *sap11* was cloned upstream of the mCherry cassette with the *rbcs2* 5′ and 3′ UTRs (U) and the first *rbcs2* intron (I). Portions of the *sap11* promoter were removed through SLiCE cloning to leave −250, −150, and −50 bp of *sap11* sequence upstream of the *sap11* TSS. (b) Flow cytometry analysis for mCherry fluorescence of 5000 transformants of the original and shortened sap11 constructs. Populations that are statistically different are indicated (a–c, Tukey's test, $p < 0.05$). (c) Putative cis-motifs (underlined) in the −150 to 0 bp region of sap11 were targeted for mutational analysis. Eight residues (bold) were replaced with either polyA (A) or polyT (T) residues to generate six *sap11*Δ*m* mutants including one in which both motif 3 and 4 were replaced (*sap11*Δ*m3–4*). (d) Flow cytometry analysis for mCherry fluorescence of 5000 transformants of the *sap11* construct compared with *sap11* motif deletion constructs.

the TC motifs are not essential in *C. reinhardtii*. The deletion of both motif 3 and 4 (*sap11Δm3* and *sap11Δm4*) resulted in significant decreases in promoter function, while deletion of motif 1 and 2 (*sap11Δm1* and *sap11Δm2*) had little effect. Interestingly, regions 3 and 4 have nearly identical reverse complement motifs (CCCATGCGA and TGCATGGG, respectively), suggesting they could be targeted by the same transcription factor. In order to determine if regions 3 and 4 were redundant, a double mutant was generated in which both regions were replaced with polyA nucleotides (*sap11Δm3–4*). This promoter functioned similarly to the individual motif 3 and 4 KOs, suggesting that motif 4 may be redundant with motif 3 or that KO of motifs 3 and 4 already eliminate any expression above background (Fig. 2d). It is important to note while this motif was essential for promoter function in *sap11*, this motif alone is not sufficient for expression as several of the non-functioning *saps* also contained this motif in a similar location (Figs. S1 & S2).

Because the CCCAT motifs had such a significant impact on *sap11* function, we set out to determine if it may be a core motif within *C. reinhardtii*. One method to identify core motifs is to identify motifs that are relatively enriched at specific locations relative to the TSS in a large number of promoters. Therefore, we analyzed the promoter regions of 4412 genes in *C. reinhardtii* (Table S4). Promoters were chosen if their 5′ UTR start sites (*C. reinhardtii* v5.5) were within 10 bp of the start site of PASA (Program to Assemble Spliced Alignments; Phytozome 10.2) assembled EST. Promoter sequences from −1000 to +500 of the 5′ UTR site were analyzed to identify motifs that are enriched in similar regions [2]. Surprisingly, the top eight motifs identified were all CCCAT-like motifs that were highly enriched only at −100 to −40 bp upstream of the TSS with a peak at −65 bp (Fig. 3a). Moreover, 10.6% (467 promoters) of all the promoters analyzed had exactly CCCATGCA sequence at this location, while 35.4% (1564 promoters) had some variation of this motif at this location (Table S5). This suggests that the CCCAT motif is a core motif within the *C. reinhardtii* promoter.

Motif sequence similarity search using TOMTOM analysis of this motif sequence revealed some homology to the cis-motif recognized by the Arabidopsis phytochrome interacting factor (PIFs; Fig. S5, Table S6; [27]). PIFs are involved in light-regulated gene expression [10]. Similarly, functional analysis of CCCAT motif-containing genes revealed enrichment in pathways that are diurnally regulated (e.g. Ribosomes, antenna proteins; Table S7). However, the CCCAT motif was found in over 1500 genes, the vast majority of which were not diurnally regulated (<5% overlap with differentially regulated genes identified in [57]). The role the CCCAT motif within the context of these native promoters remains to be determined. Interestingly, only one helix–loop–helix transcription factor (Cre14.g620850) could be identified in *C. reinhardtii* with homology to the PIF proteins in Arabidopsis, based on amino acid similarity (Table S6). It will be interesting to determine if this putative transcription factor can bind to the CCCAT motif in *C. reinhardtii*. If it does, it most likely has a unique function compared to Arabidopsis based on its target genes in *C. reinhardtii*.

2.4. C. reinhardtii promoters contain AT and TC rich motifs near TSS

CentriMo analysis of the *C. reinhardtii* promoters revealed other motifs that were enriched at specific regions relative to the TSS. Of note, AT-rich motifs appeared to peak at the TSS and then at periodic but decreasing intervals both upstream and downstream of the TSS (Fig. 3b). These intervals appeared ~130 bp apart from each other. These regions correspond to the AT-rich regions found in the top 50 genes (Fig. 1a), and when the relative GC content is analyzed in the larger genomic promoter set a similar pattern of AT-rich regions is seen (Fig. S6). Initially this periodicity suggests a relationship to nucleosome positioning. However, nucleosomes in *C. reinhardtii* protect 147 bp of DNA and typically have a period of ~170 bp [19,34]. Interestingly, this period more closely follows the period of 6 mA methylated sites around the TSS which have a period of ~134 bp [19]. However, the AT-rich sites are not located at the same

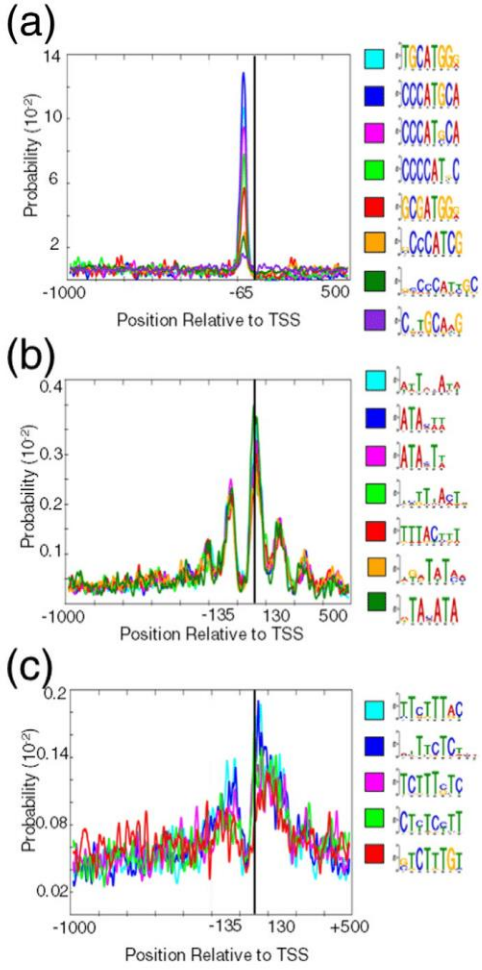

Fig. 3. Locally enriched POWRS and DREME motifs in top 4412 promoters from *C. reinhardtii* nuclear genome. EST validated promoters were analyzed with CentriMo for locally enriched motifs. Relative enrichment of motifs relative to the TSS for the top three categories of motifs is shown (a–c).

position as either the nucleosomes or the 6 mA sites. Finally, CentriMo analysis found TC rich motifs that were enriched around the TSS of *C. reinhardtii* promoters. However, their enrichment was far less significant than the CCCAT or TA rich motifs (Fig. 3c). This is consistent with the motif deletion analysis that demonstrated that this motif is not essential in the *sap11* promoter.

3. Discussion

In this study, synthetic promoters were successfully generated that were capable of driving exogenous gene expression within the *C. reinhardtii* nucleus. The *saps* generated in this study were based on native DNA motifs identified using the POWRS algorithm. Using a

stochastic method of motif placement that was based on motif location relative to the TSS in native promoters, we were able to generate *saps* that were as successful as, or better than, the best native promoters in *C. reinhardtii* [46,47]. The current best promoter for *C. reinhardtii* is a non-native promoter *ar1* that is a hybrid between two endogenous promoter regions. Our novel *saps* rely on a completely synthetic promoter backbone with a cis-regulatory motif structure informed from annotation based and experimentally derived genomic information. It should be noted that the *HSP70A* promoter acts as a transcriptional state enhancer, which increases the probability of transcription of the neighboring promoter [35]. It would be interesting to see if fusing the *HSP70A* promoter upstream our synthetic promoter further improves their function similarly to *HSP70A*'s effect on *RBCS2*. Alternatively, our promoters could also be fused with other native 5′ and 3′ UTRs, such as *psaD*, which in one study showed similar improvements over *ar1* for luciferase expression [32].

Bioinformatic analysis used to identify motifs within native promoters led to the identification of novel elements as well as information about promoter structure within the nuclear genome of *C. reinhardtii*. First, *C. reinhardtii* promoters have an AT-bias near the TSS, which is unique from other plant species studied thus far (Fig. 1a; [8,20]). This bias more than likely affects the structure of the DNA in this location and may be important for nucleosome organization or other DNA–protein interactions [21,30]. In addition to an overall AT-bias, there were also pockets of AT-rich regions upstream of the TSS, which correlated with AT-rich motifs found in the EST validated promoters (Figs. 1a & 3b). The pattern of the AT-rich regions corresponds to a similar periodic pattern of 6 mA methylation sites around the TSS, but is shifted by ~30 bp [19]. It has been suggested that the periodicity of the 6 mA sites may help establish nucleosome organization around the TSS. Therefore, the AT-bias with specific AT-rich periodic regions may work together with the 6 mA methylation site to establish nucleosome packing and encourage transcription factor and RNA polymerase binding around the TSS.

In addition to AT-rich regions, TC-rich regions were also enriched in *C. reinhardtii* promoters. This enrichment was more significant in the top 50 expressed genes compared to the genome-in whole (Table S2, Fig. 3c). This enrichment in top expressed genes is consistent with similar motifs found in *Arabidopsis* [4]. However, when this motif was removed from *sap11*, there was little loss in promoter function. It is important to note that TC motif analysis in *Arabidopsis* was only performed *in silico*. Therefore, the relative importance or function of these motifs has yet to be established *in vivo*. It is also possible that this motif is a consequence of the relative AT enrichment around the TSS and only its relative AT content is important. Since the motif was replaced with a polyA sequence, the AT content was not significantly changed. Further work is still required to rule out the relevance of the TC-rich motifs in *C. reinhardtii*.

Promoter motif deletion analysis did reveal the presence of an essential motif within the *sap11* promoter. Motif regions 3 and 4 contained nearly identical CCCAT motifs. Knock out of these motifs led to severe reduction of *sap11* function. Bioinformatic analysis further revealed that this motif is highly enriched at −65 bp upstream of the TSS of 1564 genes with 446 having the exact CCCATGCA sequence (Table S5, Fig. 3a). However, many versions of the CCCAT motif contain the conserved CATG 6 mA sequence [19]. Therefore, the CCCAT motif may function as a target for DNA methylation in its role in transcriptional regulation. While one putative *C. reinhardtii* transcription factor has been predicted to bind to the CCCAT motif based on *in silico* homology analysis, further *in vitro* and *in vivo* work is required to identify the true transcription factor partner.

The combination of bioinformatic analysis of gene structure and expression and *in vivo* testing of synthetic primers based on these analyses has proven a fruitful area of research for discovery of unknown cis elements and for use in designing strong synthetic promoters [6,31,53]. The knowledge gained in this study gives us a synthetic template to generate large promoter libraries. These libraries will be used to generate more significant data about the importance of individual motifs and overall promoter structure in *C. reinhardtii*, which will ideally enable us to generate successive rounds of engineered promoters to achieve exogenous gene expression above currently achieved levels. Large promoter libraries will also allow for the integration of multiple genes into the same host by allowing separate transgenes to be driven by unique promoters to reduce genomic rearrangements brought about by sequence specific targeting that may arise from a genome laced with identical sequences. This latter feature is particularly important in metabolic engineering, which often requires the introduction of multiple enzymes into the host organism. Finally, as we have demonstrated in this study, synthetic promoters provide a platform on which to identify motifs *in vivo* involved in transcriptional regulation in *C. reinhardtii*. In the future, this can be expanded to motifs predicted to be involved in inducible regulation such as heat shock, nickel or nitrate addition or iron-deficiency. Together these tools will represent a large step forward in the synthetic engineering of algae for the production of biofuels and bio-products.

4. Materials and methods

4.1. POWRS motif identification

The top 50 highest-expressed endogenous genes were identified based on their RNA accumulation under ambient conditions according to previously published RNA-seq data [16]. Since promoter structure is not strictly defined in *C. reinhardtii* the sequence between −1000 and +50 for the top 50 genes were analyzed using the POWRS motif identification program [13] (Phytozome 10.2, *C. reinhardtii* v5.5). All default settings on POWRS were used, except that the minimum number of sequences that a valid motif must match was lowered to ten.

4.2. Generation of synthetic promoters

Promoters were generated using random insertion of POWRS motifs, constraining positions relative to the positions of the motif clusters in the native sequences. Promoter backbones were generated to ensure similar GC content as the native promoters, including a periodic AT-rich regions (Fig. 1a). Finally, all promoters contained at least one copy of a TC rich motif around the TSS (Figs. S1 & S2). Random promoters were generated by choosing 500 random nucleotides based on the Markov model that described the native promoter GC content without periodic AT-rich regions (Table S3).

4.3. Plasmid construction

The synthetic algal promoters were synthesized as gBlocks (IDT, Coralville, IA) integrating in DNA ends that allowed cloning *via* SLiCE technology [56](Table S8). All restriction enzymes were purchased from New England Biolabs (Ipswich, MA). The pBR4 expression vector with the hygromycin B resistance gene under the control of the B-tubulin promoter and a separate cassette with the mCherry gene driven by the *ar1* promoter was used as the backbone [5,43]. pBR4 was digested with NdeI and XbaI to remove the *ar1* promoter up to end of the *RBCS2* 5'UTR and generate ends for SLiCE cloning. Synthetic promoters were cloned with the *RBCS2* 5'UTR, which was amplified with appropriate primers to allow 15 bp overhangs with the synthetic promoters as well the digested backbone (Table S8), resulting in the constructs in Fig. 1b. To rearrange *sap11* with the *hygromycin* cassette downstream of the mCherry cassette, each half of pBR4 was amplified with appropriate primers for USER cloning into the HCR1, a modified pBlueScript II (Agilent, Santa Clara, CA), as previously described [52] (Table S8). The rearranged construct was then digested with NdeI and XbaI to remove ar1 and replace it with *sap11* which was PCR amplified and SliCE cloned into the rearranged pBR4. Promoter and motif

deletions were performed by SLiCE cloning. polyA and polyT mutations were introduced using overlapping primers and PCR pieces generated were cloned into a pBR4-rearranged backbone which had been digested with EcoRI and NdeI (Table S8). All constructs were confirmed by restriction digest and sequencing.

4.4. C. reinhardtii growth and transformation

Wild-type (cc1690) *C. reinhardtii* were grown and transformed using the methods described previously using 1 μg of plasmid DNA [43]. Plasmid constructs were digested with KpnI to linearize them prior to electroporation. Transformants were first screened on TAP (Tris–acetate–phosphate)/agar plates containing 15 μg/ml hygromycin, resulting in approximately 5000 to 12,000 transformants per selection. The entire transformant pool was then collected and transferred to liquid TAP medium for two days, followed by screening on the flow cytometer.

4.5. Flow cytometry measurement of mCherry fluorescence

mCherry fluorescence was visualized by a BD LSRII flow cytometer and analyzed using FlowJo v10.0.8. The population was gated using the following strategy: the FSC and SSC parameters were obtained using a 488 nm blue laser and were used to eliminate smaller non-algal samples and clumps of algae that can be misread as a single cell. Next, the 488 nm laser using a 685LP and a 710/50 filter set was used in combination with a 405 nm violet laser and 450/50 filter to remove dead cells and remaining debris from the population. The mCherry fluorescence was then measured with a 561 nm yellow/green laser with a 600LP and 610/20 filter set. To better visualize the population, the mCherry fluorescence channel was plotted against the window created by the 405 nm laser with a 505LP and 535/30 filter set. Using the untransformed parent strain as a reference, the events containing only background fluorescence were removed from the analysis. What remained was considered single-cell, living, *C. reinhardii* that is expressing mCherry. A representative window was selected from the remaining population and the mCherry fluorescence channel was broken down into individual events, resulting in 80 to 10,000 data points.

4.6. Genomic promoter motif analysis

For whole genome promoter analysis, genome sequence and annotation for Creinhardtii_281_v5.5 was obtained from phytozome.jgi.doe.gov [39]. Annotated 5′ UTR start sites were compared to PASA assembled EST start sites. Only 4412 of the 22,892 total annotated 5′ UTR start sites were within 10 bp of a PASA EST start site and considered EST validated sites. Sequence from −1000 bp upstream to +500 bp downstream of the validated 5′ UTR start sites was analyzed for new motifs using DREME [1]. Then the promoter sequences were analyzed by CentriMo to identify POWRS or DREME motifs that are enriched in specific regions relative to the TSS [2].

Acknowledgments

This work was funded by the US Department of Energy (DE-EE0003373; SPM) and a contract from the California Energy Commission (CILMSF #500-10-039).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.algal.2016.02.011.

References

[1] T.L. Bailey, DREME: motif discovery in transcription factor ChIP-seq data, Bioinformatics 27 (2011) 1653–1659.
[2] T.L. Bailey, P. Machanick, Inferring direct DNA binding from ChIP-seq, Nucleic Acids Res. 40 (2012), e128.
[3] K.W. Berendzen, K. Stuber, K. Harter, D. Wanke, Cis-motifs upstream of the transcription and translation initiation sites are effectively revealed by their positional disequilibrium in eukaryote genomes using frequency distribution curves, BMC Bioinformatics 7 (2006) 522.
[4] V. Bernard, V. Brunaud, A. Lecharny, TC-motifs at the TATA-box expected position in plant genes: a novel class of motifs involved in the transcription regulation, BMC Genomics 11 (2010) 1–15.
[5] P. Berthold, R. Schmitt, W. Mages, An engineered *Streptomyces hygroscopicus aph 7″* gene mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*, Protist 153 (2002) 401–412.
[6] J. Blazeck, H. Alper, Promoter engineering: recent advances in controlling transcription at the most fundamental level, Biotechnol. J. 8 (2013) 46–58.
[7] J.W. Blunt, B.R. Copp, R.A. Keyzers, M.H.G. Munro, M.R. Prinsep, Marine natural products, Nat. Prod. Rep. 29 (2012) 144–222.
[8] E. Calistri, R. Livi, M. Buiatti, Evolutionary trends of GC/AT distribution patterns in promoters, Mol. Phylogenet. Evol. 60 (2011) 228–235.
[9] K.H.M. Cardozo, T. Guaratini, M.P. Barros, V.R. Falcão, A.P. Tonon, N.P. Lopes, S. Campos, M.A. Torres, A.O. Souza, P. Colepicolo, E. Pinto, Metabolites from algae with economical impact, Comp. Biochem. Physiol. Toxicol. Pharmacol. 146 (2007) 60–78.
[10] A. Castillon, H. Shen, E. Huq, Phytochrome interacting factors: central players in phytochrome-mediated light signaling networks, Trends Plant Sci. 12 (2007) 514–521.
[11] H. Cerutti, A. Johnson, N. Gillham, J. Boynton, A eubacterial gene conferring spectinomycin resistance on *Chlamydomonas reinhardtii*: integration into the nuclear genome and gene expression, Genetics 145 (1997) 97–110.
[12] J. Corchero, B. Gasser, D. Resina, W. Smith, E. Parrilli, F. Vázquez, I. Abasolo, M. Giuliani, J. Jäntti, P. Ferrer, M. Saloheimo, D. Mattanovich, S. Schwartz, M. Tutino, A. Villaverde, Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics, Biotechnol. Adv. 31 (2013) 140–153.
[13] I. Davis, C. Benninger, P. Benfey, T. Elich, POWRS: position-sensitive motif discovery, PLoS One 7 (2012), e40373.
[14] E. Díaz-Santos, M. de la Vega, M. Vila, J. Vigara, R. León, Efficiency of different heterologous promoters in the unicellular microalga *Chlamydomonas reinhardtii*, Biotechnol. Prog. **29** (2013) 319–328.
[16] W. Fang, Y. Si, S. Douglass, D. Casero, S. Merchant, M. Pellegrini, I. Ladunga, P. Liu, M. Spalding, Transcriptome-wide changes in *Chlamydomonas reinhardtii* gene expression regulated by carbon dioxide and the CO_2-concentrating mechanism regulator CIA5/CCM1, Plant Cell 24 (2012) 1876–1893.
[17] N. Fischer, J. Rochaix, The flanking regions of *PsaD* drive efficient gene expression in the nucleus of the green alga *Chlamydomonas reinhardtii*, Mol. Gen. Genomics. 265 (2001) 888–894.
[18] N. Fischer, O. Stampacchia, K. Redding, J.D. Rochaix, Selectable marker recycling in the chloroplast, Mol. Gen. Genet. 251 (1996) 373–380.
[19] Y. Fu, G.Z. Luo, K. Chen, X. Deng, M. Yu, D. Han, Z. Hao, J. Liu, X. Lu, L.C. Dore, X. Weng, Q. Ji, L. Mets, C. He, N^6-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*, Cell **161** (2015) 879–892.
[20] S. Fujimori, T. Washio, M. Tomita, GC-compositional strand bias around transcription start sites in plants and fungi, BMC Genomics 6 (2005).
[21] A.E. Gabrielian, D. Landsman, A. Bolshoy, Curved DNA in promoter sequences, In Silico Biol. 1 (1999) 183–196.
[22] D.R. Georgianna, J.H. Michael, M. Marina, W. Shuiqin, B. Kyle, J.L. Alex, H. James, M. Michael, P.M. Stephen, Production of recombinant enzymes in the marine alga *Dunaliella tertiolecta*, Algal Res. **2** (2013) 2–9.
[23] J. Gimpel, E. Specht, D. Georgianna, S. Mayfield, Advances in microalgae engineering and synthetic biology applications for biofuel production, Curr. Opin. Chem. Biol. 17 (2013) 489–495.
[24] J.A. Gimpel, S.P. Mayfield, Analysis of heterologous regulatory and coding regions in algal chloroplasts, Appl. Microbiol. Biotechnol. 97 (2013) 4499–4510.
[25] C. Griesbeck, A. Kirchmayr, Algae: an alternative to the higher plant system in gene farming, in: A. Wang, S. Ma (Eds.), Molecular Farming in Plants: Recent Advances and Future Prospects, Springer Science & Business Media, Dordrecht, Netherlands 2012, pp. 125–143.
[26] N.J. Gumpel, J.D. Rochaix, S. Purton, Studies on homologous recombination in the green alga *Chlamydomonas reinhardtii*, Curr. Genet. **26** (1994) 438–442.
[27] S. Gupta, J.A. Stamatoyannopoulos, T.L. Bailey, W.S. Noble, Quantifying similarity between motifs, Genome Biol. 8 (2007) R24.
[28] K. Hammer, I. Mijakovic, P. Jensen, Synthetic promoter libraries-tuning of gene expression, Trends Biotechnol. 24 (2006) 53–55.
[29] M.A. Haring, C.F. Beck, A promoter trap for *Chlamydomonas reinhardtii*: development of a gene cloning method using 5′ RACE-based probes, Plant J. 11 (1997) 1341–1348.
[30] A. Kanhere, M. Bansal, Structural properties of promoters: similarities and differences between prokaryotes and eukaryotes, Nucleic Acids Res. 33 (2005) 3165–3175.
[31] J. Koschmann, F. Machens, M. Becker, J. Niemeyer, J. Schulze, L. Bülow, D. Stahl, R. Hehl, Integration of bioinformatics and synthetic promoters leads to the discovery of novel elicitor-responsive cis-regulatory sequences in *Arabidopsis*, Plant Physiol. **160** (2012) 178–191.

[32] A. Kumar, V.R. Falcao, R.T. Sayre, Evaluating nuclear transgene expression systems in *Chlamydomonas reinhardtii*, Algal Res. 2 (2013) 321–332.
[33] N. Lingg, P. Zhang, Z. Song, M. Bardor, The sweet tooth of biopharmaceuticals: importance of recombinant protein glycosylation analysis, Biotechnol. J. 7 (2012) 1462–1472.
[34] M. Lodha, M. Schroda, Analysis of chromatin structure in the control regions of the *Chlamydomonas HSP70A* and *RBCS2* genes, Plant Mol. Biol. 59 (2005) 501–513.
[35] M. Lodha, M. Schulz-Raffelt, M. Schroda, A new assay for promoter analysis in *Chlamydomonas* reveals roles for heat shock elements and the TATA box in *HSP70A* promoter-mediated activation of transgene expression, Eukaryot. Cell 7 (2008) 72–176.
[36] V. Lumbreras, D. Stevens, S. Purton, Efficient foreign gene expression in *Chlamydomonas reinhardtii* mediated by an endogenous intron, Plant J. 14 (1998) 441–447.
[37] A.L. Manuell, M.V. Beligni, J.H. Elder, D.T. Siefker, M. Tran, A. Weber, T.L. McDonald, S.P. Mayfield, Robust expression of a bioactive mammalian protein in *Chlamydomonas* chloroplast, Plant Biotechnol. J. 5 (2007) 402–412.
[38] G.A. Maston, S.K. Evans, M.R. Green, Transcriptional regulatory elements in the human genome, Annu. Rev. Genomics Hum. Genet. 7 (2006) 29–59.
[39] S.S. Merchant, S.E. Prochnik, O. Vallon, E.H. Harris, S.J. Karpowicz, G.B. Witman, A. Terry, A. Salamov, L.K. Fritz-Laylin, L. Maréchal-Drouard, et al., The *Chlamydomonas* genome reveals the evolution of key animal and plant functions, Science 318 (2007) 245–250.
[40] S. Mukherji, A. van Oudenaarden, Synthetic biology: understanding biological design from synthetic circuits, Nat. Rev. Genet. 10 (2009) 859–871.
[41] M.S. Parker, T. Mock, E.V. Armbrust, Genomic insights into marine microalgae, Annu. Rev. Genet. 42 (2008) 619–645.
[42] B. Rasala, D. Barrera, J. Ng, T. Plucinak, J. Rosenberg, D. Weeks, G. Oyler, T. Peterson, F. Haerizadeh, S. Mayfield, Expanding the spectral palette of fluorescent proteins for the green microalga *Chlamydomonas reinhardtii*, Plant J. 74 (2013) 545–556.
[43] B.A. Rasala, P.A. Lee, Z.X. Shen, S.P. Briggs, M. Mendez, S.P. Mayfield, Robust expression and secretion of xylanase1 in *Chlamydomonas reinhardtii* by fusion to a selection gene and processing with the FMDV 2A peptide, PLoS One 7 (2012), e43349.
[44] S. Rosales-Mendoza, L.M.T. Paz-Maldonado, R.E. Soria-Guerra, *Chlamydomonas reinhardtii* as a viable platform for the production of recombinant proteins: current status and perspectives, Plant Cell Rep. 31 (2012) 479–494.
[45] C. Ruth, A. Glieder, Perspectives on synthetic promoters for biocatalysis and biotransformation, Chembiochem 11 (2010) 761–765.
[46] M. Schroda, C. Beck, O. Vallon, Sequence elements within an *HSP70* promoter counteract transcriptional transgene silencing in *Chlamydomonas*, Plant J. 31 (2002) 445–455.
[47] M. Schroda, D. Blöcker, C. Beck, The *HSP70A* promoter as a tool for the improved expression of transgenes in *Chlamydomonas*, Plant J. 21 (2000) 121–131.
[49] O.A. Sodeinde, K.L. Kindle, Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*, Proc. Natl. Acad. Sci. **90** (1993) 9199–9203.
[50] E. Specht, S.P. Mayfield, Synthetic oligonucleotide libraries reveal novel regulatory elements in *Chlamydomonas* chloroplast mRNAs, ACS Synth. Biol. 2 (2012) 34–46.
[51] E. Specht, S. Miyake-Stoner, S. Mayfield, Micro-algae come of age as a platform for recombinant protein production, Biotechnol. Lett. 32 (2010) 1373–1383.
[52] E.A. Specht, H.H. Nour-Eldin, K.T.D. Hoang, S.P. Mayfield, An improved ARS2-derived nuclear reporter enhances the efficiency and ease of genetic engineering in *Chlamydomonas*, Biotechnol. J. **10** (2015) 473–479.
[53] M. Venter, Synthetic promoters: genetic control through cis engineering, Trends Plant Sci. 12 (2007) 118–124.
[54] J. Wu, Z. Hu, C. Wang, S. Li, A. Lei, Efficient expression of green fluorescent protein (GFP) mediated by a chimeric promoter in *Chlamydomonas reinhardtii*, Chin. J. Oceanol. Limnol. **26** (2008) 242–247.
[55] Y.Y. Yamamoto, H. Ichida, M. Matsui, J. Obokata, T. Sakurai, M. Satou, M. Seki, K. Shinozaki, T. Abe, Identification of plant promoter constituents by analysis of local distribution of short sequences, BMC Genomics 8 (2007) 67.
[56] Y. Zhang, U. Werling, W. Edelmann, SLiCE: a novel bacterial cell extract-based DNA cloning method, Nucleic Acids Res. 40 (2012), e55.
[57] J.M. Zones, I.K. Blaby, S.S. Merchant, J.G. Umen, High-resolution profiling of a synchronized diurnal transcriptome from *Chlamydomonas reinhardtii* reveals continuous cell and metabolic differentiation, Plant Cell 10 (2015) 2743–2769.

Acknowledgements

Chapter 5, in full, is a reprint of the material as it appears in *Algal Research* 2016. Scranton, Melissa; Ostrand, Joseph; Georgianna, David; Lofgren, Shane; Li, Daphne; Ellis, Rosalie; Carruthers, David; Dräger, Andreas; Masica, David; Mayfield, Stephen. The dissertation author is the secondary investigator and author of this paper.

CHAPTER 6:

STRATEGIES FOR DESIGN AND INTERROGATION OF SECOND-GENERATION SYNTHETIC PROMOTERS IN GREEN ALGAE

Strategies for design and interrogation of second-generation synthetic promoters in green algae

Joseph T. Ostrand

The need for diversity

The landscape of the commercial market for microalgae has changed dramatically over the last decade. In 2008, conventional oil prices reached an all-time high, and the demand for fuel alternatives flooded the algal biofuels market with opportunity. As oil prices normalized, it became difficult for the pilot algal-based fuel industry to compete. Gaps in technology and scale left nearly every algae biofuel company without investors or products. In response, many algal biofuel companies have folded or changed their goals. Companies opted for alternative algae-based products that have commercial demand in a space other than fuel. Algae is being grown for feed, nutraceuticals and materials, and genetically engineered for high-value therapeutic production (1). There is also an increasing demand for natural products, utilizing very specific species for their natural abundance of desired compounds.

As these more diverse markets are developed, the requirement for industrialization of a more diverse set of algal species increases. There is now a desire for commercial optimization of an immense diversity of algal species to be grown under a wide variety of conditions. Strains which will be grown for low-value products like fuel and animal feed will be grown photosynthetically in large-scale outdoor ponds (1). These strains will need to be resistant to both biotic and abiotic challenges. The organisms will need to handle dramatic changes in temperature and light availability throughout the day, as well as the presence of predators and competition for resources.

Due to their proven success in the local environment, wild-type strains isolated from the local area may be promising candidates for industrialization(2). Other species of algae which will be cultivated for their ability to produce high-value products will be grown in closed systems under substantially more controlled conditions(3). These strains will be selected for their capacity to grow in high densities as well as their reliability because of increased costs of biomass accumulation. Each of these systems can benefit from the facile introduction of genetic elements. Wild strains in outdoor ponds can be engineered for compromised light antennae to increase light penetration and overall photosynthetic efficiency(4). Elimination of cell membrane receptors which are recognized by predators may decrease the likelihood of culture collapse due to contamination(5). Algae grown in closed systems need to maintain unnaturally dense cultures and may be stably engineered to produce novel therapeutics. We understand the ability of these techniques to improve growth under a variety of situations because of research that has been done on finite set of algal species. By developing a thorough understanding of a short list of specific organisms through basic research, clever tools have been engineered to probe the basic biology as well as to facilitate commercialization of these species. However, it has taken decades of research and thousands of publications to develop the biological understanding that we now utilize in laboratory strains, with the majority of research being done in the model species *Chlamydomonas reinhardtii*(6)Obviously, achieving this level of functional understanding for every organism we seek to grow commercially is unrealistic. Therefore, it becomes necessary to build universal tools and techniques which can be applied to a wide variety of organisms with minimal characterization.

With an increased focus on diverse algal species, it is important to develop a scheme for rapidly understanding an individual organism at a level that allows for genetic manipulation. With technological advances over the past decade, it has become increasingly simple to generate a reference genome and transcriptome for a desired organism, and these can provide a suitable dataset

for the development of a wide variety of genetic tools. Using this dataset, we can rapidly characterize endogenous elements in distantly related algal strains(7,8). Our goal is to use this information to reliably generate a library of synthetic tools for commercial strain development and systems engineering which can eventually be applied to diverse algal strains.

Strategies for synthetic genetic tool development

Gene regulation at the transcriptional level is the result of an intricate combination of DNA-binding proteins, structural changes and protein-protein interactions(9). Even in developed biological systems, these interactions have proven difficult to completely understand. Gene regulatory networks are currently challenging to interpret in silico due to precise spatial requirements of cis-motifs and the relatively small size of binding sites for trans-activating factors. Development of synthetic regulatory elements and testing their efficacy in vivo can illuminate cis-motifs in which the functional understanding is minimal(10). This elucidation will be imperative for engineering strong regulatory elements, as well as advancements of next-generation bioinformatic analysis and model improvement.

A number of studies have shown that synthetic genetic elements have the capacity to drive exogenous gene expression in a diverse set of hosts(11). Synthetic tools have been shown to drive greater expression levels than endogenous elements, and due to the non-native sequence, they are less likely to be affected by homology-based gene silencing(12). Tuning of synthetic elements can be more deliberate, and previous work has shown construction of synthetic regulatory elements that are conditionally regulated(13). Development of a suite of designer promoters will be required for systems and metabolic engineering, in which multiple genes need to be expressed at precise levels for maximum effect.

Building the blueprints for synthetic tools in *C. reinhardtii*

C. reinhardtii is an appropriate host for development of a synthetic tool building process. Thanks to the work of many hardworking scientists, *C. reinhardtii* has without a doubt the most developed genetic tools, and the most thorough transcriptional analyses of any algal species. Most of the progression of *C. reinhardtii* as a model system has come via engineering the chloroplast

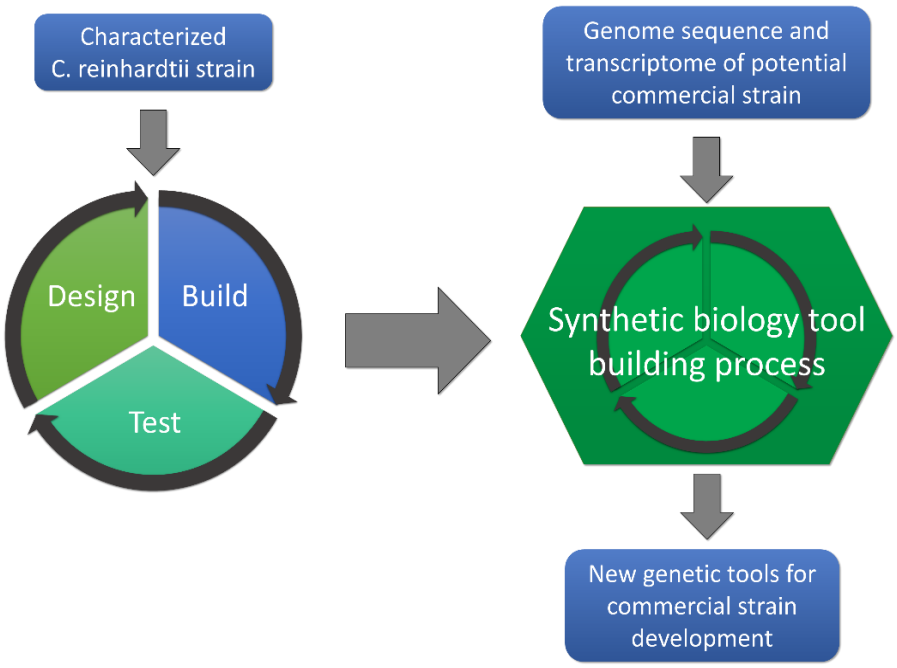

Figure 6.1. Graphical abstract. *C. reinhardtii* provides a suitable model system for constructing and testing novel synthetic tools. The concepts learned from our analyses in *C. reinhardtii* can be applied to other commercial algae strains for rapid strain development and advanced regulatory control.

genome. The chloroplast environment readily performs homologous recombination which facilitates targeted knockouts and vector introduction; this invariably has contributed to the preferential advancement of chloroplast engineering(14). Native promoters which drive photosynthetic machinery have proven incredibly effective at driving transgene expression, with heterologous protein accumulation shown to be as high as 10% of total soluble protein(14). Hybrid promoters comprised of endogenous regulatory elements and native UTRs have improved protein

accumulation in the plastid(15). A library of synthetic UTRs that was tested in the plastid was shown to be effective and has begun to decipher the complexity of gene regulation in the chloroplast (16).All of these tools have contributed to the development of *C. reinhardtii* as an industrial organism, and led to algal-expressed cancer therapeutics, an orally-available malaria vaccine, gut-active mammalian milk proteins and industrial enzymes(15,17,18). However, the chloroplast genome is fundamentally lacking in many desired attributes of a recombinant protein production system. Translated proteins are not delivered throughout the cell, compromising efforts for metabolic engineering and protein secretion systems. The chloroplast lacks advanced post-translational modifications such as glycosylation which are often required for bioactivity(19). Today, there is increased interest in development of tools for the nuclear genome.

There are major advantages to working with the nuclear genome over the chloroplast genome in *C. reinhardtii*. Multiple transit peptides have been described which direct translated proteins to specific compartments within the cell(20,21). Secretion peptides can direct heterologous proteins completely out of the cell to reduce downstream purification costs and potentially limit erroneous biomass accumulation(22). Heterologous proteins can be glycosylated, and their glycosylation patterns are much more similar to humans' than glycosylation patterns of proteins generated by other non-human production systems such as yeast(23). There are already well-described gene expression vectors using clever strategies to couple transcription of selection and expression of the gene of interest, as well as hybrid promoters comprised of elements from highly expressed endogenous genes which have been shown to be more effective than any endogenous elements alone(24). These techniques have been used to successfully express a wide variety of heterologous genes in the nuclear genome.

Development of a synthetic toolset in *C. reinhardtii* can help alleviate the low expression

levels currently plaguing the advancement of *C. reinhardtii* in the commercial recombinant protein space. The extensive history of genetic research on *C. reinhardtii* provides a suitable foundation for synthetic tool development. The lessons learned from the development of synthetic tools in *C. reinhardtii* can be applied to other commercially relevant strains for rapid domestication of wild algae. The characterization of a synthetic toolset which can be applied to other species with minimal wet-lab work will drastically lower the cost of algal strain development and facilitate the use of a diverse group of algal species to specific applications.

Building a strong synthetic promoter

Initial development of functional synthetic promoters for the *C. reinhardtii* nuclear genome was by all standards a success. By investigating the genomic sequence upstream of the most highly expressed genes using an algorithm known as POWRS (POsition-sensitive WoRd Set), short sequence motifs were identified that were conserved among those promoter regions(25,26). The POWRS algorithm determines also the positional dependence of those motifs within the given sequence region. Combining these motifs into a single synthetic promoter was shown to drive gene expression of a fluorescent protein in vivo(25). These promoters were comprised of an average of 15 sequence motifs which were placed based on their conserved location in relation to the transcriptional start site.

A major roadblock in determining promoter strength in the nuclear genome of *C. reinhardtii* is that expression vectors are inserted randomly during transformation. The location in the genome where the expression vector lands can dramatically affect the transcriptional activation of the gene of interest(27). This is caused by regulatory elements that affect the given position in the genome, as well as chromatin structuring which can affect access of trans-activating factors. Therefore, in order to compare promoter strength accurately, many individual transformation events

must be assayed for each promoter. To accommodate large sample sizes, pools comprising >5,000 transformation events for each promoter construct were analyzed by flow cytometry.

Although a functional understanding is available for many of the most highly expressed genes in *C. reinhardtii*, only the genomic sequence with some characterization as well as transcriptomic data that reveals which of these genes is highly expressed under the desired conditions is required to generate this data set.

Second-generation synthetic promoters

Although some of these first-generation synthetic promoters were shown to be the most effective nuclear promoters tested in *C. reinhardtii*(25), the accumulation of recombinant protein still lags significantly behind other systems, including even expression in the chloroplast of *C. reinhardtii*. The results from the first generation of algal synthetic promoters can be used as a reference to help guide the assembly of stronger second-generation synthetic promoters. However, the identification of 130 putative motifs from the top 500 most highly expressed genes combined with the positional dependence that dramatically affects the efficacy of many motifs yields a massive set of potential promoters with few known constrictions to guide their assembly. Because of the relatively low number of synthetic promoters tested in the first analysis, it is difficult to assign activity to specific motifs. One way that is addressed by Scranton et al. is by a comprehensive deletion assay, which identified the specific CCCAT motif as being particularly active(25). This analysis is cumbersome, and performing exhaustive promoter deletions for each of the 130 motifs is a massive undertaking. Additionally, it is possible that the motifs were simply positioned incorrectly, and that their lack of function can be attributed to a distance from the TSS. Some motifs may require a specific combination of additional motifs for their function.

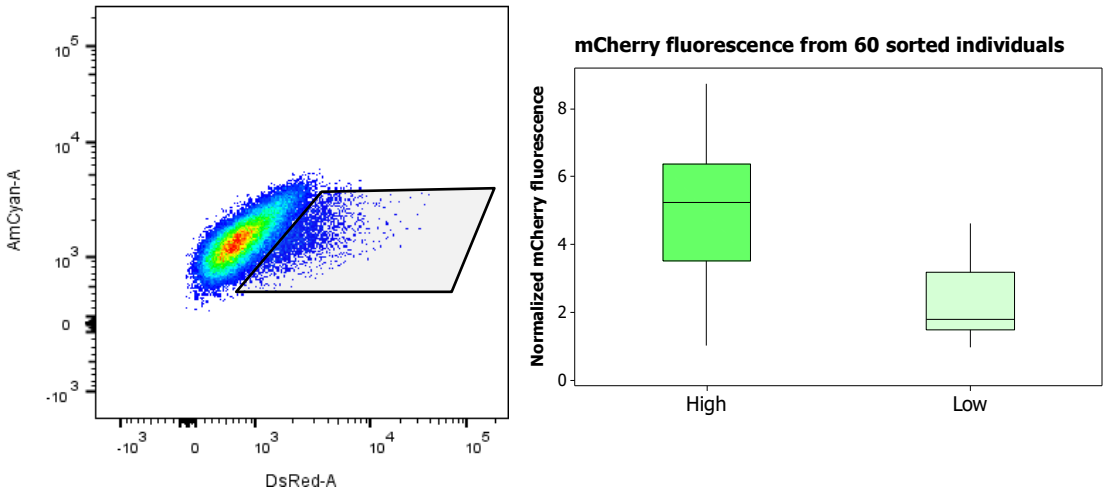

Figure 6.2. (Left) Flow cytometry analysis of a pool of 5,000+ transformants of an mCherry nuclear expression vector. Cells which fall within the gate exhibit mCherry fluorescence above background, indicating measurable accumulation due to the introduced vector. Cells which showed expression of mCherry were sorted by FACS. Two groups were sorted, one group selected only cells which displayed exceptionally high levels of mCherry fluorescence. A second group selected for cells with more moderate levels of fluorescence. (Right) 60 individual transformants were recovered from each group and assayed by fluorescent plate reader. The cells sorted from the high fluorescence pool show a conserved significant increase in mCherry fluorescence.

In an effort to develop a dataset on which synthetic promoter assembly can be constricted, a library of 1,000 unique promoters was synthesized for simultaneous analysis. In order to narrow the total variables, the POWRS algorithm was run on the upstream sequence of 4,000 native genes which were categorized into low, medium, and high expressing genes. The pool of 130 motifs from which we pulled initially was restricted to only 12 motifs which were highly represented in only highly-expressing promoters. These 12 motifs were arranged into every combination of 5 different motifs per promoter, making it much more reasonable to assign promoter activity to a specific motif than the initial 15 motifs per promoter. In addition, promoters containing 1-5 copies of a single motif arranged throughout the promoter were added to the library. The promoter region was split in half, with one half containing the 1000 unique promoter sequences, and the other half corresponding to a truncated version of sap11, the most successful promoter from round 1(25). Therefore, two libraries of 1000 unique sequences were synthesized: the 5' upstream half of the

promoter region, hereafter referred to as “Pool 1”, and the 3’ downstream half of the promoter, referred to as “Pool 2”. Together, they make up 2,000 unique promoters that can be tested simultaneously.

If the individual promoters can be properly categorized into those that are exceptional, those which are functional, and those which do not contribute to recombinant protein accumulation, analysis of the motifs present in each category will provide valuable insight into future promoter assembly. Based on preliminary data shown in Figure 2, individuals expressing a fluorescent protein can be sorted by FACS based on their fluorescence levels, strains expressing high levels of fluorescent protein seem to maintain their high expression levels. In order to assign a function to

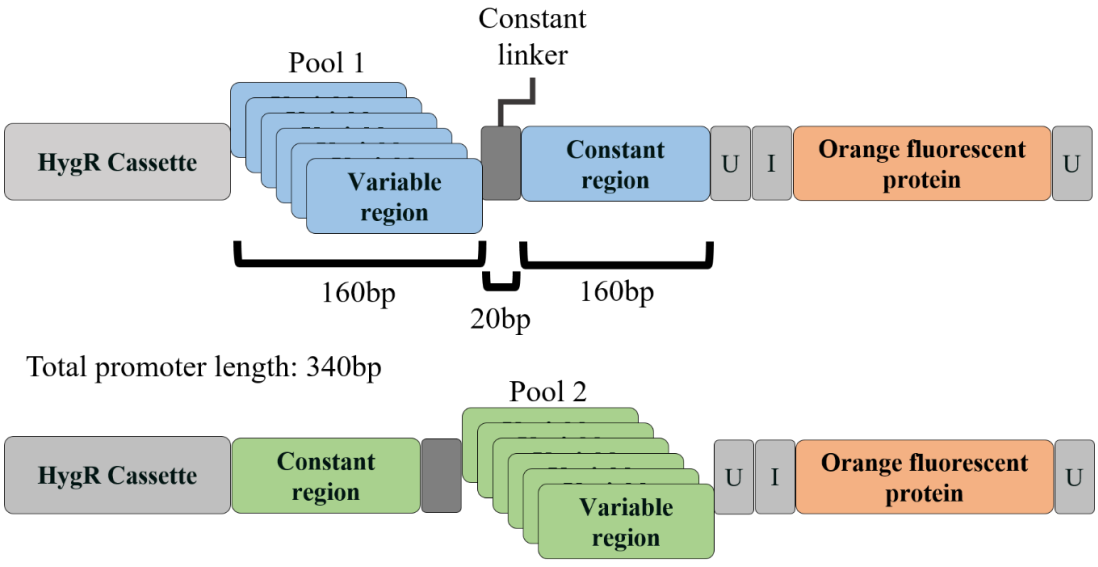

Figure 6.3. Vector design for second generation promoter library. To test the individual contribution of specific motifs to the final accumulation of a recombinant protein, synthetic promoters were designed with an average of only 5 motifs per promoter. The promoter region was split into a 5’ end and a 3’ end. Each half of the promoter will be investigated independently by combining the new synthetic promoter sequences with the corresponding region of sap11, the strongest promoter from the first round of synthetic promoter design. Two pools, each comprised of 1000 unique synthetic promoters, were synthesized and cloned into the above vector to drive expression of the *mCerianthus* orange fluorescent protein.

each promoter sequence, each pool was transformed simultaneously and sorted by FACS based on their ability to express a fluorescent protein.

Pool 1 and Pool 2 were cloned into an expression vector driving production of Orange Fluorescent Protein (OFP) (28)(Figure 3). To approximate complete representation of each sequence of the variable region in the final vector, >10,000 E. coli colonies were pooled for each vector into the final DNA prep. The collection of expression vectors containing Pool 1 and Pool 2 were electroporated into the nuclear genome of *C. reinhardtii*. For each pool, >200,000 individual transformation events were collected after selection on hygromycin plates. The individuals were combined into a liquid culture and grown mixotrophically, heterotrophically or phototrophically and analyzed by flow cytometry, as shown in Figure 4.

Individual cells which showed detectable levels of OFP were sorted for further phenotyping to verify that cell sorting effectively enriched the population for expressors of OFP. Only 1,087 individual cells were recovered from the Pool 1 library, and 1,407 were recovered from the Pool 2 library. These numbers do not provide proper coverage of the promoter libraries to comprehensively evaluate each promoter sequence. To ensure that sorting was an effective means of isolating phenotypically positive individuals, each of the sorted strains from Pool 1 and Pool 2 was independently grown in liquid culture under the same three conditions and measured for fluorescence intensity by plate reader. The fluorescence of each individual strain is shown in Figure 5, with dark boxes representing no expression, green boxes showing measurable expression above background, and white boxes showing strong expression.

Discussion

Random insertion dramatically changes population size

This library of synthetic promoter elements was designed to provide insight into the contributory effect of individual sequence motifs and probe an array of positions across genomic region upstream of the transcriptional start site. In order to establish a convincing evaluation of a promoters strength, each copy of the promoter must be analyzed across multiple insertion sites within the genome to account for positional effects. Therefore, the number of total events required to verify that (1) each promoter sequence is present in the pool of transformants, and that (2)

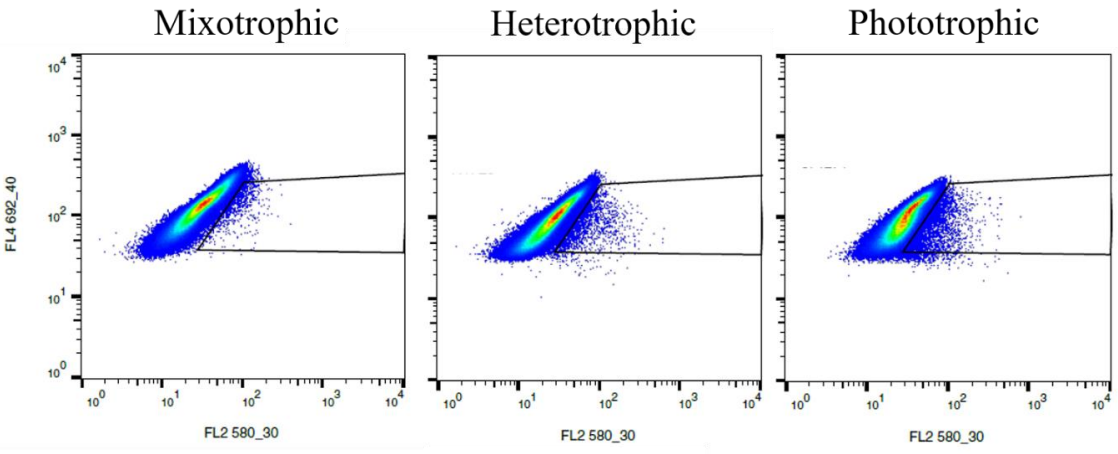

Figure 6.4. Pool 1 library grown under varying conditions. The expression vector containing Pool 1 driving OFP was transformed into the nuclear genome of *C. reinhardtii*. The pool of selected transformants was grown in three different conditions: mixotrophically, heterotrophically, and phototrophically. The events which fall in the gate represent individuals expressing OFP. Only 0.20% of cells grown mixotrophically and 0.24% of cells grown heterotrophically were shown to accumulate OFP. 0.93% of cells grown phototrophically displayed OFP accumulation.

multiple independent insertions of a given promoter sequence have made it into the transformant pool, is approximately 100x the number of sequences we are trying to investigate to obtain at least 10 insertions per promoter sequence represented. However, when transforming the expression cassette containing a functional promoter such as ar1 or sap11, only about 10% of the transformants yield a detectable phenotype(25). That means that to achieve 10 detectable transformations of each promoter sequence, closer to 1,000x the number of sequences is the necessary number of transformation events. For a pool of 1,000 unique promoters, that means 1,000,000 events at a

minimum need to be evaluated to achieve a cursory representation of the strength of every promoter in the pool.

FACS is not the optimal selection strategy

Analysis of 1,000,000 transformation events is not impossible. Flow cytometry is an effective means of measuring fluorescence levels of many individual cells, and 1,000,000 individuals is well within the scope of any flow cytometers capabilities. However, even if we could procure a sample with 1,000,000 cells that were all unique, flow cytometer analysis of 1,000,000 transformation events does not provide the information we need to achieve an individual analysis of each promoter. Instead, those 1,000,000 transformation events need to be effectively sorted for their transgene expression phenotype, recovered, and analyzed so that the sequences present in each expression profile can be determined.

While FACS can effectively sort out cells of a desired phenotype, the scale of this project requires robust selection that is not feasible when using a fluorescent protein reporter. Cells which are analyzed by flow cytometry provide a millisecond glimpse into the fluorescent properties of an individual cell. Expression of recombinant proteins varies significantly through the growth cycle, and is typically measured as a pool of individuals to average out these effects(29). Sorting of individual cells at an arbitrary point in their growth phase leads to the sorting of transiently expressing cells which may not actually show as strong of a phenotype in a monoculture. Indeed, the vast majority of sorted cells for this experiment did not have a measurable phenotype upon small scale monoculture cultivation.

To investigate the cells' phenotype by FACS and mechanically separate desired cells, the transformant pool must first be culled by antibiotic selection and grown in liquid cultures. Due to

the rarity of transformation, antibiotic selection kills about 99.99% of the starting culture(30). Removing the dead cells to simplify flow cytometry analysis is imperfect and requires outgrowth of the transformation pool, potentially removing slow-growing transformants from downstream analysis. However, this selection by antibiotic may be the optimal target for future analysis of the synthetic promoter library.

An alternative strategy for synthetic promoter library analysis

The major requirement to determine the contribution of given motifs to the ultimate phenotype of transgene expression relies on our ability to effectively categorize expression levels of a huge number of transformants. FACS was not able to robustly catalog hundreds of thousands of unique transformants into reliable phenotypic categories. However, a step in the preparation of our transformant pool for FACS analysis rapidly selected phenotypes with robust categorization for high and no expression: antibiotic resistance. Expression of an antibiotic resistance gene like Ble confers resistance to antibiotic concentrations correlated to its gene expression. Higher Ble gene expression leads to resistance to higher concentrations of zeocin. This is because Ble provides resistance by individually binding and sequestering the antibiotic as opposed to actively degrading the compound(24). A proper follow-up to this report should clone the synthetic promoter library to drive the Ble gene and use varying levels of zeocin to categorize cells based on their expression levels.

Future avenues for algal synthetic biology

Conditional and Inducible Regulatory Elements in C. reinhardtii nuclear genome

Precise control of gene expression through conditional regulatory elements is required for optimization of recombinant protein production. Triggering gene expression at the correct growth phase can decouple the biological requirement of nutrient utilization and cellular reproduction with

maximum output of a recombinant protein of interest. In many cases, expression of the desired protein may be toxic to cells at a particular life stage, and an inducible expression system is required to achieve appreciable levels of recombinant protein without crashing the culture. Multiple inducible promoters have been indentified in *C. reinhardtii* which promote protein accumulation under specific conditions. Heat shock, nickel or nitrate introduction, and iron-deficiency have all been used as abiotic signals to induce gene expression from inducible promoters. (31–35). Development of these systems required a much more detailed understanding of *C. reinhardtii* than will be afforded with other potential production species. However, there is a major abiotic factor which is easy to regulate in closed systems and may provide valuable insight into the development of conditionally active promoters across algal species: light.

For photosynthetic organisms, light and dark cycles act as major drivers of metabolism and gene expression pattern variation. During day time, green algae can utilize photosynthesis to drive the production of sugars that are then used for energy in a myriad of metabolic processes including the production of starches and sugars. During the night the cells must utilize stored energy in the form of sugars, starches, or lipids to continue metabolic activity. The switching from phototrophic to heterotropic metabolism requires large sets of genes to be switched on or off. In Chlamydomonas ~80% of the genome displays detectable periodic gene expression changes throughout a 24 hour day/night cycle (36). Unique regulatory motifs may be used to regulate these light-induced or dark-induced genes in response to light intensity. If identified, these motifs can then be utilized to drive transgene expression specifically in response to light or dark conditions. This report shows the expression of OFP driven by synthetic promoters under three separate conditions which have dramatically different metabolisms (Figure 5). However, across dozens of highly expressing strains, we rarely see a dramatic on/off phenotype. This is likely due to the fact that our library was constructed by looking at the most highly expressed genes under a single condition. Instead of categorizing motifs which were differentially conserved in light and dark regulated genes, the

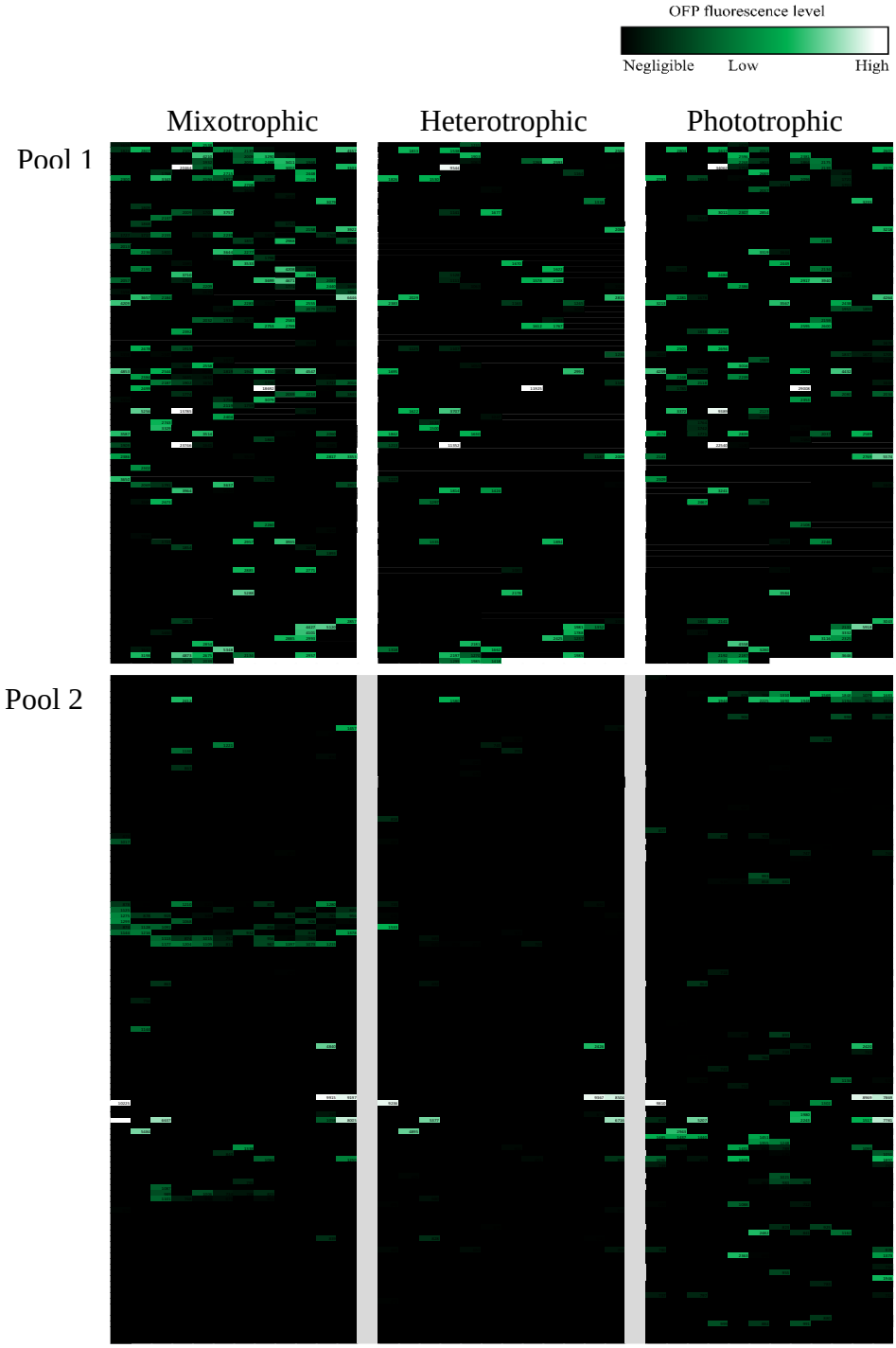

Figure 6.5. Heatmaps of sorted individuals driving OFP with synthetic promoters. These figures show the OFP flurorescence of 1,087 (Pool 1) and 1,407 (Pool 2) individuals sorted from the pool of *C. reinhardtii* transformants. Each sorted individual was grown in liquid culture under three conditions and analyzed by plate reader. The fluorescence of each culture is represented by color on the heatmap. Black cells did not show fluorescence above the wild-type background strain.

promoters contain motifs which were conserved in all highly expressed genes.

Using high resolution RNA-seq data taken from Chlamydmonas reinhardtii on a 12 hour light – 12 hour dark cycle (36)we identified genes that were differentially expressed by at least two fold between the middle of the light-period (day) and the middle of the dark-period (night) while displaying moderate to high expression levels overall during their upregulated time period. Specifically, we averaged the Reads Per Kilobase of transcript per Million mapped reads (RPKM) for each transcript during the middle 4 hours of the 12-hour light period and the middle 4 hours of the 12-hour dark period. Genes with at least a 2-fold increase in averaged read count during the light phase compared to the dark phase and an average RPKM of more than 100 were determined to be light-upregulated strong expressers. Similarly, genes with at least a 2-fold increase in average read count during the dark phase compared to the light and an average RPKM of more than 100 were determined to be dark-upregulated strong expressers. Collectively this represented 255 light-upregulated genes and 248 dark-upregulated genes. The 1000 bp region 5' from the transcriptional start site of these genes was retrieved (Phytozome 12, *Chlamydomonas reinhardtii* genome v5.5) and analyzed using the POWRS motif identification program (26). All default settings on POWRS were used and -1000 bp regions from all 17737 annotated genes in the whole genome used as the background control data set. POWRS identified 31 and 32 enriched motif clusters in the light-upregulated and dark-upregulated promoter datasets, respectively compared to promoters in the rest of the genome. Motifs enriched in the light-upregulated or dark-upregulated data sets were compared each other using the Tomtom motif comparison tool (37). Figure 6 identifies motifs unique to either the light up-regulated (A) or dark-upregulated (B) data sets. Many of the light/dark-regulated motifs are different from the motifs identified from simply looking at the highest expressed genes during logarithmic growth in the previous example. This analysis is particularly powerful because a genomic construction and light/dark transcriptomes are useful for a wide variety of species characterization analyses. Because procurement of the required dataset

for this analysis has benefits outside of promoter development, the time and money required is justifiable. The motif clusters identified here may be valuable building blocks for the future assembly of light-dependent conditional synthetic promoters in *C. reinhardtii*.

Toward synthetic circuits in *C. reinhardtii*

A major difficulty in development of advanced regulatory machinery in *Chlamydomonas reinhardtii* and algal species in general are the large gaps in native regulatory element characterization. However, a wealth of knowledge is available across the kingdom Plantae which serve as a guide to understanding the complex transcriptional regulation found in *C. reinhardtii*. One of the best-understood aspects of any regulatory system is that by encouraging an activating transcription factor to bind in a regulatory region associated with a transgene, one can increase transcript abundance and subsequent protein accumulation. Systems have been derived in *S. cerevisiae* and *E. coli* which take advantage of known DNA-binding proteins to engineer complex circuits of protein expression for a wide variety of purposes(38–40).

Transcription factor families are easily identifiable in silico and homology analysis to better-understood systems can provide a groundwork for understanding in *C. reinhardtii*. The Plant Transcription Factor Database (PTFDB) (http://planttfdb.cbi.pku.edu.cn/) has identified each family of transcription factor found in *C. reinhardtii* based on sequence homology to other plants. The PTFDB has also compiled data from across the literature to provide putative binding sites for those families of transcription factors. Transcription factor (TF) binding sites have been studied across plants through one of the following processes: ampDAP, ChIP/ChIP-seq, DAP, PBM, or SELEX. TF binding sites found in the literature that are associated with a given TF family are projected to other species to help characterize binding in a virgin system. The sequence motifs

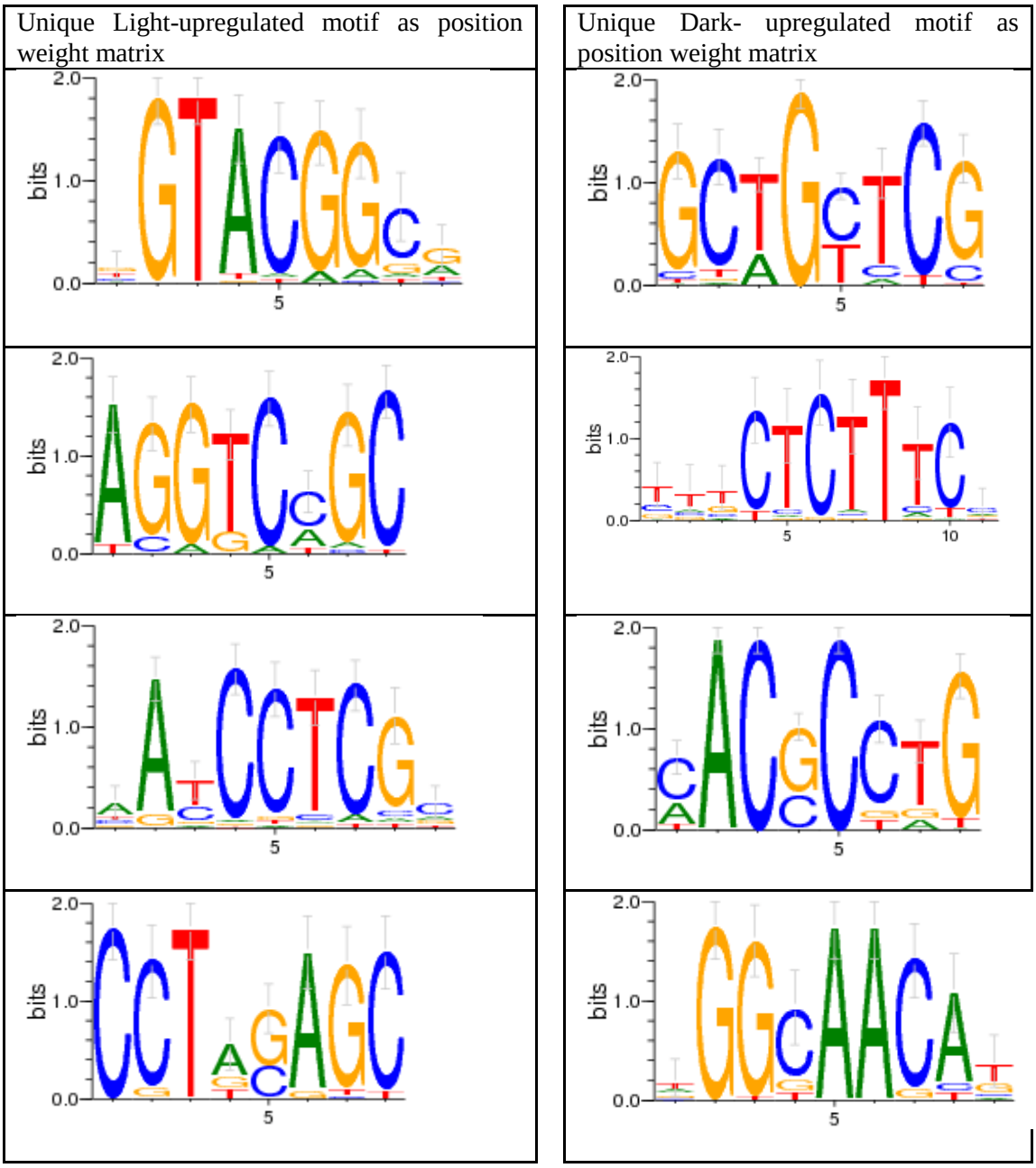

Figure 6.6. Motif clusters enriched in light (left) and dark (right) regulated genes. 255 light-upregulated genes and 248 dark-upregulated genes as determined by transcriptomic data were analyzed with the POWRS algorithm to find motif sequences which were differentially conserved among light-sensitive genes. Motifs detected from the light and dark regulated genes were compared by TOMTOM and only the motifs which were unique to the light (left) or dark (right) regulated genes are shown.

attributed to TF families found in *C. reinhardtii* are provided as position-weight matrices in Figure 7. These serve as a promising set of sequences for synthetic promoter engineering. By integrating these sequences into a novel synthetic promoter, we can project the regulation of the transgene onto one or many specific transcription factors. We know that certain transcription factors have variable function based on external stimuli (41), and as such these sequences are clear candidates for inducible promoter engineering.

In an effort to better characterize the in vivo TF/sequence cognate pairs for *C. reinhardtii*, 90 predicted transcription factors were cloned from *C. reinhardtii* cDNA into a constitutive nuclear expression construct (42). Upon characterization of their binding in a Y1H assay, a bHLH-family transcription factor (Cre02.g109700.t1.2, will be referred to as TF64) was selected for further analysis. Three strains were designed to determine if constitutive expression of a transgenic transcription factor can increase recombinant protein abundance in *C. reinhardtii*. We generated a strain which expressed high levels of TF64, one which expressed low levels of TF64, and a control strain which used the same construct to express GFP, a non-DNA binding protein. These three strains in addition to an untransformed wild-type strain were transformed with an expression cassette which drives OFP expression, which is easily detected by a fluorescent plate reader. To determine the effect of TF64 binding, the promoter associated with the OFP gene must contain binding site(s) associated with the bHLH transcription factor family (CANNTG). Conveniently, the AR1 promoter that is well-established in the field has three putative bHLH binding sites. The AR1 promoter was used to drive the expression of OFP in the TF64 expression strains, shown in Figure 8. These data indicate that presence of putative TF-binding site motifs in an expression construct when combined with their associated transcription factors can help drive recombinant protein accumulation.

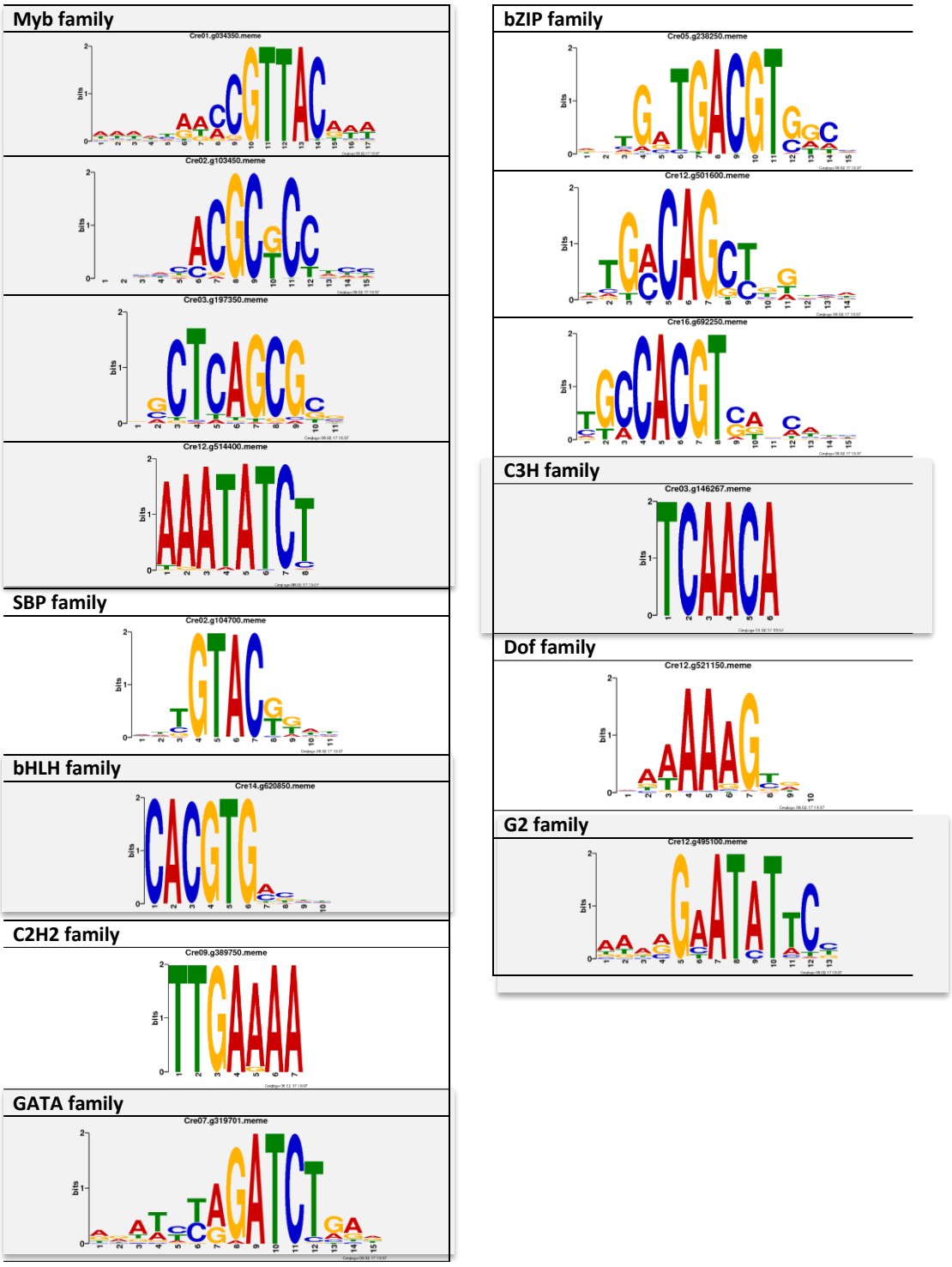

Figure 6.7 Predicted binding sites for *Chlamydomonas reinhardtii* transcription factor families as deduced by the Plant Transcription Factor Database. Letter height indicates relative frequency of nucleotides in the proposed binding sequence.

The generation of more in vivo cognate TF/site pairs based on the putative TF binding sites shown in Figure 7 will facilitate the development of more advanced promoters with the added functionality of orthogonal regulation.

These data present a bright future for synthetic elements in algal hosts. Initial success for strong promoter development has encouraged a more thorough analysis of the individual activity of the identified sequence motifs. Successful characterization of these elements will undoubtedly yield stronger second-generation promoters. In addition, we have shown by the same analysis used to generate the first round of synthetic promoters that there are conserved motifs which are unique

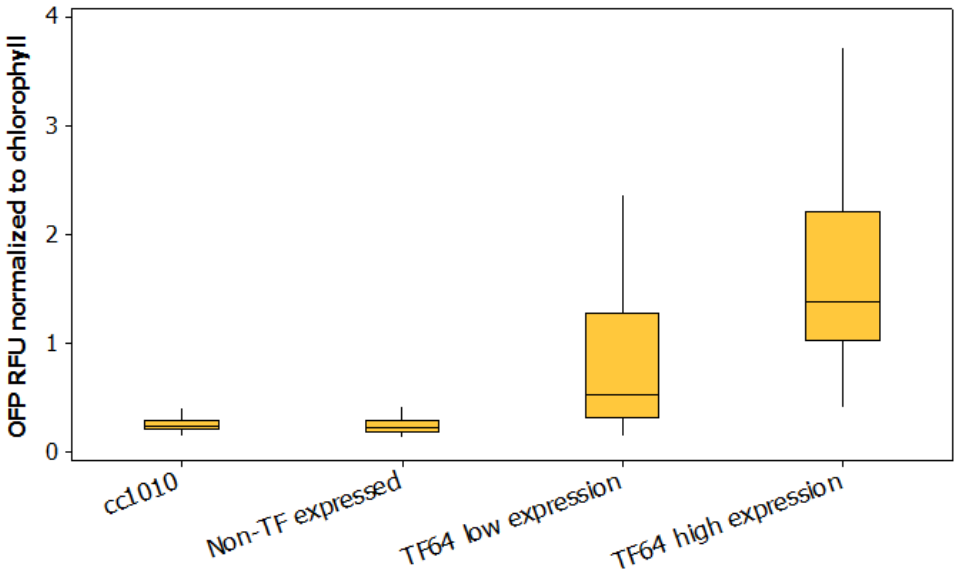

Figure 6.8. Change in exogenous gene expression when native transcription factors are over-expressed. Four strains shown above were created to have unique expression levels of an endogenous transcription factor TF64 driven by a constitutive promoter. Cc1010 is the background strain and only has natie levels of TF64. cc1010 was transformed with an expression vector for TF64 and one which expressed high levels of TF64 and one with low levels of TF64 were selected. In addition, cc1010 was transformed with the same expression vector except expressing the GFP protein which should not affect gene regulation (Non-TF expressed). Each of these strains was transformed with a separate expression vector driving OFP expression by the ar1 promoter. The ar1 promoter is predicted to have multiple binding sites for TF64. Shown above is the result of 96 individual transformants analyzed for OFP expression by a fluorescent plate reader.

to light and dark-regulated genes. These motifs may lead to the development of light or dark-inducible synthetic promoters. Finally, characterization of native transcription factors in *C. reinhardtii* has opened the door for gene circuit development and deliberate recruitment of trans-activating factors to promote exogenous gene expression. These strategies for the assembly of synthetic promoters may be the key to increasing the accumulation of recombinant protein from the nuclear genome to industrially relevant level.

References

1. Scranton MA, Ostrand JT, Fields FJ, Mayfield SP. Chlamydomonas as a model for biofuels and bio-products production. Plant J Cell Mol Biol. 2015 May;82(3):523–31.

2. Ratha SK, Prasanna R. Bioprospecting microalgae as potential sources of "green energy"--challenges and perspectives (review). Prikl Biokhim Mikrobiol. 2012 Apr;48(2):133–49.

3. Rasala BA, Mayfield SP. Photosynthetic biomanufacturing in green algae; production of recombinant proteins for industrial, nutritional, and medical uses. Photosynth Res. 2015 Mar;123(3):227–39.

4. Perrine Z, Negi S, Sayre RT. Optimization of photosynthetic light energy utilization by microalgae. Algal Res. 2012 Oct 1;1(2):134–42.

5. Day JG, Gong Y, Hu Q. Microzooplanktonic grazers – A potentially devastating threat to the commercial success of microalgal mass culture. Algal Res. 2017 Nov 1;27:356–65.

6. Harris EH. Chlamydomonas as a model organism. Annu Rev Plant Physiol Plant Mol Biol. 2001 Jun;52:363–406.

7. Jąkalski M, Takeshita K, Deblieck M, Koyanagi KO, Makałowska I, Watanabe H. Comparative genomic analysis of retrogene repertoire in two green algae Volvox carteri and *Chlamydomonas reinhardtii*. Biol Direct. 2016 Aug 4;11:35.

8. Roberts ML. The Use of Functional Genomics in Synthetic Promoter Design. Comput Biol Appl Bioinforma [Internet]. 2011 [cited 2018 Mar 21]; Available from: http://www.intechopen.com/books/computational-biology-and-applied-bioinformatics/the-use-of-functional-genomics-in-synthetic-promoter-design

9. Lelli KM, Slattery M, Mann RS. Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. Annu Rev Genet. 2012;46(1):43–68.

10. Hernandez-Garcia CM, Finer JJ. Identification and validation of promoters and cis-acting regulatory elements. Plant Sci. 2014 Mar 1;217–218:109–19.

11. Liu W, Stewart CN. Plant synthetic promoters and transcription factors. Curr Opin Biotechnol. 2016 Feb;37:36–44.

12. Chakravarthi M, Narayan JA, Subramonian N, Appunu C. Development of novel synthetic promoters for gene expression in transgenic sugarcane. 2015;12.

13. Wang R, Zhu M, Ye R, Liu Z, Zhou F, Chen H. Novel green tissue-specific synthetic promoters and *cis*-regulatory elements in rice. Sci Rep. 2015 Dec 11;5:18256.

14. Mayfield SP, Manuell AL, Chen S, Wu J, Tran M, Siefker D. *Chlamydomonas reinhardtii* chloroplasts as protein factories. Curr Opin Biotechnol. 2007 Apr;18(2):126–33.

15. Manuell AL, Beligni MV, Elder JH, Siefker DT, Tran M, Weber A. Robust expression of a bioactive mammalian protein in Chlamydomonas chloroplast. Plant Biotechnol J. 2007 May;5(3):402–12.

16. Specht EA, Mayfield SP. Synthetic oligonucleotide libraries reveal novel regulatory elements in Chlamydomonas chloroplast mRNAs. ACS Synth Biol. 2013 Jan 18;2(1):34–46.

17. Tran M, Henry RE, Siefker D, Van C, Newkirk G, Kim J. Production of anti-cancer immunotoxins in algae: ribosome inactivating proteins as fusion partners. Biotechnol Bioeng. 2013 Nov;110(11):2826–35.

18. Specht EA, Mayfield SP. Algae-based oral recombinant vaccines. Front Microbiol [Internet]. 2014 Feb 17 [cited 2018 Mar 21];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3925837/

19. Rasala BA, Mayfield SP. The microalga *Chlamydomonas reinhardtii* as a platform for the production of human protein therapeutics. Bioeng Bugs. 2011;2(1):50–4.

20. Molino JVD, Carvalho JCM de, Mayfield SP. Comparison of secretory signal peptides for heterologous protein expression in microalgae: Expanding the secretion portfolio for *Chlamydomonas reinhardtii*. PLOS ONE. 2018 Feb 6;13(2):e0192433.

21. Rasala BA, Chao S-S, Pier M, Barrera DJ, Mayfield SP. Enhanced genetic tools for engineering multigene traits into green algae. PloS One. 2014;9(4):e94028.

22. Ramos-Martinez EM, Fimognari L, Sakuragi Y. High-yield secretion of recombinant proteins from the microalga *Chlamydomonas reinhardtii*. Plant Biotechnol J. 2017 Sep;15(9):1214–24.

23. Mathieu-Rivet E, Kiefer-Meyer M-C, Vanier G, Ovide C, Burel C, Lerouge P. Protein N-glycosylation in eukaryotic microalgae and its impact on the production of nuclear expressed biopharmaceuticals. Front Plant Sci [Internet]. 2014 Jul 28 [cited 2018 Mar 22];5. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4135232/

24. Rasala BA, Lee PA, Shen Z, Briggs SP, Mendez M, Mayfield SP. Robust Expression and Secretion of Xylanase1 in *Chlamydomonas reinhardtii* by Fusion to a Selection Gene and Processing with the FMDV 2A Peptide. Stiller JW, editor. PLoS ONE. 2012 Aug 24;7(8):e43349.

25. Scranton MA, Ostrand JT, Georgianna DR, Lofgren SM, Li D, Ellis RC. Synthetic promoters capable of driving robust nuclear gene expression in the green alga *Chlamydomonas reinhardtii*. Algal Res. 2016 Apr;15:135–42.

26. Davis IW, Benninger C, Benfey PN, Elich T. POWRS: position-sensitive motif discovery. PloS One. 2012;7(7):e40373.

27. León R, Fernández E. Nuclear transformation of eukaryotic microalgae: historical overview, achievements and problems. Adv Exp Med Biol. 2007;616:1–11.

28. Ip DT-M, Wong K-B, Wan DC-C. Characterization of novel orange fluorescent protein cloned from cnidarian tube anemone Cerianthus sp. Mar Biotechnol N Y N. 2007 Aug;9(4):469–78.

29. Castruita M, Casero D, Karpowicz SJ, Kropat J, Vieler A, Hsieh SI. Systems Biology Approach in Chlamydomonas Reveals Connections between Copper Nutrition and Multiple Metabolic Steps. Plant Cell. 2011 Apr 1;23(4):1273–92.

30. Berthold P, Schmitt R, Mages W. An engineered Streptomyces hygroscopicus aph 7" gene mediates dominant resistance against hygromycin B in *Chlamydomonas reinhardtii*. Protist. 2002 Dec;153(4):401–12.

31. Fei X, Deng X. A novel Fe deficiency-responsive element (FeRE) regulates the expression of atx1 in Chlamydomonas reinharditii. Plant Cell Physiol. 2007 Oct;48(10):1496–503.

32. Ferrante P, Catalanotti C, Bonente G, Giuliano G. An Optimized, Chemically Regulated Gene Expression System for Chlamydomonas. El-Shemy HA, editor. PLoS ONE. 2008 Sep 12;3(9):e3200.

33. Ohresser M, Matagne RF, Loppes R. Expression of the arylsulphatase reporter gene under the control of the nit1 promoter in *Chlamydomonas reinhardtii*. Curr Genet. 1997 Mar;31(3):264–71.

34. Schroda M, Beck CF, Vallon O. Sequence elements within an HSP70 promoter counteract transcriptional transgene silencing in Chlamydomonas. Plant J Cell Mol Biol. 2002 Aug;31(4):445–55.

35. Wu J, Hu Z, Wang C, Li S, Lei A. Efficient expression of green fluorescent protein (GFP) mediated by a chimeric promoter in *Chlamydomonas reinhardtii*. Chin J Oceanol Limnol. 2008 Aug 1;26(3):242.

36. Zones JM, Blaby IK, Merchant SS, Umen JG. High-Resolution Profiling of a Synchronized Diurnal Transcriptome from *Chlamydomonas reinhardtii* Reveals Continuous Cell and Metabolic Differentiation. Plant Cell. 2015 Oct;27(10):2743–69.

37. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome Biol. 2007;8(2):R24.

38. Wang B, Kitney RI, Joly N, Buck M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat Commun. 2011 Oct 18;2:508.

39. Ellis T, Wang X, Collins JJ. Diversity-based, model-guided construction of synthetic gene networks with predicted functions. Nat Biotechnol. 2009 May;27(5):465–71.

40. Kotula JW, Kerns SJ, Shaket LA, Siraj L, Collins JJ, Way JC. Programmable bacteria detect and record an environmental signal in the mammalian gut. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):4838–43.

41. Riaño-Pachón DM, Corrêa LGG, Trejos-Espinosa R, Mueller-Roeber B. Green Transcription Factors: A Chlamydomonas Overview. Genetics. 2008 May;179(1):31–9.

42. Anderson MS, Muff TJ, Georgianna DR, Mayfield SP. Towards a synthetic nuclear transcription system in green algae: Characterization of *Chlamydomonas reinhardtii* nuclear transcription factors and identification of targeted promoters. Algal Res. 2017 Mar;22:47–55.

Acknowledgements

I would like to thank Melissa Scranton, Anthony Berndt, Song Cao, Stephen Espitia, Melissa Anderson, Travis Muff and Frank Fields for their contributions to the development of this chapter.

CHAPTER 7:

HIGH-THROUGHPUT SYSTEM FOR QUANTIFYING AND CHARACTERIZING HOMOLOGOUS RECOMBINATION IN *CHLAMYDOMONAS REINHARDTII*

Algal Research 31 (2018) 167–172

Contents lists available at ScienceDirect

Algal Research

journal homepage: www.elsevier.com/locate/algal

Short communication

High-throughput system for quantifying and characterizing homologous recombination in *Chlamydomonas reinhardtii*

Hussam Hassan Nour-Eldin[1], Elizabeth A. Specht[2], Joseph Ostrand, Kevin T.D. Hoang[3], Prema S. Karunanithi[4], Stephen P. Mayfield*

California Center for Algae Biotechnology, University of California San Diego, 9500 Gilman Drive, La Jolla, CA 92093, United States

ARTICLE INFO

Keywords:
Chlamydomonas reinhardtii
Nuclear genetic engineering
Homologous recombination
High-throughput screen
Split marker system
Intron homology sequence

ABSTRACT

In the green alga *Chlamydomonas reinhardtii*, introduced DNA fragments predominantly insert randomly into the nuclear genome by non-homologous recombination (NHR), often resulting in highly variable phenotypes between transformants. Homologous recombination (HR) can occur in *C. reinhardtii* but at very low frequency, and is often accompanied by insertions, deletions, and/or rearrangements at the recombination site. To benchmark the frequency and characterize the nature of HR integrations in *C. reinhardtii*, we developed a system for detecting and characterizing HR events that includes three intact markers and one split marker as well a novel design element: utilizing a long intron as the homology region. In this study we demonstrate that this system meets the following criteria: accommodates high-throughput screening; provides a high-fidelity phenotype for detecting HR without false positives from reversion or locus heterogeneity; allows and captures both single- and double-crossover HR events; reports HR and NHR rates from a single transformation; and allows characterization of imprecise recombination or rearrangement at the integration site. Using this system we reproducibly determined the HR rate in our recipient strain of *C. reinhardtii* and characterized a number of recombinants by restriction digests and sequencing of PCR amplified recombination junctions to show that both double and single crossover events were recovered and that integration occurred both via perfect and imperfect (i.e. accompanied by insertions, deletions, and rearrangements) HR. This system is valuable for systematically testing approaches for increasing HR efficiency and accuracy.

1. Introduction

Integration of linear DNA fragments into the nuclear genome of *C. reinhardtii* can occur via homologous recombination with very low frequency and is typically accompanied by insertions and/or deletions at the recombination site (referred to from hereon as imperfect HR) [1–4]. With the intent of developing methods for improving the efficiency and accuracy of HR in *C. reinhardtii*, we desired a standardized, robust, and high-throughput HR screening approach that addresses shortcomings of existing approaches. In the following, we describe its design and demonstrate its robust ability to measure consistent HR:NHR frequencies across independent transformations, and to characterize features of discrete recombination events. The properties of our approach are discussed in comparison to existing approaches.

2. Results

2.1. System design

We developed a two-step recombination detection system that allows HR and NHR frequency to be measured from the same transformation. First, a recipient strain was created using the pHR18 plasmid. From the 5′ to 3′ end, the functional part of pHR18 consists of a complete *ARS2* cassette (periplasmic arylsulfatase) [5], a complete *ARG7*

Abbreviations: HR, homologous recombination; NHR, non-homologous recombination; NHEJ, non-homologous end-joining

* Corresponding author.
E-mail addresses: huha@plen.ku.dk (H.H. Nour-Eldin), elizabeth.specht@colorado.edu (E.A. Specht), jostrand@ucsd.edu (J. Ostrand), kevin_hoang@nymc.edu (K.T.D. Hoang), pkarunan@ucdavis.edu (P.S. Karunanithi), smayfield@ucsd.edu (S.P. Mayfield).
[1] Current affiliation: DynaMo Center, Dept of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Denmark; Copenhagen Plant Science Center, Dept of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Denmark.
[2] Current affiliation: The Good Food Institute, 1380 Monroe St NW, #229 Washington, DC 20010, United States.
[3] Current affiliation: New York Medical College, School of Medicine, United States.
[4] Current affiliation: University of California Davis, College of Biological Sciences, United States.

https://doi.org/10.1016/j.algal.2018.02.005
Received 8 August 2017; Received in revised form 29 January 2018; Accepted 6 February 2018

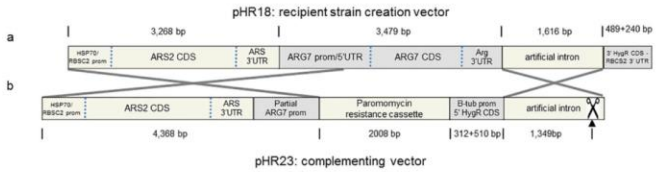

Fig. 1. Schematic of vectors comprising the HR detection system. a) pHR18 is the vector used to create the recipient strains, and integrates at random into the nuclear genome. b) pHR23 is the complementing vector designed to recombine at the pHR18 locus within the nuclear genome to confer hygromycin resistance. Diagonal lines indicate regions of homology suitable for recombination. The arrow below pHR23 indicates the *Eco*RV site, where the plasmid was linearized prior to electroporation. Sizes of individual elements is provided above and below each construct, respectively.

cassette (expressing the CDS of argininosuccinate lyase from native promoter and 3′UTR) [6], and a 1.6 kb intron (see M&M) followed by the 3′ half of a Hygromycin B phosphotransferase expression cassette (last 489 bp of Hyg CDS followed by 240 bp of RBCS2 3′UTR) (Fig. 1a). The *ARS2* cassette encodes the arylsulfatase reporter under the control of a constitutive promoter and serves as a semi-quantitative reporter for gene expression [5]. The *ARG7* cassette in pHR18 enables recovery of transformants of arginine auxotrophic strains [6]. The 1.6 kb intron was introduced to increase the length of homology between two non-functional halves of the hygromycin resistance gene (the 5′ half being located in pHR23; see below). Additionally, including an intron into a broken marker system enables capturing and characterization of recombination events that include insertions and deletions.

To screen for integration via HR, the recipient strain is transformed with pHR23. From the 5′ to 3′ end, pHR23 consists of a region identical to 4.3 kb of the 5′ end of the pHR18 vector, a complete paromomycin resistance cassette, and the 5′ half of the hygromycin resistance cassette (β-tub promoter followed by first 510 bp of Hyg CDS) followed by the intron which is also in pHR18 (Fig. 1b). Homologous recombination between the intron sequences reconstitutes a complete (intron-containing) hygromycin resistance cassette (Fig. 2a). By plating an aliquot of the pHR23 transformation on paromomycin we can estimate total number of pHR23 transformants. Comparison between the number of colonies obtained on hygromycin to those obtained on paromomycin allows calculation of an HR:NHR ratio from a single transformation. By including a region identical to the 4.3 kb at the 5′ end of the pHR18 vector, both single and double recombination events can occur and be captured (Fig. 2). By extending this region of homology into the ARG7 cassette, recombination within this region disrupts the *ARG7* cassette found in pHR18 (see below). Thus, the system is designed such that four types of integration events can be detected: 1) random integration of pHR23 results in a strain that is resistant to paramomycin (PAR^+), expresses ARS2 ($ARS2^+$), is arginine prototroph (ARG^+) but is sensitive to hygromycin (HYG^-); 2) HR mediated integration via double crossover (wherein recombination occurs in both regions of homology) results in a strain that is $ARS2^+$, ARG^-, PAR^+ and HYG^+ (Fig. 2a); 3) a single crossover within the intron after which the recombination branch migrates to the 5′ end of the pHR23 template and then recombines (non-homologously) results in a strain that is $ARS2^+$, ARG^+, PAR^+ and HYG^+ (Fig. 2b); and 4) a single crossover within the 4.3 kb 5′ homology region after which the recombination branch migrates to the 3′ end of the pHR23 template and then recombines (non-homologously) results

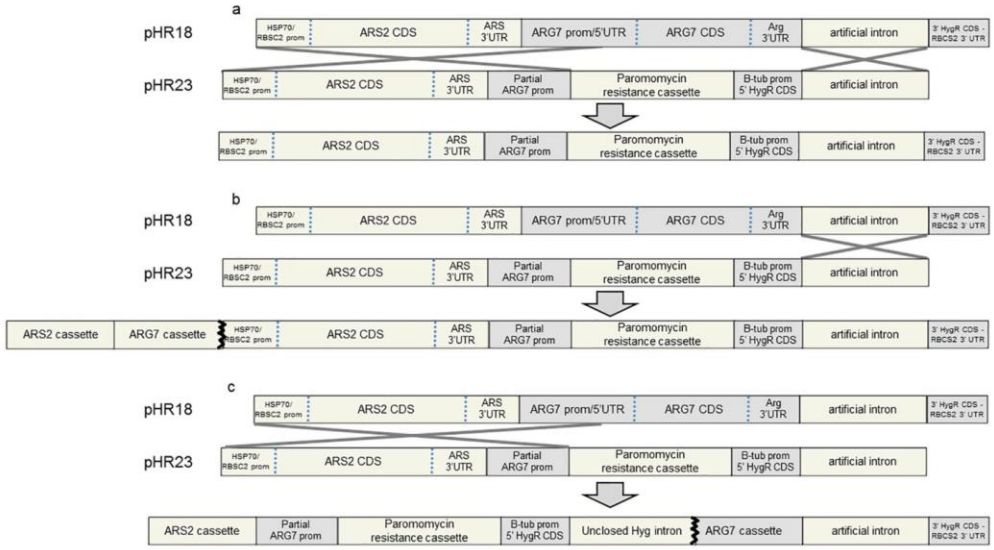

Fig. 2. Predicted outcomes for achieving hygromycin-resistant transformants. a) Schematic of the expected recombination site after a perfect double recombination event. b) Schematic of the expected recombination site after a single homologous recombination (crossover) within the 3′ homology region (artificial intron). The jagged border between the ARG7 cassette and the 5′ homology indicates that this junction is formed non-homologously after branch migration reached the end of the pHR23 template. Bacterial plasmid backbone sequence (not pictured) may also be incorporated at this site. c) Schematic of the expected recombination site after a single homologous recombination within the 5′ homology region of pHR18. Again, the jagged border indicates a region of non-homologous junction. The schematics are generally drawn to scale (1 kb = 1.8 cm). Due to space constraints the ARS and ARG7 cassettes are not this scale in the outcome schematic in b) and c). "Closed" vs "Unclosed Hyg intron" relates to whether the intron contains its closing splice site. Prior to transformation, pHR23 is linearized within the Hyg intron 265 bp upstream of the closing splice site resulting in an "unclosed intron". In pHR18, the intron is complete and contains the closing splice site.

in a strain that is ARS2$^+$, ARG$^-$, PAR$^+$ and HYG$^+$ (Fig. 2c). For the last instance, the site of linearization within pHR23 is critical (see arrow in Fig. 1b). By linearizing upstream of — and thus removing — the closing splice site in the intron in pHR23, the sequence between the 5′ and 3′ halves of the hygromycin resistance gene can be recognized as a single intron and be spliced out to yield a functional hygromycin gene. The last instance yields a reconstituted hygromycin resistance cassette with an extended intronic sequence.

2.2. Validation of experimental design

A 1.6 kb intron containing canonical splice sites (see M&M) was inserted into the hygromycin resistance gene in a putative splice site (AG:GT) [7] located at nucleotide 510 (total CDS size: 999 bp). The complete intron-containing hygromycin cassette was able to confer hygromycin resistance, indicating proper intron recognition and splicing (data not shown). pHR18 and pHR23 were each transformed separately into wild type *C. reinhardtii*. No hygromycin-resistant colonies were obtained with either vector, confirming that each half of the hygromycin cassette is non-functional on their own (data not shown). Lastly, we co-transformed a wild-type strain CC-1690 with both pHR18 and pHR23. Previously, the arg*7–8* mutant strain was rescued by transforming with two overlapping plasmids each containing the respective halves of ARG7 [3]. Here, we analyzed 312 colonies obtained on paramomycin plates from a pHR18/pHR23 co-transformation by patching onto hygromycin plates. One of the colonies grew on hygromycin. Characterization of the PCR-amplified junction by sequencing showed that the transformant was hygromycin resistant as a result of recombination within the intron homology region. A significant portion of the intron had been deleted (945 bp of the 1616 bp intron were missing). This result confirmed that the two fragments can be used to capture HR recombination events and that the intron-containing hygromycin resistance cassette constitutes a flexible broken marker suitable for also capturing and characterizing imperfect HR events.

2.3. Creation of recipient strain

To create a stable recipient strain, we transformed an arginine auxotrophic strain with pHR18 and selected on TAP plates w/o arginine [8]. Screening by PCR and arylsulfatase activity identified three colonies with intact insertion of pHR18 (B12, C6 and F5) (Fig. 3). Strain B12 was used for all subsequent experiments. Using this strain, a high-throughput strategy was developed such that thousands of transformants could be screened at once (Fig. 4).

2.4. Measuring HR frequency

The HR frequency (HR_{freq}) is expressed as the ratio of integrations via HR to total theoretical number of transformants (tot #transformants). As pHR23 contains an intact paromomycin cassette, tot #transformants were estimated by plating an aliquot on paromomycin. Transformation recovery suspensions were pooled (total vol) and an aliquot (paro vol) was plated on paromomycin while the remaining suspension (total vol – paro vol) was plated on hygromycin. Colonies on paromomycin (paro col) and hygromycin (hygro col) were counted and HR_{freq} calculated (Fig. 4):

$$\frac{\#paro\ col}{paro\ vol} \times (total\ vol - paro\ vol) = tot\ \#transformants$$

$$\frac{\#hygro\ col}{tot\ \#transformants} = HRfreq$$

In four independent experiments, we transformed the B12 strain with pHR23. Across all four experiments, the HR frequency in B12 appears relatively stable at $5.7 \pm 1.5 \times 10^{-5}$ (Table 1). > 95% of all colonies survived patching onto a 33% higher concentration of hygromycin (20 μg/ml instead of 15 μg/ml) indicating a low rate of false positives in the initial screen.

2.5. Characterization of perfect and imperfect recombination events within intron region

44 of the hygromycin resistant colonies were selected for further analysis. The recombination junction within the 1.6 kb intron was amplified by PCR using flanking primers (Fig. 5a), and the resulting fragment (1.8 kb) was blunt-end cloned into a CloneJet plasmid. In 16 of the 44 plasmids, restriction analyses with flanking restriction sites yielded two fragments (1.8 and 2.9 kb) (Fig. 5d). These fragments correspond to the CloneJet backbone and an inserted fragment of 1.8 kb. As indicated in Figs. 2a and 5a + b, this pattern indicates that integration of pHR23 into the recipient strain likely proceeded via error-free HR with crossover in at least the intron homolgy region resembling Fig. 2a and 2b. Sanger sequencing of the junction subsequently verified the integrity of the recombination in these transformants. In the remainder of transformants that were analyzed, we observed insertions, deletions, and concatemerizations within the recombination site (e.g. Fig. 5d and described in detail in Table S1). In at least three transformants, a portion of the intron was deleted (e.g. sequence R3-14 in Supplementary data). At least one other recombinant (sequence R3-07 in Supplementary data) fit the model from Fig. 2c, in which the junction between the intron and the *ARG7* cassette experienced loss of sequence on both sides prior to non-homologous end joining. In combination, these results demonstrate the flexibility provided by including a long intron into a broken marker system. Because the repair errors occurred within the intron, they were presumably spliced out during mRNA processing and thereby did not impair the functionality of the hygromycin resistance gene.

2.6. Characterization of HR within 5′ end of pHR18

To characterize recombination within the 5′ homology region, we exploited that recombination within that region could abolish functionality of the *ARG7* cassette without affecting hygromycin resistance. This homology region was, however, too long to span effectively with PCR for sequencing (> 4 kb). Instead, 28 of the 44 transformants were

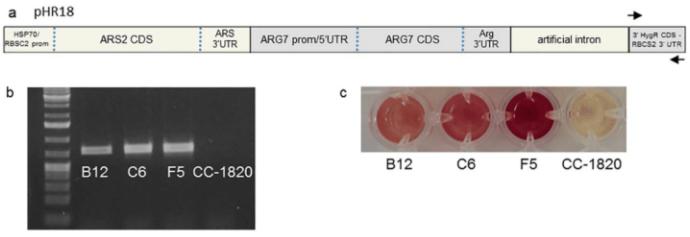

Fig. 3. Criteria for selecting recipient strains. a) Arrows indicate the location of primers annealing to pHR18, used to screen candidate algal colonies for potential full integration of the vector. b) PCR verification that the 3′ partial hygromycin resistance gene in the pHR18 vector integrated into the genome in three candidate recipient strains that were arginine-auxotrophic, using the primers indicated in (a). Parental strain CC-1820 does not contain the cassette. c) Arylsulfatase assays on the candidate recipient strains, indicating that all three strains constitutively express the ARS2 gene, unlike CC-1820.

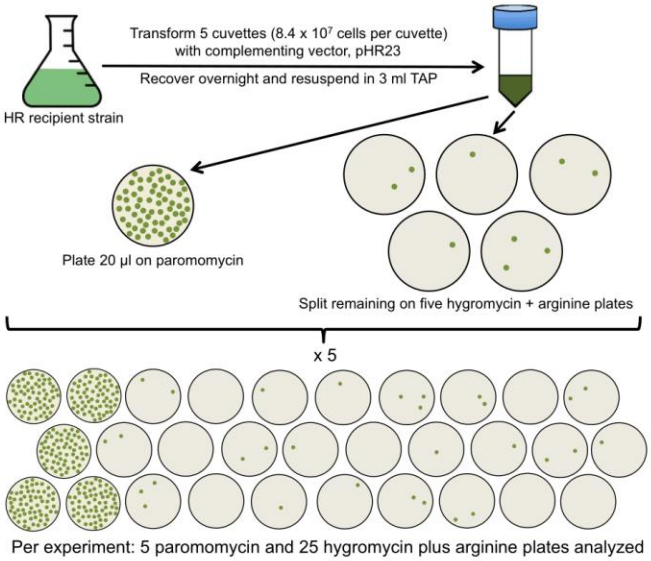

Fig. 4. Schematic of the experimental protocol for high-throughput identification of homologous recombinants. Each experiment involved 25 electroporations, handled in pools of five. Each recovery flask of five pooled transformations was resuspended in 3 ml of TAP media, and 20 µl was plated onto paromomycin, to measure the total transformation efficiency. The rest was split evenly onto five hygromycin plus arginine plates to select for homologous recombinants in which the truncated hygromycin resistance cassette was repaired by homologous recombination. Colony counts for each experiment reflect sums from five paromomycin plates and 25 hygromycin plus arginine plates.

Table 1
Reproducibility of basal HR efficiency measurement.

Experiment number	Colonies obtained on paromomycin	Total theoretical number of transformants	Colonies obtained on hygromycin	HR frequency[a]
1	1659	247,191	21	$8.5 \pm 3.6 \times 10^{-5}$
2	1929	287,421	16	$5.6 \pm 2.7 \times 10^{-5}$
3	2640	393,360	16	$4.1 \pm 2.0 \times 10^{-5}$
4	457	68,093	4	$5.9 \pm 5.8 \times 10^{-5}$
Aggregate basal homologous recombination frequency in strain B12				$5.7 \pm 1.5 \times 10^{-5}$

[a] Errors represent normal approximation of binomial 95% confidence intervals.

tested for arginine auxotrophy. 16 of these clones had mostly or entirely lost the ability to grow without supplemented arginine (Fig. 5e and f), indicating that in about half of the recombinants, the *ARG7* gene within the recipient cassette was disrupted. This does not necessarily indicate that homologous recombination occurred within the 5′ homology region, as the *ARG7* cassette could have been partially deleted during non-homologous end-joining following a single crossover in the intron (as in Fig. 2b).

3. Discussion

A number of methods for measuring the frequency of homologous recombination have been developed in *C. reinhardtii*. For example, a p5′Δ truncated NIT1 construct was used to restore nitrate prototrophy

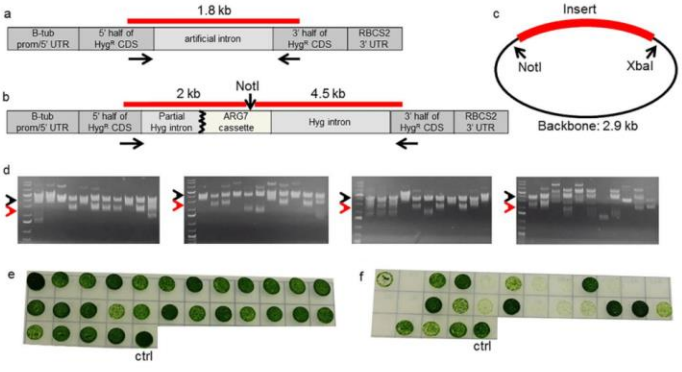

Fig. 5. Characterization of individual recombination events. a) Anticipated hygromycin cassette resulting from a double recombination event, homology-directed repair, or single recombination within the intron. Arrows below the graphics indicate primers used to amplify the recombination junction b) Anticipated cassette resulting from a single recombination event within the 5′ homology region (of pHR18) followed by non-homologous end joining. The vertical arrow indicates a *Not*I site near the 3′ end of the ARG7 cassette. c) Amplified recombination junctions inserted into the CloneJet backbone, NotI and *Xba*I sites in the vector on either side of the insert. d) Restriction digest of 44 CloneJet plasmids containing amplified recombination junctions. A transformant exhibiting precise recombination within the intron (as depicted in 5a) would have 2.9 kb (backbone) and 1.8 kb (amplified intron) fragments, indicated by black and red arrowheads. A transformant resembling 5b would have 4.5 kb, 2.9 kb, and 2 kb fragments. Additional fragments or fragments of different sizes indicate concatemerizations, rearrangements, or partial deletions. e) Growth of 28 independent hygromycin-resistant transformants on TAP media with arginine added, and f) without supplemented arginine. The last strain is a wild type arginine prototrophic control (ctrl). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to the *nit1-305* strain when grown on nitrate as sole nitrogen source [2,9], a "HR disruption cassette" was used to replace the NIT8 gene with a selection marker to generate chlorate insensitive strains [1] and arginine prototrophy has been restored in arg7–8 mutants as a selection for complementation by recombination [3,9,10]. Integration via HR in each of these approaches yields phenotypes that can be scored visually. However, they suffer from various shortcomings. Collectively, the restoration of prototrophy approaches does not measure HR and NHR from the same experiment [2,4,9,10]. In comparison, the approach wherein NIT8 is replaced does measure HR to NHR frequency directly from the same transformation event, but due to the mutagenic effect of chlorate and the ability to obtain NIT^{-} strains by mutating any of the numerous genes involved in nitrate utilization, the rate of false positives was high [1]. To avoid the ambiguity associated with targeting endogenous genes, Zorin et al. developed a fully exogenous system using a truncated paromomycin resistance gene [4]. However, their system did not report HR and NHR frequency from the same transformation, had a short length of homology, did not allow double cross over events to occur and was not compatible with insertions, deletions, or rearrangements that affected the 5′ end of their construct [4,11]. In this study, we have designed a split broken marker system that addresses these shortcomings. In designing the system, we included a long intron into the split marker system to provide flexibility at the recombination site so that unpredictable inaccurate recombination outcomes could be captured and easily examined in detail; ensure that each half of the reconstituted selectable marker is non-functional to avoid false-positive complications while at the same time providing a long region of homology between the two halves of the split marker to ensure sufficient overlap for HR to occur. In addition, we combined the intron-containing split marker system with 3 other markers to allow the total number of integration vs homologous recombination events to be estimated from the same transformation and to allow for double cross-overs to occur and easily scored.

In approximately one third of the recombinants analyzed, the recombination site within the hygromycin resistance cassette appeared exactly as predicted from a perfect crossover or homology-directed repair within the intron-homology region, i.e. with no associated insertions or deletions. In the remaining events, insertions and deletions in the homology region were observed. This confirms the flexibility provided by using an intron as the homology sequence. The length appears sufficient to allow perfect HR to occur while errors therein still allow recovery of hygromycin-resistant transformants (Fig. 2c). Based on these numbers, we estimate that the frequency of perfect HR in the B12 strain (i.e. wherein an exact replacement of one DNA segment for another occurs) is lower than estimated based solely on colony counting ($\sim 2 \times 10^{-5}$, i.e. $\sim 1/3$ of 5.7×10^{-5}). In either case, the HR frequencies presented here fall within the range of previous estimates for plant HR frequencies, reported at around 10^{-4} to 10^{-5} [2,9]. Thus, as evidenced by capturing many recombination events that included both perfect and imperfect recombination within the homology regions, and by obtaining consistent HR frequencies across several independent experiments, we believe that the B12 strain combined with the pHR23 plasmid and the protocol presented (Fig. 4), provides a robust system suitable for benchmarking an HR frequency and for subsequently measuring the effect of approaches aimed at increasing HR accuracy and efficiency.

Comparison of integration of complementing constructs at the NIT1 vs ARG7 loci, respectively suggested that HR varies depending on genome location [9]. Thus, the HR frequency measured in the B12 strain may be overestimated if the pHR18 recipient construct has been inserted in a region particularly favorable for HR or alternatively at multiple locations in the genome. Care should therefore be taken not to deduce from our results an estimate of the overall HR frequency at any given site in the *C. reinhardtii* nuclear genome. The main novel attribute of the split marker HR detection system developed in this study — the introduction of a long intron — ensured that the length of homology lays within estimated thresholds for HR to occur more efficiently [9] and ensured phenotypic flexibility at the homology region within the split marker. This feature can be applied to HR detection systems in a wide range of algal species, and even to organisms beyond algae.

4. Materials and methods

4.1. Vector design and construction

All plasmids were designed in silico using Geneious sequence analysis software (Auckland, New Zealand) and constructed seamlessly with USER cloning as described previously [12], into the USER cloning-compatible vector Hcr1 (pBluescript into which the USER cloning cassette (GCTGAGGGTTTAATATTAAGACCTCAGC) has been inserted as previously described [13]). Fragments were amplified from existing plasmids using a modified Phusion polymerase X7 [15], in 5× Phusion GC buffer (New England Biolabs) with 2 M betaine (Sigma-Aldrich) to accommodate high GC content. The 1.6 kb intron containing canonical constitutive splice sites was obtained from a list of long, native *C. reinhardtii* introns (Mark Rogers Colorado State University, personal communication; genomic locus Cre03.g190281), cloned from genomic DNA, and inserted into the hygromycin resistance gene at the location described. Plasmid DNA from candidate colonies was sequenced in full. Plasmids pHR18 and pHR23 are available through the *Chlamydomonas* Resource Center (www.chlamycollection.org).

4.2. Algal strains and growth conditions

All strains used in this work (CC-1820 and CC-1690) were obtained from the *Chlamydomonas* Resource Center, www.chlamycollection.org (St. Paul, MN). The recipient strain, B12, described in this work is also available at the *Chlamydomonas* Resource Center. All strains were grown in standard TAP media supplemented with 100 μg/ml sterile-filtered L-arginine (Sigma-Aldrich) prior to transformation. Liquid cultures were grown in constant light with shaking.

4.3. Algal transformation by electroporation, replicates, and statistical analysis

Chlamydomonas recipient strain B12 cultures were grown in TAP to a density of 3–5 × 10^6 cells/ml prior to electroporation. They were then pelleted by centrifugation at 2000 rpm for 10 min and resuspended at 3 × 10^8 cells/ml in *Chlamydomonas* transformation reagent (Life Technologies, Carlsbad, CA). For each independent replicate, 25 cuvettes were transformed in sets of five as follows: 1400 μl of resuspended cells were incubated with 5 μg of linearized DNA at room temperature for 5 min and transferred into five 4 mm electroporation cuvettes. The cuvettes were incubated on ice for 5 min prior to electroporation in a BioRad Gene Pulser with the following settings: 500 V, 50 μF capacitance, 800 Ω resistance. After electroporation, cells were recovered at room temperature for 5 min, then all five cuvettes were pooled into a 50 ml flask of TAP supplemented with 100 μg/ml of L-arginine and 40 mM sucrose. The cultures recovered for 12–16 h in constant light on a shaker. Cells were then pelleted by centrifugation (2000 rpm for 10 min) and resuspended in 3 ml TAP. From this resuspension, 20 μl was mixed with 500 μl of TAP and spread onto agar plates containing TAP medium with 20 μg/ml paromomycin to obtain a measure of the overall transformation efficiency. The remaining resuspended cells were divided evenly among five TAP plates containing 15 μg/ml hygromycin and 100 μg/ml arginine, to test for homologous recombinants at the desired locus. Colonies were typically visible on paromomycin within 4–5 days of plating and visible on hygromycin within 5–6 days, and all colonies were counted within 10 days of plating. Binomial 95% confidence intervals were calculated for each experiment using the normal approximation method, and the overall HR frequency was calculated with binomial 95% confidence intervals from all data taken in

aggregate.

4.4. Arylsulfatase assays

Colonies that were PCR-positive for the 3′ end of the recipient cassette using primers HR27 (AACAGGTGTTCCCGGAGCTGCTGCGC) and HR28 (GGTCTTAATCGCTTCAAATACGCCCAGCC) were screened for arylsulfatase activity. Colonies were inoculated into deep-well plates (Axygen Scientific) containing 0.5 ml TAP. Cultures were grown for 3–4 days to saturation and centrifuged for 5 min at 2000 rpm to pellet the cells. 100 µl of supernatant (media only, no cells) was transferred to a flat-bottom clear plate (Corning). To each well, 25 µl of N-SO_4 reaction buffer (2 M glycine-NaOH, pH 9, 50 mM imidazole, 4 mM α-naphthyl sulfate (Sigma-Aldrich)) was added, then incubated at 37 °C for 1 h. 125 µl stopping buffer (0.2 M sodium acetate, pH 4.8, 4% SDS) was added, followed by 25 µl of 10 mg/ml Fast Blue B Salt (Sigma-Aldrich). The color was allowed to develop for 1 min and read using an Infinite 200 Pro platereader (Tecan) for absorbance at 540 nm.

4.5. Sequencing and restriction analysis of recombination sites

Transformants were grown in 5 ml TAP cultures to saturation, followed by genomic DNA extraction as previously described [14]. Genomic DNA was used as template for PCR spanning the intron with primers HH235 (CTACCTGGTCATGTCGCGGATGAC) and LS330 (CAGGAGGGTGTCCACGTCAG), using Phusion GC buffer and 2 M betaine to enhance amplification. PCR-amplified fragments were gel-purified and cloned into pJet2.1 (CloneJET Kit, Thermo Scientific), sequenced via Sanger sequencing, and aligned using Geneious software. The CloneJET plasmids containing the PCR-amplified recombination junctions were analyzed by restriction digest using *Not*I-HF and *Xba*I (New England Biolabs). The CloneJET backbone contains one of each of these sites, one on each side of the PCR insert. The *ARG7* cassette contains a NotI site near the 3′ end, but these enzymes are otherwise absent from both pHR18 and pHR23.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2018.02.005.

Author contributions

HHN-E, EAS, and SPM contributed to the conception and design of the study; HHN-E, EAS, JO, KTDH, and PK acquired and analyzed data; HHN-E and EAS drafted the manuscript and all authors provided critical revisions and approved the final version of the manuscript. HHN-E (huha@plen.ku.dk), EAS (elizabeth.specht@colorado.edu) and SPM (smayfield@ucsd.edu) take responsibility for the integrity of the work as a whole, from inception to finished article.

Acknowledgements

We thank Michael Hannon and Michael Mendez for many useful discussions during the design of this system, and Adam Labadorf and Mark Rogers from the Ben-Hur lab at Colorado State University for graciously providing a compiled list of constitutively spliced *Chlamydomonas* introns. We also thank Farzad Haerizadeh and his research team at Life Technologies for their support and collaboration throughout this work. This work was supported by the Department of Energy [DE-EE0003373]; the National Science Foundation graduate research fellowship [DGE-1144086 to EAS]; and the Danish National Research Foundation DNRF99 grant and Danish Research Council for Technology and Production [09-065827/274-08-0354] to HHN-E. These funding sources had no involvement in the study design; collection, analysis, and interpretation of data; writing of this report; or the decision to publish.

Conflict of interest

The Authors declare no conflict of financial interest.

Statement of Informed Consent, Human/Animal Rights

No conflicts, informed consent, human or animal rights applicable.

References

[1] J.A. Nelson, P.A. Lefebvre, Targeted disruption of the NIT8 gene in *Chlamydomonas reinhardtii*, Mol. Cell. Biol. 15 (1995) 5762–5769.
[2] O.A. Sodeinde, K.L. Kindle, Homologous recombination in the nuclear genome of *Chlamydomonas reinhardtii*, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 9199–9203.
[3] N.J. Gumpel, J.D. Rochaix, S. Purton, Studies on homologous recombination in the green alga *Chlamydomonas reinhardtii*, Curr. Genet. 26 (1994) 438–442, http://dx.doi.org/10.1007/BF00309931.
[4] B. Zorin, P. Hegemann, I. Sizova, Nuclear-gene targeting by using single-stranded DNA avoids illegitimate DNA integration in *Chlamydomonas reinhardtii*, Eukaryot. Cell 4 (2005) 1264–1272, http://dx.doi.org/10.1128/EC.4.7.1264-1272.2005.
[5] E.A. Specht, H.H. Nour-Eldin, K.T.D. Hoang, S.P. Mayfield, An improved ARS2-derived nuclear reporter enhances the efficiency and ease of genetic engineering in *Chlamydomonas*, Biotechnol. J. 10 (2015) 473–479, http://dx.doi.org/10.1002/biot.201400172.
[6] H.H. Nour-Eldin, E.A. Specht, S.P. Mayfield, An improved ARG7-based selection cassette with highly efficient transformation rates and a small size suitable for complex expression constructs, Algal Res. 16 (2016) 383–386, http://dx.doi.org/10.1016/j.algal.2016.03.036.
[7] A. Labadorf, A. Link, M.F. Rogers, J. Thomas, A.S. Reddy, A. Ben-Hur, Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*, BMC Genomics 11 (2010) 114, http://dx.doi.org/10.1186/1471-2164-11-114.
[8] R. Loppes, A new class of arginine-requiring mutants in *Chlamydomonas reinhardtii*, Mol. Gen. Genet. MGG 104 (1969) 172–177, http://dx.doi.org/10.1007/BF00272799.
[9] A. Plecenikova, W. Mages, Ó.S. Andrésson, D. Hrossova, S. Valuchova, D. Vlcek, et al., Studies on recombination processes in two *Chlamydomonas reinhardtii* endogenous genes, NIT1 and ARG7, Protist 164 (2013) 570–582, http://dx.doi.org/10.1016/j.protis.2013.05.004.
[10] W. Mages, O. Heinrich, G. Treuner, D. Vlcek, I. Daubnerova, M. Slaninova, Complementation of the *Chlamydomonas reinhardtii* arg7–8 (arg2) point mutation by recombination with a truncated nonfunctional ARG7 gene, Protist 158 (2007) 435–446, http://dx.doi.org/10.1016/j.protis.2007.05.001.
[11] I. Sizova, A. Greiner, M. Awasthi, S. Kateriya, P. Hegemann, Nuclear gene targeting in *Chlamydomonas* using engineered zinc-finger nucleases, Plant J. 73 (2013) 873–882, http://dx.doi.org/10.1111/tpj.12066.
[12] F. Geu-Flores, H.H. Nour-Eldin, M.T. Nielsen, B.A. Halkier, USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products, Nucleic Acids Res. 35 (2007) e55, , http://dx.doi.org/10.1093/nar/gkm106.
[13] H.H. Nour-Eldin, B.G. Hansen, M.H.H. Nørholm, J.K. Jensen, B.A. Halkier, Advancing uracil-excision based cloning towards an ideal technique for cloning PCR fragments, Nucleic Acids Res. 34 (2006) e122, , http://dx.doi.org/10.1093/nar/gkl635.
[14] S.M. Newman, J.E. Boynton, N.W. Gillham, B.L. Randolph-Anderson, A.M. Johnson, E.H. Harris, Transformation of chloroplast ribosomal RNA genes in *Chlamydomonas*: molecular and genetic characterization of integration events, Genetics 126 (1990) 875–888.
[15] Nørholm, H.H. Morten, A mutant Pfu DNA polymerase designed for advanced uracil-excision DNA engineering (BMC Biotechnol.), 10 (1) (2010) 21.

Acknowledgements

Chapter 7, in full, is a reprint of material as it appears in *Algal Research* 2018. Nour-Eldin, Hussam; Specht, Elizabeth; Ostrand, Joseph; Karunanithi, Prema; Hoang, Kevin; Mayfield, Stephen. The dissertation author is the tertiary investigator and author of this paper.

CONCLUSION

The chapters in this dissertation provide a valuable foundation to the advancement of algae as an industrially relevant manufacturing platform. Application of progressive technology and techniques that we have learned from other model organisms to green algae has fostered a potential for algal products in the fields of energy, food, materials and therapeutics. For one organism to find commercial success in such a myriad of fields, we must be able tightly control its growth and production of desired products.

The recent development of novel algal-derived products for medicine and enzymes has put new focus on optimizing algal growth in closed systems. Cultivation of algae for therapeutics and high-value products requires precise control of growth conditions to maximize product yield and to comply with regulatory restrictions. The research presented in Chapter 3 provides an intricate characterization of *Chlamydomonas reinhardtii* when grown under mixotrophic conditions in fermenters and provides valuable insight on how to grow *C. reinhardtii* to the high cell densities required for economical cultivation. This research also showed that in closed systems, high cell density and energy costs limit the effect of light-driven genetic tools. This is a major hurdle for all of the best production strains of *C. reinhardtii* because of the lapse in strong genetic tools that confer transgene expression in the dark. The research reported in Chapter 4 showed that by encouraging genome shuffling via breeding and mutagenesis, strains of *C. reinhardtii* with dramatically increased levels of recombinant protein accumulation when cultured in the dark could be created. The untargeted nature of this research makes it a simple strategy to confer phenotypic variance and facilitate the production of more robust algal strains. Translational techniques like this will be necessary for the rapid development of the diverse algal species being investigated as production platforms.

Chapters 5 and 6 present a more deliberate approach to generation of robust commercial strains by analyzing endogenous regulatory elements and using them to assemble synthetic

promoters. The synthetic promoters generated were shown to drive nuclear gene expression more effectively than any previously devised promoter system in *C. reinhardtii*. Although proteins expressed by the nuclear genome are more desirable because they can be post-translationally modified and secreted, they accumulate at insufficient levels for true scalable production. Synthetic systems like those described in Chapter 6 outline how we can make strides to producing recombinant proteins to levels seen in other model organisms.

Ultimately, it is important to remember that the bio-manufacturing platforms ubiquitous today- namely yeast, *E. coli*, and mammalian cells, have become successful because of years of research and dedication. The groundwork laid by these systems has outlined the critical aspects necessary for the success of a biological production platform. The research presented in this dissertation shows significant improvement of the algal production platform on its path to becoming a ubiquitous industrial organism.