SANDIA REPORT

SAND2015-2319
Unlimited Release
April, 2015

Systems-Level Synthetic Biology for Advanced Biofuel Production

Anne M. Ruffing
Travis J. Jensen
Lucas M. Strickland
Stephen Meserole
David Tallant

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: http://www.ntis.gov/search

SAND2015-2319
Unlimited Release
April 2015

Systems-Level Synthetic Biology for Advanced Biofuel Production

Anne M. Ruffing,[1] Travis J. Jensen,[1] Lucas M. Strickland,[1] Stephen Meserole,[2] David Tallant[3]

[1]Bioenergy and Defense Technologies
[2]Materials Reliability
[3]Materials Characterization & Performance

Sandia National Laboratories
P.O. Box 5800
Albuquerque, New Mexico 87185-MS1413

Abstract

Cyanobacteria have been shown to be capable of producing a variety of advanced biofuels; however, product yields remain well below those necessary for large scale production. New genetic tools and high throughput metabolic engineering techniques are needed to optimize cyanobacterial metabolisms for enhanced biofuel production. Towards this goal, this project advances the development of a multiple promoter replacement technique for systems-level optimization of gene expression in a model cyanobacterial host: *Synechococcus* sp. PCC 7002. To realize this multiple-target approach, key capabilities were developed, including a high throughput detection method for advanced biofuels, enhanced transformation efficiency, and genetic tools for *Synechococcus* sp. PCC 7002. Moreover, several additional obstacles were identified for realization of this multiple promoter replacement technique. The techniques and tools developed in this project will help to enable future efforts in the advancement of cyanobacterial biofuels.

ACKNOWLEDGMENTS

We thank Drs. Bryan Carson and Jerilyn Timlin for graciously allowing members of this project to use equipment in the laboratories. We also thank the Biosciences management for selecting this project for funding, along with the LDRD office at Sandia National Laboratories for establishing the Early Career LDRD program. Lastly, we thank the American taxpayers for their support of government programs and members of Congress for establishing the LDRD program.

CONTENTS

FIGURES

TABLE

NOMENCLATURE

ATCC	American Type Culture Collection
ATR	attenuated total reflectance
CO_2	carbon dioxide
DNA	deoxyribonucleic acid
FACS	fluorescence activated cell sorting
FFA	free fatty acid
FTIR	Fourier transform infrared
GC/MS	gas chromatography/mass spectrometry
hGFP	hybrid green fluorescent protein
IR	infrared
MAGE	multiplexed automated genome engineering
mOrange	orange fluorescent protein
NS1	neutral site 1
NS2	neutral site 2
PCR	polymerase chain reaction
qPCR	quantitative PCR
RNA	ribonucleic acid
Ypet	yellow fluorescent protein

1. INTRODUCTION AND PROJECT OBJECTIVES

1.1. Cyanobacterial Biofuel Production

Cyanobacteria are ideal hosts for biofuel production due to their amenability for genetic manipulation and their ability to directly convert CO_2 and sunlight into drop-in replacement fuels with such genetic modification. In fact, preliminary research has demonstrated that cyanobacteria can produce biofuels and biofuel precursors including ethanol,[1, 2] isopropanol,[3] isobutanol,[4, 5] ethylene,[6] isoprene,[7] limonene,[8] bisabolene,[8] free fatty acids,[9, 10] fatty alcohols,[11] and alkanes.[11] While these studies have shown the flexibility of cyanobacterial metabolism with regard to biofuel production, biofuel productivities remain well below the levels needed for economically viable, industrial scale production.

1.2. Genetic Modification of Cyanobacteria

A main advantage of cyanobacteria relative to eukaryotic algae is their facile genetic manipulation. In fact, many of the model strains of cyanobacteria are naturally transformable, and as prokaryotes, targeted genome integration via homologous recombination is feasible.[12, 13] Despite these advantages, the tools for genetic modification of cyanobacteria are primarily derived from fundamental studies of cyanobacteria manipulation for the understanding of cyanobacterial photosynthesis and circadian rhythm.[12] These tools include promoters from Escherichia coli and cyanobacterial promoters known to drive genes involved in carbon fixation and photosynthesis. Recent studies have started to focus on developing better characterized and more flexible promoters for the advancement of cyanobacterial as a chassis organism.[14-16] These developments may enable metabolic engineers to achieve higher biofuel yields in cyanobacterial hosts; however, our incomplete understanding of the metabolism and regulatory mechanisms of cyanobacteria, particularly under natural diurnal light conditions, continues to limit our ability to optimize biofuel production in these hosts.

High-throughput genetic modification methods have recently been developed for the optimization of other host organisms, such as *E. coli*.[17] Multiplexed Automated Genome Engineering (MAGE) has proven particularly useful for the optimization of chemical synthesis in *E. coli* by targeting multiple ribosome binding sites or promoters throughout the *E. coli* genome.[18, 19] The standard application of MAGE relies on random nucleotide variation within short oligonucleotide sequences targeting specific regions in the *E. coli* genome; thus, many of the insertions with MAGE may be either neutral or even detrimental for the organism or application. MAGE therefore relies on a screening based approach to optimize the mutations for products readily detectable by high throughput approaches. As many biofuels and biofuel precursors require laborious sample preparation and analysis, it is challenge to apply a semi-random, high throughput approach such as MAGE for biofuel optimization in either *E. coli* or cyanobacterial hosts. Thus, there remains a need for high throughput genetic modification and optimization of cyanobacteria for biofuel production that cannot be met with standard *E. coli* modification techniques.

1.3. *Synechococcus* sp. PCC 7002 as Host

Synechococcus sp. PCC 7002 is an ideal candidate for development of a cyanobacterial chassis for biofuel production. The genome of *Synechococcus* sp. PCC 7002 is sequenced and available; it has demonstrated protocols for genetic manipulation; the cyanobacterium is naturally transformable; and it has a rapid doubling time of 2.6 to 4 hours.[20, 21] Moreover,

Synechococcus sp. PCC 7002 has additional properties which make it amenable for advanced biofuel production. Compared to the freshwater cyanobacterial model organisms, *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, *Synechococcus* sp. PCC 7002 has a naturally high tolerance of potential biofuel products and has demonstrated capabilities for producing high levels of free fatty acids, key precursors for many advanced biofuels.[22] Other beneficial strain traits include its high salt, light, and temperature tolerances. *Synechococcus* sp. PCC 7002 can tolerate up to 3M salt,[23] allowing for growth on seawater and other non-freshwater resources and reducing potential growth inhibition due to the elevated salinities resulting from evaporation. The strain has also demonstrated a high light tolerance, exceeding that of peak sunlight intensities; *Synechococcus* sp. PCC 7002 growth in outdoor systems will therefore not suffer from photoinhibition. Lastly, *Synechococcus* sp. PCC 7002 can tolerate temperatures up to 40°C,[23] allowing for growth and survival under the high temperatures often reached in closed photobioreactors.

1.4. Project Objectives

This project will investigate a novel method for multiple target integration in *Synechococcus* sp. PCC 7002 to optimize the host for biofuel production. Multiple genome integration cassettes will be designed to replace native genome sequences with specific target sequences. In this project, we will focus on the replacement of native promoters with characterized promoters to rationally optimize gene expression of multiple target genes throughout the genome of *Synechococcus* sp. PCC 7002. Each integration cassette will include three main components: the characterized promoter, an optical reporter (i.e. fluorescent protein) for the detection of genome integration, and flanking homologous regions targeting the native promoter to be replaced. This cassette design is shown in Figure 1. Cassettes targeting multiple promoters will be designed and synthesized, and these DNA cassettes will be simultaneously transformed with repeated rounds of transformation to generate combinatorial mutants with replaced promoters. Each promoter replacement will include the integration of an additional copy of the fluorescent protein reporter, leading to increased signal. Thus, the transformation mixture will be sorted via fluorescence activated cell sorting (FACS) based on the fluorescence intensity, and if necessary, isolated transformants may be subjected to additional rounds of targeted insertion to increase the number of targeted promoter replacements (Figure 1).

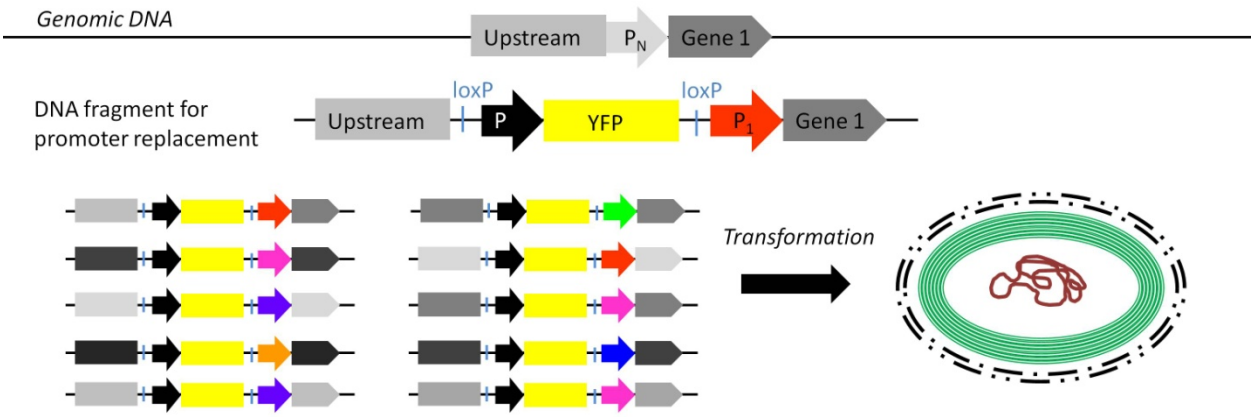

Figure 1. Schematic of proposed multiple target promoter replacement strategy. DNA integration cassette is shown with loxP sites for removal of the fluorescent reporter cassette. Multiple target DNA integration cassettes will be simultaneously transformed into *Synechococcus* sp. PCC 7002.

While the full development of this multiple target promoter replacement technique is not feasible under the time and budget limitations of this project, we will focus on several preliminary objectives to enable this technique:

1. To develop a high throughput technique for screening combinatorial mutants of *Synechococcus* sp. PCC 7002 for advanced biofuel production. While this method will focus on replacement of targeted promoters with well-characterized promoters, the number of combinatorial mutants generated by such an approach will likely require a high throughput screening approach. For example, just 10 targeted promoter replacements will generate a possible 1023 combinatorial mutants.
2. To optimize the transformation efficiency of *Synechococcus* sp. PCC 7002. While *Synechococcus* sp. PCC 7002 is naturally transformable, the transformation efficiency is reported at approximately 10^6 cells per µg of transforming DNA or 0.58% of the cell population.[24] Therefore, in order to introduce just two genome integration events, the efficiency will be only 0.34%, and for 10 targeted integration events, the efficiency will be 0.0043%.
3. To characterize fluorescent proteins and promoters in *Synechococcus* sp. PCC 7002. The proposed method for targeted multiple promoter replacement will require a fluorescent protein reporter that is readily detected and does not interfere with the native fluorescent pigments in *Synechococcus* sp. PCC 7002. Additionally, characterized promoters that interface with the native metabolic and regulatory systems in *Synechococcus* sp. PCC 7002 are required for use in the targeted integration cassettes.

Progress towards meeting these objectives are detailed in the subsequent sections of this SAND report.

1.5. Disclaimer

Due to potential copyright issues, data to be included in other journal publications currently in preparation have been excluded from this SAND report. References for these future publications are provided based on the anticipated journal for submission; however, this may be

subject to change. For updated information on publications including this data, please refer to the principal investigator's webpage: http://www.sandia.gov/bioenergy-biodefense/Ruffing.html.

2. MATERIALS AND METHODS

2.1. Materials

Materials and kits used in this project are described in our publications.[10, 24, 25] *Synechococcus* sp. PCC 7002 was obtained from the American Type Culture Collection (ATCC 27264).

2.2. Cyanobacterial Cultivation

Synechococcus sp. PCC 7002 was cultivated as described previously.[10, 25] Briefly, *Synechococcus* sp. PCC 7002 was cultivated in Innova 42R shaking incubators at 34°C, which was determined to be the optimal growth temperature under these laboratory conditions. Illumination was provided from alternating cool white and plant fluorescent lights at an intensity of 40 – 60 µmol photons m^{-2} s^{-1}. Aeration was provided by shaking at 150 rpm in 500 mL baffled Erlenmeyer flasks with ventilation caps. Strains were maintained on agar plates of medium A+ with appropriate antibiotics. Plated strains were restreaked each month to maintain growth, and cryogenic stocks were made as described previously.[26]

For transformation of *Synechococcus* sp. PCC 7002, plated stock of *Synechococcus* sp. PCC 7002 was inoculated into a 16 mm glass test tube containing 4 mL of medium A+ with ventilation cap. The inoculum was grown at 30°C and the aforementioned light and aeration conditions under continuous light for approximately 4 days. This culture was used to inoculate 100 mL of medium A+ at a dilution of 100x in a 500 mL baffled flask with ventilation cap, and the large culture was incubated for approximately 4 days at 30°C until an OD_{730} between 1.5 and 2.0 was obtained. Prior to transformation, the culture was diluted to the appropriate OD_{730}, and 1 mL was allocated into a 16 mm glass test tube with ventilation cap. 0.5 µg of linearized transforming DNA was added to each 1 mL of culture for transformation. The transformation culture was placed back in the incubator under 30°C and standard light and aeration conditions. Samples of the transformation culture (100 µL) were taken at various times (2 – 24h), diluted with medium A+, and spread on medium A+/agar plates containing 50 µg/mL of kanamycin. Transformation plates were placed under 30°C and 40 – 60 µmol photons m^{-2} s^{-1} for 4 to 7 days, until single colony growth was visible. Colonies were manually counted to determine the transformation efficiency.

2.3. Strain Development

All strains constructed in this project are described in our publication.[24]

2.4. Analytical Measurements

2.4.1. FTIR spectroscopy

Pentadecane was analyzed using an FT-IR spectrometer with SmartArk Attenuated Total Reflectance (ATR) accessory plate. For each sample, 1 mL of medium A+ or medium A+ spiked with pentadecane (69.8 µg/mL) was spotted onto the ATR plate, and the liquid was allowed to evaporate to remove water interference in the IR region. Additional water was used to rinse off salts that interfered with the alkane IR bands.

2.4.2. Nile Red and Bodipy Staining

Nile red and Bodipy staining protocols previously used for detection of fatty acids and neutral lipids (e.g., triacylglycerol) in microalgae[27] were adapted for detection of free fatty acids and alkanes in medium A+. Both Nile red (neutral red, Fisher Scientific) and Bodipy (505/515, 4,4-difluoro-1,3,5,7-tetramethyl-4-bora-3a,4a-diaza-*s*-indacene, Life Technologies) were dissolved in dimethylsulfoxide (DMSO) at concentrations described in Section 3. Dye concentration and staining time were optimized for each dye, and the sensitivity and quantitative capabilities of each dye were assessed in Section 3. Each sample consisted of 100 µL of medium A+ or medium A+ spiked with a representative FFA (palmitic acid) or alkane (pentadecane). Samples were placed in a flat, black bottom microwell plate (Costar) and analyzed using a Turner Biosystems Modulus microplate reader. Nile red and bodipy fluorescence were measured with excitation wavelengths of 525 nm and 490 nm and emission wavelengths of 580-640 nm and 510-570 nm, respectively. For each well, three technical replicate fluorescence readings were averaged for each biological sample, and each data point is the average of three biological experimental replicates. The standard deviation of the biological replicates is represented by the error bars.

2.4.3. Measurement of Genome Copy Number

Genome copy number was measured using a quantitative PCR (qPCR) method previously described.[28, 29] Primers were designed to the origin and terminus of *Synechococcus* sp. PCC 7002 for qPCR amplification of 290 and 282 bp regions, respectively (Table 1). Additional primers were designed to amplify 1,014 and 1,017 bp regions of the origin and terminus that include the smaller qPCR regions; these larger fragments were used to generate a standard calibration curve for converting the qPCR results into quantitative DNA measurements of genome copy number. The larger fragments for calibration standards were amplified using traditional PCR (95°C for 5 min, [95°C for 30 sec, 60°C for 30 sec, 68°C for 1 min] x 30, 68°C for 5 min); these fragments were quantified using PicoGreen dye (Life Technologies) and fluorescence measurements using a Nanodrop fluorometer. Origin and terminus standards were diluted and included in the qPCR measurements to generate calibration curves for determining genome copy number. For the isolation of DNA from *Synechococcus* sp. PCC 7002 cultures, 20 to 1.5 mL samples were placed into 15 mL centrifuge tubes and centrifuged at 3200 x g for 30 minutes in a Beckman Coulter Allegra X-30R centrifuge. The supernatant was discarded and the cell pellets were immediately stored at -20°C. Frozen pellets were thawed and resuspended in 0.5 mL of DNase-free water. Resuspended cells were added to 0.1 mm silica beads (Lysing Matrix B, MP Biomedical) and disrupted via vortexing for 5 min. Cellular debris and the silica beads were removed by centrifugation at 15,000 x g for 20 min in an Eppendorf microcentrifuge. Supernatant containing the released cellular DNA was diluted 1 to 20 x prior to qPCR analysis with DNase free water. qPCR was performed using SYBR Green Master Mix (Life Technologies) and the following PCR conditions in an Applied Biosystems 7500 Real-Time PCR instrument: 50°C for 2 min, 95°C for 10 min, [95°C for 15 sec, 60°C for 1 min] x 40, with standard dissociation curve. Triplicate technical replicate measurements were included for each sample, and no reverse transcriptase (RT) controls were included for both the origin and terminus reaction mixtures.

Table 1. Primers for qPCR measurement of genome copy number for *Synechococcus* sp. PCC 7002

Primer	Sequence
qOriF	GGTACGCCTCACTTCTCCAGA
qOriR	GCTGGGTATCCAAACGGTAATG
OriF	GTGACTCAAAATCCCCAATGG
OriR	GAGGTTTTCCACCGTCATCG
qTermF	GCTGGTTTAACCGATGCTGA
qTermR	GCAATGAAGTCCTGCTCGAAC
TermF	ATTTGGATACGAACAATATCACCG
TermR	CCTACGAAATAGACGGGAACAG

3. HIGH-THROUGHPUT DETECTION OF LONG-CHAIN BIOFUELS

3.1. Background and Motivation

Application of the proposed multiple-target genetic modification approach will require a high-throughput technique for characterizing biofuel production in the modified strains, as the number of possible combinatorial mutants will scale exponentially with the number of targets for genetic modification. Due to the higher energy density of long-chain biofuels, long-chain hydrocarbon metabolites, such as alkanes and free fatty acids, are ideal products for a cyanobacterial biofuel production system.[30] Unfortunately, the detection and quantification of these long-chain hydrocarbon species typically requires tedious sample preparation methods, such as lipid extraction,[31] along with long run times (minutes to hours) for sample analysis using conventional instrumentation, such as gas chromatography/mass spectrometry (GC/MS). As such traditional methods are not amenable to high-throughput screening, this study investigated alternative approaches for alkane and free fatty acid detection and quantification.

For high-throughput screening applications, spectroscopic methods of detection that do not require sample preparation are ideal. Two spectroscopic methods are capable of detecting hydrocarbons directly without significant sample preparation: Fourier transform infrared (FTIR) and Raman spectroscopies. FTIR spectra of hydrocarbons will produce strong bands in the 2800 – 3000 cm^{-1} region of the spectrum due to the C-H bonds along with weak bands at 4200 – 4400 cm^{-1} and 5500 – 6000 cm^{-1} regions.[32] Due to the overwhelming water absorption bands at 2500 – 4500 cm^{-1} and 4800 – 6200 cm^{-1}, hydrocarbon signatures are not readily detectable in aqueous solutions like cell cultures. To avoid interference from water, it is possible to spot a small volume of sample on an attenuated total reflectance (ATR) plate and allow the water to evaporate prior to detection. The time for samples spotting would slow down the throughput of this type of detection method, yet this would still be faster than traditional methods of lipid extraction and GC/MS. Raman spectroscopy is also capable of detecting hydrocarbons and does not suffer from water interference.[33] Preliminary efforts to detect pentadecane, a C15 alkane, using Raman microscopy were unsuccessful, however (data not shown).

In addition to spectroscopic detection methods, several lipophilic fluorescent dyes have been used to detect and quantify neutral lipids.[34-36] Two of the most commonly used dyes for neutral lipid detection are nile red and bodipy, and there has been some debate regarding which is the best dye for detection and quantification.[37, 38] Recently, several high-throughput screening methods for detecting bacterial hydrocarbons have been developed based on nile red staining and have been used to determine the relative quantitation of several hydrocarbon products (free fatty acids and ketones) in genetically modified strains of *E. coli*.[39, 40] For cyanobacteria, fluorescence-based high-throughput screening methods must be developed in spectra regions where there is minimal influence from the absorbance and fluorescence features of the native photosynthetic pigments. The excitation and emission spectra of nile red has been shown to be dependent on the neutral lipid moiety, with nile red stained phospholipids having excitation and emission peaks near 550 and 635 nm and nile red stained triglycerides having excitation and emission peaks near 505 and 585 nm. Meanwhile, various bodipy fluorophores have excitation peaks varying from 493 – 630 nm with corresponding emission peaks from 503 – 650 nm. As cyanobacterial phycobilisomes absorb in the range of 600 – 650 nm and chlorophyll-*a* absorbs from 400 – 470 nm and 660 – 700 nm, the 500 – 600 nm region of the visible spectrum is ideal for a cyanobacterial high-throughput screening method. Based on these spectral limitations, both nile red and bodipy based staining methods have potential for high-throughput screening of

hydrocarbon production in cyanobacteria. As these dyes are non-specific neutral lipid stains, both cell membranes and thylakoid membranes within the cyanobacteria will also contribute to the fluorescence emission signal; however, a relative quantitation of increased hydrocarbon content within the cell is feasible.

This project investigated FTIR spectroscopy, nile red staining, and bodipy staining as possible quantitative, high-throughput screening methods for the analysis of hydrocarbon production in genetically modified cyanobacteria.

3.2. Results

3.2.1. Investigation of FTIR Spectroscopy for High-Throughput Alkane Detection

A label-free, spectroscopic method of detecting long-chain alkanes in medium A+ was investigated as a non-invasive, high-throughput method of detection: FTIR spectroscopy. As water has a broad and strong absorption peak in the near-IR (3600 cm^{-1}), a SmartArk ATR accessory was used to spot samples and allow water to evaporate prior to detection by FTIR, as described in Section 2. When pentadecane was mixed with medium A+, spotted on the ATR plate, and allowed to evaporate, the components of the growth medium were found to overwhelm the alkane peaks in both the near (~ 14000 - 4000 cm^{-1}) and mid (4000 - 400 cm^{-1}) IR regions (Figure 2). When the alkane and medium A+ spot was subsequently rinsed with water and allowed to dry, the resulting spectrum included both near- and mid-IR peaks corresponding to the 69.8 µg of pentadecane.

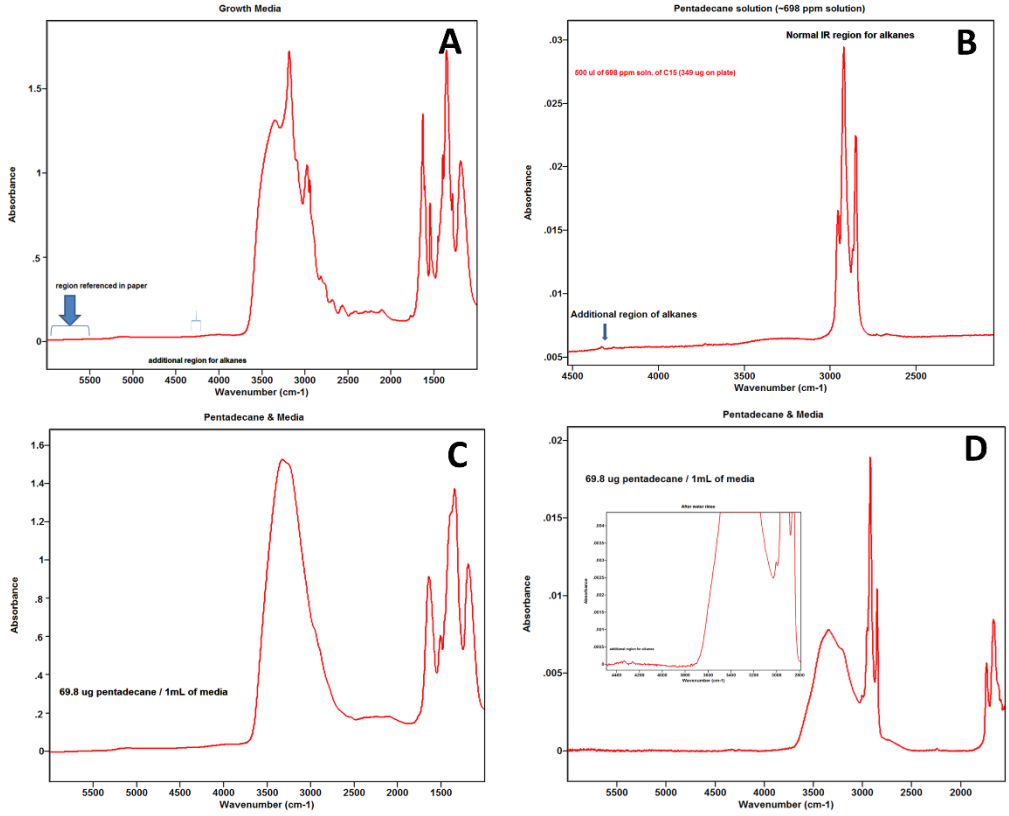

Figure 2. IR spectra of dried medium A+ (A), 698 ppm of pentadecane (B), medium A+ with 69.8 µg of pentadecane (C), and medium A+ with 69.8 µg of pentadecane after rinsing with DI water (D) with inset showing alkane signal in the 4300 – 4400 cm-1.

3.2.2. Nile Red Based Quantification of Alkanes and Free Fatty Acids

Nile red staining was investigated as a possible method for quantitative, high-throughput detection of both alkanes and free fatty acids. A model alkane (pentadecane) and model free fatty acid (palmitic acid) were used to optimize the nile red staining protocol and to evaluate the sensitivity and quantitative capability of this method. The concentration of nile red employed for pentadecane and palmitic acid staining was investigated for two backgrounds: medium A+ and a low concentration of *Synechococcus* sp. PCC 7002 (OD_{730} = 0.1). As shown in Figure 3 A and B, a nile red concentration of approximately 10 μg/mL was determined to lead to a high fluorescence signal for both background solutions and both biofuel targets. Next, the nile red incubation time was optimized for detection of 94 mM of pentadecane and 0.1 mM of palmitic acid. The optimal incubation time for pentadecane detection ranged from 5 to 10 minutes while the optimal incubation time for palmitic acid detection was approximately 5 minutes (Figure 3 C and D). There was high variability in the nile red fluorescence for pentadecane detection, and this is likely due to the low solubility of pentadecane in aqueous solution. Lastly, the sensitivity of detection with nile red was analyzed for both pentadecane and palmitic acid (Figure 3 E and F). Pentadecane was detected by nile red fluorescence at concentrations as low as 5 mM while palmitic acid was detected at a concentration of 25 μM. Due to the low solubility of pentadecane in aqueous solution, we were unable to test concentrations of pentadecane less than 5 mM; therefore, we were unable to compare the sensitivity of detection with nile red staining for these two potential biofuel products. Again, the fluorescence readings for pentadecane detection were much more variable and nonlinear compared to the readings for palmitic acid. Nile red-based detection of palmitic acid was highly linear (R^2 = 0.98 and 0.97 for nile red staining of palmitic acid in medium A+ and 7002 backgrounds, respectively), indicating that this detection method would be ideal for quantitative as well as qualitative measurement of free fatty acids.

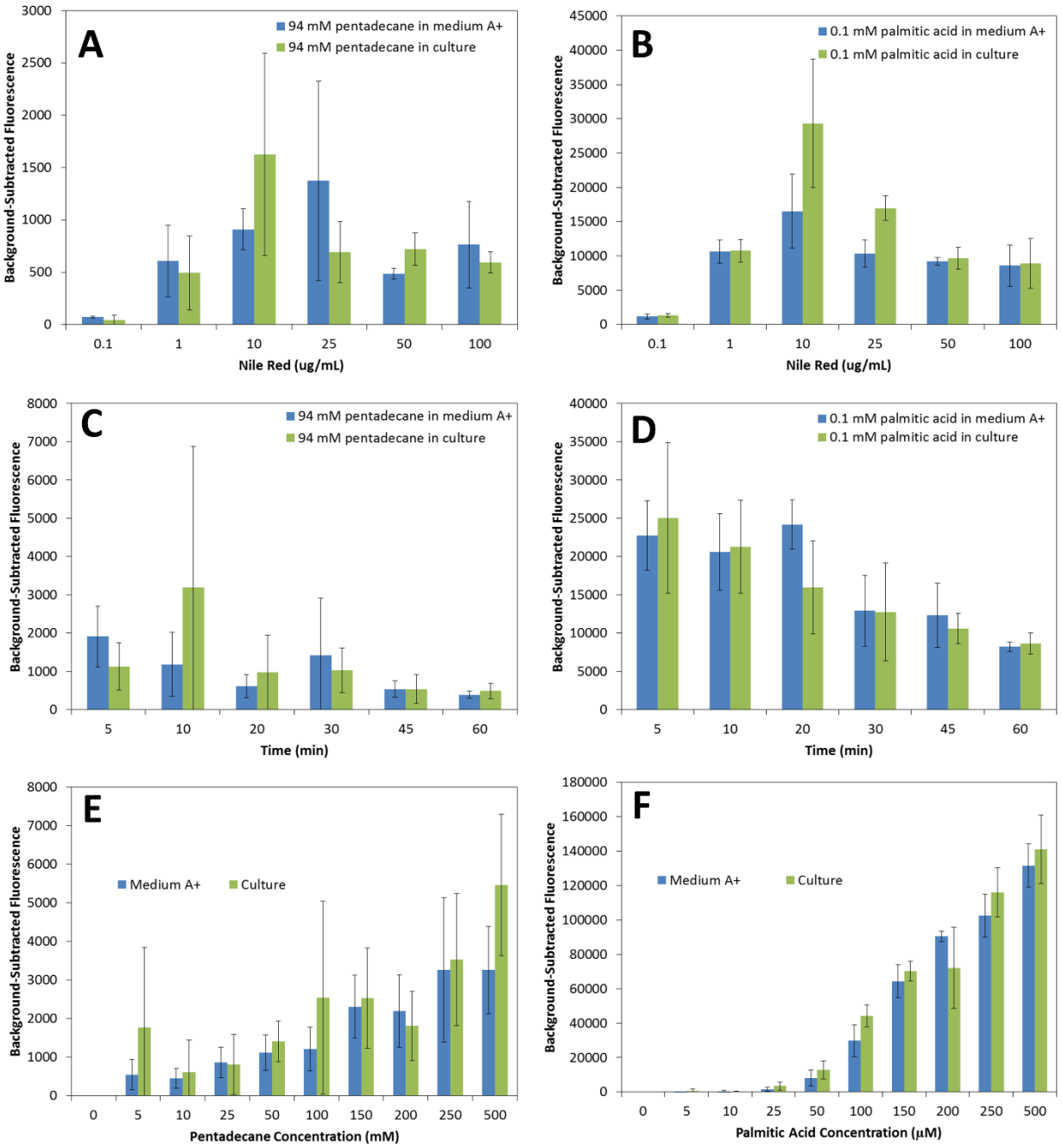

Figure 3. Optimization of nile red staining concentration (A,B) and incubation time (C,D) for medium A+ and low cell density (OD_{730} = 0.1) *Synechococcus* sp. PCC 7002 backgrounds spiked with 94 mM pentadecane (A,C) and 0.1 mM palmitic acid (B,D), and characterization of pentadecane (E) and palmitic acid (F) detection sensitivities for nile red staining. Measurements are the average of a minimum of three experimental replicates with error bars indicating the standard deviation.

3.2.3. Bodipy Based Quantification of Alkanes and Free Fatty Acids

Bodipy staining was also investigated for quantitative, high-throughput detection of alkanes and free fatty acids, again using pentadecane and palmitic acid for optimization with medium A+ and low cell density (OD_{730} = 0.1) *Synechococcus* sp. PCC 7002 culture as

backgrounds. The optimal bodipy concentration for staining and detection of 94 mM of pentadecane was approximately 500 nM while detection of 0.1 mM of palmitic acid was highest with 1 μM of bodipy stain (Figure 4 A and B). The optimal incubation time for bodipy staining ranged from 5 – 20 minutes for detection of 94 mM of pentadecane, and high fluorescence signal for bodipy staining of 0.1 mM of palmitic acid was observed for all tested incubation times: 5 – 60 minutes (Figure 4 C and D). Similar to nile red, bodipy staining was able to detect 5 mM of pentadecane, the lowest concentration of pentadecane that could be reliably mixed with aqueous solution, and 25 μM of palmitic acid (Figure 4 E and F). There was no observable change in bodipy fluorescence for 5 – 250 mM of pentadecane, and while there is some quantitative increase in bodipy fluorescence with increasing palmitic acid concentrations, the correlation was not as linear as that observed with nile red staining (R^2 = 0.93 and 0.91 for bodipy staining of palmitic acid in medium A+ and 7002 backgrounds, respectively)

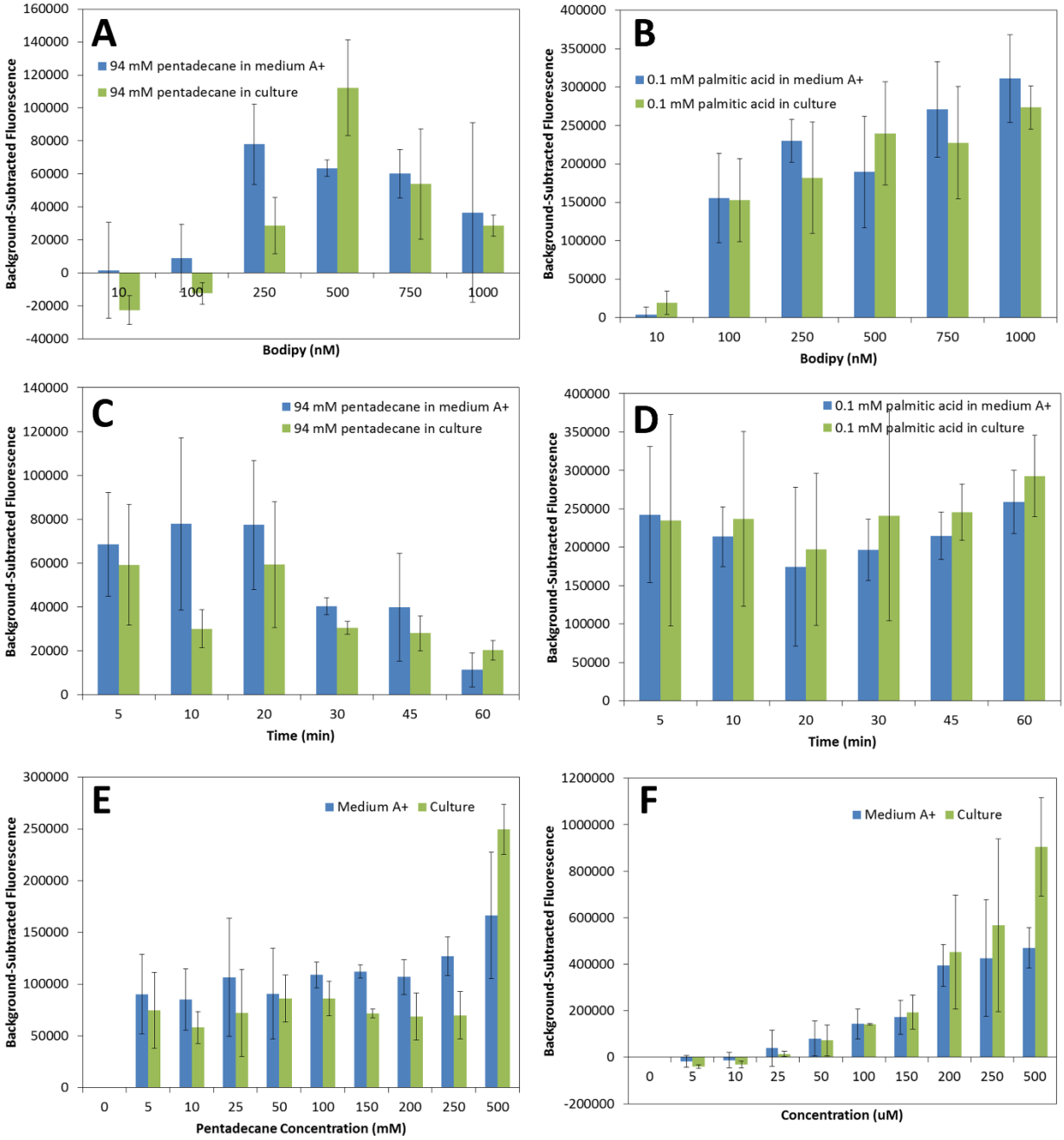

Figure 4. Optimization of bodipy staining concentration (A,B) and incubation time (C,D) for medium A+ and low cell density (OD_{730} = 0.1) *Synechococcus* sp. PCC 7002 backgrounds spiked with 94 mM pentadecane (A,C) and 0.1 mM palmitic acid (B,D), and characterization of pentadecane (E) and palmitic acid (F) detection sensitivities for bodipy staining. Measurements are the average of a minimum of three experimental replicates with error bars indicating the standard deviation.

3.3. Discussion

A high-throughput method of detecting long-chain hydrocarbon products, such as alkanes and free fatty acids, is a key step for screening biofuel-producing strains of cyanobacteria

constructed through either random or targeted strain development approaches. In this study, we investigated three approaches for high-throughput detection of both alkanes and free fatty acids. While we demonstrated that FTIR spectroscopy is capable of detecting alkanes in the cyanobacterial medium (medium A+), this detection method requires some sample preparation, including the spotting and drying of samples on an ATR plate to eliminate interference from water as well as an additional wash and evaporation step to remove interfering components from the culture medium. These sample preparation steps will hinder the rate of throughput for FTIR-based hydrocarbon detection, making this method less than ideal for high-throughput quantitation of hydrocarbon production in cyanobacterial mutants. Nile red and bodipy based detection of cyanobacterial hydrocarbons will be less chemical-specific compared to FTIR but has the potential for high throughput rates and integration with a fluorescence-activated cell sorter (FACS) for very high-throughput, single cell analysis and sorting. This study showed that both nile red and bodipy are capable of detecting alkanes and free fatty acids both in solution and in dilute cyanobacterial culture. Nile red staining proved to be more quantitative than bodipy staining for extracellular pentadecane and palmitic acid detection. Additionally, the correlation between fluorescence signal and hydrocarbon concentration was highly linear for nile red staining of palmitic acid. Meanwhile, the correlation for pentadecane was weak, suggesting that the application of neutral lipid dyes for alkane detection may not be useful for quantitative measurements. This preliminary investigation of high-throughput approaches for detecting hydrocarbon production in cyanobacterial mutants indicates that a nile red staining method will provide quantitative measurement of free fatty acids and semi-quantitative detection of alkanes and has the potential to be analyzed with FACS.

4. OPTIMIZATION OF *SYNECHOCOCCUS* SP. PCC 7002 TRANSFORMATION

4.1. Background and Motivation

In order to develop a high throughput method for targeting multiple genetic modifications based on homologous recombination, the transformation process for *Synechococcus* sp. PCC 7002 must be optimized for high efficiency. Several factors are important for transformation, including the identification of neutral integration sites in the genome, optimization of conditions for transformation, characterization of host exonuclease activity, and genome copy number (i.e., ploidy).

While *Synechococcus* sp. PCC 7002 has previously been modified with the insertion of foreign genes into its genome,[41] most efforts have utilized the desaturase B (*desB*) site for integration. DesB is important for membrane desaturation at low temperatures (22°C)[42] but is not produced under standard laboratory conditions (30-38°C). Ultimately, biofuel-producing *Synechococcus* sp. PCC 7002 will be grown under outdoor environmental conditions, which are likely to include temperatures as low as 22°C. Therefore, true neutral integration sites are needed for *Synechococcus* sp. PCC 7002. One putative neutral integration site, between two open reading frames for hypothetical proteins (SYNPCC7002_A0935 and SYNPCC7002_A0936), has been evaluated under standard laboratory conditions (37°C, 250 μmol photons m^{-2} s^{-1}) and found to be acceptable for limonene and bisabolene production.[8] However, additional neutral integration sites may be required for large or multiple foreign pathways.

Several experimental conditions have been shown to influence the transformation efficiency of genome integration via homologous recombination in Gram-negative bacteria. These include the amount of transforming DNA, whether the transforming DNA is circular or linear, the growth phase of the host cell, incubation time, temperature, agitation, and light (for photosynthetic bacteria). Initial transformation optimization determined that 0.5 μg/mL of linearized DNA with exponentially growing cells under continuous light, 30°C, agitation, and approximately 24 h of incubation are optimal for transformation of *Synechococcus* sp. PCC 7002.[43-45] Factors that remain to be explored for *Synechococcus* sp. PCC 7002 transformation include the ratio of DNA to cells, the length of homologous regions, and the influence of host exonucleases.

Lastly, recent studies in other cyanobacteria, namely *Synechococcus elongatus* PCC 7942 and *Synechocystis* sp. PCC 6803, have shown that these cyanobacterial strains possess multiple genome copies, up to 142 genome copies for exponentially growing *Synechocystis* sp. PCC 6803.[29] While *Synechococcus* sp. PCC 7002 is a model cyanobacterium, the genome copy number for this organism has yet to be reported, and the relationship between growth rate and genome copy number remains largely unexplored for cyanobacteria. Previous studies of genome copy number in *Escherichia coli* indicate that environmental conditions that affect growth may also affect genome copy number.[46] Thus, it may be possible to reduce genome copy number in *Synechococcus* sp. PCC 7002 and other cyanobacteria by utilizing non-ideal growth conditions. A low genome copy number is important for the proposed multiple-target genetic modification technique, as integration of the same construct into multiple genome copies would lead to false fluorescence signal and the isolation of homologous modified genomes would be challenging.

This chapter addresses the aforementioned factors involved in transformation of *Synechococcus* sp. PCC 7002: the identification of neutral integration sites, optimization of cell concentration, optimization of the length of homologous regions, detection of host exonuclease

activity, and measurement of genome copy number under various growth-limiting environmental conditions.

4.2. Results

4.2.1. Identification and Characterization of Neutral Integration Sites for Synechococcus sp. PCC 7002

To identify putative neutral integration sites within the genome of *Synechococcus* sp. PCC 7002, the size of intergenic, noncoding regions of the genome were analyzed. Only three intergenic, noncoding regions were found to be greater than 1 kb in size: the region between SYNPCC7002_A0932 and SYNPCC7002_A0933 (1,027 bp), the region between SYNPCC7002_A1202 and SYNPCC7002_A1203 (1,040 bp), and SYNPCC7002_A1778 and SYNPCC7002_A1779 (1,401 bp). These regions were designated neutral site 1 (NS1), neutral site 2 (NS2), and neutral site 3 (NS3), respectively. Antibiotic resistance genes were inserted into putative NS1 and NS2 to form 7002ΔNS1 and 7002ΔNS2. These neutral site mutants were shown to have no significant changes in growth or photosynthetic yield compared to the wild type.[24]

4.2.2. Optimization of Cell Concentration for Transformation of Synechococcus sp. PCC 7002

Previous studies have determined that exponentially growing cells of *Synechococcus* sp. PCC 7002 yield higher transformation efficiencies than stationary phase cells.[44] However, the ratio of cells to transforming DNA can play an important role in determining transformation efficiency. In this study, we investigated the transformation efficiency of various concentrations of *Synechococcus* sp. PCC 7002 culture with 0.5 µg/mL of linearized plasmid DNA (pSB, SpeI digested). pSB includes approximately 1 kb regions homologous to the sequences upstream and downstream of the fatty acid desaturase, *desB*, with a kanamycin resistance cassette for selection. The transformation efficiency was determined based on the number of transformants (i.e. kanamycin resistant colonies) per µg of transforming DNA. For the conditions used in this study (0.5 µg linearized pSB, 30°C, 150 rpm, 60 µmol photons m^{-2} s^{-1}), a concentration of OD_{730} = 1.0 was found to lead to the highest transformation efficiencies (Figure 5). All concentrations were determined from exponentially growing cultures of *Synechococcus* sp. PCC 7002 (OD_{730} = 0.7 – 1.0) by centrifugation at 3,900 x g for 10 min and resuspension in fresh medium A+.

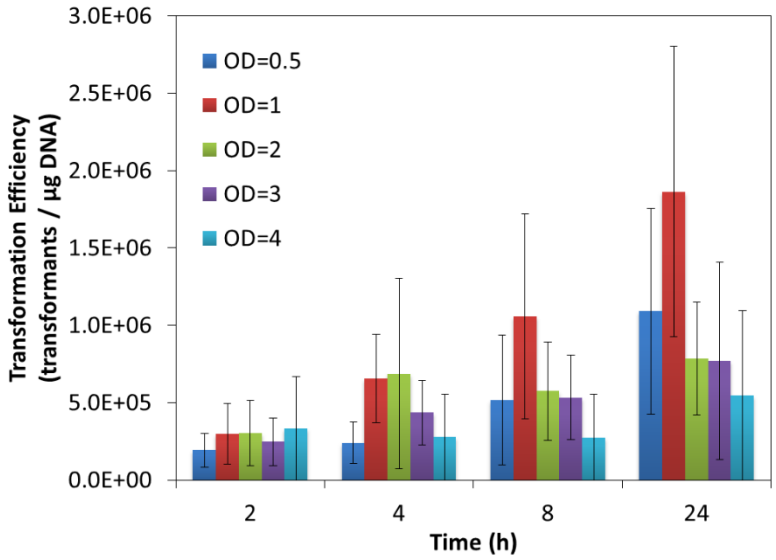

Figure 5. Transformation efficiencies of *Synechococcus* sp. PCC 7002 cultures of various concentrations based on the optical density (OD) at 730 nm with 0.5 µg of linearized pSB. The transformation efficiencies were measured at a range of time points after addition of linearized pSB. Transformation conditions = 30°C, 150 rpm, 60 µmol photons m^{-2} s^{-1}. Data are averages of three biological replicates with error bars representing the standard deviation.

4.2.3. Optimization of Length of Homologous Regions for Integration into Synechococcus sp. PCC 7002

A DNA integration cassette includes three basic components: the target gene, a selection marker (i.e. antibiotic resistance cassette), and flanking regions that are homologous to the integration region of the host genome. The length of the homologous regions for integration is often highly correlated to transformation efficiency.[47] For synthetic biology and high-throughput applications, it is often desirable to synthesize the entire DNA integration cassette; therefore, cost of this DNA synthesis can be reduced by minimizing the length of the homologous regions in the integration cassette. The relationship between transformation efficiency and the length of the homologous regions was investigated by modifying the homologous regions of pSB to include various lengths: 250 bp, 500 bp, 750 bp, 1000 bp, and 1250 bp. As expected, the transformation efficiencies correlated with the length of homologous regions, with high transformation efficiencies resulting from 1250 bp homologous regions.[44]

4.2.4. Exonuclease Activity in Synechococcus sp. PCC 7002

Host exonuclease activity is another factor affecting transformation efficiency. To determine if *Synechococcus* sp. PCC 7002 has significant exonuclease activity, the transformation efficiencies of two transforming DNA fragments were compared: (1) a PCR amplified *desB* integration cassette, consisting of 1kb flanking regions of homology and a kanamycin resistance cassette and (2) linearized pSB, consisting of the 1kb flanking regions of homology, a kanamycin resistance cassette, and protection nucleotides of 817 and 115 bp on each end of the linearized fragment. As the only difference between these two transforming DNA fragments is the addition of protection nucleotides for linearized pSB, a difference in the transformation efficiencies of these DNA fragments is evidence of exonuclease activity. As

shown in our manuscript,[24] more than a 30-fold increase in transformation efficiency of *Synechococcus* sp. PCC 7002 was observed with linearized pSB compared to the PCR *desB* integration cassette.

4.2.5. Genome Copy Number of Synechococcus sp. PCC 7002

In order to develop a method for multiple target integration in *Synechococcus* sp. PCC 7002, a low genome copy number is desirable to prevent increased signal from the same target integrating into multiple genome copies and also to ensure uniform genome sequence within a single cell. In this project, we studied the genome copy number of *Synechococcus* sp. PCC 7002 during the course of its normal growth cycle and also under non-ideal conditions (low light (LL) 15 μmol photons m^{-2} s^{-1} and low temperature (25°C and 22°C)) to determine if external conditions could be applied to control the level of ploidy. A qPCR-based method for quantifying genome copy number was employed along with flow cytometer or hemocytometer cell counts to calculate the genome copy number per cell.[28] Target regions near the origin of replication and the terminus were amplified to determine whether full or incomplete genome copies were present. From these experiments, the average genome copy number in *Synechococcus* sp. PCC 7002 varied from 1 to 22 under standard growth conditions (34°C, moderate light (ML)). As shown in Figure 6, several trends are noticeable: First, there are lower genome copy numbers during the early exponential growth phase and higher genome copy numbers during the stationary phase. Second, there is little change in genome copy number with LL or a temperature of 25°C despite the much reduced growth rate under these non-ideal conditions. Lastly, under very low temperature conditions (22°C), there is a high number of origin copies but low number of terminus copies, indicating incomplete genomes are present within *Synechococcus* sp. PCC 7002 under these conditions.

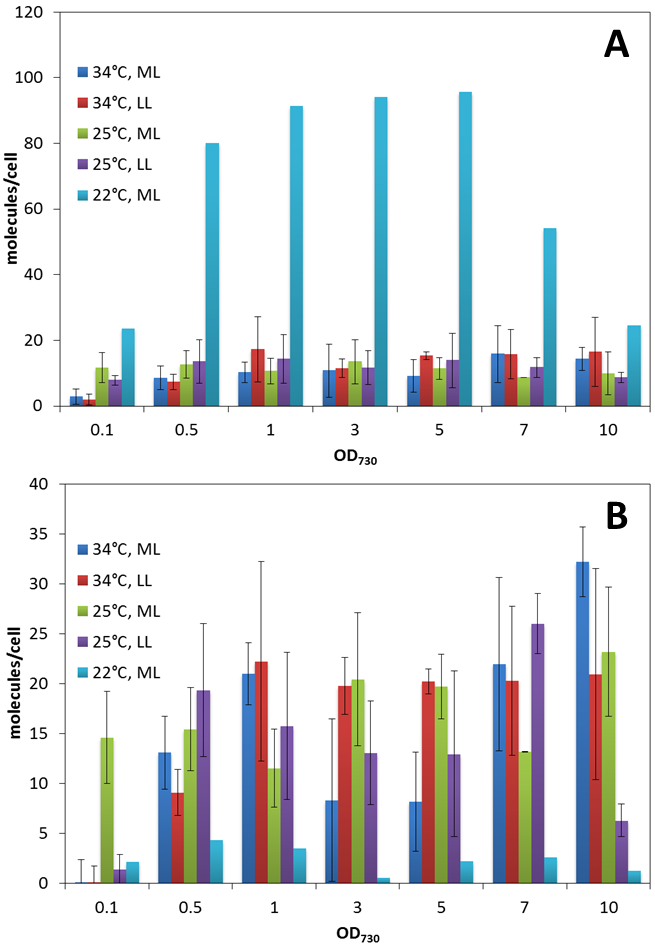

Figure 6. Genome copy number for origin (A) and terminus (B) of *Synechococcus* sp. PCC 7002 genome under various conditions: moderate light (ML) = 60 µmol photons m^{-2} s^{-1}, low light (LL) = 15 µmol photons m^{-2} s^{-1}, and temperatures of 34°C, 25°C, and 22°C. Data are averages of two biological replicates with error bars representing the standard deviation with the exception of the 22°C, ML condition of which there is only one biological replicate.

4.3. Discussion

This investigation into transformation factors for *Synechococcus* sp. PCC 7002 led to several important discoveries. First, the identification of putative neutral sites for genome integration will allow for the integration of non-native genes and pathways to utilize *Synechococcus* sp. PCC 7002 as a chassis for the production of biofuels and other chemical products. These neutral site mutants should be fully characterized under specific process conditions prior to the use of these sites for strain development, as the mutants have only been tested under standard laboratory conditions in this study. These laboratory conditions are likely different from actual process conditions which will utilize natural sunlight that is subject to diurnal variation, higher light intensities, and a broader spectral wavelength compared to fluorescent lights. Second, analysis of cell concentration during transformation revealed that cell concentration has some influence on the transformation efficiency of *Synechococcus* sp. PCC 7002, with OD_{730} = 1.0 yielding the highest transformation efficiencies for 0.5 µg of transforming DNA. However, these results also suggest that cell concentration is not the most

influential factor, as transformation efficiencies across a wide range of cell concentrations (OD_{730} = 0.5 – 4.0) are within experimental error (Figure 5). Third, the length of homologous regions required for successful genome integration in *Synechococcus* sp. PCC 7002 were determined. If minimizing the length of the homologous regions for reduced cost of DNA synthesis is a greater priority than high transformation efficiency, successful integration can be achieved with as little as 250 bp regions of homology. For high transformation efficiencies, however, homologous regions greater than 750 bp are best. The minimum length of homologous regions for integration into the *Synechococcus* sp. PCC 7002 genome is much higher than that reported for *E. coli*, which can use homologous regions as low as 23 bp.[47] It may be possible to achieve successful integration with shorter homologous regions by deleting the exonuclease in *Synechococcus* sp. PCC 7002 responsible for degrading the transformed DNA. Activity of an exonuclease was confirmed in this study; however, the gene responsible for this exonuclease activity must still be identified. An exonuclease deficient *Synechococcus* sp. PCC 7002 would be ideal for future strain development efforts but may also lead to higher mutation rates if the exonuclease deficient strain is exposed to exogenous DNA from contaminants. Lastly, high genome copy number in *Synechococcus* sp. PCC 7002 was confirmed, similar to that reported in other cyanobacterial strains.[29] While low temperature and low light conditions yielded significant reductions in the growth rate of *Synechococcus* sp. PCC 7002, the genome copy number was not significantly changed under these conditions except for a temperature of 22°C. At 22°C, a very high copy number of the origin of replication was detected along with a low copy number of the terminus region. Based on this result, it may be possible to target genes to the terminus region of *Synechococcus* sp. PCC 7002 at 22°C to avoid complications associated with polyploidy, but the transformation efficiency of *Synechococcus* sp. PCC 7002 at 22°C remains to be determined.

5. GENETIC TOOL DEVELOPMENT FOR *SYNECHOCOCCUS* SP. PCC 7002

5.1. Background and Motivation

In order to detect successful integration of the transformation targets in *Synechococcus* sp. PCC 7002, a quantitative reporter will be integrated into the transformation cassette for each target. Fluorescent proteins are ideal reporter candidates, as they have proven to be quantitative and do not require the addition of exogenous substrates or other metabolic cofactors.[48] The use of fluorescent proteins in cyanobacteria presents a unique challenge, as cyanobacterial photosynthetic pigments may absorb the excitation or emission irradiance of the fluorescent protein and the fluorescence of these natural pigments may interfere with the fluorescence of the target fluorescent protein. Despite this challenge, the use of a fluorescent protein for detection of genome integration is a promising strategy to pursue.

In addition to having a quantitative reporter for detection of genome integration, the transformation cassette will also require a characterized promoter for replacement of native promoters in the genome of *Synechococcus* sp. PCC 7002. Only a handful of promoters have been used for genetic modification of *Synechococcus* sp. PCC 7002.[8, 25, 41, 49-52] Most of these promoters are either well-characterized *Escherichia coli* promoters,[25, 50] other cyanobacterial promoters,[8, 25, 41] or native promoters of *Synechococcus* sp. PCC 7002 that are used to drive photosynthetic genes.[51, 52] A recent study also generated a promoter mutant library to produce a range of expression levels in *Synechococcus* sp. PCC 7002.[16] While these previous works have enabled the successful modification of *Synechococcus* sp. PCC 7002 for the production of biofuels and other products, they are not sufficiently characterized for the proposed multiple-target promoter optimization. First, *E. coli* promoters have been shown to yield different expression levels and regulatory characteristics in cyanobacteria.[14] Second, the few cyanobacterial promoters that have been utilized for recombinant gene expression are promoters related to photosynthesis (i.e., cpcB and psbAI);[8, 25, 49, 51, 52] thus, these promoters have high expression and are strongly correlated to environmental light conditions. The proposed promoter optimization will require characterized promoters with a range of expression levels and regulatory patterns. Third, most of these previously verified promoters were only characterized under low-level continuous light and constant temperature conditions. A biofuel-producing cyanobacterial strain will likely be cultivated under outdoor environmental conditions, which will include diurnal light cycling as well as temperature fluctuations. Lastly, the promoter modifications in our optimized strain will need to interface with the natural metabolism and regulatory mechanisms of *Synechococcus* sp. PCC 7002. Based on this requirement, natural promoters in *Synechococcus* sp. PCC 7002 are the ideal candidates for characterized promoters.

In this section, we describe the selection, expression, and characterization of fluorescent protein reporters in *Synechococcus* sp. PCC 7002 as well as native promoters for gene expression in *Synechococcus* sp. PCC 7002.

5.2. Results

5.2.1. Expression and Characterization of Fluorescent Protein Reporters in Synechococcus sp. PCC 7002

Candidate fluorescent protein reporters were selected based on minimal spectral interference with the natural absorbance and fluorescence spectra of *Synechococcus* sp. PCC

7002 (Figure 7). Few spectral features are naturally present in *Synechococcus* sp. PCC 7002 in the 480 – 600 nm region, making this region ideal for a fluorescent protein reporter. Three reporters with excitation and emission maxima in this region were selected for expression in *Synechococcus* sp. PCC 7002: green (hybrid GFP - hGFP),[24] yellow (Ypet),[53] and orange (mOrange)[54] fluorescent proteins. These fluorescent proteins were codon optimized for expression in *Synechococcus* sp. PCC 7002, and placed downstream of the *rbcL* promoter cloned from *Synechococcus* sp. PCC 7002. The excitation and emission peaks for each fluorescent protein expressed in *Synechococcus* sp. PCC 7002 were determined experimentally (Figure 8). These excitation and emission spectra match those previously reported for these proteins.[53-55] Interestingly, the excitation spectrum for hGFP expressed in *Synechococcus* sp. PCC 7002 showed some signal at 520 nm with excitation wavelengths near 400 nm; this signal may be due to chlorophyll or carotenoid fluorescence in the blue region with near-UV excitation[56, 57] or fluorescence from hGFP resulting from excitation by chlorophyll or carotenoid fluorescence emission. The fluorescence emission from the fluorescent protein reporters had little overlap with the native pigment fluorescence or absorbance in *Synechococcus* sp. PCC 7002, with the exception of hGFP which had a slight overlap with the carotenoid absorbance shoulder.[24] We also demonstrated that these fluorescent proteins can be used for fluorescence microscopy and flow cytometry interrogation of *Synechococcus* sp. PCC 7002 modified with these proteins.[24]

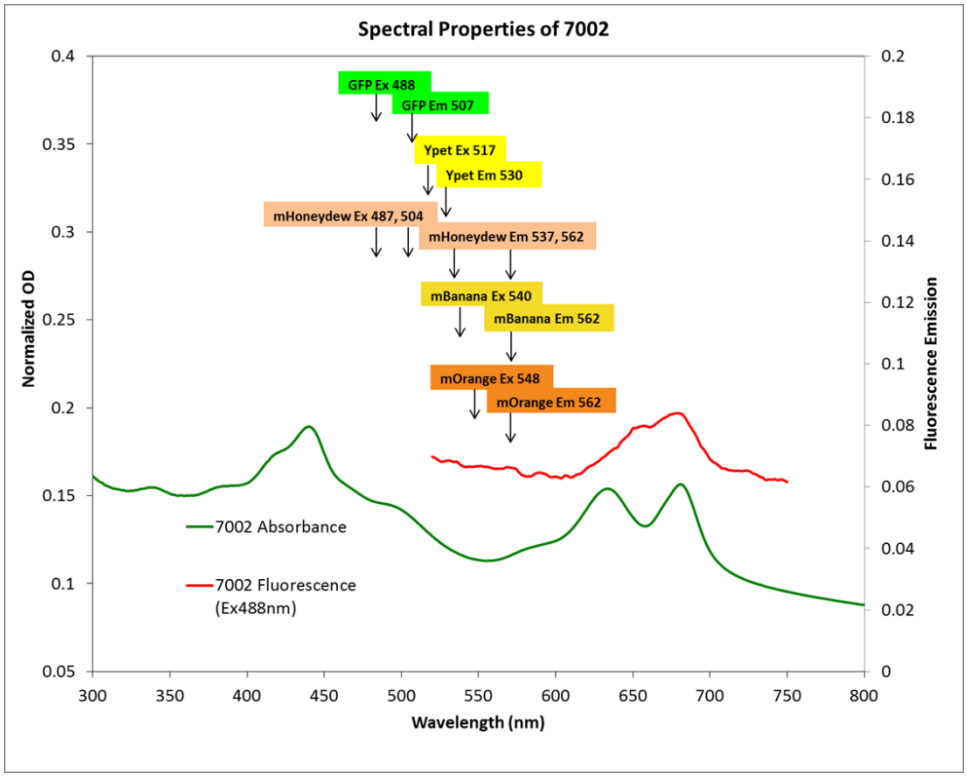

Figure 7. Spectral properties of *Synechococcus* sp. PCC 7002 with the excitation and emission maxima of candidate fluorescent proteins. Green trace = absorbance spectra of wild type *Synechococcus* sp. PCC 7002. Red trace = fluorescence emission spectra of wild type *Synechococcus* sp. PCC 7002 with excitation at 488 nm.

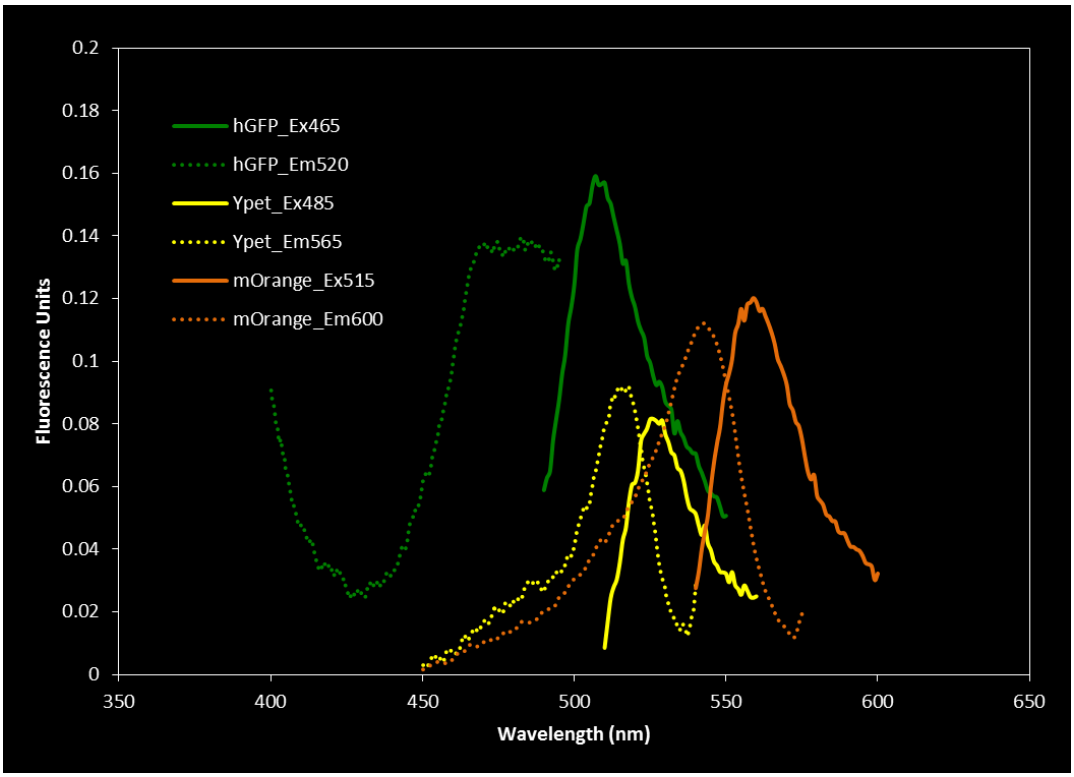

Figure 8. Fluorescence excitation (dashed lines) and emission (solid lines) spectra for fluorescent proteins expressed in *Synechococcus* sp. PCC 7002. For the excitation spectra, emission wavelengths of 520 nm, 565 nm, and 600 nm were used for hGFP, Ypet, and mOrange, respectively. For the emission spectra, excitation wavelengths of 465 nm, 485 nm, and 515 nm were used for hGFP, Ypet, and mOrange, respectively.

5.2.2. Characterization of Native Promoters as Tools for Rational Strain Development

To rationally design gene expression levels throughout the genome of *Synechococcus* sp. PCC 7002, characterized promoters that can readily integrate with native regulatory mechanisms are needed. Towards this goal, we selected native promoters with varying expression levels (10^{-5} – 10^{-2} counts per total read count) and regulatory patterns (constitutive, linear phase, and stationary phase) based on RNA-seq data for *Synechococcus* sp. PCC 7002 available in the literature (Figure 9).[21] A total of 24 promoters were selected: 8 promoters for each regulatory pattern with 2 promoters at each expression level. For each candidate promoter, 500 bp of sequence upstream of the target gene was cloned upstream of the Ypet reporter. Each promoter-Ypet construct was integrated into the *Synechococcus* sp. PCC 7002 genome at neutral integration site 2 (NS2), and Ypet fluorescence was monitored in the modified strains over the course of a 3 week growth period under both continuous light and diurnal conditions. In general, the level of expression observed with the promoter-Ypet mutants correlated with the level of expression reported by RNA-seq.[21, 24] However, promoters A2579 and A2520 had low expression levels from RNA-seq (10^{-3} and 10^{-5} counts per total count), but very high Ypet expression (10 to 60-fold higher fluorescence compared to wild type).[24] Additionally, there was very little agreement between regulatory patters of the RNA-seq and Ypet expression data. While cyanobacterial genes have been shown to oscillate with diurnal light conditions due to circadian rhythm,[58] Ypet expression under diurnal light conditions only varied for a few select promoters compared to Ypet expression under continuous light.[24] These characterized promoters, with

varying expression level and regulation, may be used to optimize gene expression in *Synechococcus* sp. PCC 7002.

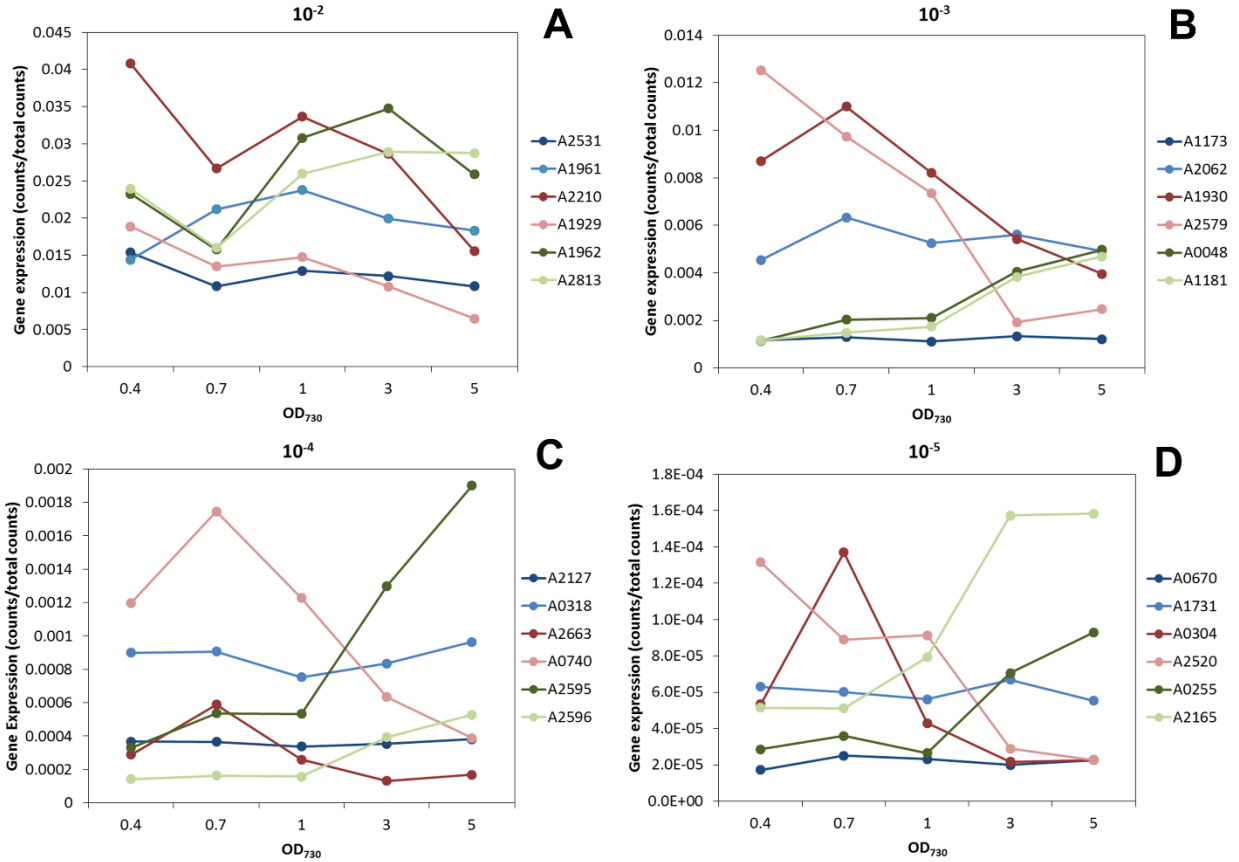

Figure 9. Gene expression levels in *Synechococcus* sp. PCC 7002 during growth measured by a previous RNA-seq experiment.[21] (A) High expression (10^{-2} counts/total counts); (B) moderate expression (10^{-3} counts/total counts); (C) low expression (10^{-4} counts/total counts); and (D) weak expression (10^{-5} counts/total counts) with constitutive expression in blue, linear phase expression in red, and stationary phase expression in green.

5.3. Discussion

The genetic tools developed in this study will advance the development of *Synechococcus* sp. PCC 7002 as a chassis organism for biofuel and chemical syntheses. The fluorescent protein reporters will enable facile analysis of gene expression in this cyanobacterial host, along with fluorescence microscopy to identify subcellular localization of proteins fused with fluorescent proteins. One limitation of the fluorescent proteins in this study is that they may persist in the cell long after mRNA transcript levels have been reduced due to the natural half-life of fluorescent proteins.[59] To overcome this limitation, future studies for gene expression analysis may include protein degradation tags, such as those previously described,[15] to obtain more accurate correlation between fluorescent protein and mRNA transcript levels. In addition to these traditional applications of fluorescent proteins, the fluorescent reporters may be applied for the detection of genome integration of DNA cassettes for promoter replacement, as proposed in this study. After transformation and integration of the DNA cassettes containing the fluorescent

protein, the transformants may be sorted by fluorescence activated cell sorting (FACS) to separate cells with varying numbers of targeted promoter replacements. In addition to the fluorescent reporters characterized in this study, the characterized promoters from *Synechococcus* sp. PCC 7002 may be used as tools for designing synthetic pathways and optimizing expression in this host. An unexpected result of this study was the differences detected between Ypet promoter expression and previous RNA-seq results. These differences may result from variability between the two experiments, including genetic drift, different environmental conditions, and limitations in promoter sequence. The *Synechococcus* sp. PCC 7002 strain employed in the RNA-seq study[21] was a strain that has been maintained in Donald Bryant's lab for approximately 30 years, while the strain of *Synechococcus* sp. PCC 7002 used in this study was purchased from the American Type Culture Collection (ATCC), which was deposited in either the 1970s or 1980s. As previous studies have shown,[60] there can be considerable genetic drift in laboratory strains maintained over long periods of time compared to those stored under cryogenic or lyophilized conditions. Additionally, there are slight differences in the environmental conditions employed in the two studies: the RNA-seq study utilized a temperature of 38°C and illumination at 250 μmol photons m^{-2} s^{-1} while this study used a temperature of 34°C (determined to be the optimal growth temperature for this strain)[25] and alternating cool white and plant fluorescent lights at 40 – 60 μmol photons m^{-2} s^{-1}. These environmental conditions may also impact the regulation of gene expression in *Synechococcus* sp. PCC 7002. DNA sequences upstream of 500 bp and downstream of the start codon have been shown to participate in the regulation of gene expression in prokaryotes.[61] Therefore, by limiting the promoter region to 500 bp upstream of the target gene, some key regulatory elements may be missing in the promoter-Ypet constructs compared to the native gene expression measured by RNA-seq, or DNA supercoiling and secondary DNA structure may be absent in the promoter-Ypet constructs. Translational differences resulting from different ribosome binding sites in the promoter region may also contribute to differences between the RNA-seq results and Ypet expression. Lastly, antisense RNA targeting the native gene transcripts may be responsible for the lower transcript levels in the RNA-seq experiment compared to the Ypet expression study.[62] While the exact reason for this discrepancy remains to be determined, the apparent inconsistency between RNA-seq and Ypet expression results highlights the importance of testing gene expression under conditions that are identical to the final process conditions.

6. CONCLUSIONS AND FUTURE DIRECTIONS

There are several notable scientific contributions resulting from this project. First, a high throughput method for the detection of alkanes and FFAs using the neutral lipid dye, nile red, was developed. Second, several tools and methods were developed for the advancement of *Synechococcus* sp. PCC 7002 as a cyanobacterial chassis for biofuel and chemical syntheses. Several putative neutral integration sites were identified in the genome of *Synechococcus* sp. PCC 7002 for the integration of foreign genes and pathways, and no notable effect on cell growth or photosynthetic efficiency was observed under standard laboratory conditions for the neutral site mutants. Both the cell concentration and the length of homologous regions in the integration cassette were optimized for efficient transformation in *Synechococcus* sp. PCC 7002, and exonucleases were found to be active in this host. Future efforts to identify the active exonucleases may lead to a more efficient *Synechococcus* sp. PCC 7002 host for strain development via genetic engineering. Lastly, three fluorescent protein reporters and twenty-four promoters were developed and characterized as tools for the interrogation and modification of *Synechococcus* sp. PCC 7002.

Despite these advancements, several issues remain to be overcome in order to realize the proposed multiple target promoter replacement method. The polyploidy nature of *Synechococcus* sp. PCC 7002 is challenging for several reasons. First, a single target may integrate into multiple copies of the genome, leading to the production of a higher fluorescence signal. This may result in false positive identification of multiple integration events. Second, and perhaps more importantly, there is no selective pressure for the isolation of identical genome copies. Traditional genetic modification efforts in cyanobacteria involve antibiotic selection, which enables the isolation of identical genome copies containing the antibiotic resistance cassette after several rounds of cultivation.[25] With a fluorescent protein reporter as the selection marker for genome integration, isolation of identical genome copies containing the fluorescent reporter cassette would require multiple rounds of FACS. This additional requirement for isolation of identical genome copies is a significant drawback, yet the method may still be feasible with repeated rounds of transformation and FACS. Another challenge of the proposed multiple target promoter replacement is reduced transformation efficiency associated with multiple copies of the fluorescent protein reporter gene. With each additional genome integration event, another copy of the fluorescent protein reporter would be present and would interact with other DNA integration cassettes containing the same fluorescent protein reporter to effectively reduce the transformation efficiency. Possible solutions to this challenge include the use of different fluorescent protein reporters (hGFP, Ypet, and mOrange) and the use of different codon preferences for copies of the same fluorescent protein. As hGFP, Ypet, and mOrange are all derived from the green fluorescent protein from *Aequorea victoria*, there are only slight changes in the DNA sequence of these genes, and therefore, homologous binding is still likely to occur. By changing codon usage, however, significant variation in the DNA sequence of the fluorescent protein genes is possible. Lastly, the addition of a recombinase such as Cre or FLP[63, 64] and loxP or FRT recombination sites flanking the fluorescent protein will allow for the removal of the fluorescent protein reporter after integration, allowing for additional rounds of mutagenesis. If these challenges can be overcome, the proposed multiple target promoter replacement technique is likely to succeed.

The development of a metabolic network for in silico analysis of possible promoter replacements with the characterized promoters developed in this project will also be a key focus

area for future research. As the genome sequence is available for *Synechococcus* sp. PCC 7002 as well as RNA-seq data,[21, 23] a metabolic network for promoter replacement may be developed from this preliminary information. Additional refinement of the network model will be based on experimental data of targeted promoter replacements with characterized promoters. The development of this *Synechococcus* sp. PCC 7002 metabolic network will complement the multiple target promoter replacement and help to more efficiently optimize *Synechococcus* sp. PCC 7002 as a biofuel and chemical production host.

7. REFERENCES

1. Deng, M.-D. and J.R. Coleman, *Ethanol Synthesis by Genetic Engineering in Cyanobacteria.* Applied and Environmental Microbiology, 1999. 65(2): p. 523-528.

2. Gao, Z., et al., *Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria.* Energy & Environmental Science, 2012. 5(12): p. 9857-9865.

3. Kusakabe, T., et al., *Engineering a synthetic pathway in cyanobacteria for isopropanol production directly from carbon dioxide and light.* Metabolic Engineering, 2013. 20(0): p. 101-108.

4. Atsumi, S., W. Higashide, and J.C. Liao, *Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde.* Nat Biotech, 2009. 27(12): p. 1177-1180.

5. Lan, E.I. and J.C. Liao, *Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide.* Metabolic engineering, 2011. 13(4): p. 353-363.

6. Ungerer, J., et al., *Sustained photosynthetic conversion of CO2 to ethylene in recombinant cyanobacterium Synechocystis 6803.* Energy & Environmental Science, 2012. 5(10): p. 8998-9006.

7. Lindberg, P., S. Park, and A. Melis, *Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism.* Metabolic Engineering, 2010. 12(1): p. 70-79.

8. Davies, F.K., et al., *Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of Synechococcus sp. PCC 7002.* Frontiers in Bioengineering and Biotechnology, 2014. 2: p. 21.

9. Liu, X., J. Sheng, and R. Curtiss III, *Fatty acid production in genetically modified cyanobacteria.* Proceedings of the National Academy of Sciences, 2011. 108(17): p. 6899-6904.

10. Ruffing, A.M. and H.D.T. Jones, *Physiological effects of free fatty acid production in genetically engineered Synechococcus elongatus PCC 7942.* Biotechnology and Bioengineering, 2012. 109(9): p. 2190-2199.

11. Tan, X., et al., *Photosynthesis driven conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria.* Metabolic Engineering, 2011. 13(2): p. 169-176.

12. Koksharova, O. and C. Wolk, *Genetic tools for cyanobacteria.* Applied Microbiology and Biotechnology, 2002. 58(2): p. 123-137.

13. Porter, R.D., *Transformation in Cyanobacteria.* Critical Reviews in Microbiology, 1986. 13(2): p. 111-132.

14. Camsund, D., T. Heidorn, and P. Lindblad, *Design and analysis of LacI-repressed promoters and DNA-looping in a cyanobacterium.* Journal of Biological Engineering, 2014. 8(4).

15. Huang, H.-H., et al., *Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology.* Nucleic Acids Research, 2010.

16. Markley, A.L., et al., *Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002.* ACS Synthetic Biology, 2014.

17. Jeong, J., et al., *Genome-scale genetic engineering in Escherichia coli.* Biotechnology Advances, 2013. 31(6): p. 804-810.

18. Wang, H.H., et al., *Programming cells by multiplex genome engineering and accelerated evolution.* Nature, 2009. 460(7257): p. 894-898.

19. Wang, H.H., et al., *Genome-scale promoter engineering by coselection MAGE.* Nat Meth, 2012. 9(6): p. 591-593.

20. Bryant, D.A. *Synechococcus sp. PCC 7002: a Robust, Versatile, and Cosmopolitan Cyanobacterial Platform for Biofuels Production*. Available from: http://bioenergy.asu.edu/cb&p/abstracts/Bryant-abstract.html.

21. Ludwig, M. and D.A. Bryant, *Acclimation of the global transcriptome of the cyanobacterium Synechococcus sp. strain PCC 7002 to nutrient limitations and different nitrogen sources.* Frontiers in Microbiology, 2012. 3.

22. Ruffing, A.M. and C.A. Trahan, *Biofuel toxicity and mechanisms of biofuel tolerance in three model cyanobacteria.* Algal Research, 2014. 5: p. 121-132.

23. Ludwig, M. and D.A. Bryant, *Synechococcus sp. Strain PCC 7002 Transcriptome: Acclimation to Temperature, Salinity, Oxidative Stress, and Mixotrophic Growth Conditions.* Frontiers in Microbiology, 2012. 3: p. 354.

24. Ruffing, A.M., L.M. Strickland, and T.J. Jensen, *Genetic Tools and Methods for Advancement of Synechococcus sp. PCC 7002 as a Cyanobacterial Chassis.* Microbial Cell Factories, 2015. TBD.

25. Ruffing, A.M., *Improved Free Fatty Acid Production in Cyanobacteria with Synechococcus sp. PCC 7002 as Host.* Frontiers in Bioengineering and Biotechnology, 2014. 2.

26. Rastoll, M.J., et al., *The development of a cryopreservation method suitable for a large cyanobacteria collection.* Journal of Applied Phycology, 2013. 25(5): p. 1483-1493.

27. Rumin, J., et al., *The use of fluorescent Nile red and BODIPY for lipid measurement in microalgae.* Biotechnology for Biofuels, 2015. 8: p. 42.

28. Breuert, S., et al., *Regulated Polyploidy in Halophilic Archaea.* PLoS ONE, 2006. 1(1): p. e92.

29. Griese, M., C. Lange, and J. Soppa, *Ploidy in cyanobacteria.* FEMS Microbiology Letters, 2011. 323(2): p. 124-131.

30. Lu, X., *A perspective: Photosynthetic production of fatty acid-based biofuels in genetically engineered cyanobacteria.* Biotechnology Advances, 2010. 28(6): p. 742-746.

31. Jensen, S.K., *Improved Bligh and Dyer extraction procedure.* Lipid Technology, 2008. 20(12): p. 280-281.

32. Azizian, H. and J.G. Kramer, *A rapid method for the quantification of fatty acids in fats and oils with emphasis on trans fatty acids using fourier transform near infrared spectroscopy (FT-NIR).* Lipids, 2005. 40(8): p. 855-867.

33. Larsson, K., *Conformation-dependent features in the raman spectra of simple lipids.* Chemistry and Physics of Lipids, 1973. 10(2): p. 165-176.

34. Cooksey, K.E., et al., *Fluorometric determination of the neutral lipid content of microalgal cells using Nile Red.* Journal of Microbiological Methods, 1987. 6(6): p. 333-345.

35. Cooper, M.S., et al., *Visualizing "green oil" in live algal cells.* Journal of Bioscience and Bioengineering, 2010. 109(2): p. 198-201.

36. Elsey, D., et al., *Fluorescent measurement of microalgal neutral lipids.* Journal of Microbiological Methods, 2007. 68(3): p. 639-642.

37. De la Hoz Siegler, H., et al., *Improving the reliability of fluorescence-based neutral lipid content measurements in microalgal cultures.* Algal Research, 2012. 1(2): p. 176-184.

38. Govender, T., et al., *BODIPY staining, an alternative to the Nile Red fluorescence method for the evaluation of intracellular lipids in microalgae.* Bioresource Technology, 2012. 114(0): p. 507-511.

39. Hoover, S.W., et al., *Isolation of improved free fatty acid overproducing strains of Escherichia coli via nile red based high-throughput screening.* Environmental Progress & Sustainable Energy, 2012. 31(1): p. 17-23.

40. Pinzon, N.M., et al., *Nile Red Detection of Bacterial Hydrocarbons and Ketones in a High-Throughput Format.* mBio, 2011. 2(4).

41. Boyanapalli, R., et al., *Luminescent Whole-Cell Cyanobacterial Bioreporter for Measuring Fe Availability in Diverse Marine Environments.* Applied and Environmental Microbiology, 2007. 73(3): p. 1019-1024.

42. Sakamoto, T., et al., *Low-temperature-induced desaturation of fatty acids and expression of desaturase genes in the cyanobacterium Synechococcus sp. PCC 7002.* FEMS Microbiology Letters, 1997. 152(2): p. 313-320.

43. Essich, E., S.E. Stevens, and R.D. Porter, *Chromosomal transformation in the cyanobacterium Agmenellum quadruplicatum.* Journal of Bacteriology, 1990. 172(4): p. 1916-1922.

44. Stevens, S.E. and R.D. Porter, *Transformation in Agmenellum quadruplicatum.* Proceedings of the National Academy of Sciences of the United States of America, 1980. 77(10): p. 6052-6056.

45. Xu, Y., et al., *Expression of Genes in Cyanobacteria: Adaptation of Endogenous Plasmids as Platforms for High-Level Gene Expression in Synechococcus sp. PCC 7002*, in *Photosynthesis Research Protocols*, R. Carpentier, Editor 2011, Humana Press. p. 273-293.

46. Pecoraro, V., et al., *Quantification of Ploidy in Proteobacteria Revealed the Existence of Monoploid, (Mero-)Oligoploid and Polyploid Species.* PLoS ONE, 2011. 6(1): p. e16392.

47. Shen, P. and H.V. Huang, *HOMOLOGOUS RECOMBINATION IN ESCHERICHIA COLI: DEPENDENCE ON SUBSTRATE LENGTH AND HOMOLOGY.* Genetics, 1986. 112(3): p. 441-457.

48. Shaner, N.C., *Chapter 6 - Fluorescent proteins for quantitative microscopy: Important properties and practical evaluation*, in *Methods in Cell Biology*, C.W. Jennifer and W. Torsten, Editors. 2014, Academic Press. p. 95-111.

49. Angsuthanasombat, C. and S. Panyim, *Biosynthesis of 130-kilodalton mosquito larvicide in the cyanobacterium Agmenellum quadruplicatum PR-6.* Applied and Environmental Microbiology, 1989. 55(9): p. 2428-2430.

50. Kawata, Y., et al., *Expression of salmon growth hormone in the cyanobacteriumAgmenellum quadruplicatum.* Biotechnology Letters, 1991. 13(12): p. 851-856.

51. Murphy, R.C. and S.E. Stevens, *Cloning and expression of the cryIVD gene of Bacillus thuringiensis subsp. israelensis in the cyanobacterium Agmenellum quadruplicatum PR-6 and its resulting larvicidal activity.* Applied and Environmental Microbiology, 1992. 58(5): p. 1650-1655.

52. Stevens, S.E., Jr., et al., *A genetically engineered mosquitocidal cyanobacterium.* Journal of Applied Phycology, 1994. 6(2): p. 187-197.

53. Nguyen, A.W. and P.S. Daugherty, *Evolutionary optimization of fluorescent proteins for intracellular FRET.* Nat Biotech, 2005. 23(3): p. 355-360.

54. Shaner, N.C., et al., *Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein.* Nat Biotech, 2004. 22(12): p. 1567-1572.

55. Shaner, N.C., P.A. Steinbach, and R.Y. Tsien, *A guide to choosing fluorescent proteins.* Nat Meth, 2005. 2(12): p. 905-909.

56. Meyer, S., et al., *UV□induced blue□green and far□red fluorescence along wheat leaves: a potential signature of leaf ageing.* Journal of Experimental Botany, 2003. 54(383): p. 757-769.

57. Merzlyak, M.N., T.B. Melø, and K.R. Naqvi, *Effect of anthocyanins, carotenoids, and flavonols on chlorophyll fluorescence excitation spectra in apple fruit: signature analysis, assessment, modelling, and relevance to photoprotection.* Journal of Experimental Botany, 2008. 59(2): p. 349-359.

58. Liu, Y., et al., *Circadian orchestration of gene expression in cyanobacteria.* Genes & Development, 1995. 9(12): p. 1469-1478.

59. Corish, P. and C. Tyler-Smith, *Attenuation of green fluorescent protein half-life in mammalian cells.* Protein Engineering, 1999. 12(12): p. 1035-1040.

60. Brand, J.J. and K.R. Diller, *Application and theory of algal cryopreservation.* Nova Hedwigia, 2004. 79(1-2): p. 175-189.

61. van Hijum, S.A.F.T., M.H. Medema, and O.P. Kuipers, *Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation.* Microbiology and Molecular Biology Reviews, 2009. 73(3): p. 481-509.

62. Beck, C., et al., *Daily Expression Pattern of Protein-Encoding Genes and Small Noncoding RNAs in Synechocystis sp. Strain PCC 6803.* Applied and Environmental Microbiology, 2014. 80(17): p. 5195-5206.

63. Lambert, J.M., R.S. Bongers, and M. Kleerebezem, *Cre-lox-Based System for Multiple Gene Deletions and Selectable-Marker Removal in Lactobacillus plantarum.* Applied and Environmental Microbiology, 2007. 73(4): p. 1126-1135.

64. Tan, X., et al., *Application of the FLP/FRT recombination system in cyanobacteria for construction of markerless mutants.* Applied Microbiology and Biotechnology, 2013. 97(14): p. 6373-6382.

DISTRIBUTION

1	MS0895	Reynolds M. Salerno	8630
1	MS1413	James P. Carney	8631
1	MS9291	Blake Simmons	8610
1	MS9292	Paula Imbro	8600
1	MS0899	Technical Library	9536 (electronic copy)
1	MS0359	D. Chavez, LDRD Office	1911


Sandia National Laboratories