TEMPERATURE-REGULATED CONTROL OF TOXIC GENES IN THE METHYLERYTHRITOL PHOSPHATE PATHWAY OF *SYNECHOCOCCUS* SP. PCC 7002 CYANOBACTERIA

By Meghan M. Raebel

Cyanobacteria are microalgae that have the potential to become a frontrunner for clean energy and renewable bioproducts. These green bacteria efficiently harvest sunlight and atmospheric CO_2, to produce a variety of organic compounds, including isoprene, a precursor for synthetic rubber, pharmaceuticals, and Bio-Jet fuel. Cyanobacteria, such as the marine strain, *Synechococcus* sp. PCC 7002, replicate quickly, can be grown in wastewaters, and need not compete with food crops for arable land. Additionally, cyanobacteria contain the methyl erythritol phosphate (MEP) pathway, by which isoprene can be produced.

Our group has engineered isoprene production in *Synechococcus* cyanobacteria (US patent 9,382,554) and we are working to enhance this technology toward commercialization. One strategy for increased isoprene yields involves introducing optimized genes for the MEP pathway, but some of these genes or their products may be toxic and prevent cyanobacterial growth. Thus, the objective of this research is to develop a means of controlling such genes so that they are expressed only after the cyanobacterial culture has reached a high density. One means of control is through genetic promoters and regulators. One such regulator is *cI857-pR*, a temperature-regulated repressor-promoter from a bacterial virus. We believe this will provide a means to regulate gene expression and determine whether introduced genes are toxic. In this thesis project, I have created genetic fusions of this regulator to genes within the MEP pathway, and have demonstrated an increase in isoprene production up to three times the rate observed in a control strain. Understanding how these gene products and their regulation works in the cyanobacteria will inform us how to more effectively insert enhanced genes for the MEP pathway, control their activity, and develop strains with enhanced isoprene yields.

TEMPERATURE-REGULATED CONTROL OF TOXIC GENES IN THE METHYLERYTHRITOL PHOSPHATE PATHWAY OF *SYNECHOCOCCUS* SP. PCC 7002 CYANOBACTERIA

by

Meghan M. Raebel

A Thesis Submitted
In Partial Fulfillment of the Requirements
For the Degree of

Master of Science - Biology

Microbiology

at

The University of Wisconsin Oshkosh
Oshkosh WI 54901-8621

March, 2017

COMMITTEE APPROVAL

Advisor

10/17/2017 Date Approved

Member

17 Oct 2017 Date Approved

Member

10/17/17 Date Approved

PROVOST AND VICE CHANCELLOR

Oct. 24, 2017
Date Approved

FORMAT APPROVAL

4/7/17
Date Approved

This thesis is dedicated to Shelley, William, and Michael. I love you all.

ACKNOWLEDGEMENTS

To my advisor, Dr. Toivo Kallas, I would like to extend my profound appreciation. His dedication, mentorship, and support through my time at the University of Wisconsin-Oshkosh will never be forgotten. I am also grateful for his belief in my potential to succeed as a graduate student despite my research inexperience and for providing me the opportunity to be a part of the projects in his lab and at Algoma Algal Biotechnology. I would also like to thank my committee members Dr. Eric Matson and Dr. Lisa Dorn for their valued input. One invaluable resource that was irreplaceable in the lab is Matt Nelson. He was *always* able to take any inane question, request or comment and return with a response. Without his help, this thesis would not have come to fruition. I would like to extend my gratitude to the other students within the lab: Ola Aremu and Rhiannon Carr. Ola was my first resource in the lab, and his help will never be forgotten. Rhiannon, my friend and lab mate, was always there to lend an ear or a hand or even just company. Without her there would have been far too many tubes forgotten in the thermocycler. Finally, to my family: thank you for believing and supporting me, even if didn't completely understand what I was doing.

I cannot thank any of you enough.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

STATEMENT OF PROBLEM

In December 2015, at the Paris Climate Conference, 195 countries (including but not limited to: China, North Korea, Russia, and the United States) adopted the first ever universal, legally binding global climate deal. The agreement stipulates a global action plan to put the world on track to avoid dangerous climate change by limiting global warming to well below 2°C (*1*). The agreement is due for enforcement in 2020—at the time of writing, three years from now. Other aims of the convention are increasing the ability to adapt to the adverse impacts of climate change and foster climate resilience and low greenhouse gas emissions development, in a manner that does not threaten food production and making finance flows consistent with a pathway towards low greenhouse gas emissions and climate-resilient development (*2*). Despite the recent withdrawal of the U.S., by the Trump administration, from the Paris Treaty, most of the world, including many U.S. states and cities are moving forward with efforts to mitigate climate change.

In order to meet these goals, drastic changes to our current fuel consumption and carbon emissions must be made. One way that this can be achieved is through the use of alternative fuels. A biofuel is a fuel that is manufactured through contemporary biological processes, such as agriculture and anaerobic digestion, rather than a fuel that has been produced by geological processes such as those involved in the formation of fossil fuels, such as coal and petroleum, from prehistoric biological matter. Biofuels can be derived directly from plants, or indirectly from agricultural, commercial, domestic, and/or

industrial wastes (*3*). Renewable biofuels typically require carbon fixation, such as those that occur in plants or microalgae through the process of photosynthesis.

There are several generations of biofuels, each one more advanced than the other. "First generation," or conventional biofuels are made from sugar, starch, or vegetable oil. The three types of first generation biofuels used commercially are biodiesel, ethanol, and biogas of which large quantities have been manufactured so far and for which the production process is considered established technology (*4*). Biodiesel biofuels are an alternative of diesel and are produced through transesterification of vegetable oils and residual oils and fats. With minor engine modifications they can serve as a full substitute as well (*5*). Bioethanol are biofuels that are a substitute of gasoline, and can be a full substitute for gasoline in vehicles that are equipped in the correct manner (*6*). It is derived from sugar or starch through fermentation, but also microalgae have been engineered for efficient, photosynthetic production of bioethanol (*7*) . The feedstock that makes up a majority of bioethanol fuels is corn. Corn is a valuable feedstock for ethanol because it contains a large amount of carbohydrates (*8*). Biogas, or biomethane, is a fuel that can be used in gasoline vehicles with slight adaptations. It can be produced through anaerobic digestion of liquid manure and other digestible feedstock. At present, biodiesel, bioethanol and biogas are produced from commodities that are also used for food (*4*). The problem with first generation fuels, however, is that they are available in limited volumes that do not make them significant alternatives for petroleum.

Second generation biofuels, on the other hand, are from forest and crop residues, energy crops and municipal and construction waste, and will arguably reduce net carbon emission, increase energy efficiency and reduce energy dependency, potentially overcoming the limitations of first-generation biofuels (*9*). Second generation biofuels are produced from biomass in a more sustainable manner, which is carbon neutral in terms of its impact on CO_2 concentrations. In the context of biofuel production, the term 'plant biomass' refers largely to lignocellulose material as this makes up the majority of the cheap and abundant nonfood materials available from plants. Second generation biofuels are made from lignocellulosic biomass or woody crops, agricultural residues or waste, which makes it harder to extract the required fuel. A series of physical and chemical treatments are typically required to convert lignocellulosic biomass to liquid fuels suitable for transportation (*10*, *11*). The large disadvantage of second-generation biofuels, however, is low energy efficiency of production processes.

The newest generation of biofuels is the third generation. The term third generation biofuel has only recently entered the mainstream as it refers to biofuel derived from algae. Previously, algae were lumped in with second generation biofuels. Microalgae capture carbon from atmospheric CO_2 and many have high growth rates, doubling their biomass in as little as a few hours, and can grow in salty or waste water. Oil production from microalgae per unit of cultivated area can be up to 300 times greater than that of oil plants (*12*).

Algae are particularly useful when considering fuels, as their benefits far outweigh their costs. Additionally, disadvantages with first and second generation

biofuels can be overcome bearing in mind that algae, which can produce large volumes of biofuels, on much smaller areas, as a viable, alternative energy resource. At the present time, several companies are attempting to reduce costs to make algae production viable. In the United States alone there have been nearly 1,600 publications yearly regarding microalgal biofuel technologies (*13*).

The primary benefit that algae present is enhanced energy yields, when compared to current, conventional biofuels. These traditional biofuels require large swaths of land in order to be cultivated (for crop plants such as wheat, barley, corn, potato, sugarcane, and sugar beet). This land usage can be significantly reduced through the use of algae as a feedstock. Available data and predictions indicate that the production rate (L/ha) from microalgae (91%) can be much higher than of other feedstock such as palm (3%), coconut (1.5%), jatropha (1.2%), avocado (1.4%) and rapeseed/canola (1%) (Figure 1) (*14*). Furthermore, algae are superior solar energy convertors. Compared to terrestrial crops, algae's solar energy conversion efficiency is 3%-9%, versus the theoretical maximum for switch grass (2.4%) or maize (3.7%) (*15*). Moreover, algae are able to grow year round in bioreactors, whereas traditional crops for biofuel production are subject to seasons and soil conditions.

Likewise, microalgae's requirements for water are more amenable than those of land plants for second-generation biofuels. The natural water source for marine algae is seawater, which is ample. Water could also come from nutrient-laden agricultural wastewater, while terrestrial biofuel crops require freshwater (*15*). The water source need not be close by, as required by most terrestrial crops, as transporting water will quickly

become cost prohibitive. Mass cultivation of algae is currently done at coastal locations in open ponds and interior locations in desert climate using nutrient- supplemented seawater.

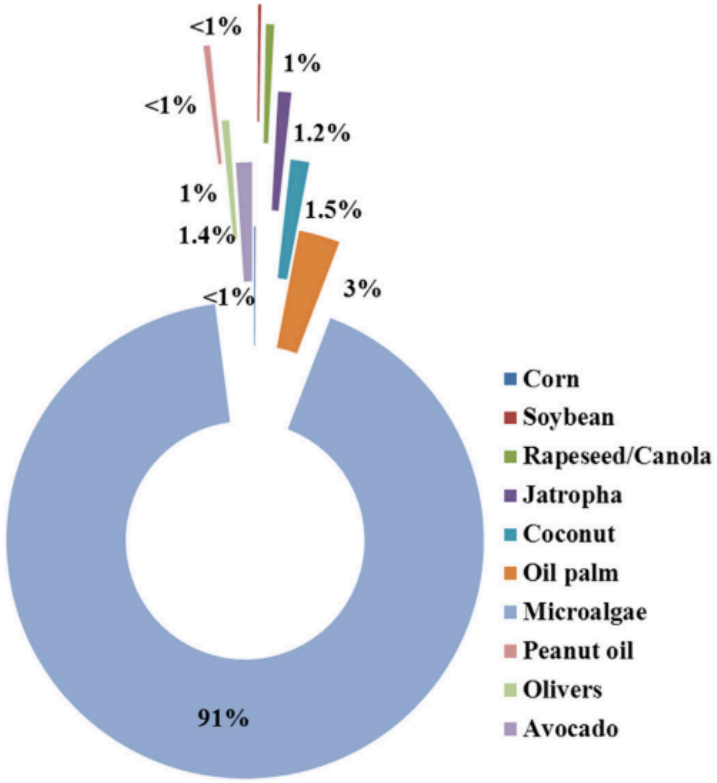

Figure 1. Biofuel production rates. Predicted rates of biofuel production (L/ha) from microalgae relative to other sources: The present scenario (*14*).

Microalgae are better than terrestrial plants for biofuel production for several additional reasons. Microalgae do not require extensive structural biopolymers essential for higher plant growth in terrestrial environments (*15*), The resulting absence of cellulose, hemi-cellulose, and lignin eliminates the need for pretreatments needed to

breakdown lignocellulose structures, and lignin cannot, at this time, be transformed easily into transportation fuels) (*16*). This advantage alone could save money through a simplified process.

Microalgae, by their nature, have extensive metabolic diversity. There exist an estimated 200,000- 800,000 algal species, of which about 50,000 species have been described (*17*). Most microalgae species produce lipids, carotenoids, antioxidants, fatty acids, enzyme polymers, peptides, toxins, and sterols (*17–20*). This diversity allows selection of species that are adapted for growth in locally available aquifers, or have morphological features that allow cost-effective harvesting (*15*). Furthermore, many microalgal species are already well characterized, with genomes sequenced and annotated. Some of these species have already been genetically investigated or modified toward identifying promising targets for genetic modifications and optimized bioenergy production (*21*).

Finally, these organisms can also generate valuable byproducts. Not only can these organisms be utilized for oil production, but a significant fraction of the residual biomass following lipid and carbohydrate extraction is protein. These byproducts can be directed toward secondary markets, where significant income could be derived through the sale of the protein for animal feed (*15*). Harvesting biomass is not, however, the only method by which third-generation biofuels can be produced – the genetic tractability of cyanobacteria lends them to the direct synthesis of a host of energy-dense compounds suitable for use as transport fuels.

BACKGROUND

CYANOBACTERIA AS A PLATFORM FOR METABOLIC ENGINEERING

Particularly in genetic engineering, the traditional microbial workhorse has been *E. coli*. It's rise and exalted status in biology stems from how easily and rapidly it can be cultured and modified genetically. Hardy, non-pathogenic, and versatile strains that grow quickly on many different nutrients can be isolated from virtually any human (*22*). These traits made *E. coli* important within microbiology. However, while *E. coli* is useful in generating pharmaceutical products like insulin, erythropoietin (for treatment of anemia), and human growth hormone, there are other microbial platforms that are equally useful for industrial applications (*e.g.* biofuel production) (*23*, *24*).

One particular microbial platform that is practical for 'greener' technologies is cyanobacteria. These organisms are just as metabolically flexible as *E. coli*, but have the added benefit of being photosynthetically active (more so than plants). Cyanobacteria are photosynthetic microalgae that have the potential to become the frontrunner for clean energy. These green bacteria can efficiently harvest sunlight, atmospheric CO_2, and wastewater to produce a variety of organic compounds, including petroleum-replacement feedstocks, pharmaceuticals, and a variety of rubbers (*25*, *26*). Furthermore, their minimal environmental demands ensure that they are relatively economical to cultivate and do not compete for valuable farm land, as do many of the current biofuel organisms, such as corn for ethanol production (*27*). In fact, cyanobacteria can be cultivated in some of the least hospitable environments, allowing previously barren locations to serve as a bioenergy farms. Many cyanobacterial species, including *Synechococcus* sp. PCC 7002,

Synechococcus elongatus PCC 7942, *Synechocystis sp*. PCC 6803, and *Thermosynechococcus elongatus* are naturally competent to take up exogenous DNA, which is even more convenient for genetic modifications than in *E. coli*, which must be made competent for DNA uptake by chemical or electrical inducement (*28–32*).

E. coli, on the other hand, is known for its rapid doubling time of ~20 minutes. Thus *E. coli* grows much more rapidly than cyanobacteria. However, the recently rediscovered *Synechococcus elongatus* UTEX 2973 strain, comes close to the rate of *E. coli* grown on minimal medium, with a two-hour doubling time under optimal conditions. The next, known fastest cyanobacterium, *Synechococcus* sp. PCC 7002, doubles in approximately four hours, and *Synechocystis sp*. PCC 6803 is even slower at more than six and a half hours (*33*). This slow growth, however, may not be as limiting as it appears. For industrial applications, slower growth may be preferred, as less cell shading and lower cell density may increase product yields relative to cell mass.

Terpenoid Biosynthesis in Cyanobacteria

Given that cyanobacteria are able to fix carbon from the atmosphere, their use in the expanding field of alternative biofuels is of particular interest. They require less water for growth compared to corn, and can utilize seawater or nutrient-laden agricultural wastewater for growth, where terrestrial plants for biofuels require freshwater. Furthermore, cyanobacteria require significantly less land to grow (on land that need not be arable), because of their high photosynthetic efficiency, whereas crops that produce ethanol or cellulosic ethanol require significant land use (*15*). Therefore, cyanobacteria are better suited to produce biofuel, as they are inherently more efficient solar collectors,

use less land, can be converted into liquid fuels using simpler technologies, and have additional uses that fossil fuels do not provide. Algal biomass, for example, can be cultivated to have a high protein and oil content, which can be used to produce animal feeds. In addition, microalgal biomass, which is rich in micronutrients, is already used for dietary supplements to advance human health (*34*).

Finally, cyanobacteria are able to produce a variety of useful compounds in sufficient quantities to compete with conventional petrochemicals without causing the cultures to crash. Both ethanol and 1-butanol have been produced in limited quantities in *Synechocystis* PCC 6803 and *Synechococcus* PCC 7942, respectively, which provides a route to replacing conventional gasoline, if production can be boosted substantially. The variety of fuels that can be obtained from raw petroleum sources is not, however, limited to gasoline and diesel – aviation fuels are composed of high-density petrochemicals like kerosene – and likewise the range of biofuels produced by cyanobacteria should not be limited to alcohols and fatty acids (*25*, *35*, *36*).

Recently, our lab group has engineered *Synechococcus* sp. PCC 7002 cyanobacteria to produce isoprene. Isoprene, or 2-methyl-1,3-butadiene, is a common organic compound with the formula $CH_2=C(CH_3)-CH=CH_2$ (Figure 2). Isoprene is a volatile hydrophobic compound and passes through cellular membranes allowing it separate from the cyanobacterial biomass and then be captured in the headspace of a bioreactor or sealed container for collection (*37*). It is a colorless gas that is particularly significant, as this compound is a precursor for many terpene bioproucts and high-density fuels (*38*). It takes seven gallons of crude oil to produce a gallon of petroleum-based

isoprene, which makes isoprene a relatively high-value chemical (*4*). In fact, when converted into forms like limonene or pinene (dimers of isoprene), the combustible energy is nearly identical to that of aviation fuel (41.9 versus 42.1 MJ kg^{-1}), but without the carbon footprint of petrochemically derived fuels (*39*). Bio-based isoprene is potentially more economical than that produced from petroleum and does not have the same, adverse environmental impacts. Global demand for isoprene yearly amounts to 800,000 metric tons, with 70 percent in the tire industry, 20 percent for manufacturing adhesives and the remainder for medical and personal supplies.

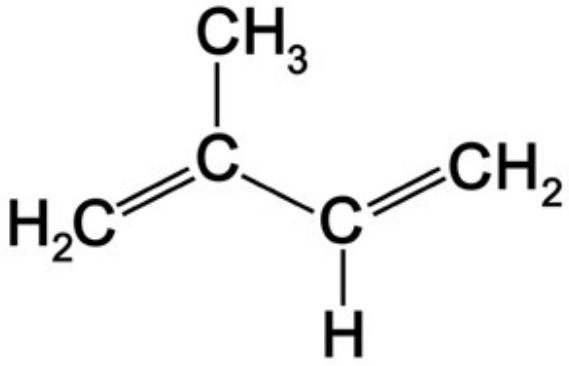

Figure 2. Isoprene. The chemical structure of isoprene.

Isoprenoids (or terpenoids) are a large family of compounds useful in biological functions such as hormone-based signaling, electron transport, regulation of transcriptional and post-translational processes, protein degradation and glycoprotein biosynthesis (*40*). Isoprenoid biosynthesis in living organisms is required for the production of compounds such as carotenoids, vitamins, quinones, chlorophyll, and the primary constituents of essential oils in plants (*41*). Terpenoids or isoprenoids are compounds of terpenes modified by chemical oxidation and contain functional groups.

Terpenes are composed of isoprene units and vary in structure from linear to ring arrangements. The length of these terpenoids can vary from one isoprene unit (making isoprene) to over eight isoprene units (making up natural rubber).

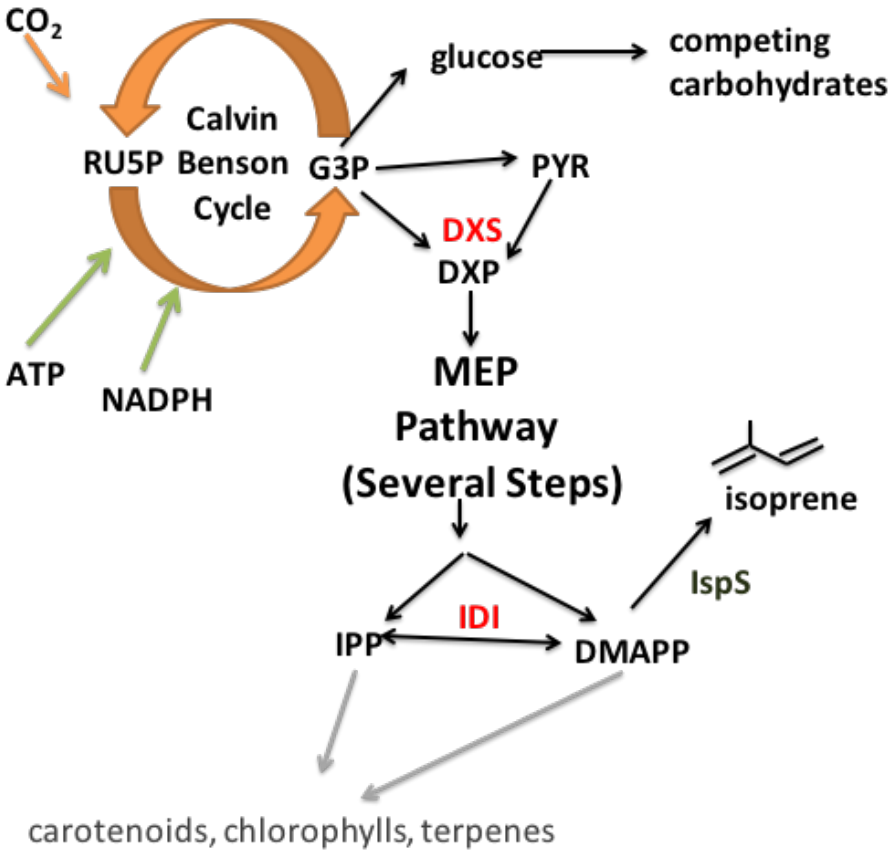

Figure 3. Methyl erythritol phosphate (MEP) pathway and its modification to produce isoprene and terpenes. In several enzymatic steps, the MEP pathway converts glyderaldehyde-3-phosphate (G3P) and pyruvate (PYR) to isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are precursors for natively-produced terpenoids such as carotenoids and chlorophylls, as well as substrates for additional terpene-producing enzymes such as isoprene synthase (IspS) for isoprene production. The IPP-DMAPP isomerase (IDI, shown in red) manages the balance between IPP and DMAPP, and has been shown to increase isoprene production in *IspS*-containing strains of *Synechococcus* sp. PCC 7002. The deoxy-d-xylulose-5-phosphate synthase (DXS) enzyme (shown in red) catalyzes a rate-limiting step, but may aggregate if overproduced within the cell, causing cell death, and is one enzyme that we seek to control by means of a temperature-controlled regulator.

Two different terpenoid biosynthesis pathways are known: first, the mevalonate acid (MVA) pathway (*42*) found in archaea, eukaryotes and some bacteria; second, the non-mevalonate or 2C-methyl-D-erythritol-4-phosphate (MEP) pathway (Figure 3) in cyanobacteria, algae, eubacteria and plant chloroplasts (*43*). In this study, I examined the MEP pathway for terpenoid biosynthesis in *Synechococcus* sp. PCC 7002 cyanobacteria.

Cyanobacteria can funnel atmospheric carbon captured by photosynthesis into the methyl erythritol phosphate (MEP) pathway. This is a metabolic pathway within *Synechococcus* beginning with the enzyme DXP synthetase (DXS) and several other enzymes to generate the end-products, isopentyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), which are precursors for important cellular metabolites such as chlorophyll and carotenoids as well as isoprenoids. Cyanobacteria lack the key isoprene synthase (IspS) enzyme for converting DMAPP to isoprene (*44*). However, isoprene synthesis in cyanobacteria has been achieved by the introduction of optimized genes for isoprene synthase (IspS) from *Poplar* and *Kudzu* plant species (*45*).

Temperature Sensitive Regulation

It is not clear yet how the MEP pathway in cyanobacteria is regulated, but this regulation may involve the accumulation of products in the pathway, effector molecules, transcriptional control, (*40*) or a variety of these mechanisms. Studies involving transgenic plants have shown that overexpression or antisense repression of the DXS enzyme resulted in corresponding increases or decreases in accumulation of isoprenoid

products (*46*), suggesting that DXS catalyzes a rate-limiting step for IPP and DMAPP synthesis and isoprenoid final products.

Increasing the yield of isoprene, however, does come at a cost. With the introduction of foreign genes into the MEP pathway to further increase isoprene yields, we find that strains of *Synechococcus* may not persist and cannot be propagated. Our lab group has evidence that indicates that this premature culture crash may be due to the newly introduced genes within the MEP pathway being over expressed, and as a result, an overproduction of protein. This can lead to protein aggregation within the cells, which could be the reason behind the culture crash. Thus, one objective of the thesis research was to develop a means of controlling the expression of these genes within the MEP pathway. One means of control is through genetic promoters and regulators. One regulator of interest is cI857, a temperature-sensitive regulator from *E. coli* bacteriophage (bacterial virus) lambda (λ) (*47*). Such controlled promoters can relieve cells of metabolic burdens such as continuous isoprene production. More so, such regulatable promoters provide a means of control for isoprene production over a broad range of environmental conditions. At low temperatures, e.g. below 30°C, the cI857 repressor protein is synthesized and prevents transcription by binding of the pR promoter region in the phage lambda DNA. At higher temperatures, e.g. at 40°C, the cI857 repressor protein begins to denature (unfold), allowing transcription to proceed.

The expression of genes controlled by the cI857 repressor is controlled through the binding of a repressor dimer to three operator sites in the phage DNA. The cI857 repressor protein is made up of two domains, the amino-terminal domain which binds to

a 17 base pair operator site, and the carboxyl-terminal domain. The repressor dimerizes (by contacts in the carboxyl domain), and sits within the major groove of the DNA (*48*). In the phage lambda genome, there is a second protein, Cro, which interacts with the operator site as well. In this study I removed all of the *Cro* gene, save for the first 22 amino acids, so the full Cro protein was not created. Previous evidence (*49*) indicates that retaining the first half of the Cro protein to create a fusion with target proteins may assist in enhanced protein production (Matthew Nelson, unpublished data).

The relationship between the operator regions and the promoter regions is complex, and is illustrated in Figure 4. Each operator region is 17 base pairs of similar but not identical sequences, of which there are three: O_R1, O_R2, and O_R3. The promoter regions for *Cro* and the cI857 repressor protein overlap, wherein the repressor's promoter region (P_{RM}) overlaps completely with O_R3 and partially with O_R2. *Cro*'s promoter region (P_R) then overlaps partially with O_R2 and completely with O_R1. As a result, RNA polymerase will never sit on P_{RM} or P_R as long as the cI857 is bound to its operator sites. An important difference between the promoter P_R and P_{RM} is that RNA polymerase binds and begins transcription at the former without the aid of any positive regulatory protein. In contrast, RNA polymerase works efficiently at P_{RM} only if helped by an activator protein, a role that is also played by the cI857 repressor (*48*). Temperature sensitivity in the cI857 repressor has occurred because of a small mutation in the amino-terminal portion of the repressor (*50*).

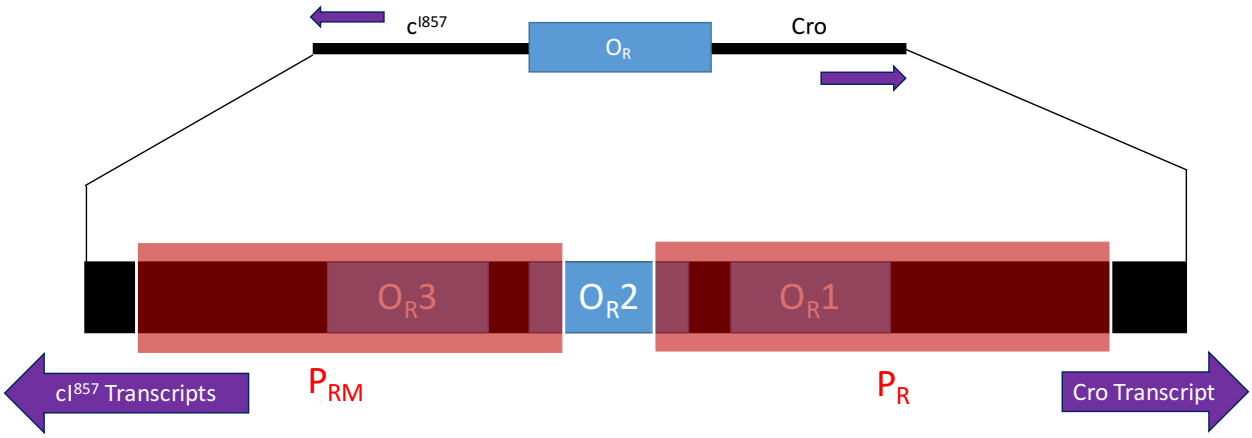

Figure 4. cI857, Cro and the operators. Shown is a short segment of the phage λ genomic DNA. The back-to-back promoters (P_{RM} and P_R) send polymerase traveling in opposite directions, leftward to transcribe the repressor gene (*cI857*) and rightward to transcribe the *cro* gene. The tripartite operators (O_R) overlap the two promoters. Each of the three pairs of the operator is called an operator site (Figure modified from **(*48*)**).

The plasmid used in this study was pGH276 (Figure 5). It contains genes for the cI857 repressor, Cro and an ampicillin antibiotic-resistance marker. The cI857 repressor protein on this plasmid is thermolabile, and shows progressively decreasing activity at temperatures above 33°C. At 30°C, however, the repressor is fully functional in phage lambda and *E. coli* (*47*).

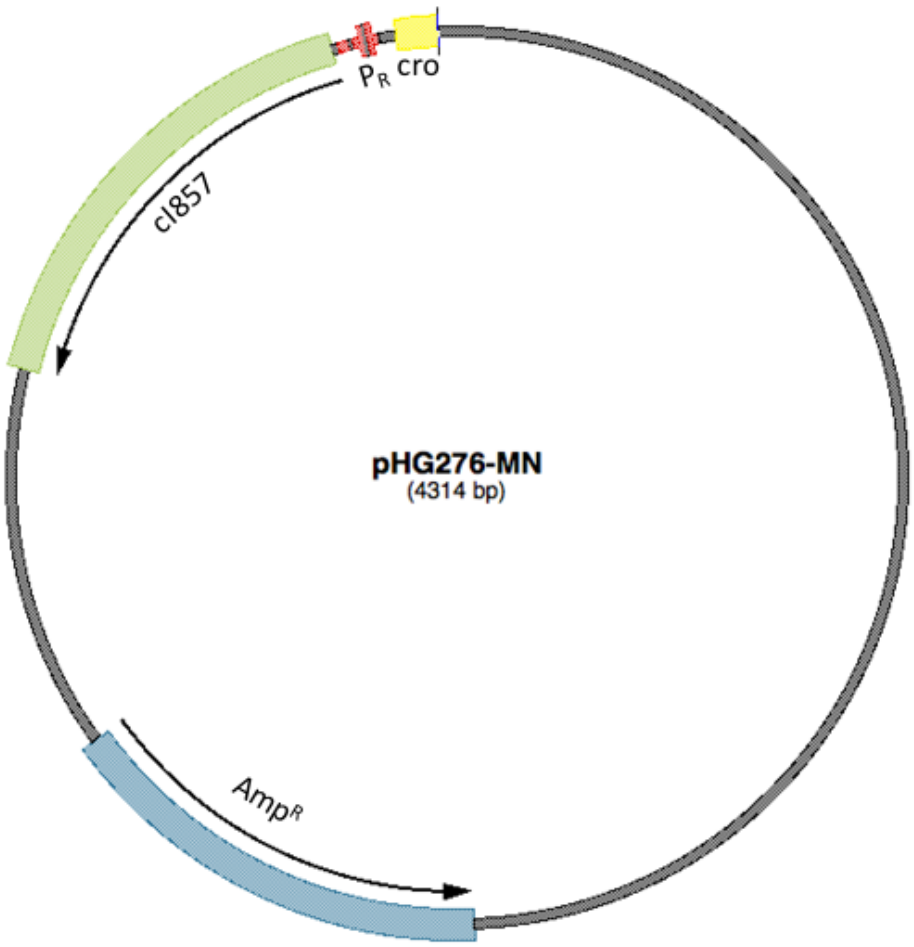

Figure 5. Plasmid pHG276. This 4314 base pair plasmid contains genes for the cI857 repressor (green), and ampicillin antibiotic-resistance cassette (teal). The *cro* gene is included as well (yellow). Only the first 22 codons of *cro* were used in this study. The P_R promoter is enlarged region between cI857 and *cro*, the alternating red and black regions indicate the operator sites.

I postulated that the cI857 repressor and pR promoter will provide a controllable means of regulating gene expression in *Synechococcus* cyanobacteria and enable me to determine whether gene products of the MEP pathway are toxic when overexpressed. Understanding how these gene products work within the cyanobacteria will inform us

how to more effectively insert enhanced genes for the MEP pathway and control their activity.

Previous attempts at introducing optimized genes for isoprene synthesis into *Synechococcus* sp. PCC 7002 involved targeting these genes to the high-copy plasmid, pAQ1. Declining isoprene yields observed in such strains after several days of growth (data not shown) may be a consequence of plasmid loss or the result of metabolic burden by multiple copies of these genes, highly expressed from plasmid pAQ1. Furthermore, any attempts made in this thesis to target genes to the pAQ1 plasmid resulted in failure of *Synechococcus* to stably maintain the transgenes. Therefore, because of these and other attempts by our lab group, it became clear that it will be advantageous to express transgenes in the host chromosome rather than on plasmids, as plasmid loss appears to be a major cause of reduced heterologous gene expression in these cyanobacteria.

Previous studies in our group and by others has shown that chromosomal targeting of transgenes leads to complete replacement of target sites via segregation and thus stable integration of transgenes (generally quite quickly as well) without the need for selective pressure through the addition of adjacent antibiotic resistance genes (*45*). Although, at least initially, a selectable marker is used. In this study I targeted transgenes to a neutral insertion site designated NISFD (located between open reading frames SYNPCC7002_RS04670 and SYNPCC7002_RS04675) (*51*). Previous work inserting transgenes into this location displayed nearly instantaneous segregation of the chromosome, with no obvious detrimental effects on growth. However, introduction of

toxic transgenes would still be expected to yield either incomplete segregrants or deletion or complete loss of the transgenes.

PROJECT SCOPE AND OBJECTIVES

I therefore proposed the following hypotheses: First, the bacteriophage lambda cI857 repressor will function in *Synechococcus* sp. PCC 7002 in a temperature-dependent manner. Second, the cI857 repressor will regulate the expression of potentially toxic MEP genes in a temperature-dependent manner leading to increased carbon flow and isoprene production. To test these hypotheses, the cI857-pR repressor-promoter were first tested by linking these to a gene (*yfp*) for the yellow fluorescent protein (YFP), to determine the repressor-promoter function in *Synechococcus* 7002 with respect to temperature. Once this was established, the *yfp* gene was removed and replaced with several *DXS* and other MEP gene constructs (Figure 6) such that these were all placed under control of the cI857-pR repressor-promoter. Construct 1602-B encodes the first 22 amino acids of *Cro*, followed by a gap of ~100 bp, followed by *DXS*, *IspF* and *IspH* genes. Construct 1605-A encodes the first 22 amino acids of Cro fused in-frame with DXS. Construct 1605-B encodes DXS on it own, no Cro.

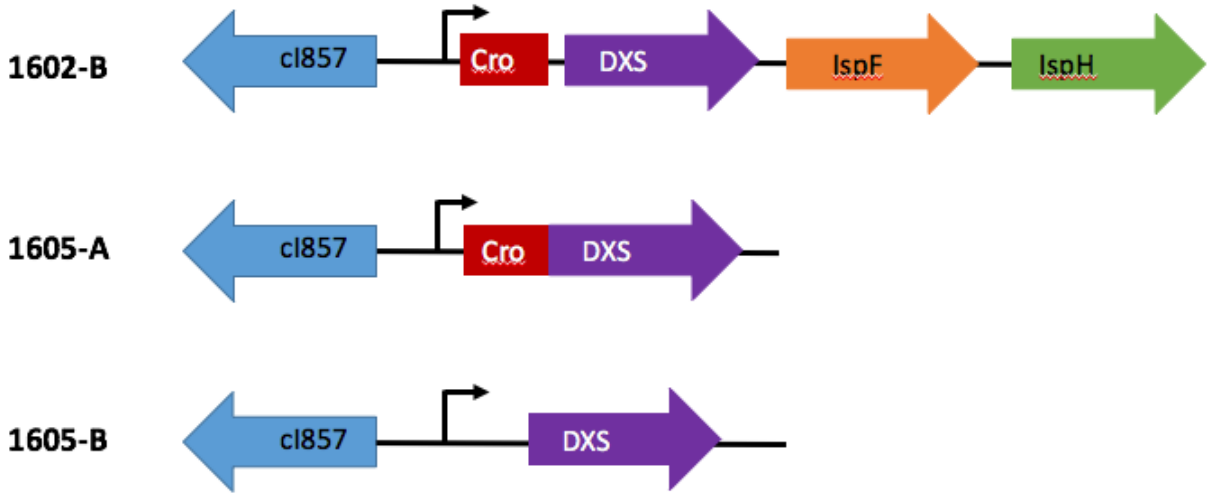

Figure 6. Genetic constructs for MEP pathway *DXS*, *IspF*, and *IspH* genes under control of the cI857-pR repressor-promoter. All of these gene sets are controlled by the cI857 temperature-sensitive repressor and pR promoter (indicated by arrow). All of the *Synechococcus* 7002 transformants tested with these genes also carried codon-optimized *IspS-IDI* genes (as in *Synechococcus* 7002 strain 1310) for isoprene synthesis.

In this thesis, I describe 1) the successful construction of three recombinant plasmids carrying copies of various combinations of genes for DXS and MEP pathway IspF and IspH enzymes all under control of the temperature-responsive cI857-pR repressor-promoter as shown in Figure 6, and 2) three transgenic, segregant *Synechococcus* sp. PCC 7002 strains that carry these gene sets targeted to a neutral site within the genome. These genes were introduced into *Synechococcus* 7002 1310 strains that already carried optimized poplar *IspS-IDI* genes targeted to the *PetJ2* region of the chromosome. This generated 7002::1310/*cI857-DXS* transgenic strains, in which these strains produced isoprene under temperature-regulated control. At 40ºC the

7002::1310/cI857-DXS (no fusion, no Cro) produced isoprene at three times the rate of strains with IspS-IDI genes alone.

CHAPTER II: MATERIALS AND METHODS

STRAINS AND CULTURE MAINTENANCE

The wild type (WT) *Synechococcus* sp. PCC 7002 (hereafter *Synechococcus 7002*), is a euryhaline, gram-negative, rod shaped, unicellular prokaryotic organism, which was isolated from the Magueys, marine mud flats of Puerto Rico (*52*). *Synechococcus* 7002 grows well under photoautotrophic conditions and requires vitamin B_{12} for growth. *Synechococcus* 7002 is capable of growth over a wide range of sodium chloride (NaCl) concentrations and is extremely tolerant of high-light irradiation (*29*, *53*). Under optimal conditions (37 – 39°C, 1 – 3 % (v/v) CO_2 in air, and saturating irradiation of ~250 μmol photons $m^{-2}s^{-2}$), *Synechococcus* 7002 has a doubling time of approximately 3.5 hours, one of the fastest among cyanobacteria. *Synechococcus* 7002 is naturally competent for DNA uptake (*29*) and its genome is completely sequenced (available from the National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov/).

The DH5α *Escherichia coli* strain was used to propagate the recombinant plasmids. Its genotypic characteristics include the following: F-, *lacZ-β* fragment, *recA*-, *endA*-, *gyrA96* (nalidixic acid resistance). DH5α forms blue colonies on medium containing Isopropyl β-D-1-thiogalactopyranoside (IPTG) and X-gal, when it carries a plasmid that carries the *lacZ-α* fragment. DH5α was grown on either Luria-Bertani (LB), rich medium or Minimal Salts/glucose/thiamine medium (*54*).

The *Synechococcus* transformant strains that were available previously or generated in this study and their characteristic genotypes are shown in Table 1.

Table 1. *Synechococcus* sp. PCC 7002 strains used in this study

Strains	Genotypes & Characteristics	Source
7002 Wild Type	*Synechococcus* sp. PCC 7002, control strain	Pasteur Culture Collections
700::1310, ΔpetJ2	*petJ2::IspS-IDI*-Sp^R/Sm^R Transformant strain carrying *IspS-IDI* genes inserted into the chromosomal *petJ2* gene. Carries resistance to Sm/Sp (streptomycin and spectinomycin).	Olalekan R. Aremu, UWO (*45*)
7002::pAQ-YFP	*pAQ1::cpcB-YFP*- Sp^R/Sm^R Transformant strain carrying the *YFP* gene inserted into plasmid pAQ1. Carries resistance to Sm/Sp	Yu Xu, Pennsylvania State University
7002::1602-A	*NISFD::cI857-YFP*- Km^R Transformant strain carrying a temperature controlled *YFP* gene inserted into the chromosomal neutral site, SYNPCC7002_RS04670 and SYNPCC7002_RS04675. Carries resistances to Km (kanamycin).	This study
7002::1602-B	*NISFD::cI857-DXS-IspF-IspH*- Km^R Transformant strain carrying temperature controlled *DXS, IspF and IspH* genes targeted to the chromosomal neutral site, NISFD. Carries resistance to Km.	This study
7002::1605-A	*NISFD::cI857-CroDXS Fusion*- Km^R Transformant strain carrying a temperature controlled *DXS* gene fused to the first 22 codons of *Cro*. Carries resistance to Km.	This study
7002::1605-B	*NISFD::cI857-DXS*- Km^R Transformant strain carrying temperature controlled *DXS* gene. Carries resistance to Km.	This study
7002::1602-B/1310	*NISFD::cI857-DXS-IspF-IspH/IspS-IDI*- Km^R/ Sp^R/Sm^R Transformant strain carrying temperature controlled *DXS, IspF and IspH* genes targeted to a chromosomal neutral site and *IspS-IDI* genes targeted to the *PetJ2* chromosomal site. Carries resistance to Km and Sm/Sp.	This study

7002::1605-A/1310	*NISFD::cI857-CroDXS Fusion/IspS-IDI*- Km^R/ Sp^R/Sm^R Transformant strain carrying a temperature controlled *DXS* gene fused to the first 22 codons of *Cro*, and *IspS-IDI* genes targeted to the *PetJ2* chromosomal site. Carries resistance to Km and Sm/Sp.	This study
7002::1605-B/1310	*NISFD::cI857-DXS/IspS-IDI*- Km^R/ Sp^R/Sm^R Transformant strain carrying a temperature controlled *DXS* gene targeted to a neutral site, and *IspS-IDI* genes targeted to the *PetJ2* chromosomal site. Carries resistance to Km and Sm/Sp.	This study

GROWTH AND MAINTENANCE OF STOCK CULTURES

Long term stock cultures of WT and mutant *Synechococcus* 7002 were grown and propagated in liquid A(D7) and solid A(D7) media containing the appropriate antibiotics (Appendix A). These cultures were kept at room temperature, under cool white fluorescent lamps ~80 μmol photons $m^{-2}s^{-2}$. Long term liquid cultures were either kept under the same conditions, with slow shaking with an orbital shaker, or kept as frozen stocks by growing to ~3 OD in 3mL cultures, centrifuged to 500uL, and adding DMSO to a 7% concentration.

Fresh stocks in preparation for experiments were grown with shaking (~100rpms) on an orbital shaker in a Percival incubator (Percival scientific, Iowa, USA) with controlled growth conditions (at 39°C and light intensity of ~200 μmol photons $m^{-2}s^{-2}$).

Stock cultures of WT and mutant strains were inspected for contamination by

streaking samples onto Luria-Bertani (LB) agar, followed by overnight incubation at 37°C. Subsequently, cultures that were free of contamination were used as inocula for growth experiments.

Optical Density and Culture Growth

An Agilent 8453 UV-visible spectrophotometer was used for optical density (OD) measurements of cultures at 750 nm in 1.0 cm light-path cuvettes. Before cell density measurements are taken, double-distilled (or Milli-Q) water (ddH_2O) was used as a blank. Then, 1.0 mL of culture samples were drawn from culture flasks from which 100 µL were transferred into the clean cuvette for OD measurements.

DNA Manipulations

Quantification of DNA

An Agilent 8453 UV-visible spectrophotometer was used according to the manufacturer's recommendation. DNA samples were diluted into distilled water. DNA concentrations were calculated relative to a net absorbance of 1.0 (A260nm-A320nm) = 50 µg/mL double-stranded DNA.

Gel Electrophoresis

DNA samples were run in gels using Tris-Borate-EDTA buffer (see Appendix B). Typically, 5.0 µL of PCR or enzyme digestion products were mixed with 2 µL of 1x loading buffer and 1µL of Syber Green dye (1/1000 dilution) and loaded onto an 0.7% (w/v) agarose gel. 5 µL of 1 kb NEB DNA ladder (containing 1 µL, 100 µg of ladder

stock, 2 µL of 10 mM Tris-HCl, 0.1 mM EDTA pH 7.5 buffer, and 2 µL of gel loading dye) was loaded as a standard, and stained with 1 µL of added Syber Green. The gel apparatus was supplied with a constant voltage of 95 volts for 50 minutes. After electrophoresis, gels were scanned and recorded using either a Bio-Rad FX® scanner or Gel Doc imager®.

POLYMERASE CHAIN REACTION (PCR)

Master mixes were prepared based on 1x PCR reactions. A typical 1x reaction contained: 35 µL nuclease-free water, 10 µL HF Phusion DNA Polymerase Buffer (NEB), 200 µM deoxynucleoside triphosphates (dNTPs), 0.2 µM of both a forward and reverse primer, double stranded DNA template, and 0.5 µL HF Phusion DNA polymerase (2U/µL) (NEB) all to a final volume of 50 µL. One (1.0) ng of plasmid DNA or 100 – 200 ng of genomic DNA was used as template DNA for PCR. Reactions were typically run in an Eppendorf gradient thermalcycler, with the cycle profile: 98°C for 2 minutes initial denaturation, followed by 30 cycles of 98°C for 30 seconds, 30 seconds at optimal annealing temperature for the primer pair (see Table 2), 72°C extension (based on estimated DNA synthesis rate of 40 seconds per 1000 bp), and a final extension of 10 minutes at 72°C.

Table 2. PCR primers used in this study.

Primers	Nucleotide sequence (5'-3' direction)	Tm	Amplicon
NISFDup-R	CAGGCGTCACTGTAAGCTG	60ºC	Flanking regions for homologous recombination on plasmid pOSH1409
KpnI-T7-F	GGATCCGGCTGCTAACAAAG	60ºC	T7 terminator region and the kanamycin resistance cassette on pOSH1409
pOSH1301-NISFD-R	CTTTGTTAGCAGCCGGATCC//ctctggacatctccctcaag	62ºC	DXS, IspF, and IspH on pOSH1301
pOSH1301-DXS-F	GAATTCAGGAGCTAGAACTGG	62ºC	DXS on pOSH1301
cI857-NISFD-F	CAGCTTACAGTGACGCCTG//tgcaggtgatgattatcagc	60ºC	cI857 on pGH276
cI857-R-1301	CCAGTTCTAGCTCCTGAATTC//cggatctttagctgtcttgg	60ºC	cI857 on pGH276
eyfp-F	ATGGTGAGCAAGGGCGAGGA	62ºC	YFP on pAQ-YFP
YFP-NISFD-R	CTTTGTTAGCAGCCGGATCC	62ºC	YFP on pAQ-YFP
cI857-R	TCCTCGCCCTTGCTCACCAT//cggatctttagctgtcttgg	60ºC	cI857 on pGH276
pUC57-NISFDup-F	GTACTGAGAGTGCAC//CAGTCGAAAGTCAATATTCG	60ºC	Flank Up to Dn on
pUC57-NISFDdn-R	GATTACGCCA//CAAGAAGTTAGAGGTGGAC	60ºC	Flank Up to Dn on pOSH1602 & pOSH 1605
cI857cro-DXS-F	ccaagacagctaaagatccg//GACCTGTTAAGTATTCAAGATCC	60ºC	DXS w/ Cro on pOSH1605-A
cI857cro-R	cggatctttagctgtcttgg	60ºC	DXS w/ Cro on pOSH1605-A
cI857-RBS-DXS-F	gtactaaggaggttgt//ATGGACCTGTTAAGTATTCAAGATC	60ºC	DXS w/o Cro on pOSH1605-B
cI857-RBS-R	acaacctccttagtacatgc	60ºC	DXS w/o Cro on pOSH1605-B

The melting temperatures, T_m, represent estimated temperatures at which a primer-template duplex is at the halfway point of dissociation. Annealing temperatures for PCR reactions were set near the lowest T_m of a primer pair. The nucleotide sequences after the

double slash in lowercase represent primer overhangs to allow annealing of overlapping complementary regions of component DNA fragments in Gibson Assembly reactions.

Plasmid Extraction

The plasmid extraction process (usually Promega Wizard SV reagents) proceeded as directed by the manufacturer except in the final step, where samples were eluted into 50 µL of 10T/0.1E to improve DNA concentration and long-term stability. Additional, minor, variance occurred when supplies of Promega silica-membrane minicolumns and column wash solution were exhausted and replaced with compatible EconoSpin columns and DNA wash buffer made from manufacturer documentation (see Appendix B). Plasmids used this in this study are listed in Table 3.

Genomic DNA Extraction

Genomic DNA was extracted from cells via a fast, ethanol-based method developed by Matthew Nelson. A small volume of cells (~5mL of stationary-phase culture) were pelleted firmly (e.g. 5 minutes at max speed), with the resulting supernatant discarded. Pelleted cells were suspended in 200µL of 35% ethanol and transferred, if necessary, to a 1.5mL microcentrifuge tube. This cell suspension was incubated in a 65+ºC water bath until the cells lysed (~10 minutes). The lysed suspension was spun down at max speed for 5 minutes. The supernatant (cell lysate containing free genomic DNA) was pipetted into a clean, sterile tube, and stored at -20ºC.

Table 3. Key plasmids used and constructed in this study

Plasmid	Characteristics	Source
pAQ1-Ex-YFP	Carries *YFP* gene, with resistance to Sm/Sp. Targeted to the pAQ1 plasmid in *Synechococcus* sp. PCC 7002.	Xu & Bryant (*55*)
pOSH1310	Carries *IspS-IDI*, with resistance to Sm/Sp. Targeted to the *PetJ2* region of the chromosome in *Synechococcus* sp. PCC 7002.	Ola Aremu, UWO (*45*)
pOSH1511-A	Carries *bPinS* genes, with resistance to Km. Targeted to the neutral site NISFD in *Synechococcus* sp. PCC 7002.	Rhiannon Carr, UWO
pOSH1602-A	Carries temperature regulated *YFP*, with resistance to Km. Targeted to the neutral site NISFD in *Synechococcus* sp. PCC 7002.	This study
pOSH1602-B	Carries temperature regulated *DXS, IspF* and *IspH*, with resistance to Km. Targeted to the neutral site NISFD in *Synechococcus* sp. PCC 7002.	This study
pOSH1605-A	Carries temperature regulated *DXS*-cro fusion, with resistance to Km. Targeted to the neutral site NISFD in *Synechococcus* sp. PCC 7002.	This study
pOSH1605-B	Carries temperature regulated *DXS*, with resistance to Km. Targeted to the neutral site NISFD in *Synechococcus* sp. PCC 7002.	This study

NUCLEIC ACID QUANTIFICATION

As when measuring culture density, to quantity DNA or RNA, an Agilent 8453 UV-visible spectrophotometer was employed against a ddH_2O blank. Samples were diluted into ddH_2O by a factor of 20 (or 30 in the case of very concentrated samples), and the net absorbance at 260nm ($A_{260nm} - A_{320nm}$) was measured. DNA concentration in

ng/μL was calculated according to the formula: ($A_{260} - A_{320}$) of 1.0 = 50 ng/μL for double-stranded DNA and 33 ng/μL for single-stranded RNA.

DNA Fragment Purification

DNA products produced by PCR were purified using one of two methods: the Promega Wizard SV Gel & PCR Clean-Up System or the Zymo Clean & Concentrator-5 reagent kits. Both column-based kits can purify DNA fragments directly from a reaction tube, but the Promega kit is also designed for purification of DNA fragments from gel slices, in cases where PCR amplification produces multiple products. For the purification of cleanly-amplified fragments, however, the Zymo kit was preferred for both the rapidity of its protocol and for the smaller elution volume, which allowed for more-concentrated DNA samples. In general, fragments purified with the Promega kit were eluted in 50 μL 10T/.1E, and those purified with the Zymo kit were eluted in 20 μL 10T/.1E buffer.

Assembly of Recombinant Plasmids

To generate recombinant plasmids carrying desired genetic constructs, the component fragments were first amplified from genomic DNA or plasmid DNA templates by PCR. To generate DNA constructs for targeting genes to *Synechcoccus* chromosomal sites, the construct is composed of an upstream chromosomal gene fragment, the cI857-pR regulator-promoter, the gene that the cI857-pR regulator controls, an antibiotic resistance gene cassette, and a downstream chromosomal gene fragment. To facilitate replacement of wild type copies of gene(s) of interest by homologous

recombination, flanking DNA fragments of ~500 – 800 bp in size with homology to the upstream and downstream regions of gene of interest were used to target linear DNA constructs to the neutral insertion site between SYNPCC7002_RS04670 and SYNPCC7002_RS04675 genes as described by Fiona Davies (*51*).

The recombinant plasmids constructed in this study used a segment of plasmid pOSH1511-A as a plasmid backbone fragment. This was initially amplified from plasmid pOSH1108 (see appendices for illustrations of all plasmids used). The pOSH1511-A DNA fragment carries both an origin of replication and an ampicillin resistance gene cassette (Ap^R), and a kanamycin resistance gene cassette (Km^R). The cI857-pR repressor-promoter was amplified from plasmid pHG276, while the MEP pathway genes (*DXS, IspF* and *IspH*) were amplified from plasmid pOSH1301.

The amplification of DNA fragments by PCR was achieved with primers shown in Table 2. Lastly, an in-vitro, Gibson Assembly technique (*56*), was used to assemble the component DNA fragments into a recombinant circularized plasmid. The assembled DNA (recombinant plasmid) was used as a template for PCR to generate the linear fragment used for introduction into and genetic transformation of *Synechococcus*.

The Gibson Assembly reaction allows for assembly of multiple overlapping DNA fragments in-vitro. The Gibson Assembly Master Mix contains 1) An exonuclease that chews back 5’ ends to create exposed 3’ single-stranded overhangs that anneal to fragments that share complementary overlap regions, 2) A DNA polymerase that fills gaps within each annealed fragment, and 3) A DNA ligase that seals the nicks in the assembled DNA (*56*).

The concentration of each DNA fragment used in the Gibson assembly reaction was calculated as recommended by NEB. Thus: concentration (pmol) = [weight (ng) x 1000] / [fragment size (bp) x 650 Daltons]. Each DNA component (at 0.1 - 0.2 pmol) was added to 10 μL Gibson assembly mix followed by addition of nuclease-free water for a final volume of 20 μL. The reaction mixture was incubated at 50°C for 30 to 60 minutes (depending on the number of DNA fragments in the reaction) followed by inactivation at 98°C for 2 minutes.

Restriction Enzyme Digestion

To ensure that there was no carry-over of parent plasmid DNA after PCR and before Gibson assembly, the PCR constructs were digested using restriction enzymes that cut only at unique restriction sites. The entire PCR reaction was digested using 1μL (20 units) of *DpnI* restriction enzyme [NEB]. The digestion reactions were incubated at 37°C for 2 hours, followed by heat inactivation at 65°C for 20 minutes. Successful digestion was confirmed by agarose gel electrophoresis.

Along with PCR confirmation, plasmids were digested with *Nde*I restriction enzyme [NEB] or *Pst*I restriction enzyme [NEB] as a secondary confirmation of correct assembly. The digestion reactions were incubated at 37°C for 2 hours, followed by heat inactivation at 65°C for 20 minutes. Successful digestion was confirmed by agarose gel electrophoresis. If the observed banding pattern correlated with the expected pattern, the plasmid was used as the template for transformation of *Synechococcus.*

TRANSFORMATION PROCEDURES

ESCHERICHIA COLI

Electrocompetent cells were prepared as described in the Appendix. For transformation of recombinant DNA constructs prepared by Gibson Assembly, 2 μL of Gibson Assembly reaction product containing a desired recombinant plasmid construct (~20 ng of plasmid DNA) was added to 40 μL electrocompetent *E. coli* DH5α cells. The tube containing the mixture was incubated on ice for 1 minute. The cell suspension was transferred into a cold 0.2 cm gap electroporation cuvette and placed into a pre-chilled electroporation chamber in a BioRad Gene Pulser device set to 25 μF constant capacitance and 200 Ω resistance, and pulsed once at 2 KV (~4.5 ms). Then, 0.5 mL S.O.C. medium (see Appendix A) was immediately added to the cuvette as a wash and the cell suspension transferred into a sterile 15 mL culture tube. An additional 0.5 mL SOC was used to rinse the cuvette. The cell suspension was incubated at 37°C with shaking (~100 rpm) for 1 hour. Then, 100 μL of the cell suspension was spread onto LB plates containing antibiotic (as determined by the antibiotic resistance gene cassette) at appropriate concentrations to select for recombinant plasmids. Transformation plates were allowed to dry under sterile conditions after inoculation with bacteria and then incubated overnight at 37°C.

SYNECHOCOCCUS SP. PCC 7002

Being naturally competent to exogenous DNA, *Synechococcus* PCC 7002 need not be prepared specifically for transformation. However, to promote high efficiency of

transformation, the culture should be in the exponential phase of growth. To this end, on the day prior to the addition of DNA, a 2 mL culture of cells was diluted from the stock to an OD_{750nm} of 0.2 in A medium (with antibiotics when appropriate). Such cultures were grown at 39°C with shaking, light intensity of ~200μmol photons m^{-2} s^{-1}, under ~3% CO_2 (these conditions are hereafter designated as those in the "high-CO_2 incubator") for 16-24 hours to until cells had grown to an OD_{750nm} of 1 to 2. Following this growth, ~45 μL of linear DNA amplified by PCR was added directly to the culture; cultures were allowed to rest at a slant under lighted conditions for 30-60 minutes prior to being returned to the high-CO_2 incubator for 16-24 additional hours of growth with shaking.

To select for transformants, cultures incubated with DNA were plated onto A(D7) agar medium (1.2% Difco® Bacto® agar) with appropriate antibiotic selection. These plates were grown in the high-CO_2 incubator with shading (<150 μmol photons m^{-2} s^{-1}) until colonies were observed (5-9 days), then for 3-7 more days without shading. Colonies were randomly selected and patched onto fresh A(D7) + plates, with growth in the high-CO_2 incubator for an additional 3-7 days.

VERIFICATION OF TRANSGENE UPTAKE IN *SYNECHOCOCCUS* SP.PCC 7002

A sterile loop was used to transfer a sample of cells from individual patches to 200 μL of a 35% ethanol solution, from which genomic DNA was extracted by the method described above. This gDNA extract was used as a template for PCR, which amplified the target region using the forward primer from the upstream flanking region and the reverse primer from the downstream flanking region. As a control, wild-type

gDNA was also amplified from the same primers. Uptake of the transformation construct was verified by running gel electrophoresis on the PCR products, with transformants displaying an amplicon consistent in size with the linear fragment used for transformation, versus the wild-type fragment. Segregated transformants, which have integrated the construct into all copies of the transformation region, are indicated by the exclusive presence of construct-length amplicons.

CONSTRUCTION OF *SYNECHOCOCCUS* SP. PCC 7002 MUTANT STRAINS

Synechococcus 7002 mutant strains were constructed starting from a wild type host strain. These strains were constructed by a genetic engineering strategy that targeted *YFP* or optimized MEP pathway genes, under control of the cI857-pR repressor-promoter, to the NISFD neutral site. All of these DNA constructs, abbreviated *YFP* (720 bp in size), *DXS* (1900 bp in size), and *DXS-IspF-IspH* (3580 bp in size), were flanked by the NISFD neutral site upstream and downstream regions and carried a kanamycin-resistant gene cassette for selection. Directly upstream of these target genes was the cI857-pR repressor-promoter. In this study, I created three strains that carry optimized MEP genes. 1) A *DXS-Cro* fusion. This strain has the first 22 codons of the *Cro* gene fused to DXS, which lacks an ATG start codon. 2) *DXS* alone and 3) A *DXS-IspF-IspH Cro* hybrid. This strain has the first 22 codons of the *Cro* gene, and following a 100 base pair gap, the then the beginning of *DXS*, with its own start codon (Figure 6).

For each of the DNA constructs, the entire region including the flanking sequences, cI857-pR repressor-promoter, regulated gene(s), and Km-selection tag, was

amplified by PCR using the pUC57-NISFDupF/pUC57-NISFDdn-R primer pair (Table 2). Purified DNA constructs generated from PCR were used to transform wild type *Synechococcus* 7002 with targeting to the NISFD neutral size in the chromosome.

STRAIN TESTING

GROWTH EXPERIMENTS

To test whether transgenic *Synechococcus* strains showed long lag phases or other negative metabolic effects, growth experiments were conducted for each strain. After genetic confirmation of transgene integration and segregation by PCR, sterile A (D7) medium, e.g. 200 mL, containing appropriate antibiotics was inoculated with a fresh overnight stock culture to 0.01 OD_{750nm}. Cultures were incubated at with 3% CO_2 under a light intensity of ~ 500 µmol photons m^{-2} s^{-2} in a Phenometrics Environmental Photobioreactor (ePBR) until the culture reached an optical density (OD_{750nm}) of ~2.0. Culture vessels were maintained at a constant temperature of either 40°C or 25°C with constant stirring using a magnetic stir bar. The ePBR automatically recorded optical density (at 940nm) every minute until the conclusion of the experiment. OD units from the ePBR at 940 nm were converted to OD_{750nm} values per 1.0 cm light path.

TURBIDOSTAT CULTURE CONDITIONS

For turbidostat cultures, an ePBR was modified by connecting two tubes. One for fresh medium inflow to the culture vessel and another from the culture vessel to a waste container. These tubes were threaded through peristaltic pump that controlled flow. The culture vessel contained 200 mL of either Minimal Salts or A(D7) medium (with

appropriate antibiotics) depending on whether *E. coli* or *Synechococcus* was being grown. For *Synechococcus* cultures, 3 mL of an overnight, room temperature inoculum was added to the culture vessel, for growth under a light intensity of ~500 μmol photons m^{-2} s^{-2}, and constant stirring with a magnetic stir bar.

When testing *E. coli*, the culture was bubbled with ambient air, kept at a constant turbidity of 0.4 (O.D. 940) and the temperature was increased every six hours from 24°C to 40°C. Before each temperature increase, 1 mL of the culture was removed and stored at 20°C. Once all the samples were collected, they were imaged for YFP fluorescence with a fluorescence emission spectrometer. When testing *Synechococcus* strains, the culture was bubbled with 3% CO_2 in air, pH was maintained at 7.5 by regulated CO_2 inflow, and the culture was kept at a constant turbidity of 0.2 ePBR units (OD_{940nm}). The vessel was exposed to irradiation of ~500 μmol photons m^{-2} s^{-2}. The culture started at a temperature of 28°C and was increased two degrees every twelve hours. As before, 1 mL of the culture was removed and stored at 20°C before each temperature increase.

Fluorescent Measurements

Fluorescence was measured in one of three ways. First, for *E. coli* on agar plates, a simple UV hand lamp was used to determine fluorescence. Second, for *Synechococcus* 7002 on agar plates, a BioRad Molecular Imager® FX Pro Plus with external laser was used to image the plates (with 488 nm excitation and emission centered at 530 nm +/- 10nm). This allowed YFP fluorescence to be clearly distinguished form phycobiliprotein and chlorophyll fluorescence emissions in the cyanobacteria. Third, to determine YFP

expression levels within both *E.coli* and *Synechococcus* as a function of temperature, a Photon Technology International (PTI) fluorescence emission spectrophotometer was used. To a 1.0 cm quartz cuvette, 2.0 mL of cell sample (*E. coli* or *Synechococcus*) at 1.0 OD_{750nm} was added. The excitation wavelength was 475 nm, emission was scanned from 500-550 nm. The program FelixX32 was used to capture and analyze these data.

PRIMERS AND PROBES FOR QUANTITATIVE PCR (QPCR)

A TaqMan method was used for qPCR (*57*). Primer and probe designs were based on the following parameters; for primers: annealing temperature, T_m = 57- 61°C with difference in primer pair T_m not exceeding 2° C, a GC content of 45-60% with a G or C nucleotide at the 3' position of the primers and a primer length of about 15-22 bp. For probes; Tm = 67-70°C, a maximum T_m difference of 5-6°C between primer and probe, GC content of 45-60%, probe length 23-27 bp. For probes, guanine was avoided at the 5' end. The primers and probe (shown in Table 4) were designed to assess *YFP* gene copy number in both the control strain where YFP is constitutively expressed and the temperature dependent YFP strain. The probe was designed to carry a fluorescent 5-carboxyfluorescein dye (5-FAM) at the 5' end and a Black-hole quencher (BHQ-1) dye at the 3' end.

Table 4. Probe and primers designed for qPCR

Gene	Type	Sequence
YFP	YFP-Taq Probe	[6~FAM]-ATCGACTTCAAGGAGGACGGCAAC-[BHQa~6FAM]
	YFP-F Primer	GAACCGCATCGAGCTGAA
	YFP-R Primer	GCTTGTCGGCCATGATATAG

PRIMER-PROBE VALIDATION AND EFFICIENCY DETERMINATION

Normalization of results obtained from the described primer-probe set requires the validation of such primers and probes for qPCR experiments. Reaction volumes were scaled down to 20μL. Plasmid pAQ1EX-YFP (*55*) template DNA was serially diluted 10 fold over 5 $\log_{10}$ values (from 10 ng/μL to 0.1 pg/μL). Thus, for the primer-probe validation, five reaction tubes were set up. The reaction master mix was set up thus: 10 μL of 2x Bull's Eye Taqman PCR master mix, 1 μL of both forward and reverse primer (final concentration 1 μM), 0.25 μL of corresponding probe (final concentration 200 nM), and nuclease free water for a final volume of 20 μL. Reactions were run in an Applied Biosystems (ABI Step OneTM) real-time PCR thermalcycler using a profile of 95°C for 10 minutes for DNA polymerase activation, followed by denaturation of double-stranded DNA at 95°C for 15 seconds, and primer annealing and extension at 60°C for 1 minute, repeated for 40 cycles.

The validation reactions were set up using the ABI 'Quantitation Standard Curve' program with threshold cycle values automatically set to 0.05 by the software. The cycle

to threshold (C_T) values obtained for each primer-probe dilution were plotted as a function of the $\log_{10}$ concentration of the starting template. The efficiency of the primer-probe set was determined from the slope of the line on this graph. The PCR Amplification Efficiency is determined by (E – 1) x 100 % where E, the Amplification Efficiency = $10^{-1/slope}$ (ref----). 100% amplification efficiency gives a slope of -3.33. This is because $\log_2 = 3.33_{\log 10}$, or $2^{-3.33} = 0.1$, thus 3.33 additional cycles are needed to amplify 1/10th the starting material to the same threshold level. A PCR efficiency of 92% or greater was considered acceptable for direct comparisons.

qPCR Gene Copy Number Studies

The reaction components and cycling profile for these experiments were similar to those described above for primer-probe efficiency tests. However, the method of analysis used was comparative C_T (or ΔC_T). The means of cycle to threshold (C_T) values of three technical replicates for each sample (or biological replicate) were used as the cycle to threshold value for each biological replicate. DNA templates for qPCR were provided by the cells, where 1 uL of cells at 0.05 OD_{750nm} were added to the reaction mix. The threshold values were manually adjusted to 0.5 for all samples.

Isoprene Capture and Quantification by Gas Chromatography

Once segregated transgenic cultures of *Synechococcus* PCC 7002 were obtained, it was necessary to ascertain the functionality of the novel enzymes by assessing isoprene production. To culture the transgenic strains, 50 μM of sodium bicarbonate were added to 250 mL Boston bottles containing 200 mL of culture in A (D7) medium, with the

appropriate antibiotics. These were incubated with stirring in an aquarium held at 38°C or 20°C. Cultures were diluted to an optical density (OD_{750nm}) of 0.03 at the start of these experiments and the bottles sealed so that no gas escaped. Cultures were grown for a 24-48 hours, to an approximate optical density of 0.5 (OD_{750nm}). After this period, 1 mL of the headspace gas was tested by gas chromatography-mass spectroscopy (GC-FID/PID for the presence of isoprene.

Gas chromatography was performed with a Hewlett Packard (Agilent) 5890 Series II Gas Chromatograph equipped with flame ionization (FID) and photoionization (PID) detectors and a Restek PLOT column. The oven temperature was 150°C, the isoprene retention time was 9 minutes, with an 11 minute run time, and the column head pressure was set to 12 PSI. Quantification of isoprene peaks was calculated from the program Instrument 1. Isoprene concentrations were determined relative to a three-point standard curve with points at 1300ppm, 130ppm and 6.5ppm. Standards were prepared by adding 2.5ul of liquid isoprene to a clean 250 mL Boston bottle (Bottle 1, 1300ppm). A syringe was into Bottle 1 and carefully plunged up and down 6-8 times to mix the contents of the bottle. Without removing the syringe 12.5 mL was extracted from Bottle 1 and administered to a second bottle (Bottle 2, 130ppm). For the low end standard 12.5mL of gas was extracted from Bottle 2 as described and add to Bottle 3 (6.5ppm).

CHAPTER III: RESULTS

CHARACTERIZATION OF THE CI857-PR REPRESSOR-PROMOTER

CHARACTERIZATION IN *E. COLI*

To create the *NISFD-cI857-YFP* plasmid (pOSH1602-A), three separate DNA fragments were amplified by PCR to create the final product. The *YFP* gene was amplified from the pAQ-Ex1-YFP plasmid using the eyfp-F and YFP-NISFD-R primers. The temperature sensitive cI857-pR repressor-promoter and *Cro* region were amplified from the plasmid pHF276 using the primers cI857-NISFD-F and cI857-R. The final piece, which contains the NISFD, neutral-site targeting regions for integration into the *Synechococcus* PCC 7002 chromosome, and the antibiotic resistance cassette, was amplified from the plasmid pOSH1511-A using the primers NISFDup-R and KpnI-T7-F (Figure 7-A). These fragments were used in a Gibson Assembly reaction to generate the plasmid pOSH1602-A (Figure 7-B).

E. coli transformed with plasmid pOSH1602-A and grown on separate agar plates at two different temperatures, room temperature (~21°C) and 39°C, displayed temperature-sensitive YFP fluorescence upon excitation under a UV lamp (Figure 8-A). Closer inspection of these plates with a fluorescence scanner (488 nm excitation, 530 nm +/ 10 nm emission filter) revealed specific YFP fluorescence from E. coli carrying the pOSH1602-A (cI857-pR-YFP) plasmid when grown at 40°C, with no detectable fluorescence when grown at 28°C. (Figure 8-B).

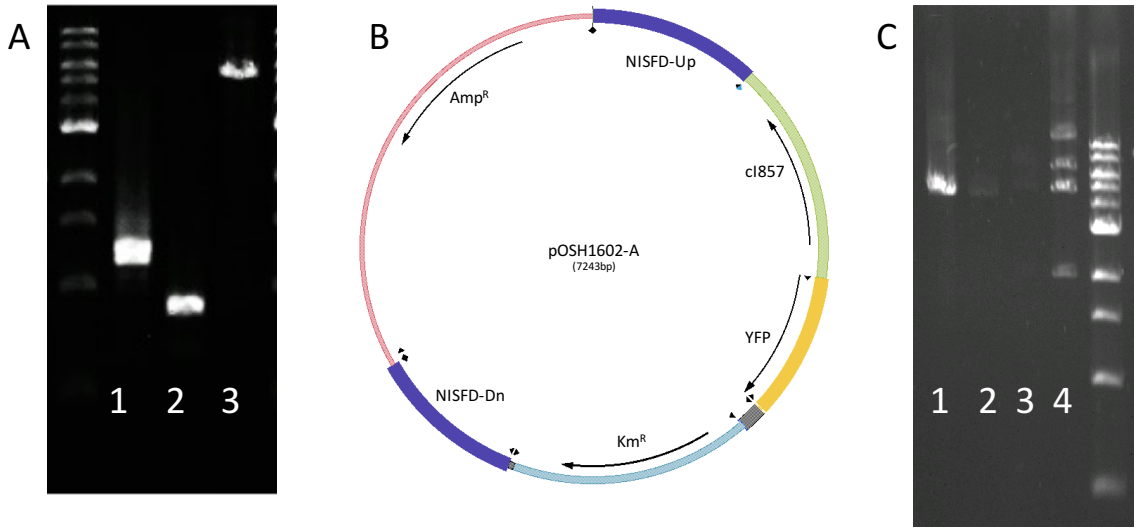

Figure 7. Plasmid pOSH1602-A, cI857-pR-YFP-NISFD construction. A. PCR amplification of the *cI857-pR* repressor-promoter (lane one, 1100 bp), *YFP* (lane 2, 700 bp), and the *NISFD* neutral-site plasmid backbone (lane 3, 5400bp) DNA fragments, respectively. **B**. Plasmid pOSH1602-A for targeting the *YFP* gene, under control of the temperature-sensitive *cI857-pR* repressor-promoter, to neutral-site NISFD in the *Synechococcus* 7002 chromosome. **C.** Verification of pOSH1602-A plasmid assembly: PCR amplification of a region containing the *YFP* gene fused to the *cI857-pR* repressor-promoter from NISFD-Up to NISFD-Dn (Lane 1 ~4800bp).

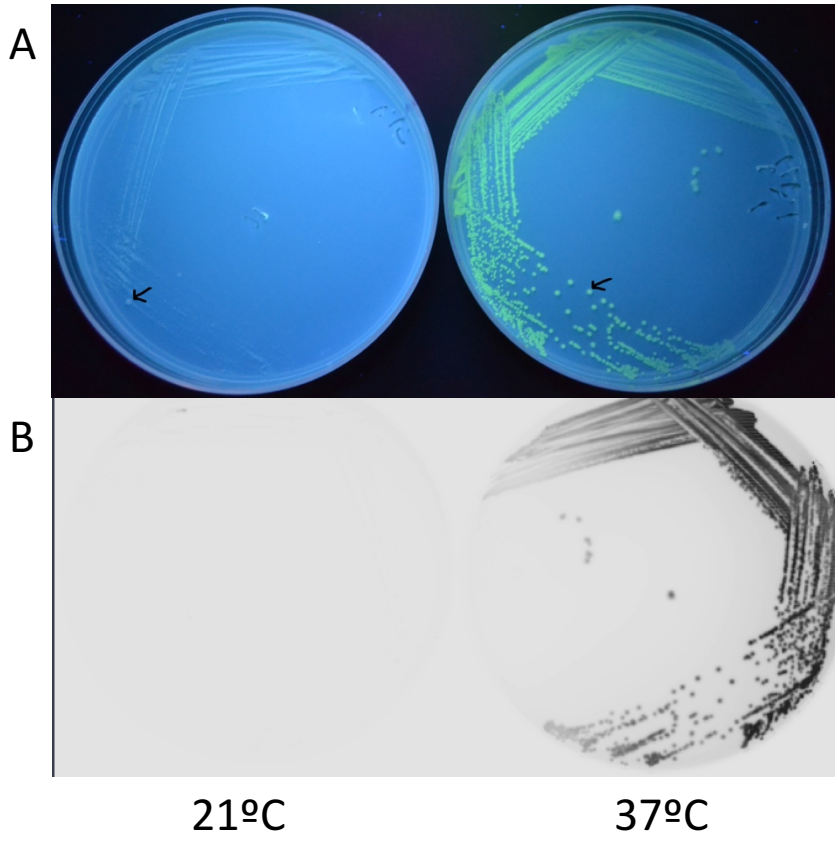

Figure 8. Temperature-dependent YFP fluorescence in *E. coli* carrying plasmid pOSH1602-A (*cI857-pR-YFP*). A. UV light excitation. *E. coli* (pOSH1602-A) grown on LB/Km agar at 21°C (left) and 37°C (right). At 25°C, the cI857 repressor binds to the pR promoter region and prevents *YFP* transcription. At 37°C, the repressor has denatured, allowing *YFP* gene expression and YFP protein synthesis. Arrows indicate colonies of approximately the same size. **B.** Fluorescence scans (excitation 475 nm, emission 530 nm +/- 10 nm). *E. coli* (pOSH1602-A) grown on LB/Km agar at 21°C (left) and 37°C (right), as above. Specific YFP fluorescence is clearly evident at 39°C but undetectable at 21°C.

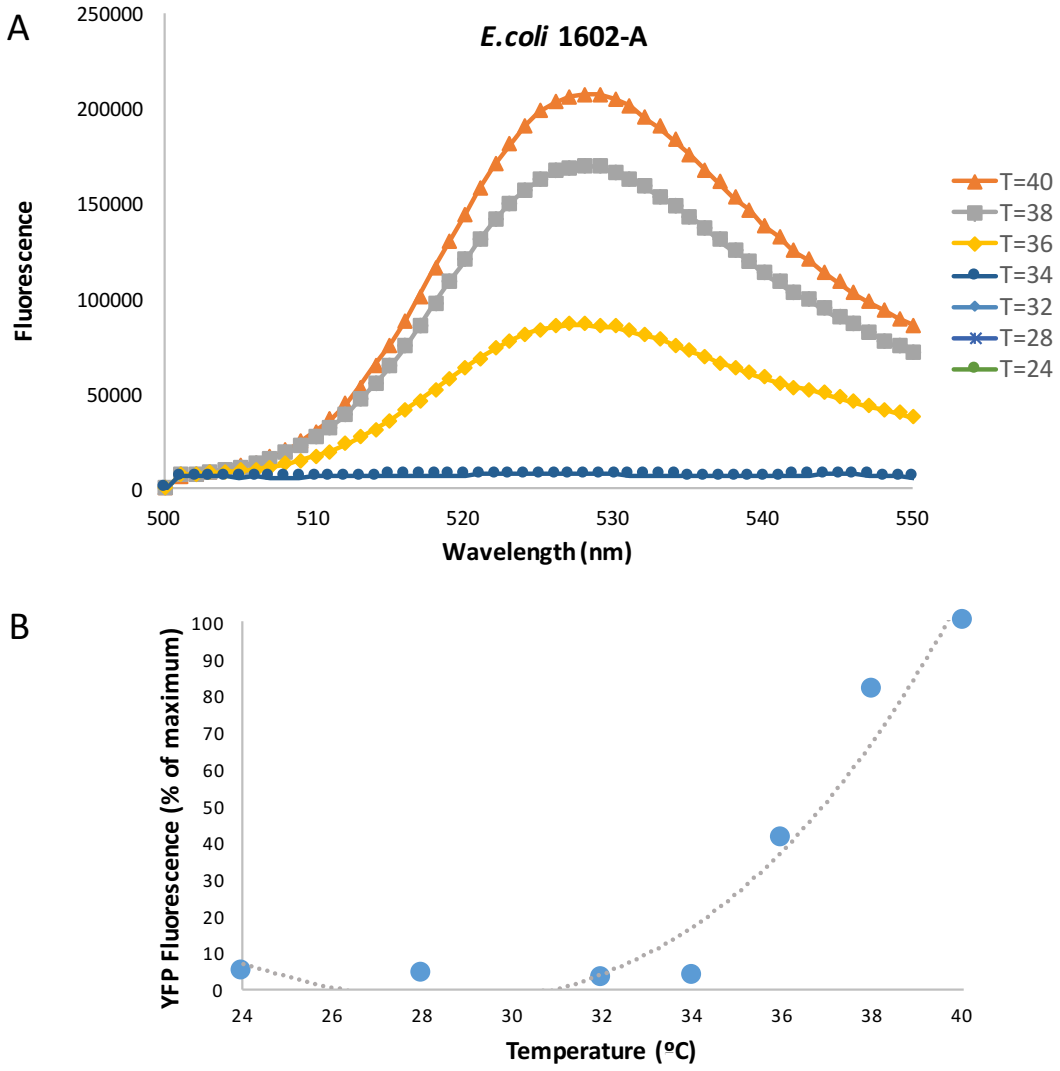

Figure 9. YFP Expression in *E. coli* (pOSH1602-A, *cI857-pR-YFP*) as a function of temperature. A. YFP fluorescence emission scans (excitation 475 nm). Between 24°C and 34°C, YFP expression remained essentially undetectable. When the temperature was increased past 34°C, YFP expression increased dramatically up to a temperature of at least 40°C. **B.** Data from panel A, plotted as percentage of the maximum YFP signal at 40°C as a function of growth temperature. Data in both panels were collected from a turbidostat culture of *E. coli* (pOSH1602-A) maintained at a constant cell density of 1.0 ePBR units (~0.4 OD_{750nm} cm^{-1})

To determine the dynamic range and response of the *cI857-pR* regulator, whether steep or gradual, transgenic *E. coli* (pOSH1602-A, *cI857-pR-YFP*) was tested over a series of temperatures at a constant cell density maintained in a turbidostat culture (Figure 9-A). As Figure 9 illustrates, the cI857 repressor kept the *YFP* gene turned off until the culture temperature reached ~34°C. YFP fluorescence increased sharply after 34°C and kept increasing until at least 40°C (Figure 9-B).

CHARACTERIZATION IN *SYNECHOCOCCUS* SP. PCC 7002

Once plasmid pOSH1602-A (*cI857-pR-YFP*) was assembled, confirmed, and tested in *E. coli*, it was introduced into *Synechococcus* sp. PCC 7002. Primers pUC57-NISFDup-F and pUC57-NISFDdn-R were used to amplify the region between NISFD-Up and NISFD-Dn (Figure 7 B) to generate a linear DNA fragment carrying the *cI857-pR-YFP* gene. This was introduced into *Synechococcus* by transformation with integration into the NISFD neutral site as shown in Figure 10.

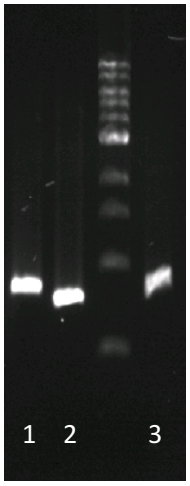

Figure 10. Verification of *cI857-pR-YFP* integration into the *Synechococcus* 7002 chromosome. Lanes 1-3: The *YFP* gene (~700 bp) was amplified from several separate sources of genomic DNA extracted from *Synechococcus* 7002 (1602-A, *cI857-pR-YFP*) transformants. Size standards (NEB 1 kb ladder) are shown between lanes 2 and 3.

Synechococcus 7002 (1602-A) transformed with the *cI857-pR-YFP* genes from plasmid pOSH1602-A were grown on separate A(D7)-Sm/Sp agar plates at room temperature (~21°C) and 39°C. Fluorescence emission scans of these plates showed strong YFP fluorescence from the plate incubated at 39°C and no detectable fluorescence at 21°C (Figure 11). The insets in this Figure show comparable chlorophyll fluorescence from these plates and thus comparable cell density.

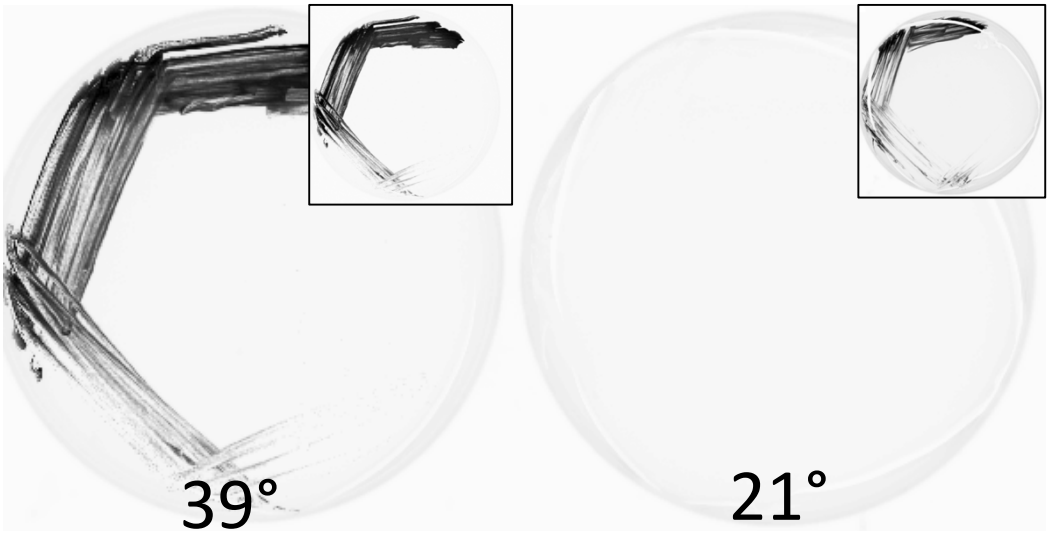

Figure 11. Temperature-dependent *YFP* gene expression in *Synechococcus* (1602-A). *Synechococcus* (1602-A, *cI857-pR-YFP)* grown on A(D7)-Sm/Sp agar medium at 21°C and 39°C. At 39°C, the cI857 repressor had denatured, allowing *YFP* gene expression and YFP protein synthesis. Excitation 488 nm, emission 530 nm +/- 10 nm. Inset: Chlorophyll fluorescence, excitation 614nm, emission 680 nm +/- 5 nm.

To determine the dynamic range and response of the cI857-pR repressor-promoter in *Synechococcus* (1602-A), the transgenic cyanobacteria were grown in a turbidostat culture to maintain constant cell density and tested over a series of temperatures (Figure 12-A). As this figure illustrates, the cI857 repressor kept the *YFP* gene largely turned off until the temperature reached ~32°C. YFP fluorescence increased gradually from 32°C to at least 38°C (Figure 12-B). This increase was somewhat more gradual than in *E. coli* (1602-A) that carry the same *cI857-pR-YFP* gene set (Figure 9B). In addition, *Synechococcus* (1602-A) showed a higher basal level of *YFP* expression, at ~15% of the maximum detected at 38°C, than did *E. coli* (1602-A). This suggests that the cI857 regulator may be somewhat 'leaky' in *Synechococcus* 7002, i.e. it may not completely repress transcription from the pR promoter.

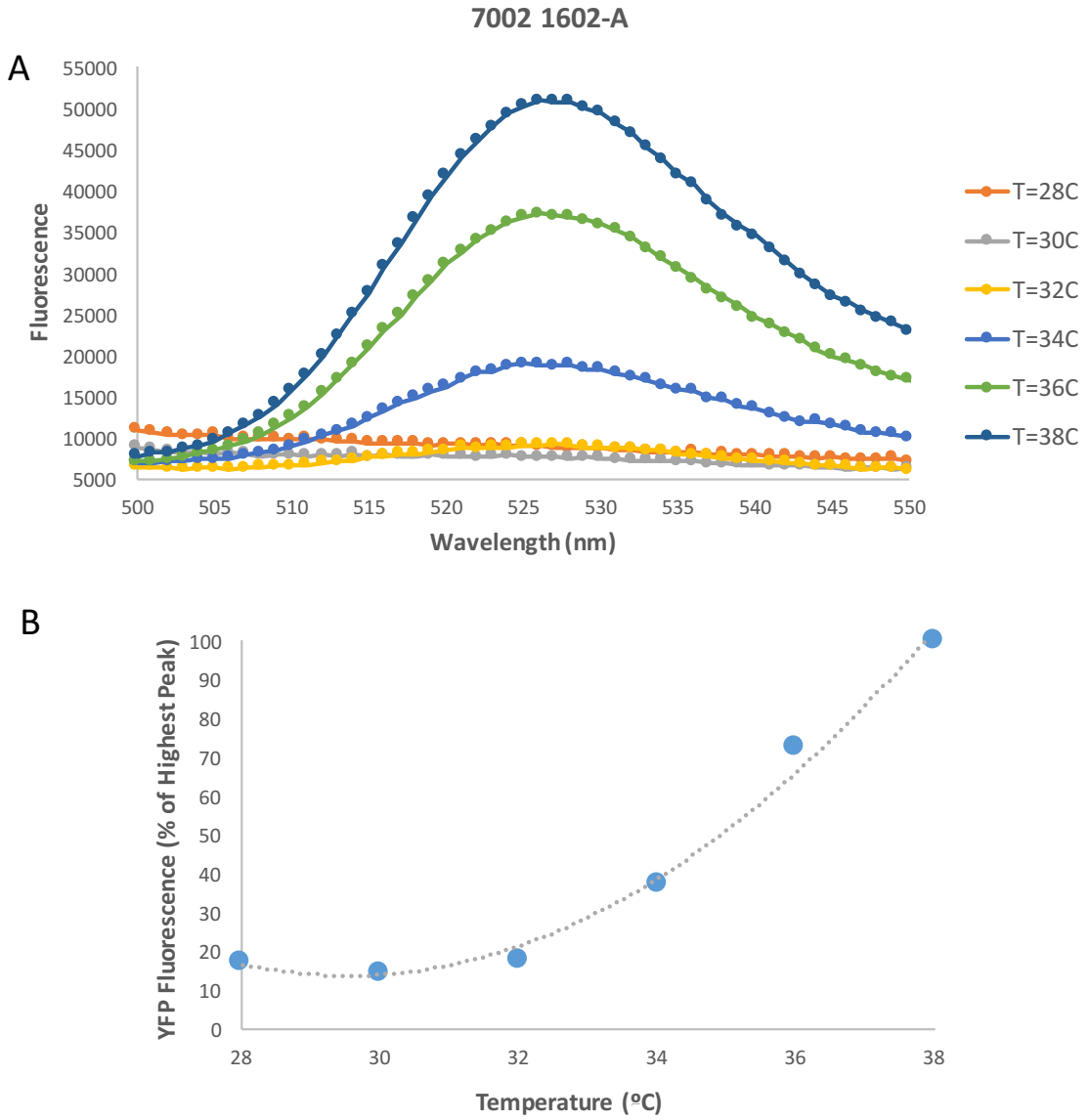

Figure 12. Temperature-regulated YFP expression in *Synechococcus* 7002 (1602-A, *cI857-pR-YFP*). A. Fluorescence emission spectra as a function of temperature (excitation 475 nm). At temperatures above 34°C, YFP expression increased linearly. **B.** Data from panel A, plotted as percentage of the maximum YFP signal at 38°C as a function of growth temperature. Data in both panels were collected from a turbidostat culture of *Synechococcus* (pOSH1602-A) maintained at a constant cell density of 1.0 ePBR units (~0.4 OD_{750nm} cm^{-1}).

Having demonstrated temperature-regulated control of gene expression via the cI857-pR repressor-promoter, a following question of interest was the strength of the induced *pR* promoter relative to the strong, *Synechocystis* PCC 6803 *PcpcB* promoter (*55*) that our group has previously used for transgene expression in *Synechococcus* 7002. Our previous attempts to increase isoprene production in *Synechococcus* 7002 through elevated expression of *DXS* and other MEP pathway genes involved placing these under control of the 6803 *PcpcB* promoter and other strong, constitutive promoters. These experiments failed, we believe because of these continuously expressed strong promoters leading to accumulation of toxic levels of the overexpressed enzymes or of the metabolites that they produce. As one test of this hypothesis, I examined *YFP* expression levels in *Synechococcus* 7002 from plasmid pAQEX1-YFP, where *YFP* is expressed from the 6803 *PcpcB* promoter, and in *Synechococcus* (1602-A), where *YFP* is expressed from the *cI857-pR* repressor-regulator from the chromosomal NISFD site. As illustrated in Figure 13, maximal *YFP* expression from the *PcpcB* promoter ('control YFP') is about 16-fold higher than from the induced *pR* promoter ('cI857-YFP'). However, it should be noted that based on qPCR studies, there are about four-times the *YFP* gene copies in *Synechococcus* (pAQEX1-YFP), which carries the plasmid-borne *PcpcB-YFP* construct, relative to *Synechococcus* (1602-A), which carries the *cI857-pR-YFP* construct (see Appendix D). Thus the *PcpcB* promoter may be only about 4-times stronger than the fully-induced *pR* promoter.

In terms of gene regulation, *YFP* expression from the *PcpcB* promoter increased only ~1.3 fold at 38°C relative to 28°C (Figure 13), which can be attributed to increased metabolic activity at the higher temperature. In contrast, *YFP* expression from the *cI857-pR* regulator increased by ~3-fold at 38°C relative to 28°C, which can be attributed in large part to the denaturation of the cI857 repressor at the higher temperature. Therefore, both because of lower overall promoter strength and transcriptional regulation, the *cI857-pR* regulator appears to be well suited for testing increased *DXS* and MEP pathway gene expression for elevated isoprene synthesis.

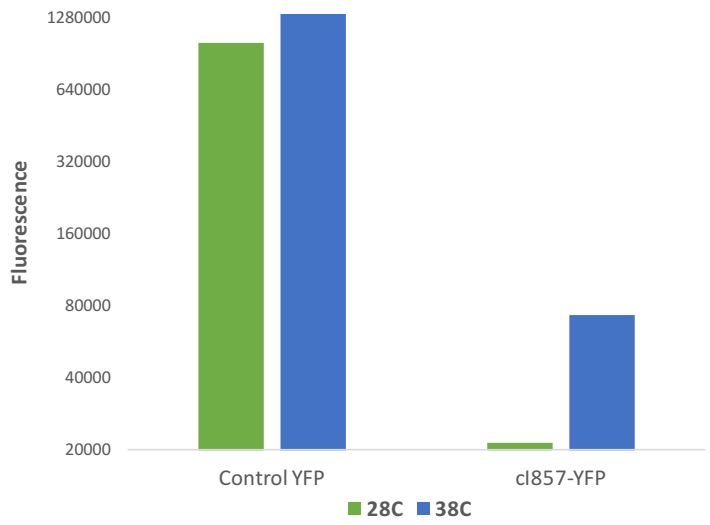

Figure 13. *YFP* expression from the *PcpcB* promoter and *cI857-pR* repressor-promoter in *Synechococcus* 7002. Comparison of *YFP* expression from the constitutive *PcpcB* promoter in *Synechococcus* (pAQEx1-YFP) relative to expression in *Synechococcus* (1602-A) from the *cI857-pR-YFP* genes in chromosomal-site NISFD. The *Synechococcus* strains were grown at a constant temperature of 28°C or 38°C until a predetermined optical density of 1.0 OD_{750nm} was reached and then measured for YFP fluorescence by fluorescence scanning (475 nm excitation, 530 +/- 10 nm emission).

CONSTRUCTION OF *SYNECHOCOCCUS* STRAINS WITH REGULATED MEP PATHWAY GENES FOR ENHANCED ISOPRENE PRODUCTION

To test the *cI857-pR* repressor-promoter for regulated *DXS* and MEP pathway gene expression for enhanced isoprene production, three separate transgenic strains were created (Figure 14-A). One strain (1605-A) carries a direct fusion of part of the bacteriophage lambda *Cro* gene to the *DXS* gene. In this strain, the first 22 codons of *Cro* are joined in-frame with an ATG-less *DXS* gene. To create this strain, primers cI857cro-DXS-F and cI857cro-R amplified the *DXS* gene and primers NISFDup-R and KpnI-T7-F amplified the backbone of the pOSH1602-A plasmid (Figure 14-B). The second strain (1605-B) is similar to the first, but the *Cro* gene is completely eliminated. The *DXS* gene retains its ATG start codon, and no fusion is made. To create this strain, primers cI857-RBS-DXS-F and cI857-RBS-R amplified the *DXS* gene, and primers NISFDup-R and KpnI-T7-F amplified the backbone of plasmid pOSH1602-A (Figure 14-C). The final transgenic strain (1602-B) is different: it contains added *DXS, IspF*, and *IspH* genes. This third strain includes the first 22 amino acids of *Cro*, but because of an error in the initial primer design, there is a non-coding region of ~100 bp between *Cro* and the start of *DXS* (Each with its own ATG start codon). To create this strain, primers pOSH1301-DXS-F and pOSH1301-NISFD-R amplified the *DXS, IspF* and *IspH* genes, primers cI857-NISFD-F and cI857-R-1301 amplified the cI857-pR region from plasmid pHG276, and primers NISFDup-R and KpnI-T7-F amplified the backbone of plasmid pOSH1511-A (Figure 14-D).

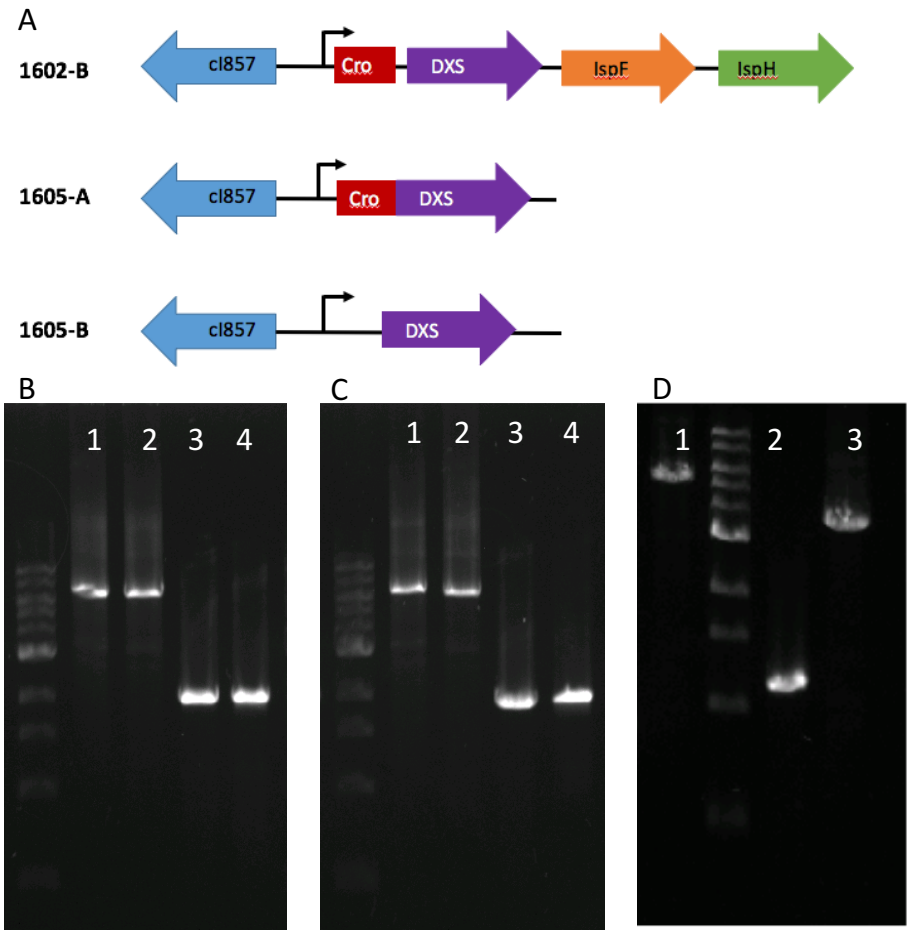

Figure 14. *Synechococcus* strains carrying temperature-regulated *DXS* and MEP pathway genes for isoprene production. **A.** All *DXS* gene sets are controlled by the cI857-pR repressor-promoter (promoter sites indicated by arrows) and all *Synechococcus* 7002 transformants carried *IspS-IDI* (1310) genes for isoprene synthesis. **B.** DNA segments for assembly of plasmid pOSH1605-A: Lanes 1 and 2: PCR amplification of the pOSH1602-A backbone (**Figure** 7) — includes everything but the *YFP* gene (~6500bp). Lanes 3 and 4: PCR amplification of the *DXS* gene from plasmid pOSH1301 (~1900bp). **C.** DNA segments for assembly of plasmid pOSH1605-B: Lanes 1 and 2: PCR amplification of the pOSH1602-A backbone —includes everything but the *YFP* gene (~6500bp). Lanes 3 and 4: PCR amplification of the *DXS* gene from pOSH1301 (~1900bp). **D.** DNA segments for assembly of plasmid pOSH1602-B: Lane 1: PCR amplification from plasmid pOSH1511-A of an amplicon containing the NISFD targeting regions, and Ap^R and Km^R genes (~5400bp). Lane 2: PCR amplification of the cI857-pR repressor-promoter from plasmid pHG276 (~1100bp). Lane 3: PCR amplification of the *DXS, IspF*, and *IspH* genes from plasmid pOSH1301 (~3500bp).

Once plasmid pOSH1602-A was created and tested, the *YFP* gene was removed and replaced with the *DXS* and MEP pathway genes. This method allowed for a secondary screening method — *E. coli* strains that had lost the YFP fluorescence were selected for the second step of PCR screening to determine whether the transgenes had been correctly inserted. Once correctly assembled in *E. coli*, the *DXS* and/or MEP genes and flanking regions from these plasmids were introduced into *Synechococcus sp*. PCC 7002 by physiological transformation. Once in *Synechococcus*, however, the *DXS* and MEP pathway genes alone do not produce isoprene. To do so requires the addition of an isoprene synthase gene (*IspS*) and for efficient isoprene production, and IDI isomerase gene (*IDI*). This set of genes was targeted to the *Synechococcus* chromosome at the *petJ2* site, and are known to result in isoprene production at rates of at least 1.5 mg of isoprene gDW^{-1} 12 h day^{-1}. Successful integration of *DXS* and/or MEP pathway genes from plasmids pOSH1605-A, pOSH1605-B, and pOSH1602-B) into the NISFD neutral site in *Synechococcus* 7002, together with integration of the *IspS-IDI* genes into *petJ2* is shown in Figure 15. Strains that had integrated both sets of genes, with complete segregational loss of the corresponding wild type regions, were selected for further testing.

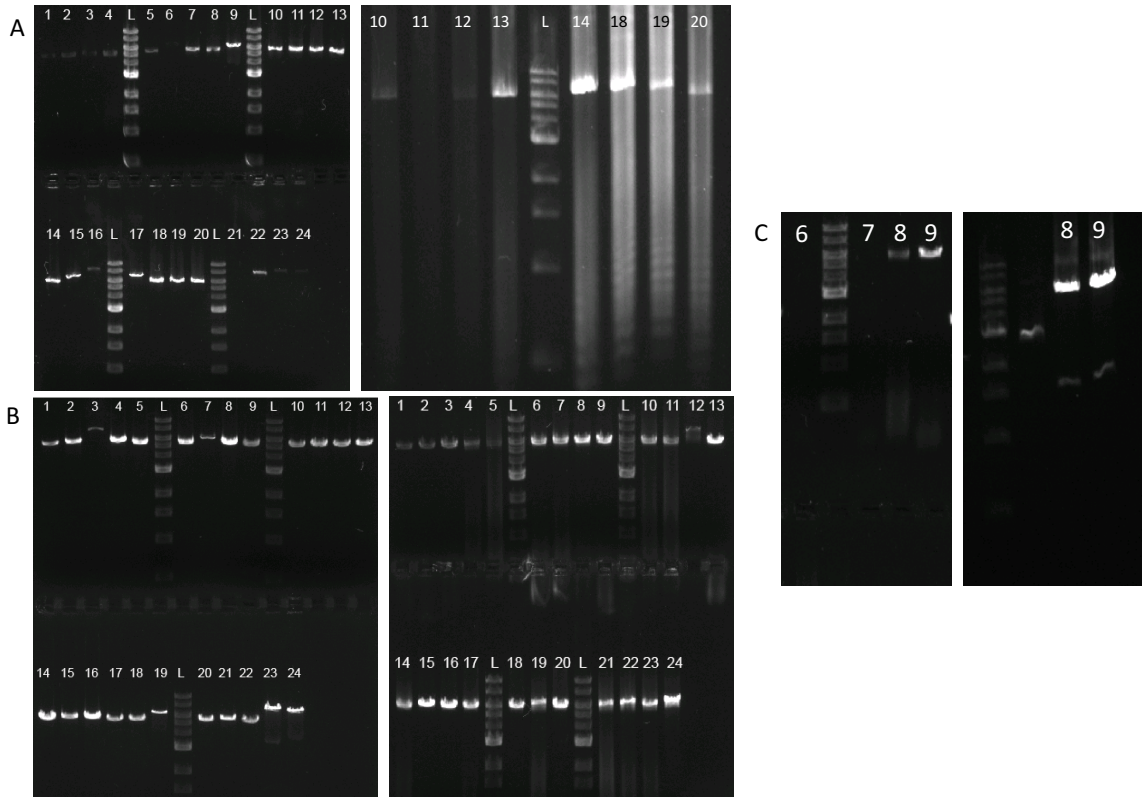

Figure 15. Integration of *DXS*, MEP pathway, and *IspS-IDI* genes into *Synechococcus* 7002 chromosomal sites. A. *Synechococcus* 7002::1310/1605-A. Left gel: PCR screening for chromosomally integrated *IspS-IDI* genes (~5000 bp). Strains 10-14 show strong amplification bands at the expected size. Right gel: PCR screening for the *cI857-pR-CroDXS* integration cassette (~5960 bp). Strain 14 was used for subsequent GC/MS tests. **B.** *Synechococcus* 7002::1310/1605-B. Left gel: PCR screening for the *cI857-pR-RBS-DXS* integration cassette (~5900 bp). Right gel: PCR screening for chromosomally integrated *IspS-IDI* genes (~5000 bp). Strains 14-24 show strong amplification at the expected size. **C.** *Synechococcus* 7002::1310/1602-B. Left gel: PCR screening for the *cI857-pR-DXS-IspF-IspH* integration cassette (~7900 bp). Right Gel: PCR screening for chromosomally integrated *IspS-IDI* genes (~5000 bp). Strain 8 was used for subsequent GC tests. Lanes in all panels show size standards (10 kB DNA ladder)

ISOPRENE PRODUCTION IN *SYNECHOCOCCUS* CARRYING REGULATED MEP PATHWAY AND ISOPRENE SYNTHESIS GENES

One concern in creating isoprene-producing *Synechococcus* strains that carry *DXS* genes is the effect of the overproduced DXS protein on growth and viability. Previous attempts to introduce the *DXS* gene (expressed from a strong, constitutive promoter) into either *E. coli* or *Synechococcus* were problematic. *E. coli* strains that carried *DXS* could not grow above 30°C, and *Synechococcus* transformants could not be obtained that carried intact, unaltered copies of *DXS*. To test the impact of introduced *DXS* and MEP pathway genes controlled by the *cI857-pR* repressor-promoter, growth experiments were conducted with the *Synechococcus* strains that had incorporated the *cI857-pR-DXS* and/or MEP gene cassettes together with the *IspS-IDI* genes for isoprene production. (Figure 16).

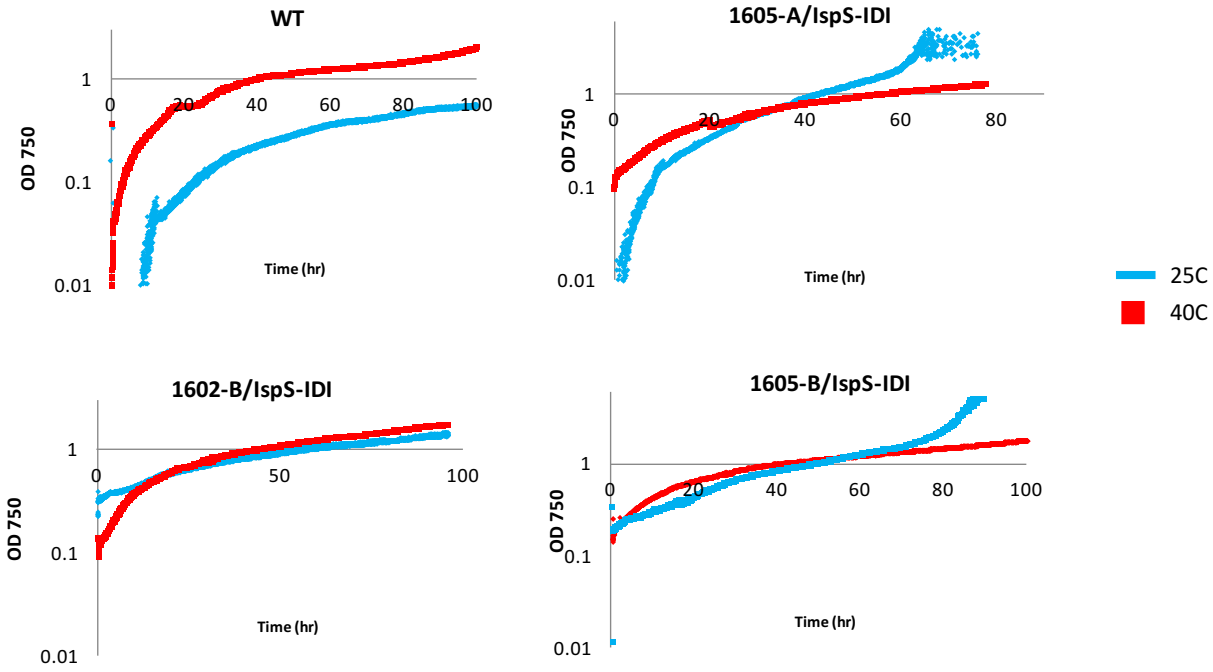

Figure 16. Growth of transgenic *Synechococcus* strains carrying *DXS*, MEP, and *ISP-IDI* genes at low and high temperatures. A. Wild Type *Synechococcus* 7002. **B.** *Synechococcus* (1605-A/1310, *cI857-pR-CroDXS*). Because of instrument failure, this experiment was terminated at 80 hours. **C.** *Synechococcus* (1605-B/1310, *cI857-pR-DXS*) growth. **D.** *Synechococcus* (1602-B/1310, *cI857-pR-Cro-DXS-IspF-IspH*). All of the *Synechococcus* strains carried 1310, *IspS-IDI* genes integrated into the *petJ2* site. Growth experiments were conducted in ePBR photobioreactors in batch mode at. ePBR OD units at 940 nm were converted to OD units at 750 nm cm^{-1}. For comparisons, growth rates were compared at cell densities of ~0.2 OD_{750nm} as shown in Table 5.

Wild type *Synechococcus* 7002 grew faster and to a higher cell density at 40°C than at 25°C. The transgenic isoprene-producing strains, however, behaved differently. The 7002::1605-A/1310 (*cI857-pR-CroDXS*) strain grow at similar rates at 25°C and 40°C until around 40 hours, where the 25°C cells suddenly outgrew those at 40°C. A similar pattern was also seen in the 7002::1605-B/1310 (*cI857-pR-DXS*) strain. In

contrast, the 7002::1602-B/1310 (*cI857-pR-Cro-DXS-IspF-IspH*) strain, grew in a manner more similar to the wild type — where the cells grown at 40°C outpaced those at 25°C. Table 5 shows a comparison of the maximum exponential growth rates recorded from these strains at 25°C and 40°C in these experiments.

Table 5. Maximum exponential growth rates

Strain	Temp (°C)	Doubling Time (hrs)
1605-A/*IspS-IDI* (*cI857-pR-CroDXS*)	25 40	2.8 5.8
1605-B/*IspS-IDI* (*cI857-pR-DXS*)	25 40	16.1 3.1
1602-B/*IspS-IDI* (*cI857-pR-Cro-DXS-IspF-IspH*)	25 40	19.0 3.0
1310/*IspS-IDI*	25 40	13.8 2.0

To determine the impact of the added *DXS* and/or MEP pathway genes on isoprene production in the transgenic *Synechococcus* strains, isoprene production was measured in these strains after growth in sealed 250ikik mL bottles at 25°C and 40°C over a 24-hour period. Headspace gas was collected and measured by gas chromatograph (Table 6).

Table 6. Isoprene production in transgenic *Synechococcus* at 25°C and 40°C

Strain	Temp (°C)	Isoprene, mg gDW^{-1} $12h^{-1}$
1605-A/*IspS-IDI* (*cI857-pR-CroDXS*)	25 40	0.58 (0.09) 2.03 (0.7)
1605-B/*IspS-IDI* (*cI857-pR-DXS*)	25 40	0.3 (0.1) 2.92 (0.5)
1602-B/*IspS-IDI* (*cI857-pR-Cro-DXS-IspF-IspH*)	25 40	0.97 (0.2) 1.36 (0.1)
1310/*IspS-IDI*	25 40	0.38 (0.03) 1.37 (0.05)

Cultures were incubated with stirring under continuous illumination (~200 µmol photons m^{-2} s^{-1}). Values shown are means and standard deviations of 3 biological replicates.

The *Synechococcus::*1310 (*IspS-IDI*) strain does not contain any additional *DXS* or MEP pathway genes, and increased isoprene production at 40°C relative to 25°C is the result of higher metabolic and enzyme activity at the higher temperature. The other three strains, however, carry *DXS* and MEP pathway genes under control of the *cI857-pR* temperature-sensitive repressor-promoter, which is inactive at 25°C, but active at 40°C. *Synechococcus* 7002::1605-A (*cI857-pR-CroDXS*) exhibited a 3.5-fold increase in isoprene production at 40°C relative to 25°C, and about a 2-fold greater rate in at 40°C relative to strain 1310 (*IspS-IDI* alone) (Figure 17). The 7002::1605-B (*cI857-pR-DXS*) strain displayed a 10-fold increase in isoprene production at 40°C relative to 25°C, and a 3-fold greater rate at 40°C relative to strain 1310 (*IspS-IDI* alone) (significantly different at $P<0.05$) (Figure 17). The 7002::1602-B (*cI857-pR-Cro-DXS-IspF-IspH*) strain showed little difference in isoprene production at high and low temperatures, and essentially no

difference in isoprene production at 40°C relative to strain 1310 (*IspS-IDI* alone). Strain 1602-B showed an isoprene production rate at 25°C about 2-fold higher than in any of the other strains, but this may be an anomaly.

All of the strains produced more isoprene at 40°C than at 25°C (although only marginally so for strain 1602-B) as expected because of higher enzyme activity at the higher temperature. The increase that can be attributed to temperature is ~3.6-fold shown in strain 1310 (*IspS-IDI* alone) (Table 4 and Figure 17). Significantly, expression of the temperature-regulated *DXS* gene at 40°C further increased isoprene production beyond that from temperature alone. This is seen most clearly in strain 1605-B (*cI857-pR-DXS*) where the rate at 25°C was similar to that in strain 1310 (*IspS-IDI* alone), but increased overall by ~10-fold at 40°C to reach a rate more that 2-fold higher than in strain 1310.

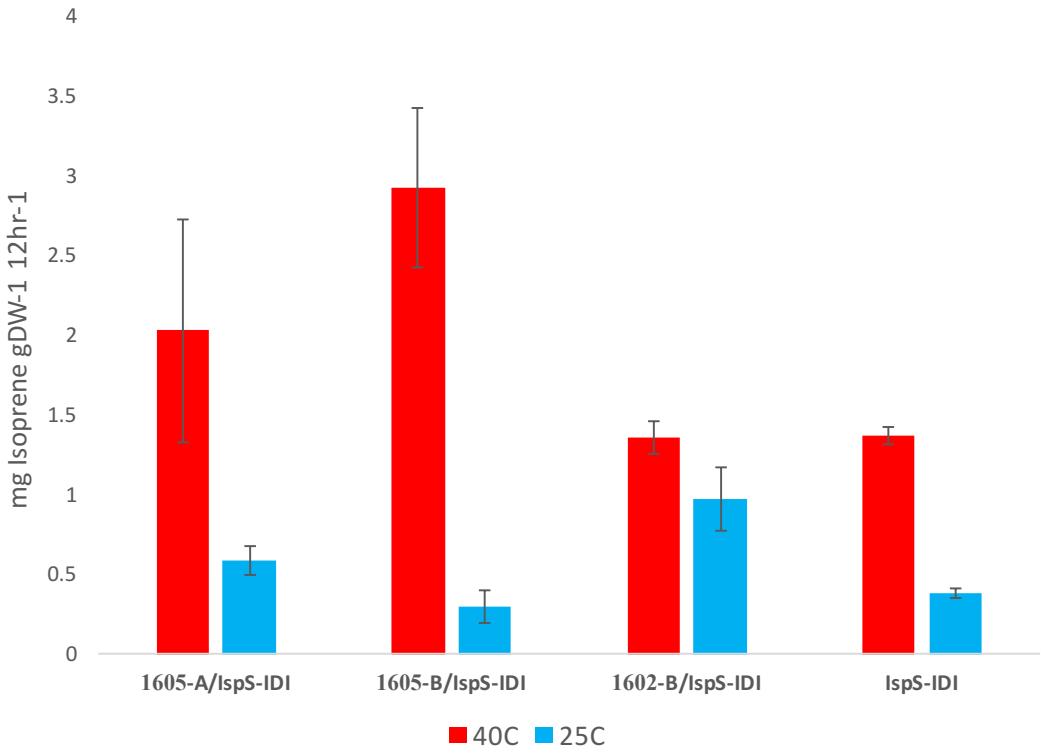

Figure 17. Isoprene production as a function of temperature in transgenic *Synechococcus* carrying the *cI857-pR* repressor-promoter. Isoprene production was measured after a 24 hour incubation period in four transgenic strains (as described in Table 4) at 40°C (red) and 25°C (blue), ~200mol m^{-2} s^{-1} light intensity. Data shown are the means of 3 biological replicates and error bars represent standard deviations.

CHAPTER IV: DISCUSSION

In recent years, photosynthetic organisms such as cyanobacteria and algae have been identified as potentially viable biofactories for bioproducts and biofuels production. The ability of cyanobacteria to harness abundant solar energy and atmospheric carbon dioxide through photosynthesis makes them particularly attractive as model organisms for generation of 'green fuels' (or environmentally-friendly fuels). Cyanobacteria have higher photosynthetic efficiencies than terrestrial plants, (*15*) are amenable to genetic engineering, can grow to high densities in photobioreactors (culture chambers), and have a much faster growth rate than plants. All of these features make them ideal as host organisms for synthesis of terpenoids and industrially-relevant biochemicals. Despite these advantages and recent developments in algal biofuels research, the challenge of achieving significant biofuels production for commercial exploitation is still a major concern for researchers and as such, different strategies for circumventing these challenges have been proposed.

CHARACTERIZATION OF THE *cI857-pR* REPRESSOR-PROMOTER: TEMPERATURE-SENSITIVE YFP PRODUCTION

In this thesis work, I have investigated the *cI857-pR*, temperature-regulated repressor-promoter, initially discovered in bacteriophage lambda and used extensively in *E. coli* (*47*, *48*, *50*), as a way to control the expression of potentially toxic methylerythritol-phosphate (MEP) pathway genes for increased isoprene production in cyanobacteria. The *cI857-pR* repressor-promoter has been used previously in cyanobacteria (*58*, *59*) but not to my knowledge in *Synechococcus* PCC 7002 nor for

regulation of an important set of bioproduction genes. My initial characterization of its utility in *Synechococcus* 7002 was done by linking the *cI857-pR* regulator to a yellow fluorescent protein (YFP) reporter gene as shown in Figure 7. I found, as expected, that the *cI857-pR* regulator worked remarkably well in controlling YFP expression in *E. coli*. When grown in a photobioreactor in turbidostat mode (Figure 9), the YFP expression was remarkably temperature-sensitive. YFP was not expressed, or expressed only at basal levels, at temperatures of 28°C to 34°C, above which, YFP fluorescence increased exponentially until at least 40°C. In *E. coli*, the *cI857-pR* regulator acted in a gradual fashion (rather than as an immediate, on-off switch) as the temperature increased. As the culture warmed, transcription from the *pR* promoter increased, consistent with gradually increasing denaturation of the *cI857* repressor.

When integrated into the *Synechococcus sp.* PCC 7002 chromosome, the temperature-sensitive *cI857-pR-YFP* construct acted in a similar, but not identical manner to its behavior in *E. coli*. Again, YFP expression was temperature sensitive (Figure 12). but YFP fluorescence was not as strong as in *E. coli*. In *E. coli*, YFP expression from the *cI857-pR* regulator was about four times that of its expression in *Synechococcus*. However, this in itself may not be a limitation. When this project was first begun, the primary objective was to regulate the production of the MEP pathway DXS enzyme, which resulted in cell death when overproduced. Having a regulator that is not as strong as the constitutive *Synechocystis* PCC 6803 *PcpcB* promoter may be advantageous

because the *PcpcB* promoter was too strong for productive expression of the DXS protein.

Temperature regulated Isoprene Production

Toward improving terpenoid (or bioproduct) yields in cyanobacteria, a suite of genetic engineering strategies will be required toward enhancing photosynthetic efficiency, optimizing carbon utilization, and partitioning to increase carbon flow to desired biosynthetic pathways. These strategies may include inactivating competing carbon pathways to allow carbon to be redirected to a desired pathway, expressing specific or complete suites of heterologous genes for alternative pathways, and targeting of such heterologous genes to chromosomal sites for improved genetic stability. Isoprene is a volatile, hydrophobic compound with a global market demand of 770,000 metric tons per year estimated at 3 billion dollars annually (*60*), and an important precursor for industrial and biofuels applications. Matthew Nelson in our lab has successfully engineered isoprene production in *Synechococcus* sp. PCC 7002 cyanobacteria, but isoprene yields need to be significantly improved to realize the dream of commercial production.

In this research work, I hypothesized that the insertion of a temperature sensitive regulator to control the expression of potentially toxic gene products, such as DXS and other enzymes of the MEP pathway, would significantly improve isoprene production in *Synechococcus* sp. PCC 7002. This hypothesis was premised on previous experiments in which the *DXS* gene under the control of a strong constitutive promoter (*PcpcB* from the cyanobacterium *Synechocystis* sp. PCC 6803) resulted in *Synechococcus* cultures that

crashed, or *E. coli* cultures that only grew at low temperatures. We postulated that the constitutive *PcpcB* promoter was too strong, and that a temperature-sensitive regulator would enable controlled gene expression by which increased isoprene production could be realized. Data presented here support this hypothesis. These data show that expression of DXS via the temperature-regulated *cI857-pR* repressor-promoter indeed increased isoprene production in *Synechococcus* 7002 that carry isoprene synthase (*IspS*) and IDI-isomerase (*IDI*) genes. However, not all temperature-regulated *DXS* gene sets, in combination with other genes, produced similar amounts of isoprene (Figure 17).

The *Synechoccous sp.* PCC 7002 strain that produced the most isoprene was 7002::1605-B/1310 (*cI857-pR-DXS, IspS-IDI*), a strain that carries *IspS-IDI* genes, and a temperature-controlled *DXS* gene alone, with no fused codons from the bacteriophage lambda *Cro* gene (Figure 14). The 1605-B/1310 strain exhibited temperature-sensitive isoprene production: with 0.3 mg of isoprene produced (per gDW^{-1} 12 h^{-1}) at 25°C, and about ten times that (3 mg gDW^{-1} 12 h^{-1}) at 40°C (Table 6). This strain produced about three times more isoprene at 40°C than the base *Synechococcus* 7002::1310 (*IspS-IDI*) strain. The 1605-B/1310 (*cI857-pR-DXS/IspS-IDI*) strain was, by far, the best in terms of both dynamic range of temperature-regulated isoprene production and total yield of isoprene produced. Note that isoprene production rates are presented here per gram dry weight (gDW) of cells per 12 hours to reflect production during a 12h hour daylight period. If photobioreactors driven by artificial (e.g. LED) light become practical, then daily production rates (per 24 hours) would be twice those presented here.

The other two *Synechococcus* strains that carry *DXS* constructs regulated by *cI857-pR*, did not perform as well. The 7002::1605-A/1310 (*cI857-pR-CroDXS*/*IspS-IDI*) strain, which carries a *Cro-DXS* gene fusion (with the first 22 codons of *Cro* joined in-frame with *DXS*), only produced ~2 mg isoprene gDW^{-1} 12 h^{-1} at 40°C, which is somewhat more than in the 7002::1310 (*IspS-IDI*) control, but not significantly so at $p<0.05$. Also because the standard deviation in the 1605-A data is large, it is not yet possible to conclude that isoprene production in this strain was significantly lower (at $p<0.05$) than the ~3 mg gDW^{-1} 12 h^{-1} observed in strain 1605-B (Figure 17). Nonetheless, it appears that the rate is lower in strain 1605-A/1310 (*cI857-pR-CroDXS*/*IspS-IDI*) than in 1605-B/1310 (*cI857-pR-DXS/IspS-IDI*), and raises questions as to why this might be. Often protein fusions stabilize and increase the activity of overproduced proteins (*61–64*)and indeed the first successful demonstration of the bacteriophage lambda *cI857-pR* regulator in *E. coli* involved fusion of the first 21 codons of *Cro* with the *lacZ* gene for β-galactosidase (*47*). In *E. coli*, this fusion resulted in a 1500-fold dynamic range of expression, from repressed to fully induced. The first demonstration of the *cI857-pR* regulator in cyanobacteria also involved a fusion of part of the *Cro* gene with *lacZ* (*59*). As the only difference between the 1605-B and 1605-A strains is the first 22 codons of *Cro* joined in-frame with the *DXS* gene to generate a Cro-DXS fusion protein in 1605-A , this suggests that the fusion may hinder DXS protein folding or activity, rather than enhancing the performance of this enzyme in *Synechococcus* 7002. The Cro protein is 60 amino acids in length (*48*), and the addition of nearly a third of the protein to DXS appears to be disadvantageous.

As to why the 7002::1602-B/1310 (*cI857-pR-Cro-DXS-IspF-IspH/IspS-IDI*) strain did not perform as well 1605-B/1310 is not clear. This strain carries *DXS*, *IspF* and *IspH* genes, as well as the first 22 codons of *Cro*, but lacks any fusion between *Cro* and *DXS.* Given that there are extra MEP pathway genes in the 1602-B strain (*IspF* and *IspH*), one might expect this strain to produce the most isoprene, but at 40°C it performed at the same level as the control 7002::1310 (*IspS-IDI*) base strain (Table 6 and Figure 17). Furthermore and surprisingly, whereas the other *DXS* strains, as well as the 7002::1310 control, all showed significant increases in isoprene production from 25°C to 40°C, the 7002::1602-B/1310 strain showed only a small increase at the higher temperature.

Without mass spectrometry of immunological data to assess protein content, it is not possible to determine the impact of the overexpressed MEP pathway DXS, IspF, and IspH protein on the expression level of these proteins in the 1602-B/1310 strain. Possibly the ~100 bp of DNA between the end of the *Cro* gene fragment and the beginning of the *DXS* gene, or the expressed fragment of the Cro protein, may have caused a change at the protein level that rendered the DXS enzyme inert. This could result from an interaction of the Cro protein fragment with DXS, possibly improper 'read-through' translation from the Cro fragment, or toxic protein aggregation or toxic metabolite pools generated by overproduction of the IspF or IspH proteins.

The growth data shown in Figure 16 may help elucidate some of these questions. The wild type strain, grown at 40°C is expected to grow faster than at 25°C because of increased metabolic activity at the higher temperature. When the 1605-A/1310 (*cI857-*

pR-CroDXS/IspS-IDI) and 1605-B/1310 (*cI857-pR-DXS/IspS-IDI*) strains were grown at 40°C and 25°C, the pattern changed. In both of these strains, after 60 to 80 hours, the cells at 25°C began to grow much more rapidly and reached a much higher cell density than the cells at 40°C. This, suggests that in these strains, growth or incubation at 40°C causes a metabolic strain. This may be consistent with the idea that isoprene production at 40°C, or increased expression or activity of the overexpressed DXS enzyme, or perhaps accumulation of toxic metabolites that it produces, results in a stress condition that impedes growth.

The growth data for the 1602-B/1310 (*cI857-pR-Cro-DXS-IspF-IspH/IspS-IDI*) strain tells a different story. This strain's growth curve is more similar to that of the wild type, at least with respect to somewhat more rapid growth at 40°C than at 25°C. This suggests, that unlike the 1605-B/1310 and 1605-A/1310 strains, 1602-B/1310 is not under the same metabolic stress at 40°C. This may arise from its overall, lower rate of isoprene production at the high temperature, or perhaps because it does not effectively express the added DXS enzyme and thus does not experience its possible, toxic impacts. It should be noted, however, that strain 1602-B/1310 does not grow as rapidly as the wild type at 40°C and thus apparently still experiences certain metabolic stresses that may arise from expression of the added isoprene synthase or MEP pathway genes.

FUTURE WORK

To increase isoprene yield in *Synechococcus* 7002 or other cyanobacteria, insertion of genes that increase carbon flow through the MEP pathway or the inactivation

of genes for competing pathways should result in greater isoprene production. However, metabolism is complex and these simple predictions do not always yield the expected outcomes. In the current study, the strain that produced the most isoprene, 1605-B/1310 (*cI857-pR-DXS/IspS-IDI*), had only the solidarity *DXS* gene under control of the *cI857-pR* regulator (along with the isoprene synthase and IDI-isomerase, *IspS-IDI* genes). To further increase the isoprene that is produced, the addition of supplementary, MEP pathway genes would appear to be beneficial. While the added *IspF* and *IspH* genes in the 7002::1602-B/1310 (*cI857-pR-Cro-DXS-IspF-IspH/IspS-IDI*) strain did not appear to enhance isoprene production, this particular gene set may not be constructed optimally, as discussed above. In principle, as long as added genes do not result in toxic aggregates of overproduced proteins or the accumulation of toxic metabolite pools, the expression of additional MEP pathway enzymes should increase carbon flow through this pathway leading to increased isoprene production.

Our lab has already genetically engineered several *Synechococcus* strains that could assist in this endeavor as well. One knockout strain, 7002::*ΔGlgA1A2-IspS-IDI* has resulted in a 20% increase in isoprene production (*45*). This strain was created under the hypothesis that the inactivation of a competing carbon pathway for glycogen biosynthesis would significantly improve isoprene production in *Synechococcus* 7002. This hypothesis was premised on the assumption that if *Synechococcus* cannot store excess carbon as glycogen, then this carbon will be channeled into other pathways including the MEP pathway for isoprene production. There are other strains as well, such as a ΔGlgC knockout (which prevents synthesis of both glycogen and glucosyl-glycerol(ate)), have

shown at least a twofold increase in production (*65*, *66*). The goal behind the ΔGlgC strain is the same: to redirect majority of the fixed carbon from biomass production into production of the desired product.

Another method for increased isoprene output is the use of further or improved, regulatable promoters. The identification of regulatable promoters will allow for controlled isoprene production in *Synechococcus* 7002 and thus, relieve cells of the extra metabolic burden of continuous isoprene production or continuous production of proteins that do not increase growth and do not provide any immediate selective advantage for the host organism. Continuous isoprene production itself may not impose much of a metabolic burden because it can serve as a pathway for 'overflow metabolism' that relieves buildup of excess reductant (*67*), and indeed, our group has demonstrated continuous, stable production of isoprene in *Synechococcus* 7002 for up to month. Regardless, regulatable promoters are clearly beneficial, as shown here, and have many potential uses for production of other high-impact biochemical products in the future and provide a means of control for bioproduction over a broad range of environmental conditions. Another regulatable promoter of interest is from a Ni-inducible lysis system that has been used in *Synechocystis sp*. PCC 6803 (*35*). This system comprises of a nickel-sensing/responding system (*nrsR-nrsS-nrsP*) that could prove useful for increased isoprene production.

Finally, with respect to our current temperature-regulated, *cI857-pR*, repressor-promoter, one means of increased production is through modification of the promoter itself. Previous work though modification of promoter regions have shown increased

output (*68*). In *Synechococcus* sp. PCC 7002, for example, Pfleger *et al.* developed two constitutive promoter libraries based on modifications to P_{cpcb} in PCC 7002, which contain promoters with a range of strengths spanning 3 orders of magnitude. These promoter libraries were used to construct an IPTG induction system based on a promoter of cyanobacterial origin. Further modifications to this system resulted in stronger repression and higher intrinsic expression levels. The resulting IPTG induction system was capable of a 48 ± 7 fold increase of YFP expression.

Endogenous metal responsive promoters have been developed for use as induction systems with better than 100-fold dynamic range (*69*). For example, the promoter for the *Synechocystis* sp. PCC 6803 gene, *coaA*, was induced 500-fold by 6 μ M Co2+ (*30*), and P_{smt} from *Synechococcus elongatus* PCC 7942 was induced 300-fold by 2 μ M Zn2+(*70*). In these and in our own system, the goal is tight regulation (fully off when repressed) and a wide, strong range of response. To create such promoters random mutagenesis of regulator-promoter regions has been used to create libraries of mutant regulators and promoters as well as ribosomal binding sites. Markley et al. (*68*), were able to create two constitutive promoter libraries in *Synechococcus* PCC 7002 that contain promoters with a range of strengths spanning three orders of magnitude. These promoter libraries were used to construct an induction system based on the Isopropyl β-D-1-thiogalactopyranoside (IPTG) -inducible *E. coli lac* promoter that functioned well over a large expression range in the cyanobacteria. Iterative modifications to this system

resulted in stronger repression and higher intrinsic expression levels. The resulting IPTG induction system was capable of a 48-fold increase of *YFP* expression (*68*).

CONCLUSION

This research has successfully demonstrated several important proofs-of-concept toward the development of enhanced, isoprene-producing *Synechococcus* sp. PCC 7002 cyanobacteria for commercial applications. First, a genetic construct was created that linked the temperature-sensitive bacteriophage lambda *cI857-pR* repressor-promoter to a gene fusion consisting of the first 22 codons of the lambda *Cro* gene to a yellow fluorescent protein (YFP) gene. This was done under the belief that the resulting Cro-YFP protein fusion would result in increased protein stability and higher protein expression compared to a non-fusion, YFP protein. Work presented here established that this regulator worked very well in *Synechococcus sp.* PCC 7002 in a temperature dependent manner, keeping YFP expression 'off' at temperatures below 32°C with expression turned 'on' and gradually increasing above 32°C. Once temperature-dependent gene expression was confirmed in *Synechococcus* 7002, the *YFP* gene was removed and replaced with combinations of the deoxyxylulose-synthase (*DXS*) and other methylerythritol phosphate (MEP) pathway gene. Of the three different variations tested, the best isoprene producer was the one that lacked a protein fusion of the first 22 amino acids of Cro with DXS and carried only the *DXS* gene (controlled by the *cI857-pR* regulator) in addition to isoprene synthase and IDI isomerase (*IspS-IDI*) genes. This *Synechococcus* 7002::1605-B/1310 (*cI857-pR-DXS/IspS-IDI*) strain produced over three times the amount of isoprene compared to the control, base strain 7002::1310 (*IspS-IDI*)

that carries only the *IspS-IDI* genes. Overall this research has further demonstrated the amenability of *Synechococcus sp.* PCC 7002 as a model organism for genetic engineering and an excellent platform for algal bioproducts and biofuels development. This work has contributed toward improved photosynthetic isoprene production in *Synechococcus* cyanobacteria but further genetic engineering approaches will be needed to create a 'super-strain' strain for an economically sustainable algal biotechnology industry.

APPENDIX A: MICROBIAL MEDIA

LB Media for *E. coli*

Into approximately 900mL of ddH_2O, dissolve 10 grams of tryptone, 5 grams of yeast extract, and 10 grams of NaCl. Adjust the pH to 7.5 with NaOH, then fill to one liter. Autoclave at 121°C for 15 minutes to sterilize, then store at room temperature. If plates are required, add 15 grams of agar prior to pH adjustment; if antibiotics are necessary, they should be added post-sterilization, once medium has cooled to approximately 50°C, and prior to pouring. One liter of media pours approximately 30 plates, which should be stored at 4°C once solidified.

S.O.C. Media for Electrotransformed *E. coli*

As it is a rich medium used primarily for electotransformation, S.O.C. (and its precursor, S.O.B.) is primarily manufactured in small batches. 100mL may be produced by dissolving 0.5 grams of yeast extract, 2 grams of tryptone, 0.06 grams of NaCl, and 0.02 grams of KCl in approximately 90mL of ddH_2O. A magnesium component is also required, which may be either added prior to sterilization, as 0.24 grams of $MgSO_4$, or as 500µL of a separately-sterilized 2M $MgCl_2$ solution after autoclaving at 121°C for 15 minutes and allowing to cool to approximately 60°C. Bringing this solution to 100mL total volume completes S.O.B. medium; completing S.O.C. medium requires an additional 2mL of filter-sterilized 20% glucose solution prior to filling the volume to 100mL with ddH_2O. Media may be stored at room temperature in small volumes, or at 4°C to reduce the risk of contamination.

A(D7) Media for *Synechococcus sp.* PCC 7002

Table A-1. Initial salt solution for A(D7) growth medium.

Salt	*Amount per liter (g)*	*Final Concentration (mM)*
NaCl	17.53	300
KCl	0.6	8.0
$NaNO_3$	1.02	12.0
$MgSO_4$ / $MgSO_4 \cdot 7H_2O$	2.44 / 5.0	20.0
$CaCl_2$	0.3	2.5

Into approximately 990mL of ddH_2O, dissolve the salts shown in Table A-1, plus 1.0mL each of a 15mM FeEDTA solution (0.62 grams $FeEDTA \cdot 2H_2O$ in 100mL ddH_2O, sterilized) and 1000x D7 micronutrients (see Table A-2 for details). Prepare also a 1.0M Tris-HCl solution (pH 8.2) and a 1.0M KH_2PO_4 solution. Autoclave all solutions at 121°C for 15 minutes, then allow to cool to approximately 50°C. To the salt base, add 8.3mL of the 1.0M Tris-HCl and 370µL of the 1.0M KH_2PO_4 solutions, as well as 100µL of 80mg/L vitamin B_{12} (methylcobalamin) solution. Store at room temperature, or at 4°C if supplemented with antibiotics.

To make A(D7) plates, first produce a 2x salt solution by dissolving the salts in Table A-1, plus 1.3 grams of anhydrous sodium thiosulfate (2.0 grams pentahydrate), into approximately 490mL ddH_2O, then add 1.0mL each of 15mM FeEDTA solution and 1000X D7 micronutrients. Separately dissolve 12 grams of Difco® Bacto® agar in 500mL ddH_2O (VWR and Acros agars have been found to have a negative impact on cyanobacterial growth). Autoclave the salt and agar solutions separately at 121°C for 15

minutes; cool the salt base rapidly to 50°C using a cool water bath, then add 8.3mL of 1.0M Tris-HCl (pH 8.2), 370µL of 1.0M KH_2PO_4, 100µL of 80mg/L vitamin B_{12}, and any necessary antibiotics, mixing well. Allow the agar solution to cool to 50°C before adding it to the salt solution, mixing briefly before pouring. One liter pours 20-25 thick plates, which should be stored at 4°C once solidified.

Table A-2. D7 micronutrient formulation, as described by Arnon et al **(*71*)**; the Buzby et al **(*72*)** recipe may vary slightly.

Trace Metal	*Molar Mass (g/mol)*	*Amount (g/L)*
H_3BO_3	61.83	2.86
$MnCl_2 \cdot 4H_2O$	197.91	1.81
$ZnSO_4 \cdot 7H_2O$	287.54	0.22
$Na_2MoO_4 \cdot 2H_2O$	241.95	1.26
$CuSO_4 \cdot 5H_2O$	249.68	0.079
$NaVO_3$	121.9	0.239
$CoCl_2 \cdot 6H_2O$	237.93	0.04

APPENDIX B: BUFFERS AND REAGENTS

Antibiotics

Stock solutions of all antibiotics were made in 1000x (w/v) solutions, relative to working concentrations. Batches of antibiotic stocks were prepared by dissolving the salt weight described in Table B-1 in ddH_2O to a total volume of 5mL; these suspensions were sterilized using a clean syringe attached to a 0.2µm-pore filter, dispensing into pre-sterilized 1.5mL screw-cap tubes. Antibiotics were stored at -20°C.

Table B-1. Stock concentrations of antibiotics used in this study.

Antibiotic	*Stock Concentration*	*Salt per 5mL (g)*
Ampicillin	150mg/mL	0.750
Streptomycin	50mg/mL	0.250
Spectinomycin	50mg/mL	0.250
Kanamycin	50mg/mL	0.250

TBE Buffer (5X) for Gel Electrophoresis

Into ~850mL ddH_2O, dissolve 54 grams of Tris base and 27.5 grams of boric acid, then add 20mL of 0.5M EDTA solution before adding ddH_2O as necessary to 1 liter of total volume. Store in a sealed bottle at room temperature, diluting to 1X as needed. A crystalline precipitate may form over time – if the bottle has been sealed well to prevent evaporation, the fluid may be heated to return this precipitate to solution.

10T/0.1E

Make 1.0M Tris-HCl (pH 7.5; 121.14g Tris base per liter) and 0.5M EDTA (186.12g Na_2EDTA per liter); autoclave separately along with ddH_2O (121°C for 15 minutes). Combine 300µL 1.0M Tris-HCl with 6µL 0.5M EDTA in approximately

29mL sterile ddH_2O, then fill to 30mL – this produces a 10mM Tris/0.1mM EDTA solution suitable for long-term storage of DNA. Aliquot into small volumes and store at room temperature.

0.5 M Sodium EDTA (Disodium Ethylenediaminetetraacetate) [pH=8.0]

Add 148.9 grams of sodium EDTA salt to 600 mL sterile Milli-Q H2O in an oven baked Wheaton bottle with a magnetic stir bar. Adjust pH and final volume to 8.0 and 800 mL respectively. Add DEPC to 0.1% final concentration and stir overnight with cap loosely fit. After overnight stirring, autoclave and store at room temperature.

0.5 M Tris (Tris Hydroxymethylaminomethane)

Add 15.15 grams Tris Base to an oven-baked flask containing 250 mL sterile Milli-Q H2O. Autoclave and store at room temperature.

Wizard SV Miniprep Column Wash Solution

To 19ml sterile ddH_2O, add 4.4µL of 0.5M EDTA, 450µL of 1.0M Tris-HCl (pH 7.5) and 0.32g potassium acetate; fill to 20mL. Alternatively, to 15mL sterile ddH_2O, add Tris-HCl and EDTA as described, plus 3.256mL of a 1.0M potassium acetate solution; fill to 20mL. To this 20mL stock, add 35mL of 95% ethanol to produce the working wash solution.

Vitamin B_{12} (Methylcobalamin)

Prepare stock solution at 80mg/mL in sterile ddH_2O, then sterilize using a clean syringe screwed into a 0.2µm-pore filter. Store in sterile microfuge tubes at -20°C, protected from light.

APPENDIX C: PROTOCOLS

Preparation of Electrocompetent *E. coli* Cells

Inoculate 4mL of LB with a single, isolated colony of *E. coli* (a ΔrecA strain, such as DH5α or Scarab Xpress, is recommended) and grow overnight at 37°C with moderately vigorous shaking. Use this overnight culture to inoculate 400mL LB in a large (>1L) flask, which should be grown at 37°C with vigorous shaking until an optical density (OD_{750}) of 0.5-0.7 is reached (should take 3-4 hours). While this culture is growing, prepare and chill: two 250mL sterilized bottles, two 50mL sterilized tubes, 700mL 10% glycerol divided into 400mL/200mL/100mL aliquots, and 30 pre-labelled tubes. Once the 400mL *E. coli* culture has reached the desired optical density, chill rapidly using an ice-water bath and divide into the two 250mL bottles (balance by weight) – from this point forward, the cells should be kept as cold as possible at all times.

Centrifuge in a refrigerated centrifuge (4°C) at 4000Xg for 5 minutes, then remove as much supernatant as possible (pour off, then use a transfer pipette to remove the remainder). Resuspend the cells in 400mL (200mL per pellet) ice-cold 10% glycerol, balancing and centrifuging as before (the pellets may become slippery, requiring delicacy in handling and possibly multiple spins in the centrifuge to avoid losing cells). Remove supernatant and repeat the centrifugation process with 200mL ice-cold 10% glycerol. Resuspend both pellets in the same ~90mL ice-cold 10% glycerol and divide between the two 50mL tubes before centrifuging as above. Remove the supernatant and perform a final resuspension of all cells into the same 1.2mL ice-cold 10% glycerol – the resulting suspension should be an opaque, milky off-white (~10^{10} cells/mL). Distribute 40μL

aliquots into the prepared tubes and store at -80°C (flash-freezing in liquid nitrogen or a dry-ice/ethanol bath is optional).

Freezer Storage of *E. coli* and *Synechococcus* strains

Cells may be cryopreserved in 7% DMSO suspensions. For *E. coli*, this is accomplished by dispensing 500µL of a dense overnight culture of the target strain atop 35µL DMSO in a pre-labelled 1.5mL screw-cap tube and mixing gently prior to freezing at -80°C (flash-freezing is not usually necessary, but may help with delicate strains). For *Synechococcus*, pellet 2mL of exponential-phase culture of the target strain, then resuspend in 500µL fresh A(D7) medium before transferring to a pre-labelled 1.5mL screw-cap tube; add 35µL DMSO and mix gently before freezing at -80°C. DMSO is stored at -20°C, requiring brief heating at 37°C to thaw for us.

APPENDIX D: qPCR DATA

qPCR was used to determine the relative copy number of the *Synechococcus* chromosome to the pAQ1 plasmid. Because there are general comparisons between the constitutively expressed YFP (located on the plasmid) and the temperature sensitive YFP (located on the chromosome), copy number needed to be determined. Through this experiment, it was determined that the plasmid is expressed approximately 4 times as much as the chromosome.

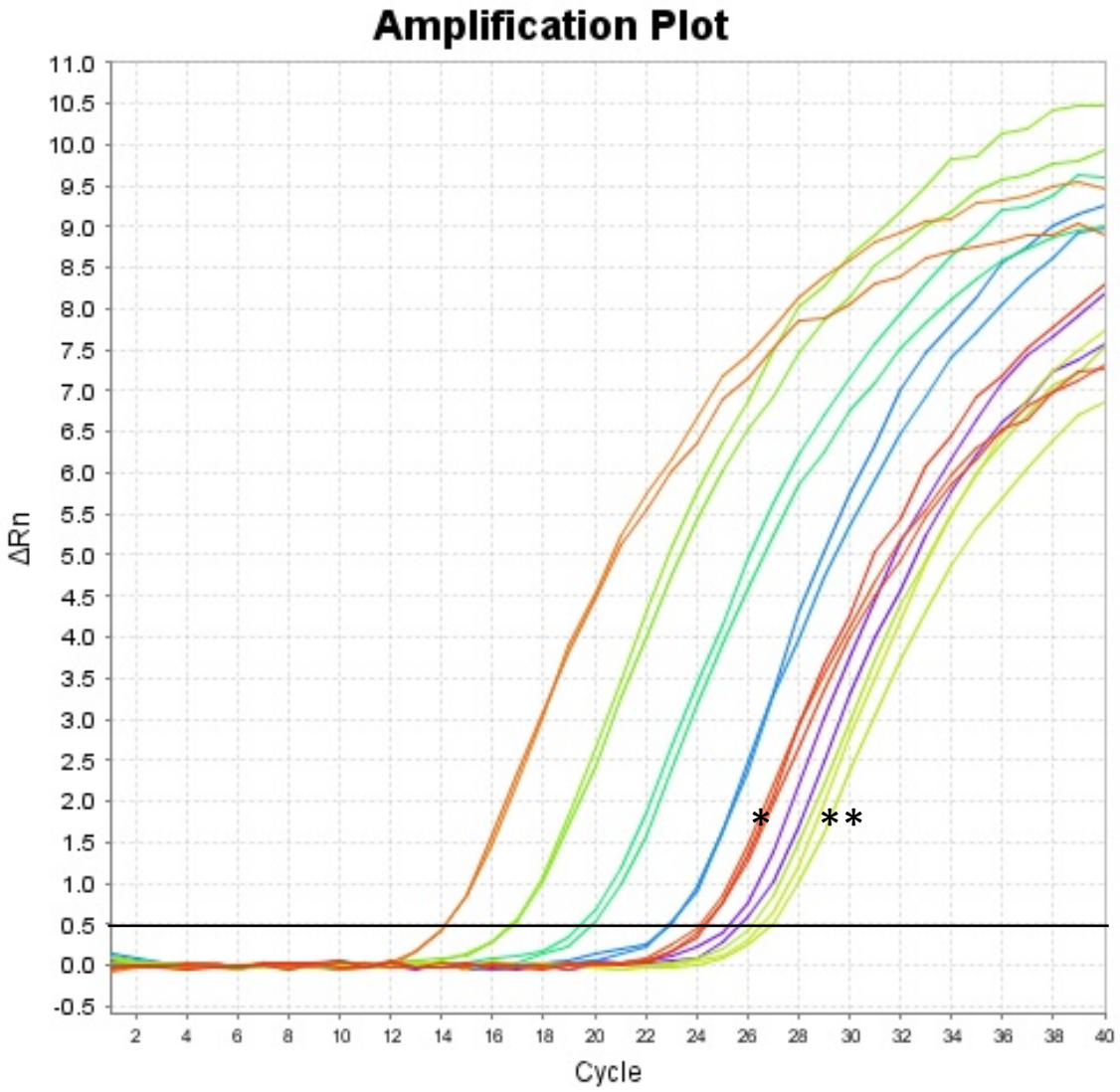

Figure 18. YFP amplification qPCR plot. Fresh cells of the same optical density (0.05@ OD 750) from 7002::pAQ-YFP and 7002::NISFD-cI857-YFP were added directly to the qPCR reaction mix. The plot shows fluorescence signal intensity as a function of PCR amplification cycle. Relative DNA quantities are determined by the number of cycles taken to reach a signal threshold set at 0.5 (heavy black line). The set of red amplifications is 7002::pAQ-YFP (single asterisk, targeted to the pAQ plasmid), and the set of goldenrod is 7002::NISFD-cI857-YFP (double asterisk, targeted to the chromosome).

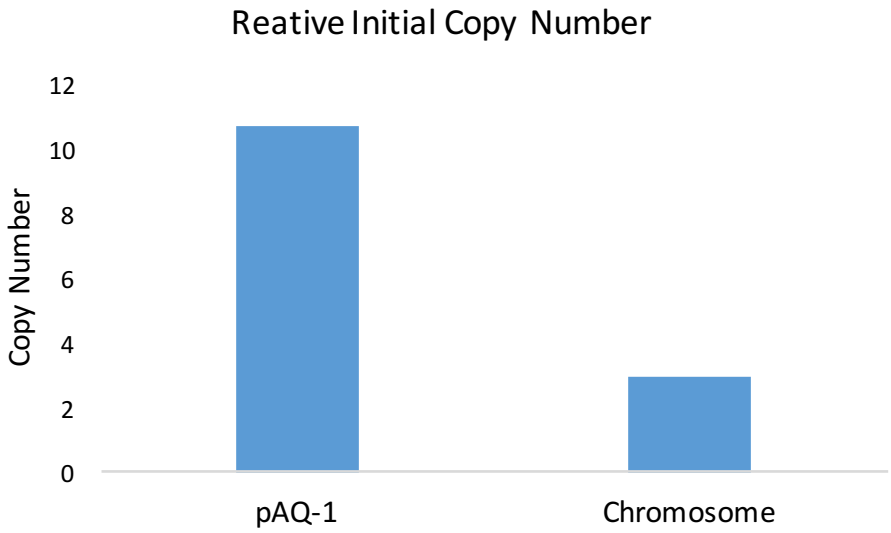

Figure 19. Relative initial copy number of genetic elements in *Synechococcus* sp. PCC 7002. Through qPCR it was determined that the pAQ-1 plasmid was approximately four times more abundant compared to the chromosome.

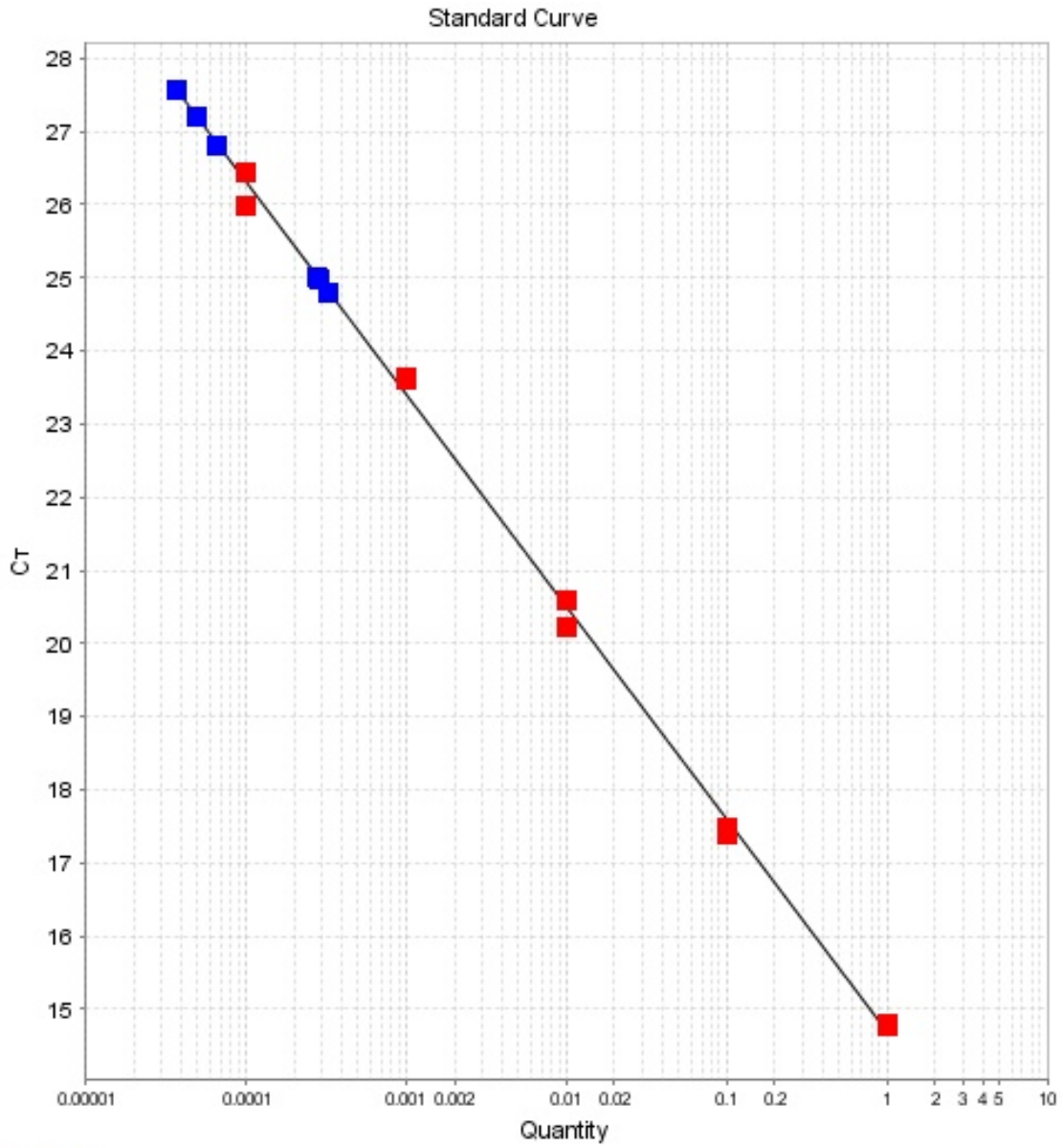

Figure 20. qPCR standard curve for YFP. The plot shows the cycle to threshold (set at 0.5) as a function of the concentration of the standards (in red). The standards are known concentrations of the plasmid pAQ-YFP. The slope of the standard line is -2.902.

APPENDIX E: SEQUENCE DATA

Below are the sequence data for each of the engineered plasmids. The sequenced area began in the middle of the cI857, reading through the fusion (or no fusion) site, to the DXS (or in 1602-A's case, YFP) gene. This sequencing was also done in reverse. Below, is the sequence data in four parts. The contig is the assembled sequence, in forward and reverse. The expected sequence is what we initially designed our primers on, and the following two lines are the actual sequence data, labeled by primer. The sequencing of this particular region was chosen to ensure that no modifications or alterations had been made in the engineering process. As illustrated below, no major changes (insertions, deletions, etc.) were discovered.

1602-A (cI857-YFP).

```
                  10         20         30         40         50
Contig     TA3AAA3CAC CTGACCGCTA TCCC5TGATC AGTTT5TTGA AGGTAAACTC
EXPECTED   TAAAAA-CAC CTGACCGCTA TCCC-TGATC AGTTTCTTGA AGGTAAACTC
YFP-F      --NAAAACAC CTGACCGCTA TCCCCTGATC AGTTT-TTGA AGGTAAACTC
YFP-R      ---------- ---------- ---------- ---------- ----------

                  60         70         80         90        100
Contig     ATCACCCCCA AGTCTGGCTA TGCAGAAATC ACCTGGCTCA ACAGCCTGCT
EXPECTED   ATCACCCCCA AGTCTGGCTA TGCAGAAATC ACCTGGCTCA ACAGCCTGCT
YFP-F      ATCACCCCCA AGTCTGGCTA TGCAGAAATC ACCTGGCTCA ACAGCCTGCT
YFP-R      ---------- ---------- ---------- ---------- ----------

                 110        120        130        140        150
Contig     CAGGGTCAAC GAGAATTAAC ATTCCGTCAG GAAAGCTTGG CTTG2AGCCT
EXPECTED   CAGGGTCAAC GAGAATTAAC ATTCCGTCAG GAAAGCTTGG CTTGGAGCCT
YFP-F      CAGGGTCAAC GAGAATTAAC ATTCCGTCAG GAAAGCTTGG CTTG-AGCCT
YFP-R      ---------- ---------- ---------- ---------- ----------

                 160        170        180        190        200
Contig     GTTGGTGCGG TCATGGAATT ACCTTCAACC TCAAGCCAGA ATGCAGAATC
EXPECTED   GTTGGTGCGG TCATGGAATT ACCTTCAACC TCAAGCCAGA ATGCAGAATC
YFP-F      GTTGGTGCGG TCATGGAATT ACCTTCAACC TCAAGCCAGA ATGCAGAATC
YFP-R      ---------- ---------- ---------- ---------- ----------

                 210        220        230        240        250
Contig     ACTGGCTTTT TTGGTTGTGC TTACCCATCT CTCCGCATCA CCTTTGGTAA
EXPECTED   ACTGGCTTTT TTGGTTGTGC TTACCCATCT CTCCGCATCA CCTTTGGTAA
YFP-F      ACTGGCTTTT TTGGTTGTGC TTACCCATCT CTCCGCATCA CCTTTGGTAA
YFP-R      ---------- ---------- ---------- ---------- ----------

                 260        270        280        290        300
Contig     AGGTTCTAAG CTYAGGTGAG AACATCCCTG CCTGAACATG AGAAAAAACA
EXPECTED   AGGTTCTAAG CTCAGGTGAG AACATCCCTG CCTGAACATG AGAAAAAACA
YFP-F      AGGTTCTAAG CTTAGGTGAG AACATCCCTG CCTGAACATG AGAAAAAACA
YFP-R      ---------- ---------- ---------- ---------- ----------

                 310        320        330        340        350
Contig     GGGTACTCAT ACTCACTTCT AAGTGACGGC TGCATACTAA CCGCTTCATA
EXPECTED   GGGTACTCAT ACTCACTTCT AAGTGACGGC TGCATACTAA CCGCTTCATA
```

```
YFP-F     GGGTACTCAT ACTCACTTCT AAGTGACGGC TGCATACTAA CCGCTTCATA
YFP-R     ---------- ---------- ---------- ---------- ----------

                360        370        380        390        400
Contig    CATCTCGTAG ATTTCTCTGG CGATTGAAGG GCTAAATTCT TCAACGCTAA
EXPECTED  CATCTCGTAG ATTTCTCTGG CGATTGAAGG GCTAAATTCT TCAACGCTAA
YFP-F     CATCTCGTAG ATTTCTCTGG CGATTGAAGG GCTAAATTCT TCAACGCTAA
YFP-R     ---------- ---------- ---------- ---------- ----------

                410        420        430        440        450
Contig    CTTTGAGAAT TTTTGYAAGC AATGCGGCGT TATAAGCATT TAATGCATTG
EXPECTED  CTTTGAGAAT TTTTGCAAGC AATGCGGCGT TATAAGCATT TAATGCATTG
YFP-F     CTTTGAGAAT TTTTGTAAGC AATGCGGCGT TATAAGCATT TAATGCATTG
YFP-R     ---------- ---------- ---------- ---------- ----------

                460        470        480        490        500
Contig    ATGCCATTAA ATAAAGCACC AACGCCTGAC TGCCCCATCC CCATCTTGTC
EXPECTED  ATGCCATTAA ATAAAGCACC AACGCCTGAC TGCCCCATCC CCATCTTGTC
YFP-F     ATGCCATTAA ATAAAGCACC AACGCCTGAC TGCCCCATCC CCATCTTGTC
YFP-R     ---------- ---------- ---------- ---------- ----------

                510        520        530        540        550
Contig    TGCGACAGAT TCCTGGGATA AGCCAAGTTC ATTTTTCTTT TTTTCATAAA
EXPECTED  TGCGACAGAT TCCTGGGATA AGCCAAGTTC ATTTTTCTTT TTTTCATAAA
YFP-F     TGCGACAGAT TCCTGGGATA AGCCAAGTTC ATTTTTCTTT TTTTCATAAA
YFP-R     ---------- ---------- ---------- ---------- ----------

                560        570        580        590        600
Contig    TTGCTTTAAG GCGACGTGCG TCCTCAAGCT GCTCTTGTGT TAATGGTTTC
EXPECTED  TTGCTTTAAG GCGACGTGCG TCCTCAAGCT GCTCTTGTGT TAATGGTTTC
YFP-F     TTGCTTTAAG GCGACGTGCG TCCTCAAGCT GCTCTTGTGT TAATGGTTTC
YFP-R     ---------- ---------- ---------- ---------- ----------

                610        620        630        640        650
Contig    TTTTTTGTGC TYWT4A5GTT A33TCT3T5A CCGCA3GGGA TAAATATCTA
EXPECTED  TTTTTTGTGC TCAT-ACGTT AAATCTATCA CCGCAAGGGA TAAATATCTA
YFP-F     TTTTTTGTGC TCAT-ACGTT AAATCTATCA CCGCAAGGGA TAAATATCTA
YFP-R     ---------C TTTTTA-GTT A--TCT-T-A CCGCA-GGGA TAAATATCTA

                660        670        680        690        700
Contig    ACACCGTGCG TGTTGACTAT TTTACCTCTG GCGGTGATAA TGGTTGCATG
EXPECTED  ACACCGTGCG TGTTGACTAT TTTACCTCTG GCGGTGATAA TGGTTGCATG
YFP-F     ACACCGTGCG TGTTGACTAT TTTACCTCTG GCGGTGATAA TGGTTGCATG
YFP-R     ACACCGTGCG TGTTGACTAT TTTACCTCTG GCGGTGATAA TGGTTGCATG

                710        720        730        740        750
Contig    TACTAAGGAG GTTGTATGGA ACAACGCATA ACCCTGAAAG ATTATGCAAT
EXPECTED  TACTAAGGAG GTTGTATGGA ACAACGCATA ACCCTGAAAG ATTATGCAAT
YFP-F     TACTAAGGAG GTTGTATGGA ACAACGCATA ACCCTGAAAG ATTATGCAAT
YFP-R     TACTAAGGAG GTTGTATGGA ACAACGCATA ACCCTGAAAG ATTATGCAAT

                760        770        780        790        800
```

```
Contig    GCGCTTTGGG CAAACCAAGA CAGCTAAAGA TCCGATGGTG AGCAAGGGCG
EXPECTED  GCGCTTTGGG CAAACCAAGA CAGCTAAAGA TCCGATGGTG AGCAAGGGCG
YFP-F     GCGCTTTGGG CAAACCAAGA CAGCTAAAGA TCCGATGGTG AGCAAGGGCG
YFP-R     GCGCTTTGGG CAAACCAAGA CAGCTAAAGA TCCGATGGTG AGCAAGGGCG

                810        820        830        840        850
Contig    AGGAGCTGTT CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC
EXPECTED  AGGAGCTGTT CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC
YFP-F     AGGAGCTGTT CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC
YFP-R     AGGAGCTGTT CACCGGGGTG GTGCCCATCC TGGTCGAGCT GGACGGCGAC

                860        870        880        890        900
Contig    GTAAACGGCC ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC
EXPECTED  GTAAACGGCC ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC
YFP-F     GTAAACGGCC ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC
YFP-R     GTAAACGGCC ACAAGTTCAG CGTGTCCGGC GAGGGCGAGG GCGATGCCAC

                910        920        930        940        950
Contig    CTACGGCAAG CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG
EXPECTED  CTACGGCAAG CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG
YFP-F     CTACGGCAAG CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG
YFP-R     CTACGGCAAG CTGACCCTGA AGTTCATCTG CACCACCGGC AAGCTGCCCG

                960        970        980        990       1000
Contig    TGCCCTGGCC CACCCTCGTG ACCACCTTCG GCTACGGCCT GCA22T5GC5
EXPECTED  TGCCCTGGCC CACCCTCGTG ACCACCTTCG GCTACGGCCT GCAG-T-GC-
YFP-F     TGCCCTGGCC CACCCTCGTG ACCACCTTCG GCTACGGCCT GCANGTCGCC
YFP-R     TGCCCTGGCC CACCCTCGTG ACCACCTTCG GCTACGGCCT GCAG-T-GC-

               1010       1020       1030       1040       1050
Contig    4233T4CGCC CGCTACCCCG ACCACATGAA GCAGCACGAC TTCTTCAAGT
EXPECTED  T---T-CGCC CGCTACCCCG ACCACATGAA GCAGCACGAC TTCTTCAAGT
YFP-F     NGAAT 1002 ......... ..........  .......... ..........
YFP-R     T---T-CGCC CGCTACCCCG ACCACATGAA G 1018 GAC TTCTTCAAGT

               1060       1070       1080       1090       1100
Contig    CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC
EXPECTED  CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC
YFP-F     .......... .......... .......... .......... ..........
YFP-R     CCGCCATGCC CGAAGGCTAC GTCCAGGAGC GCACCATCTT CTTCAAGGAC

               1110       1120       1130       1140       1150
Contig    GACGGCAACT ACAAGACCCG CGCCGAGGTG ARGTTCGAGG GCGACACCCT
EXPECTED  GACGGCAACT ACAAGACCCG CGCCGAGGTG AAGTTCGAGG GCGACACCCT
YFP-F     .......... .......... .......... .......... ..........
YFP-R     GACGGCAACT ACAAGACCCG CGCCGAGGTG AGGTTCGAGG GCGACACCCT

               1160       1170       1180       1190       1200
Contig    GGTGAACCGC ATCGAGCTGA AGGGCATCGA CTTCAAGGAG GACGGCAACA
EXPECTED  GGTGAACCGC ATCGAGCTGA AGGGCATCGA CTTCAAGGAG GACGGCAACA
YFP-F     .......... .......... .......... .......... ..........
YFP-R     GGTGAACCGC ATCGAGCTGA AGGGCATCGA CTTCAAGGAG GACGGCAACA
```

```
                1210       1220       1230       1240       1250
Contig      TCCTGGGGCA CAAGCTGGAG TACAACTACA ACAGCCACAA CGTCTATATC
EXPECTED    TCCTGGGGCA CAAGCTGGAG TACAACTACA ACAGCCACAA CGTCTATATC
YFP-F       .......... .......... .......... .......... ..........
YFP-R       TCCTGGGGCA CAAGCTGGAG TACAACTACA ACAGCCACAA CGTCTATATC

                1260       1270       1280       1290       1300
Contig      ATGGCCGACA AGCAGAAGAA CGGCATCAAG GTGAACTTCA AGATCCGCCA
EXPECTED    ATGGCCGACA AGCAGAAGAA CGGCATCAAG GTGAACTTCA AGATCCGCCA
YFP-F       .......... .......... .......... .......... ..........
YFP-R       ATGGCCGACA AGCAGAAGAA CGGCATCAAG GTGAACTTCA AGATCCGCCA

                1310       1320       1330       1340       1350
Contig      CAACATCGAG GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA
EXPECTED    CAACATCGAG GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA
YFP-F       .......... .......... .......... .......... ..........
YFP-R       CAACATCGAG GACGGCAGCG TGCAGCTCGC CGACCACTAC CAGCAGAACA

                1360       1370       1380       1390       1400
Contig      CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC
EXPECTED    CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC
YFP-F       .......... .......... .......... .......... ..........
YFP-R       CCCCCATCGG CGACGGCCCC GTGCTGCTGC CCGACAACCA CTACCTGAGC

                1410       1420       1430       1440       1450
Contig      TACCAGTCCG CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT
EXPECTED    TACCAGTCCG CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT
YFP-F       .......... .......... .......... .......... ..........
YFP-R       TACCAGTCCG CCCTGAGCAA AGACCCCAAC GAGAAGCGCG ATCACATGGT

                1460       1470       1480       1490       1500
Contig      CCTGCTGGAG TTCGTGACCG CCGCCGGGAT CACTCTCGGC ATGGACGAGC
EXPECTED    CCTGCTGGAG TTCGTGACCG CCGCCGGGAT CACTCTCGGC ATGGACGAGC
YFP-F       .......... .......... .......... .......... ..........
YFP-R       CCTGCTGGAG TTCGTGACCG CCGCCGGGAT CACTCTCGGC ATGGACGAGC

                1510       1520       1530       1540       1550
Contig      TGTACAAGTA ASGAKS2CGG TRC44422C3 GGATCCGGCT GCTAACAAAG
EXPECTED    TGTACAAGTA AGGATC-CGG TAC-----C- GGATCCGGCT GCTAACAAAG
YFP-F       .......... .......... .......... .......... ..........
YFP-R       TGTACAAGTA ACGAGGGCGG TGCTTTGGCA GGATCCGGCT GCTAACAAAG

                1560       1570       1580       1590       1600
Contig      CCCGAAAGGA A2CTGARTTG GCTGCTGCCA CCGCTGARCA ATAACTAGCA
EXPECTED    CCCGAAAGGA AGCTGAGTTG GCTGCTGCCA CCGCTGAGCA ATAACTAGCA
YFP-F       .......... .......... .......... .......... ..........
YFP-R       CCCGAAAGGA ANCTGAATTG GCTGCTGCCA CCGCTGAACA ATAACTAGCA

                1610       1620       1630       1640
Contig      TAA5CCC5TT GGGGCCTCTA AACGGGTCTT GAGGGGTTTN  1609
EXPECTED    TAACCCC-TT GGGGCCTCTA AACGGGTCTT GAGGGGTTT 1621
```

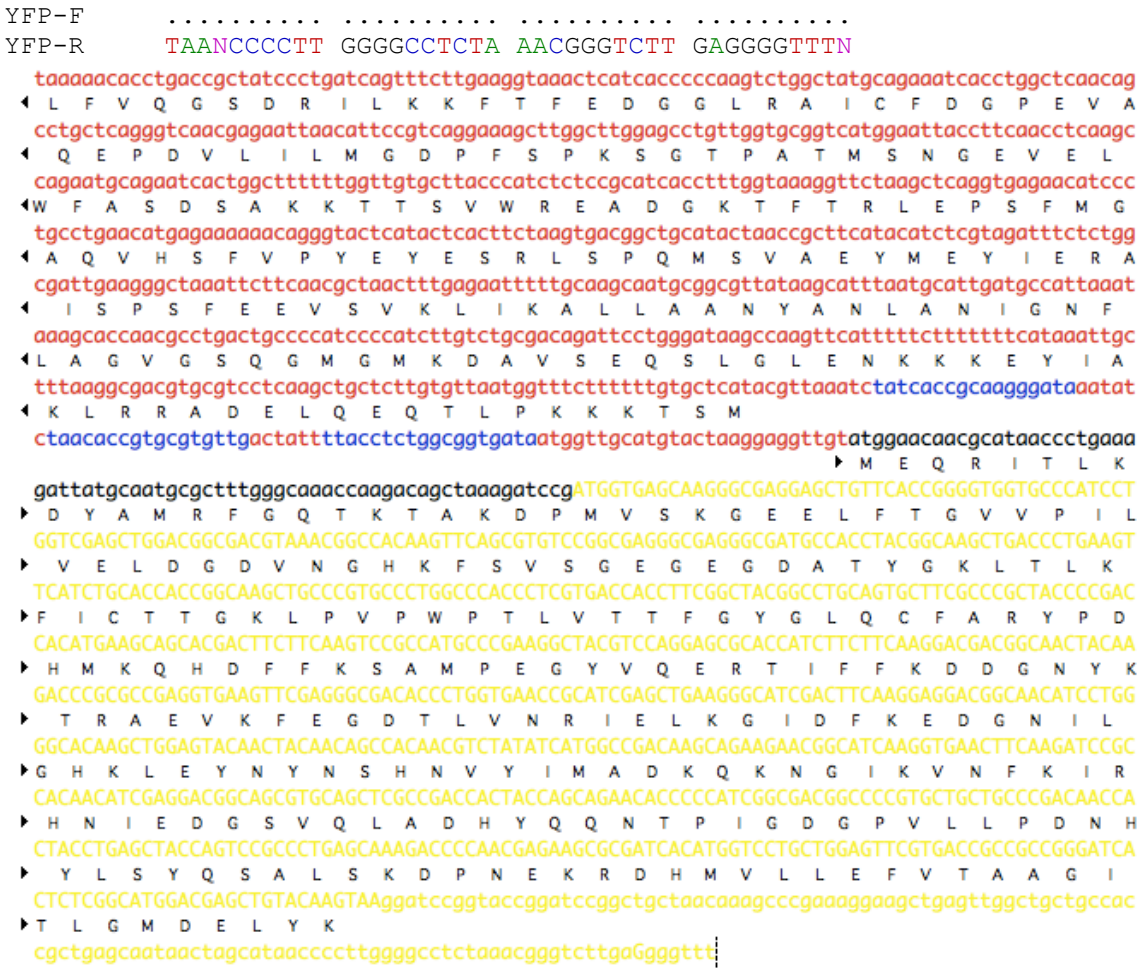

Figure 21. Assembled 1602-A annotated contig with open reading frames and single letter amino acids. In red is cI857, blue are the three operator regions, black the first 22 amino acids of Cro, and finally YFP in yellow.

1605-A. (cI857-Cro-Fusion-DXS).

Contig alignment against expected sequence:

```
                 10         20         30         40         50
Contig     TCCCTTTATT TTTGCTGCGG TAAGTCGCAT AAAAACCATT CTTCATAATT
EXPECTED   TCCCTTTATT TTTGCTGCGG TAAGTCGCAT AAAAACCATT CTTCATAATT
1605—A-F   -----TTATT TTTGCTGCGG TAAGTCGCAT AAAAACCATT CTTCATAATT
1605-A-R   ---------- ---------- ---------- ---------- ----------

                 60         70         80         90         100
Contig     CAATCCATTT ACTATGTTAT GTTCTGAGGG GAG4GAAAAT TCCCCTAATT
EXPECTED   CAATCCATTT ACTATGTTAT GTTCTGAGGG GAGTGAAAAT TCCCCTAATT
1605-A-F   CAATCCATTT ACTATGTTAT GTTCTGAGGG GAGNGAAAAT TCCCCTAATT
1605-A-R   ---------- ---------- ---------- ---------- ----------

                110         120         130         140         150
Contig     CGATGAAGAT TCTTGCTCAA TTGTTATCAG CTATGCGCCG ACCAGAACAC
EXPECTED   CGATGAAGAT TCTTGCTCAA TTGTTATCAG CTATGCGCCG ACCAGAACAC
1605-A-F   CGATGAAGAT TCTTGCTCAA TTGTTATCAG CTATGCGCCG ACCAGAACAC
1605-A-R   ---------- ---------- ---------- ---------- ----------

                160         170         180         190         200
Contig     CTTGCCGATC AGCCAAACGT CTCTTCAGGC CACTGACTAG CGATAACTTT
EXPECTED   CTTGCCGATC AGCCAAACGT CTCTTCAGGC CACTGACTAG CGATAACTTT
1605-A-F   CTTGCCGATC AGCCAAACGT CTCTTCAGGC CACTGACTAG CGATAACTTT
1605-A-R   ---------- ---------- ---------- ---------- ----------

                210         220         230         240         250
Contig     CCCCACAACG GAACAACTCT CATTGCATGG GATCATTGGG TACTGTGGGT
EXPECTED   CCCCACAACG GAACAACTCT CATTGCATGG GATCATTGGG TACTGTGGGT
1605-A-F   CCCCACAACG GAACAACTCT CATTGCATGG GATCATTGGG TACTGTGGGT
1605-A-R   ---------- ---------- ---------- ---------- ----------

                260         270         280         290         300
Contig     TTAGTGGTTG TAAAAACACC TGACCGCTAT CCCTGATCAG TTTCTTGAAG
EXPECTED   TTAGTGGTTG TAAAAACACC TGACCGCTAT CCCTGATCAG TTTCTTGAAG
1605-A-F   TTAGTGGTTG TAAAAACACC TGACCGCTAT CCCTGATCAG TTTCTTGAAG
1605-A-R   ---------- ---------- ---------- ---------- ----------

                310         320         330         340         350
Contig     GTAAACTCAT CACCCCCAAG TCTGGCTATG CAGAAATCAC CTGGCTCAAC
EXPECTED   GTAAACTCAT CACCCCCAAG TCTGGCTATG CAGAAATCAC CTGGCTCAAC
1605-A-F   GTAAACTCAT CACCCCCAAG TCTGGCTATG CAGAAATCAC CTGGCTCAAC
1605-A-R   ---------- ---------- ---------- ---------- ----------

                360         370         380         390         400
Contig     AGCCTGCTCA GGGTCAACGA GAATTAACAT TCCGTCAGGA AAGCTTGGCT
EXPECTED   AGCCTGCTCA GGGTCAACGA GAATTAACAT TCCGTCAGGA AAGCTTGGCT
1605-A-F   AGCCTGCTCA GGGTCAACGA GAATTAACAT TCCGTCAGGA AAGCTTGGCT
1605-A-R   ---------- ---------- ---------- ---------- ----------

                410         420         430         440         450
Contig     TGGAGCCTGT TGGTGCGGTC ATGGAATTAC CTTCAACCTC AAGCCAGAAT
```

```
EXPECTED  TGGAGCCTGT TGGTGCGGTC ATGGAATTAC CTTCAACCTC AAGCCAGAAT
1605-A-F  TGGAGCCTGT TGGTGCGGTC ATGGAATTAC CTTCAACCTC AAGCCAGAAT
1605-A-R  ---------- ---------- ---------- ---------- ----------

                460        470        480        490        500
Contig    GCAGAATCAC TGGCTTTTTT GGTTGTGCTT ACCCATCTCT CCGCATCACC
EXPECTED  GCAGAATCAC TGGCTTTTTT GGTTGTGCTT ACCCATCTCT CCGCATCACC
1605-A-F  GCAGAATCAC TGGCTTTTTT GGTTGTGCTT ACCCATCTCT CCGCATCACC
1605-A-R  ---------- ---------- ---------- ---------- ----------

                510        520        530        540        550
Contig    TTTGGTAAAG GTTCTAAGCT YAGGTGAGAA CATCCCTGCC TGAACATGAG
EXPECTED  TTTGGTAAAG GTTCTAAGCT CAGGTGAGAA CATCCCTGCC TGAACATGAG
1605-A-F  TTTGGTAAAG GTTCTAAGCT TAGGTGAGAA CATCCCTGCC TGAACATGAG
1605-A-R  ---------- ---------- ---------- ---------- ----------

                560        570        580        590        600
Contig    AAAAAACAGG GTACTCATAC TCACTTCTAA GTGACGGCTG CATACTAACC
EXPECTED  AAAAAACAGG GTACTCATAC TCACTTCTAA GTGACGGCTG CATACTAACC
1605-A-F  AAAAAACAGG GTACTCATAC TCACTTCTAA GTGACGGCTG CATACTAACC
1605-A-R  ---------- ---------- ---------- ---------- ----------

                610        620        630        640        650
Contig    GCTTCATACA TCTCGTAGAT TTCTCTGGCG ATTGAAGGGC TAAATTCTTC
EXPECTED  GCTTCATACA TCTCGTAGAT TTCTCTGGCG ATTGAAGGGC TAAATTCTTC
1605-A-F  GCTTCATACA TCTCGTAGAT TTCTCTGGCG ATTGAAGGGC TAAATTCTTC
1605-A-R  ---------- ---------- ---------- ---------- ----------

                660        670        680        690        700
Contig    AACGCTAACT TTGAGAATTT TTGYAAGCAA TGCGGCGTTA TAAGCATTTA
EXPECTED  AACGCTAACT TTGAGAATTT TTGCAAGCAA TGCGGCGTTA TAAGCATTTA
1605-A-F  AACGCTAACT TTGAGAATTT TTGTAAGCAA TGCGGCGTTA TAAGCATTTA
1605-A-R  ---------- ---------- ---------- ---------- ----------

                710        720        730        740        750
Contig    ATGCATTGAT GCCATTAAAT AAAGCACCAA CGCCTGACTG CCCCATCCCC
EXPECTED  ATGCATTGAT GCCATTAAAT AAAGCACCAA CGCCTGACTG CCCCATCCCC
1605-A-F  ATGCATTGAT GCCATTAAAT AAAGCACCAA CGCCTGACTG CCCCATCCCC
1605-A-R  ---------- ---------- ---------- ---------- ----------

                760        770        780        790        800
Contig    ATCTTGTCTG CGACAGATTC CTGGGATAAG CCAA2T4C3T TTT4CTTTTT
EXPECTED  ATCTTGTCTG CGACAGATTC CTGGGATAAG CCAAGTTCAT TTTTCTTTTT
1605-A-F  ATCTTGTCTG CGACAGATTC CTGGGATAAG CCAAGTTCAT TTTTCTTTTT
1605-A-R  ---------- ---------- ---------- --AA-T-C-T TTT-CTTTTT

                810        820        830        840        850
Contig    T4CATAA3TT GCTTTA3GGC GACGTGCGTC CTCAAGCTGC TCTTGTGTTA
EXPECTED  TTCATAAATT GCTTTAAGGC GACGTGCGTC CTCAAGCTGC TCTTGTGTTA
1605-A-F  TTCATAAATT GCTTTAAGGC GACGTGCGTC CTCAAGCTGC TCTTGTGTTA
1605-A-R  T-CATAA-TT GCTTTA-GGC GACGTGCGTC CTCAAGCTGC TCTTGTGTTA
```

```
                860        870        880        890        900
Contig    ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG CAAGGGATAA
EXPECTED  ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG CAAGGGATAA
1605-A-F  ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG CAAGGGATAA
1605-A-R  ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG CAAGGGATAA

                910        920        930        940        950
Contig    ATATCTAACA CCGTGCGTGT TGACTATTTT ACCTCTGGCG GTGATAATGG
EXPECTED  ATATCTAACA CCGTGCGTGT TGACTATTTT ACCTCTGGCG GTGATAATGG
1605-A-F  ATATCTAACA CCGTGCGTGT TGACTATTTT ACCTCTGGCG GTGATAATGG
1605-A-R  ATATCTAACA CCGTGCGTGT TGACTATTTT ACCTCTGGCG GTGATAATGG

                960        970        980        990       1000
Contig    TTGCATGTAC TAAGGAGGTT GTATGGAACA ACGCATAACC CTGAAAGATT
EXPECTED  TTGCATGTAC TAAGGAGGTT GTATGGAACA ACGCATAACC CTGAAAGATT
1605-A-F  TTGCATGTAC TAAGGAGGTT GTATGGAACA ACGCATAACC CTGAAAGATT
1605-A-R  TTGCATGTAC TAAGGAGGTT GTATGGAACA ACGCATAACC CTGAAAGATT

               1010       1020       1030       1040       1050
Contig    ATGCAATGCG CTT4GG2CAA MC35A3GA5A GCTAAAGAT5 5G2ACCTGTT
EXPECTED  ATGCAATGCG CTTTGGGCAA AC-CAAGACA GCTAAAGAT- -G-ACCTGTT
1605-A-F  ATGCAATGCG CTT-GG-CAA CCANA 1022 ......... ..........
1605-A-R  ATGCAATGCG CTTTGGGCAA AC-CAA 1018 TAAAGATC CGGACCTGTT

               1060       1070       1080       1090       1100
Contig    AAGTATTCAA GATCCGAGTT TTTTAAAGAA GATGTCCATT GAGCAACTCG
EXPECTED  AAGTATTCAA GATCCGAGTT TTTTAAAGAA GATGTCCATT GAGCAACTCG
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  AAGTATTCAA GATCCGAGTT TTTTAAAGAA GATGTCCATT GAGCAACTCG

               1110       1120       1130       1140       1150
Contig    AGGAACTCTC TGAAGAAATT CGCAATTTTC TCATCACCAG TCTCAGCGCG
EXPECTED  AGGAACTCTC TGAAGAAATT CGCAATTTTC TCATCACCAG TCTCAGCGCG
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  AGGAACTCTC TGAAGAAATT CGCAATTTTC TCATCACCAG TCTCAGCGCG

               1160       1170       1180       1190       1200
Contig    TCGGGAGGAC ATATTGGGCC GAATCTGGGC GTGGTCGAAT TAACAATTGC
EXPECTED  TCGGGAGGAC ATATTGGGCC GAATCTGGGC GTGGTCGAAT TAACAATTGC
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  TCGGGAGGAC ATATTGGGCC GAATCTGGGC GTGGTCGAAT TAACAATTGC

               1210       1220       1230       1240       1250
Contig    CTTGCACAAA GAATTTGACA GCCCCAAAGA CAAATTTCTG TGGGACGTCG
EXPECTED  CTTGCACAAA GAATTTGACA GCCCCAAAGA CAAATTTCTG TGGGACGTCG
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  CTTGCACAAA GAATTTGACA GCCCCAAAGA CAAATTTCTG TGGGACGTCG

               1260       1270       1280       1290       1300
Contig    GCCACCAGTC GTATGTCCAC AAATTGCTTA CCGGCCGTGG GAAAGAATTT
EXPECTED  GCCACCAGTC GTATGTCCAC AAATTGCTTA CCGGCCGTGG GAAAGAATTT
1605-A-F  .......... .......... .......... .......... ..........
```

```
1605-A-R  GCCACCAGTC GTATGTCCAC AAATTGCTTA CCGGCCGTGG GAAAGAATTT

                1310       1320       1330       1340       1350
Contig    GAAACTCTGC GCCAATACAA AGGGTTGTGC GGGTTCCCTA AACGTAGCGA
EXPECTED  GAAACTCTGC GCCAATACAA AGGGTTGTGC GGGTTCCCTA AACGTAGCGA
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  GAAACTCTGC GCCAATACAA AGGGTTGTGC GGGTTCCCTA AACGTAGCGA

                1360       1370       1380       1390       1400
Contig    AAGTGAACAT GATGTGTGGG AAACGGGCCA TAGTTCCACA AGTTTATCCG
EXPECTED  AAGTGAACAT GATGTGTGGG AAACGGGCCA TAGTTCCACA AGTTTATCCG
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  AAGTGAACAT GATGTGTGGG AAACGGGCCA TAGTTCCACA AGTTTATCCG

                1410       1420       1430       1440       1450
Contig    GGGCGATGGG TATGGCTGCC GCCCGAGACA TTAAAGGCTC GAAAGAATAC
EXPECTED  GGGCGATGGG TATGGCTGCC GCCCGAGACA TTAAAGGCTC GAAAGAATAC
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  GGGCGATGGG TATGGCTGCC GCCCGAGACA TTAAAGGCTC GAAAGAATAC

                1460       1470       1480       1490       1500
Contig    ATCATCCCCA TTATTGGTGA CGGTGCGTTA ACCGGCGGCA TGGCCTTAGA
EXPECTED  ATCATCCCCA TTATTGGTGA CGGTGCGTTA ACCGGCGGCA TGGCCTTAGA
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  ATCATCCCCA TTATTGGTGA CGGTGCGTTA ACCGGCGGCA TGGCCTTAGA

                1510       1520       1530       1540       1550
Contig    GGCGCTCAAC CACATTGGCG ACGAGAAGAA AGATATGATC GTGATCCTGA
EXPECTED  GGCGCTCAAC CACATTGGCG ACGAGAAGAA AGATATGATC GTGATCCTGA
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  GGCGCTCAAC CACATTGGCG ACGAGAAGAA AGATATGATC GTGATCCTGA

                1560       1570       1580       1590       1600
Contig    ATGATAATGA AATGTCCATC GCGCCCAATG TCGGAGCTAT TCACTCCATG
EXPECTED  ATGATAATGA AATGTCCATC GCGCCCAATG TCGGAGCTAT TCACTCCATG
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  ATGATAATGA AATGTCCATC GCGCCCAATG TCGGAGCTAT TCACTCCATG

                1610       1620       1630       1640       1650
Contig    CTGGGGCGCC TTCGGACAGC GGGCAAATAT CAATGGGTGA AA3GATGAAC
EXPECTED  CTGGGGCGCC TTCGGACAGC GGGCAAATAT CAATGGGTGA AA-GATGAAC
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  CTGGGGCGCC TTCGGACAGC GGGCAAATAT CAATGGGTGA AAAGATGAAC

                1660       1670       1680       1690       1700
Contig    TGGAATACTT GTTTAAACGC ATCCCGGCTG TTGGGGG2CA AATTGGCGGC
EXPECTED  TGGAATACTT GTTTAAACGC ATCCCGGCTG TTGGGGG-CA AATTGGCGGC
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  TGGAATACTT GTTTAAACGC ATCCCGGCTG TTGGGGGGCA AATTGGCGGC

                1710       1720       1730       1740       1750
Contig    GACCGCTGAG CGTATTAAAG ATAGTCTGAA GTACATGCTC GTGTCTGGAA
```

```
EXPECTED  GACCGCTGAG CGTATTAAAG ATAGTCTGAA GTACATGCTC GTGTCTGGAA
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  GACCGCTGAG CGTATTAAAG ATAGTCTGAA GTACATGCTC GTGTCTGGAA

               1760       1770       1780       1790       1800
Contig    TGTTTTT4CG AA2AACTCGG CTTTACCTAC CTGGG5CC5G GTTGATGGCC
EXPECTED  TGTTTTT-CG AAGAACTCGG CTTTACCTAC CTGGGCCC-G GTTGATGGCC
1605-A-F  .......... .......... .......... .......... ..........
1605-A-R  TGTTTTTTCG AANAACTCGG CTTTACCTAC CTGGGNCCCG GTTGATGGCC

               1810
Contig    3CTC4TWTCA  1785
EXPECTED  ACTCTTATCA  1802
1605-A-F  ..........
1605-A-R  NCTCNTTT..
```

Figure 22. Assembled 1605-A annotated contig with open reading frames and single letter amino acids. In red is cI857, blue are the three operator regions, black the first 22 amino acids of Cro, and finally DXS in green.

1605-B (cI857-DXS).

Contig alignment against expected sequence:

```
                10         20         30         40         50
Contig    GATTATCAGC CAGCAG3GAA TTAAGGAAAA CAGACAGGTT TATTGAGCGC
EXPECTED  GATTATCAGC CAGCAGAGAA TTAAGGAAAA CAGACAGGTT TATTGAGCGC
1605-B-F  ---------C CAGCAGNGAA TTAAGGAAAA CAGACAGGTT TATTGAGCGC
1605-B-R  ---------- ---------- ---------- ---------- ----------

                60         70         80         90        100
Contig    TTATCTTTCC C4TTATTTTT GCTGCGGTAA GTCGCATAAA AACCATTCTT
EXPECTED  TTATCTTTCC CTTTATTTTT GCTGCGGTAA GTCGCATAAA AACCATTCTT
1605-B-F  TTATCTTTCC CNTTATTTTT GCTGCGGTAA GTCGCATAAA AACCATTCTT
1605-B-R  ---------- ---------- ---------- ---------- ----------

               110        120        130        140        150
Contig    CATAATTCAA TCCATTTACT ATGTTATGTT CTGAGGGGAG 4GAAAATTCC
EXPECTED  CATAATTCAA TCCATTTACT ATGTTATGTT CTGAGGGGAG TGAAAATTCC
1605-B-F  CATAATTCAA TCCATTTACT ATGTTATGTT CTGAGGGGAG NGAAAATTCC
1605-B-R  ---------- ---------- ---------- ---------- ----------

               160        170        180        190        200
Contig    CCTAA3TTCG ATGAAGATTC TTGCTCAATT GTTATCAGCT ATGCGCCGAC
EXPECTED  CCTAA-TTCG ATGAAGATTC TTGCTCAATT GTTATCAGCT ATGCGCCGAC
1605-B-F  CCTAAATTCG ATGAAGATTC TTGCTCAATT GTTATCAGCT ATGCGCCGAC
1605-B-R  ---------- ---------- ---------- ---------- ----------

               210        220        230        240        250
Contig    CAGAACACCT TGCCGATCAG CCAAACGTCT CTTCAGGCCA CTGACTAGCG
EXPECTED  CAGAACACCT TGCCGATCAG CCAAACGTCT CTTCAGGCCA CTGACTAGCG
1605-B-F  CAGAACACCT TGCCGATCAG CCAAACGTCT CTTCAGGCCA CTGACTAGCG
1605-B-R  ---------- ---------- ---------- ---------- ----------

               260        270        280        290        300
Contig    ATAACTTTCC CCACAACGGA ACAACTCTCA TTGCATGGGA TCATTGGGTA
EXPECTED  ATAACTTTCC CCACAACGGA ACAACTCTCA TTGCATGGGA TCATTGGGTA
1605-B-F  ATAACTTTCC CCACAACGGA ACAACTCTCA TTGCATGGGA TCATTGGGTA
1605-B-R  ---------- ---------- ---------- ---------- ----------

               310        320        330        340        350
Contig    CTGTGGGTTT AGTGGTTGTA AAAACACCTG ACCGCTATCC CTGATCAGTT
EXPECTED  CTGTGGGTTT AGTGGTTGTA AAAACACCTG ACCGCTATCC CTGATCAGTT
1605-B-F  CTGTGGGTTT AGTGGTTGTA AAAACACCTG ACCGCTATCC CTGATCAGTT
1605-B-R  ---------- ---------- ---------- ---------- ----------

               360        370        380        390        400
Contig    TCTTGAAGGT AAACTCATCA CCCCCAAGTC TGGCTATGCA GAAATCACCT
EXPECTED  TCTTGAAGGT AAACTCATCA CCCCCAAGTC TGGCTATGCA GAAATCACCT
1605-B-F  TCTTGAAGGT AAACTCATCA CCCCCAAGTC TGGCTATGCA GAAATCACCT
1605-B-R  ---------- ---------- ---------- ---------- ----------

               410        420        430        440        450
Contig    GGCTCAACAG CCTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA
```

```
EXPECTED  GGCTCAACAG CCTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA
1605-B-F  GGCTCAACAG CCTGCTCAGG GTCAACGAGA ATTAACATTC CGTCAGGAAA
1605-B-R  ---------- ---------- ---------- ---------- ----------

                460        470        480        490        500
Contig    GCTTGGCTTG GAGCCTGTTG GTGCGGTCAT GGAATTACCT TCAACCTCAA
EXPECTED  GCTTGGCTTG GAGCCTGTTG GTGCGGTCAT GGAATTACCT TCAACCTCAA
1605-B-F  GCTTGGCTTG GAGCCTGTTG GTGCGGTCAT GGAATTACCT TCAACCTCAA
1605-B-R  ---------- ---------- ---------- ---------- ----------

                510        520        530        540        550
Contig    GCCAGAATGC AGAATCACTG GCTTTTTTGG TTGTGCTTAC CCATCTCTCC
EXPECTED  GCCAGAATGC AGAATCACTG GCTTTTTTGG TTGTGCTTAC CCATCTCTCC
1605-B-F  GCCAGAATGC AGAATCACTG GCTTTTTTGG TTGTGCTTAC CCATCTCTCC
1605-B-R  ---------- ---------- ---------- ---------- ----------

                560        570        580        590        600
Contig    GCATCACCTT TGGTAAAGGT TCTAAGCTYA GGTGAGAACA TCCCTGCCTG
EXPECTED  GCATCACCTT TGGTAAAGGT TCTAAGCTCA GGTGAGAACA TCCCTGCCTG
1605-B-F  GCATCACCTT TGGTAAAGGT TCTAAGCTTA GGTGAGAACA TCCCTGCCTG
1605-B-R  ---------- ---------- ---------- ---------- ----------

                610        620        630        640        650
Contig    AACATGAGAA AAAACAGGGT ACTCATACTC ACTTCTAAGT GACGGCTGCA
EXPECTED  AACATGAGAA AAAACAGGGT ACTCATACTC ACTTCTAAGT GACGGCTGCA
1605-B-F  AACATGAGAA AAAACAGGGT ACTCATACTC ACTTCTAAGT GACGGCTGCA
1605-B-R  ---------- ---------- ---------- ---------- ----------

                660        670        680        690        700
Contig    TACTAACCGC TTCATACATC TCGTAGATTT CTCTGGCGAT TGAAGGGCTA
EXPECTED  TACTAACCGC TTCATACATC TCGTAGATTT CTCTGGCGAT TGAAGGGCTA
1605-B-F  TACTAACCGC TTCATACATC TCGTAGATTT CTCTGGCGAT TGAAGGGCTA
1605-B-R  ---------- ---------- ---------- ---------- ----------

                710        720        730        740        750
Contig    AATTCTTCAA CGCTAACTTT GAGAATTTTT GYAAGCAATG CGGCGTTATA
EXPECTED  AATTCTTCAA CGCTAACTTT GAGAATTTTT GCAAGCAATG CGGCGTTATA
1605-B-F  AATTCTTCAA CGCTAACTTT GAGAATTTTT GTAAGCAATG CGGCGTTATA
1605-B-R  ---------- ---------- ---------- ---------- ----------

                760        770        780        790        800
Contig    AGCATTTAAT GCATTGATGC CATTAAATAA AGCACCAACG CCTGACTGCC
EXPECTED  AGCATTTAAT GCATTGATGC CATTAAATAA AGCACCAACG CCTGACTGCC
1605-B-F  AGCATTTAAT GCATTGATGC CATTAAATAA AGCACCAACG CCTGACTGCC
1605-B-R  ---------- ---------- ---------- ---------- ----------

                810        820        830        840        850
Contig    CCATCCCCAT CTTGTCTGCG ACAGATTCCT GGGATAAGCC AAGTTCA3T5
EXPECTED  CCATCCCCAT CTTGTCTGCG ACAGATTCCT GGGATAAGCC AAGTTCA-T-
1605-B-F  CCATCCCCAT CTTGTCTGCG ACAGATTCCT GGGATAAGCC AAGTTCA-T-
1605-B-R  ---------- ---------- ---------- ---------- ------AATC
```

```
                860         870         880         890         900
Contig    TTTTCTTTTT T4CATAAATT GCTTTA3GGC GACGTGCGTC CTCAAGCTGC
EXPECTED  TTTTCTTTTT TTCATAAATT GCTTTAAGGC GACGTGCGTC CTCAAGCTGC
1605-B-F  TTTTCTTTTT TTCATAAATT GCTTTAAGGC GACGTGCGTC CTCAAGCTGC
1605-B-R  TTTTCTTTTT T-CATAAATT GCTTTA-GGC GACGTGCGTC CTCAAGCTGC

                910         920         930         940         950
Contig    TCTTGTGTTA ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG
EXPECTED  TCTTGTGTTA ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG
1605-B-F  TCTTGTGTTA ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG
1605-B-R  TCTTGTGTTA ATGGTTTCTT TTTTGTGCTC ATACGTTAAA TCTATCACCG

                960         970         980         990        1000
Contig    CAAGGGATAA ATATCTAA55 ACCGTGCGTG TTGACTATTT TACCTCTGGC
EXPECTED  CAAGGGATAA ATATCTAAC- ACCGTGCGTG TTGACTATTT TACCTCTGGC
1605-B-F  CAAGGGATAA ATATCTAANC ACCGTGCGTG TTGACTATTT TACCTCTGGC
1605-B-R  CAAGGGATAA ATATCTAAC- ACCGTGCGTG TTGACTATTT TACCTCTGGC

               1010        1020        1030        1040        1050
Contig    GGTGATAATG GT4GCAN3TG TACTAAGGAG GTTGTATGGA CCTGTTAAGT
EXPECTED  GGTGATAATG GTTGCA--TG TACTAAGGAG GTTGTATGGA CCTGTTAAGT
1605-B-F  GGTGATAAT 1006 ANA.. .......... .......... ..........
1605-B-R  GGTGATAATG GTTGCA--TG TACTAAGGAG GTTGT 1030 CTGTTAAGT

               1060        1070        1080        1090        1100
Contig    ATTCAAGATC CGAGTTTTTT AAAGAAGATG TCCATTGAGC AACTCGAGGA
EXPECTED  ATTCAAGATC CGAGTTTTTT AAAGAAGATG TCCATTGAGC AACTCGAGGA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  ATTCAAGATC CGAGTTTTTT AAAGAAGATG TCCATTGAGC AACTCGAGGA

               1110        1120        1130        1140        1150
Contig    ACTCTCTGAA GAAATTCGCA ATTTTCTCAT CACCAGTCTC AGCGCGTCGG
EXPECTED  ACTCTCTGAA GAAATTCGCA ATTTTCTCAT CACCAGTCTC AGCGCGTCGG
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  ACTCTCTGAA GAAATTCGCA ATTTTCTCAT CACCAGTCTC AGCGCGTCGG

               1160        1170        1180        1190        1200
Contig    GAGGACATAT TGGGCCGAAT CTGGGCGTGG TCGAATTAAC AATTGCCTTG
EXPECTED  GAGGACATAT TGGGCCGAAT CTGGGCGTGG TCGAATTAAC AATTGCCTTG
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  GAGGACATAT TGGGCCGAAT CTGGGCGTGG TCGAATTAAC AATTGCCTTG

               1210        1220        1230        1240        1250
Contig    CACAAAGAAT TTGACAGCCC CAAAGACAAA TTTCTGTGGG ACGTCGGCCA
EXPECTED  CACAAAGAAT TTGACAGCCC CAAAGACAAA TTTCTGTGGG ACGTCGGCCA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  CACAAAGAAT TTGACAGCCC CAAAGACAAA TTTCTGTGGG ACGTCGGCCA

               1260        1270        1280        1290        1300
Contig    CCAGTCGTAT GTCCACAAAT TGCTTACCGG CCGTGGGAAA GAATTTGAAA
EXPECTED  CCAGTCGTAT GTCCACAAAT TGCTTACCGG CCGTGGGAAA GAATTTGAAA
1605-B-F  .......... .......... .......... .......... ..........
```

```
1605-B-R  CCAGTCGTAT GTCCACAAAT TGCTTACCGG CCGTGGGAAA GAATTTGAAA

                1310       1320       1330       1340       1350
Contig    CTCTGCGCCA ATACAAAGGG TTGTGCGGGT TCCCTAAACG TAGCGAAAGT
EXPECTED  CTCTGCGCCA ATACAAAGGG TTGTGCGGGT TCCCTAAACG TAGCGAAAGT
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  CTCTGCGCCA ATACAAAGGG TTGTGCGGGT TCCCTAAACG TAGCGAAAGT

                1360       1370       1380       1390       1400
Contig    GAACATGATG TGTGGGAAAC GGGCCATAGT TCCACAAGTT TATCCGGGGC
EXPECTED  GAACATGATG TGTGGGAAAC GGGCCATAGT TCCACAAGTT TATCCGGGGC
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  GAACATGATG TGTGGGAAAC GGGCCATAGT TCCACAAGTT TATCCGGGGC

                1410       1420       1430       1440       1450
Contig    GATGGGTATG GCTGCCGCCC GAGACATTAA AGGCTCGAAA GAATACATCA
EXPECTED  GATGGGTATG GCTGCCGCCC GAGACATTAA AGGCTCGAAA GAATACATCA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  GATGGGTATG GCTGCCGCCC GAGACATTAA AGGCTCGAAA GAATACATCA

                1460       1470       1480       1490       1500
Contig    TCCCCATTAT TGGTGACGGT GCGTTAACCG GCGGCATGGC CTTAGAGGCG
EXPECTED  TCCCCATTAT TGGTGACGGT GCGTTAACCG GCGGCATGGC CTTAGAGGCG
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TCCCCATTAT TGGTGACGGT GCGTTAACCG GCGGCATGGC CTTAGAGGCG

                1510       1520       1530       1540       1550
Contig    CTCAACCACA TTGGCGACGA GAAGAAAGAT ATGATCGTGA TCCTGAATGA
EXPECTED  CTCAACCACA TTGGCGACGA GAAGAAAGAT ATGATCGTGA TCCTGAATGA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  CTCAACCACA TTGGCGACGA GAAGAAAGAT ATGATCGTGA TCCTGAATGA

                1560       1570       1580       1590       1600
Contig    TAATGAAATG TCCATCGCGC CCAATGTCGG AGCTATTCAC TCCATGCTGG
EXPECTED  TAATGAAATG TCCATCGCGC CCAATGTCGG AGCTATTCAC TCCATGCTGG
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TAATGAAATG TCCATCGCGC CCAATGTCGG AGCTATTCAC TCCATGCTGG

                1610       1620       1630       1640       1650
Contig    GGCGCCTTCG GACAGCGGGC AAATATCAAT GGGTGAAAGA TGAACTGGAA
EXPECTED  GGCGCCTTCG GACAGCGGGC AAATATCAAT GGGTGAAAGA TGAACTGGAA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  GGCGCCTTCG GACAGCGGGC AAATATCAAT GGGTGAAAGA TGAACTGGAA

                1660       1670       1680       1690       1700
Contig    TACTTGTTTA AACGCATCCC GGCTGTTGGG GGCAAATTGG CGGCGACCGC
EXPECTED  TACTTGTTTA AACGCATCCC GGCTGTTGGG GGCAAATTGG CGGCGACCGC
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TACTTGTTTA AACGCATCCC GGCTGTTGGG GGCAAATTGG CGGCGACCGC

                1710       1720       1730       1740       1750
Contig    TGAGCGTATT AAAGATAGTC TGAAGTACAT GCTCGTGTCT GGAATGTTTT
```

```
EXPECTED  TGAGCGTATT AAAGATAGTC TGAAGTACAT GCTCGTGTCT GGAATGTTTT
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TGAGCGTATT AAAGATAGTC TGAAGTACAT GCTCGTGTCT GGAATGTTTT

               1760       1770       1780       1790       1800
Contig    TCGAAGAACT CGGCTTTACC TACCTGGGCC C5GGTTGATG GCCACTCTTA
EXPECTED  TCGAAGAACT CGGCTTTACC TACCTGGGCC C-GGTTGATG GCCACTCTTA
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TCGAAGAACT CGGCTTTACC TACCTGGGCC CCGGTTGATG GCCACTCTTA

               1810       1820       1830       1840       1850
Contig    TCACGAATTG TTTGAAAACC 5TGCAGTATG CAAAGAAAAC TAAGGGGCCC
EXPECTED  TCACGAATTG TTTGAAAACC -TGCAGTATG CAAAGAAAAC TAAGGGGCCC
1605-B-F  .......... .......... .......... .......... ..........
1605-B-R  TCACGAATTG TTTGAAAACC CTGCAGTATG CAAAGAAAAC TAAGGGGCCC

               1860       1870       1880
Contig    GTGC5TCTTG C3CGTCATTA CCAA2AA2GG 2AAS 1865
EXPECTED  GTGC-TCTTG CACGTCATTA CCAAGAAGGG -AAS 1874
1605-B-F  .......... .......... .......... ....
1605-B-R  GTGCCTCTTG CNCGTCATTA CCAANAANGG G...
```

Figure 23. Assembled 1605-B annotated contig with open reading frames and single letter amino acids. In red is cI857, blue are the three operator regions, and finally DXS in green.

1602-B (cI857-Cro-DXS-IspF-IspH).

```
                    10         20         30         40         50
Contig     TTGTT3TCAG CTATGCGCCG ACCAGAACAC CTTGCCGATC AGCCAAACGT
EXPECTED   TTGTTATCAG CTATGCGCCG ACCAGAACAC CTTGCCGATC AGCCAAACGT
1602-B-F   ---TT-TCAG CTATGCGCCG ACCAGAACAC CTTGCCGATC AGCCAAACGT
1602-B-R   ---------- ---------- ---------- ---------- ----------

                    60         70         80         90         100
```

```
Contig    CT5TTCAGGC CACTGACTAG CGATAACTTT CCCC5AC3AC GGAACAACTC
EXPECTED  CTCTTCAGGC CACTGACTAG CGATAACTTT CCCC-ACAAC GGAACAACTC
1602-B-F  CTNTTCAGGC CACTGACTAG CGATAACTTT CCCCCACNAC GGAACAACTC
1602-B-R  ---------- ---------- ---------- ---------- ----------

                110        120        130        140        150
Contig    TCATTGCATG G2ATCATTGG GTACTGTGGG TTTAGTGGTT GTAAAAACAC
EXPECTED  TCATTGCATG GGATCATTGG GTACTGTGGG TTTAGTGGTT GTAAAAACAC
1602-B-F  TCATTGCATG GNATCATTGG GTACTGTGGG TTTAGTGGTT GTAAAAACAC
1602-B-R  ---------- ---------- ---------- ---------- ----------

                160        170        180        190        200
Contig    CTGACCGCTA TCCCTGATCA GTTTCTTGAA GGTAAACTCA TCACCCCCAA
EXPECTED  CTGACCGCTA TCCCTGATCA GTTTCTTGAA GGTAAACTCA TCACCCCCAA
1602-B-F  CTGACCGCTA TCCCTGATCA GTTTCTTGAA GGTAAACTCA TCACCCCCAA
1602-B-R  ---------- ---------- ---------- ---------- ----------

                210        220        230        240        250
Contig    GTCTGGCTAT GCAGAAATCA CCTGGCTCAA CAGCCTGCTC AGGGTCAACG
EXPECTED  GTCTGGCTAT GCAGAAATCA CCTGGCTCAA CAGCCTGCTC AGGGTCAACG
1602-B-F  GTCTGGCTAT GCAGAAATCA CCTGGCTCAA CAGCCTGCTC AGGGTCAACG
1602-B-R  ---------- ---------- ---------- ---------- ----------

                260        270        280        290        300
Contig    AGAATTAACA TTCCGTCAGG AAAGCTTGGC TTGGAGCCTG TTGGTGCGGT
EXPECTED  AGAATTAACA TTCCGTCAGG AAAGCTTGGC TTGGAGCCTG TTGGTGCGGT
1602-B-F  AGAATTAACA TTCCGTCAGG AAAGCTTGGC TTGGAGCCTG TTGGTGCGGT
1602-B-R  ---------- ---------- ---------- ---------- ----------

                310        320        330        340        350
Contig    CATGGAATTA CCTTCAACCT CAAGCCAGAA TGCAGAATCA CTGGCTTTTT
EXPECTED  CATGGAATTA CCTTCAACCT CAAGCCAGAA TGCAGAATCA CTGGCTTTTT
1602-B-F  CATGGAATTA CCTTCAACCT CAAGCCAGAA TGCAGAATCA CTGGCTTTTT
1602-B-R  ---------- ---------- ---------- ---------- ----------

                360        370        380        390        400
Contig    TGGTTGTGCT TACCCATCTC TCCGCATCAC CTTTGGTAAA GGTTCTAAGC
EXPECTED  TGGTTGTGCT TACCCATCTC TCCGCATCAC CTTTGGTAAA GGTTCTAAGC
1602-B-F  TGGTTGTGCT TACCCATCTC TCCGCATCAC CTTTGGTAAA GGTTCTAAGC
1602-B-R  ---------- ---------- ---------- ---------- ----------

                410        420        430        440        450
Contig    TYAGGTGAGA ACATCCCTGC CTGAACATGA GAAAAAACAG GGTACTCATA
EXPECTED  TCAGGTGAGA ACATCCCTGC CTGAACATGA GAAAAAACAG GGTACTCATA
1602-B-F  TTAGGTGAGA ACATCCCTGC CTGAACATGA GAAAAAACAG GGTACTCATA
1602-B-R  ---------- ---------- ---------- ---------- ----------

                460        470        480        490        500
Contig    CTCACTTCTA AGTGACGGCT GCATACTAAC CGCTTCATAC ATCTCGTAGA
EXPECTED  CTCACTTCTA AGTGACGGCT GCATACTAAC CGCTTCATAC ATCTCGTAGA
1602-B-F  CTCACTTCTA AGTGACGGCT GCATACTAAC CGCTTCATAC ATCTCGTAGA
1602-B-R  ---------- ---------- ---------- ---------- ----------
```

```
                510        520        530        540        550
Contig    TTTCTCTGGC GATTGAAGGG CTAAATTCTT CAACGCTAAC TTTGAGAATT
EXPECTED  TTTCTCTGGC GATTGAAGGG CTAAATTCTT CAACGCTAAC TTTGAGAATT
1602-B-F  TTTCTCTGGC GATTGAAGGG CTAAATTCTT CAACGCTAAC TTTGAGAATT
1602-B-R  ---------- ---------- ---------- ---------- ----------

                560        570        580        590        600
Contig    TTTGYAAGCA ATGCGGCGTT ATAAGCATTT AATGCATTGA TGCCATTAAA
EXPECTED  TTTGCAAGCA ATGCGGCGTT ATAAGCATTT AATGCATTGA TGCCATTAAA
1602-B-F  TTTGTAAGCA ATGCGGCGTT ATAAGCATTT AATGCATTGA TGCCATTAAA
1602-B-R  ---------- ---------- ---------- ---------- ----------

                610        620        630        640        650
Contig    TAAAGCACCA ACGCCTGACT GCCCCATCCC CATCTTGTCT GCGACAGATT
EXPECTED  TAAAGCACCA ACGCCTGACT GCCCCATCCC CATCTTGTCT GCGACAGATT
1602-B-F  TAAAGCACCA ACGCCTGACT GCCCCATCCC CATCTTGTCT GCGACAGATT
1602-B-R  ---------- ---------- ---------- ---------- ----------

                660        670        680        690        700
Contig    CCTGGGATAA GCCAA2T4C3 TTTT45TTTT TT4C3TAAAT TGCTTTA3GG
EXPECTED  CCTGGGATAA GCCAAGTTCA TTTTTCTTTT TTTCATAAAT TGCTTTAAGG
1602-B-F  CCTGGGATAA GCCAAGTTCA TTTTTCTTTT TTTCATAAAT TGCTTTAAGG
1602-B-R  ---------- ---AA-T-C- TTTTN-TTTT TT-C-TAAAT TGCTTTA-GG

                710        720        730        740        750
Contig    CGACGTGCGT CCTCAAGCTG CTCTTGTGTT AATGGTTTCT TTTTTGTGCT
EXPECTED  CGACGTGCGT CCTCAAGCTG CTCTTGTGTT AATGGTTTCT TTTTTGTGCT
1602-B-F  CGACGTGCGT CCTCAAGCTG CTCTTGTGTT AATGGTTTCT TTTTTGTGCT
1602-B-R  CGACGTGCGT CCTCAAGCTG CTCTTGTGTT AATGGTTTCT TTTTTGTGCT

                760        770        780        790        800
Contig    CATACGTTAA ATCTATCACC GCAAGGGATA AATATCTAAC ACCGTGCGTG
EXPECTED  CATACGTTAA ATCTATCACC GCAAGGGATA AATATCTAAC ACCGTGCGTG
1602-B-F  CATACGTTAA ATCTATCACC GCAAGGGATA AATATCTAAC ACCGTGCGTG
1602-B-R  CATACGTTAA ATCTATCACC GCAAGGGATA AATATCTAAC ACCGTGCGTG

                810        820        830        840        850
Contig    TTGACTATTT TACCTCTGGC GGTGATAATG GTTGCATGTA CTAAGGAGGT
EXPECTED  TTGACTATTT TACCTCTGGC GGTGATAATG GTTGCATGTA CTAAGGAGGT
1602-B-F  TTGACTATTT TACCTCTGGC GGTGATAATG GTTGCATGTA CTAAGGAGGT
1602-B-R  TTGACTATTT TACCTCTGGC GGTGATAATG GTTGCATGTA CTAAGGAGGT

                860        870        880        890        900
Contig    TGTATGGAAC AACGCATAAC CCTGAAAGAT TATGCAATGC GCTTTGGGCA
EXPECTED  TGTATGGAAC AACGCATAAC CCTGAAAGAT TATGCAATGC GCTTTGGGCA
1602-B-F  TGTATGGAAC AACGCATAAC CCTGAAAGAT TATGCAATGC GCTTTGGGCA
1602-B-R  TGTATGGAAC AACGCATAAC CCTGAAAGAT TATGCAATGC GCTTTGGGCA

                910        920        930        940        950
Contig    AACCAAGACA GCTAAAGATC CGGAATTCAG GAGCTAGAAC TGGTCAGGGC
EXPECTED  AACCAAGACA GCTAAAGATC CGGAATTCAG GAGCTAGAAC TGGTCAGGGC
```

```
1602-B-F  AACCAAGACA GCTAAAGATC CGGAATTCAG GAGCTAGAAC TGGTCAGGGC
1602-B-R  AACCAAGACA GCTAAAGATC CGGAATTCAG GAGCTAGAAC TGGTCAGGGC

               960        970        980        990       1000
Contig    TGGGGCAATT TTTAATTATT GTTACGCAGG TCTTGCCTAG GGGGGGGGAG
EXPECTED  TGGGGCAATT TTTAATTATT GTTACGCAGG TCTTGCCTAG GGGGGGGGAG
1602-B-F  TGGGGCAATT TTTAATTATT GTTACGCAGG TCTTGCCTAG GGGGGGGGAG
1602-B-R  TGGGGCAATT TTTAATTATT GTTACGCAGG TCTTGCCTAG GGGGGGGGAG

              1010       1020       1030       1040       1050
Contig    GCCGTATTAT CTTCTAGTGA TGTT4GC4GA AAACGC54GA 3G2AGAA4AA
EXPECTED  GCCGTATTAT CTTCTAGTGA TGTTTGCTGA AAACGCCTGA AGGAGAATAA
1602-B-F  GCCGTATTAT CTTCTAGTGA TGTT-GC-GA AAACGCN-GA -G-AGA 1039 .
1602-B-R  GCCGTATTAT CTTCTAGTGA TGTTTGCTGA AAACGCC 1030 AGAATAA

              1060       1070       1080       1090       1100
Contig    CATATGGACC TGTTAAGTAT TCAAGATCCG AGTTTTTTAA AGAAGATGTC
EXPECTED  CATATGGACC TGTTAAGTAT TCAAGATCCG AGTTTTTTAA AGAAGATGTC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  CATATGGACC TGTTAAGTAT TCAAGATCCG AGTTTTTTAA AGAAGATGTC

              1110       1120       1130       1140       1150
Contig    CATTGAGCAA CTCGAGGAAC TCTCTGAAGA AATTCGCAAT TTTCTCATCA
EXPECTED  CATTGAGCAA CTCGAGGAAC TCTCTGAAGA AATTCGCAAT TTTCTCATCA
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  CATTGAGCAA CTCGAGGAAC TCTCTGAAGA AATTCGCAAT TTTCTCATCA

              1160       1170       1180       1190       1200
Contig    CCAGTCTCAG CGCGTCGGGA GGACATATTG GGCCGAATCT GGGCGTGGTC
EXPECTED  CCAGTCTCAG CGCGTCGGGA GGACATATTG GGCCGAATCT GGGCGTGGTC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  CCAGTCTCAG CGCGTCGGGA GGACATATTG GGCCGAATCT GGGCGTGGTC

              1210       1220       1230       1240       1250
Contig    GAATTAACAA TTGCCTTGCA CAAAGAATTT GACAGCCCCA AAGACAAATT
EXPECTED  GAATTAACAA TTGCCTTGCA CAAAGAATTT GACAGCCCCA AAGACAAATT
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  GAATTAACAA TTGCCTTGCA CAAAGAATTT GACAGCCCCA AAGACAAATT

              1260       1270       1280       1290       1300
Contig    TCTGTGGGAC GTCGGCCACC AGTCGTATGT CCACAAATTG CTTACCGGCC
EXPECTED  TCTGTGGGAC GTCGGCCACC AGTCGTATGT CCACAAATTG CTTACCGGCC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  TCTGTGGGAC GTCGGCCACC AGTCGTATGT CCACAAATTG CTTACCGGCC

              1310       1320       1330       1340       1350
Contig    GTGGGAAAGA ATTTGAAACT CTGCGCCAAT ACAAAGGGTT GTGCGGGTTC
EXPECTED  GTGGGAAAGA ATTTGAAACT CTGCGCCAAT ACAAAGGGTT GTGCGGGTTC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  GTGGGAAAGA ATTTGAAACT CTGCGCCAAT ACAAAGGGTT GTGCGGGTTC

              1360       1370       1380       1390       1400
```

```
Contig    CCTAAACGTA GCGAAAGTGA ACATGATGTG TGGGAAACGG GCCATAGTTC
EXPECTED  CCTAAACGTA GCGAAAGTGA ACATGATGTG TGGGAAACGG GCCATAGTTC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  CCTAAACGTA GCGAAAGTGA ACATGATGTG TGGGAAACGG GCCATAGTTC

               1410       1420       1430       1440       1450
Contig    CACAAGTTTA TCCGGGGCGA TGGGTATGGC TGCCGCCCGA GACATTAAAG
EXPECTED  CACAAGTTTA TCCGGGGCGA TGGGTATGGC TGCCGCCCGA GACATTAAAG
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  CACAAGTTTA TCCGGGGCGA TGGGTATGGC TGCCGCCCGA GACATTAAAG

               1460       1470       1480       1490       1500
Contig    GCTCGAAAGA ATACATCATC CCCATTATTG GTGACGGTGC GTTAACCGGC
EXPECTED  GCTCGAAAGA ATACATCATC CCCATTATTG GTGACGGTGC GTTAACCGGC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  GCTCGAAAGA ATACATCATC CCCATTATTG GTGACGGTGC GTTAACCGGC

               1510       1520       1530       1540       1550
Contig    GGCATGGCCT TAGAGGCGCT CAACCACATT GGC233CGAG AAGAAAGATA
EXPECTED  GGCATGGCCT TAGAGGCGCT CAACCACATT GGCGA-CGAG AAGAAAGATA
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  GGCATGGCCT TAGAGGCGCT CAACCACATT GGCNNACGAG AAGAAAGATA

               1560       1570       1580       1590       1600
Contig    TGATCGTGAT CCTGAATGAT AATGAAATGT CCATC2CGCC CAATGTCGG2
EXPECTED  TGATCGTGAT CCTGAATGAT AATGAAATGT CCATCGCGCC CAATGTCGG-
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  TGATCGTGAT CCTGAATGAT AATGAAATGT CCATCNCGCC CAATGTCGGG

               1610       1620       1630       1640       1650
Contig    A2CTATTC3C 5TCCATGCT2 GGG2CSCCTT CGGACAGCGG GCAAATATCA
EXPECTED  AGCTATTCAC -TCCATGCTG GGG-CGCCTT CGGACAGCGG GCAAATATCA
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  ANCTATTCNC CTCCATGCTN GGGGCCCCTT CGGACAGCGG GCAAATATCA

               1660       1670       1680       1690       1700
Contig    ATGGGKG3AA 32A44GAA55 TGGAA44ACT TGTTT3AACS C53TCCC2GC
EXPECTED  ATGGGTGAAA -GAT-GAAC- TGGAAT-ACT TGTTTAAACG C-ATCCCGGC
1602-B-F  .......... .......... .......... .......... ..........
1602-B-R  ATGGGGGNAA ANANTGAANC TGGAANTACT TGTTTNAACC CCNTCCCNGC


Contig    T 1658
EXPECTED  T 1691
1602-B-F  .
1602-B-R  .
```

Figure 24. Assembled 1602-B annotated contig with open reading frames and single letter amino acids. In red is cI857, blue are the three operator regions, in cyan is the ~100 bp of extra DNA, and finally DXS in green.

REFERENCES

1. Paris Agreement - European Commission. *Eur. Com.*, (available at http://ec.europa.eu/clima/policies/international/negotiations/paris/index_en.htm).

2. International Energy Agency, *World energy outlook 2010.* (International Energy Agency, Paris, 2010).

3. What is biofuel? definition and meaning. *BusinessDictionary.com*, (available at http://www.businessdictionary.com/definition/biofuel.html).

4. S. N. Naik, V. V. Goud, P. K. Rout, A. K. Dalai, Production of first and second generation biofuels: A comprehensive review. *Renew. Sustain. Energy Rev.* **14**, 578–597 (2010).

5. H. Omidvarborna, A. Kumar, D.-S. Kim, Characterization of particulate matter emitted from transit buses fueled with B20 in idle modes. *J. Environ. Chem. Eng.* **2**, 2335–2342 (2014).

6. Y. Chisti, Biodiesel from microalgae beats bioethanol. *Trends Biotechnol.* **26**, 126–131 (2008).

7. J. Dexter, P. Armshaw, C. Sheahan, J. T. Pembroke, The state of autotrophic ethanol production in Cyanobacteria. *J. Appl. Microbiol.* **119**, 11–24 (2015).

8. M. A. Martin, First generation biofuels compete. *New Biotechnol.* **27**, 596–608 (2010).

9. B. Antizar-Ladislao, J. L. Turrion-Gomez, Second-generation biofuels and local bioenergy systems. *Biofuels Bioprod. Biorefining*. **2**, 455–469 (2008).

10. J. A. Ramirez, R. J. Brown, T. J. Rainey, A Review of Hydrothermal Liquefaction Bio-Crude Properties and Prospects for Upgrading to Transportation Fuels. *Energies*. **8**, 6765–6794 (2015).

11. Svandyk, The potential and challenges of drop-in fuels (members only) | IEA Bioenergy Task 39 – Commercializing Liquid Biofuels, (available at http://task39.org/2014/01/the-potential-and-challenges-of-drop-in-fuels-members-only/).

12. V. Plata, V. Kafarov, N. Moreno, Optimization of Third Generation Biofuels Production: Biodiesel from Microalgae Oil by Homogeneous Transesterification. *Chem. Eng. Trans.* **21** (2010), doi:10.3303/CET1021201.

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

14. J. P. Maity, J. Bundschuh, C.-Y. Chen, P. Bhattacharya, Microalgae for third generation biofuel production, mitigation of greenhouse gas emissions and wastewater treatment: Present and future perspectives – A mini review. *Energy*. **78**, 104–113 (2014).

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

16. L. R. Lynd *et al.*, How biotech can transform biofuels. *Nat. Biotechnol.* **26**, 169–172 (2008).

17. A. Richmond, *Handbook of Microalgal Culture: Applied Phycology and Biotechnology, Second Edition* (2013; https://www.researchgate.net/publication/277691223_Handbook_of_Microalgal_Culture_Applied_Phycology_and_Biotechnology_Second_Edition).

18. R. I, R. K. R, M. T, and B. F, Biodiesel from microalgae: A critical evaluation from laboratory to large scale production. *ResearchGate*. **103**, 444–467 (2013).

19. E. Suali, R. Sarbatly, Conversion of microalgae to biofuel. *ResearchGate*. **16**, 4316–4342 (2012).

20. S. Abad, X. Turon, Valorization of biodiesel derived glycerol as a carbon source to obtain added-value metabolites: Focus on polyunsaturated fatty acids. *Biotechnol. Adv.* **30**, 733–741 (2012).

21. A. R. Grossman *et al.*, Novel metabolism in Chlamydomonas through the lens of genomics. *Curr. Opin. Plant Biol.* **10**, 190–198 (2007).

22. Z. D. Blount, The unexhausted potential of E. coli. *eLife*. **4**, e05826 (2015).

23. H. Alper, G. Stephanopoulos, Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* **7**, 715–723 (2009).

24. S. K. Lee, H. Chou, T. S. Ham, T. S. Lee, J. D. Keasling, Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Curr. Opin. Biotechnol.* **19**, 556–563 (2008).

25. E. I. Lan, J. C. Liao, Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. *Metab. Eng.* **13**, 353–363 (2011).

26. *Nature's Chemicals* (https://global.oup.com/academic/product/natures-chemicals-9780199603022).

27. N. Quintana, F. Van der Kooy, M. D. Van de Rhee, G. P. Voshol, R. Verpoorte, Renewable energy from Cyanobacteria: energy production optimization by metabolic pathway engineering. *Appl. Microbiol. Biotechnol.* **91**, 471–490 (2011).

28. *Photosynthesis Research Protocols | Robert Carpentier | Springer* (http://www.springer.com/us/book/9781607619246).

29. N.-U. Frigaard, Y. Sakuragi, D. Bryant, in *Photosynthesis Research Protocols*, R. Carpentier, Ed. (Humana Press, 2004; http://dx.doi.org/10.1385/1-59259-799-8%3A325), *Methods In Molecular Biology*^{TM}, pp. 325–340.

30. F. Guerrero, V. Carbonell, M. Cossu, D. Correddu, P. R. Jones, Ethylene Synthesis and Regulated Expression of Recombinant Protein in Synechocystis sp. PCC 6803. *PLOS ONE*. **7**, e50470 (2012).

31. C. K. Holtman *et al.*, High-throughput functional analysis of the Synechococcus elongatus PCC 7942 genome. *DNA Res.* **12**, 103–115 (2005).

32. M. Iwai, H. Katoh, M. Katayama, M. Ikeuchi, Improved genetic transformation of the thermophilic cyanobacterium, Thermosynechococcus elongatus BP-1. *Plant Cell Physiol.* **45**, 171–175 (2004).

33. J. Yu *et al.*, Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO2. *Sci. Rep.* **5**, 8132 (2015).

34. J. A. Stamey, D. M. Shepherd, M. J. de Veth, B. A. Corl, Use of algae or algal oil rich in n-3 fatty acids as a feed supplement for dairy cattle. *J. Dairy Sci.* **95**, 5269–5275 (2012).

35. X. Liu, R. Curtiss, Nickel-inducible lysis system in Synechocystis sp. PCC 6803. *Proc. Natl. Acad. Sci.* **106**, 21550–21554 (2009).

36. J. Dexter, P. Fu, Metabolic engineering of cyanobacteria for ethanol production. *Energy Environ. Sci.* **2**, 857–864 (2009).

37. F. K. Bentley, A. Melis, Diffusion-based process for carbon dioxide uptake and isoprene emission in gaseous/aqueous two-phase photobioreactors by photosynthetic microorganisms. *Biotechnol. Bioeng.* **109**, 100–109 (2012).

38. L. Brennan, P. Owende, Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. *Renew. Sustain. Energy Rev.* **14**, 557–577 (2010).

39. H. A. Meylemans, R. L. Quintana, B. G. Harvey, Efficient conversion of pure and mixed terpene feedstocks to high density fuels. *Fuel*. **97**, 560–568 (2012).

40. W. N. Hunter, The Non-mevalonate Pathway of Isoprenoid Precursor Biosynthesis. *J. Biol. Chem.* **282**, 21573–21577 (2007).

41. Y. V. Ershov, R. R. Gantt, J. Francis X. Cunningham, E. Gantt, Isoprenoid Biosynthesis in Synechocystis sp. Strain PCC6803 Is Stimulated by Compounds of the Pentose Phosphate Cycle but Not by Pyruvate or Deoxyxylulose-5-Phosphate. *J. Bacteriol.* **184**, 5045–5051 (2002).

42. F. K. Bentley, A. Zurbriggen, A. Melis, Heterologous Expression of the Mevalonic Acid Pathway in Cyanobacteria Enhances Endogenous Carbon Partitioning to Isoprene. *Mol. Plant*. **7**, 71–86 (2014).

43. M. Rohmer, M. Rohmer, The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants†. *Nat. Prod. Rep.* **16**, 565–574 (1999).

44. N. Pade *et al.*, Insights into isoprene production using the cyanobacterium Synechocystis sp. PCC 6803. *Biotechnol. Biofuels*. **9** (2016), doi:10.1186/s13068-016-0503-4.

45. O. Aremu, Inactivation of competing pathways and increased gene copies for enhanced isoprene production in Synechococcous sp. PCC 7002 cyanobacteria (2015).

46. J. M. Estévez, A. Cantero, A. Reindl, S. Reichler, P. León, 1-Deoxy-D-xylulose-5-phosphate synthase, a limiting enzyme for plastidic isoprenoid biosynthesis in plants. *J. Biol. Chem.* **276**, 22901–22909 (2001).

47. G. S. A. B. Stewart, S. Lubinsky-Mink, J. Kuhn, pHG276: A multiple cloning site pBR322 copy number vector expressing a functional lacα peptide from the bacteriophage lambda PR promoter. *Plasmid*. **15**, 182–190 (1986).

48. M. Ptashne, *A Genetic Switch: Phage Lambda Revisited* (CSHL Press, 2004).

49. A. Thill *et al.*, Cytotoxicity of CeO2 Nanoparticles for Escherichia coli. Physico-Chemical Insight of the Cytotoxicity Mechanism. *Environ. Sci. Technol.* **40**, 6151–6156 (2006).

50. C. O. Pabo, R. T. Sauer, J. M. Sturtevant, M. Ptashne, The lambda repressor contains two domains. *Proc. Natl. Acad. Sci.* **76**, 1608–1612 (1979).

51. F. K. Davies, V. H. Work, A. S. Beliaev, M. C. Posewitz, Engineering Limonene and Bisabolene Production in Wild Type and a Glycogen-Deficient Mutant of Synechococcus sp. PCC 7002. *Front. Bioeng. Biotechnol.* **2**, 21 (2014).

52. M. Ludwig, D. A. Bryant, Synechococcus sp. Strain PCC 7002 Transcriptome: Acclimation to Temperature, Salinity, Oxidative Stress, and Mixotrophic Growth Conditions. *Front. Microbiol.* **3** (2012), doi:10.3389/fmicb.2012.00354.

53. C. T. Nomura, T. Sakamoto, D. A. Bryant, Roles for heme–copper oxidases in extreme high-light and oxidative stress response in the cyanobacterium Synechococcus sp. PCC 7002. *Arch. Microbiol.* **185**, 471–479 (2006).

54. H. Malke, Jeffrey H. Miller, A Short Course in Bacterial Genetics – A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria. Cold Spring Harbor 1992. Cold Spring Harbor Laboratory Press. ISBN: 0–87969-349–5. *J. Basic Microbiol.* **33**, 278–278 (1993).

55. Y. Xu *et al.*, Expression of genes in cyanobacteria: adaptation of endogenous plasmids as platforms for high-level gene expression in Synechococcus sp. PCC 7002. *Methods Mol. Biol. Clifton NJ.* **684**, 273–293 (2011).

56. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*. **6**, 343–345 (2009).

57. J. Zangl, Real Time PCR Protocol (2010).

58. P. Marraccini, S. Bulteau, C. Cassier-Chauvat, P. Mermet-Bouvier, F. Chauvat, A conjugative plasmid vector for promoter analysis in several cyanobacteria of the genera Synechococcus and Synechocystis. *Plant Mol. Biol.* **23**, 905–909 (1993).

59. P. Mermet-Bouvier, F. Chauvat, A conditional expression vector for the cyanobacteriaSynechocystis sp. strains PCC6803 and PCC6714 orSynechococcus sp. strains PCC7942 and PCC6301. *Curr. Microbiol.* **28**, 145–148 (1994).

60. $2.96 Billion Isoprene Market 2017 - Global Forecast to 2021 - Research and Markets. *Oil Gas 360* (2017), (available at https://www.oilandgas360.com/2-96-billion-isoprene-market-2017-global-forecast-to-2021-research-and-markets/).

61. J. M. Heuckmann *et al.*, Differential Protein Stability and ALK Inhibitor Sensitivity of EML4-ALK Fusion Variants. *Clin. Cancer Res.* **18**, 4682–4690 (2012).

62. P. Sojikul, N. Buehner, H. S. Mason, A plant signal peptide–hepatitis B surface antigen fusion protein with enhanced stability and immunogenicity expressed in plant cells. *Proc. Natl. Acad. Sci.* **100**, 2209–2214 (2003).

63. E. Soltes-Rak, D. J. Kushner, D. D. Williams, J. R. Coleman, Effect of promoter modification on mosquitocidal cryIVB gene expression in Synechococcus sp. strain PCC 7942. *Appl. Environ. Microbiol.* **59**, 2404–2410 (1993).

64. M. Sugita, M. Sugiura, The existence of eukaryotic ribonucleoprotein consensus sequence-type RNA-binding proteins in a prokaryote, Synechococcus 6301. *Nucleic Acids Res.* **22**, 25–31 (1994).

65. K. Namakoshi, T. Nakajima, K. Yoshikawa, Y. Toya, H. Shimizu, Combinatorial deletions of glgC and phaCE enhance ethanol production in Synechocystis sp. PCC 6803. *J. Biotechnol.* **239**, 13–19 (2016).

66. A. D. van der Woude, S. A. Angermayr, V. Puthan Veetil, A. Osnato, K. J. Hellingwerf, Carbon sink removal: Increased photosynthetic production of lactic acid by Synechocystis sp. PCC6803 in a glycogen storage mutant. *J. Biotechnol.* **184**, 100–102 (2014).

67. R. L. Burnap, Systems and Photosystems: Cellular Limits of Autotrophic Productivity in Cyanobacteria. *Front. Bioeng. Biotechnol.* **3** (2015), doi:10.3389/fbioe.2015.00001.

68. A. L. Markley, M. B. Begemann, R. E. Clarke, G. C. Gordon, B. F. Pfleger, Synthetic Biology Toolbox for Controlling Gene Expression in the Cyanobacterium Synechococcus sp. strain PCC 7002. *ACS Synth. Biol.* **4**, 595–603 (2015).

69. B. M. Berla *et al.*, Synthetic biology of cyanobacteria: unique challenges and opportunities. *Front. Microbiol.* **4**, 246 (2013).

70. J. L. Erbe, A. C. Adams, K. B. Taylor, L. M. Hall, Cyanobacteria carrying an smt-lux transcriptional fusion as biosensors for the detection of heavy metal cations. *J. Ind. Microbiol.* **17**, 80–83 (1996).

71. D. I. Arnon, B. D. McSwain, H. Y. Tsujimoto, K. Wada, Photochemical activity and components of membrane preparations from blue-green algae. I. Coexistence of two photosystems in relation to chlorophyll a and removal of phycocyanin. *Biochim. Biophys. Acta BBA - Bioenerg.* **357**, 231–245 (1974).

72. J. S. Buzby, R. D. Porter, S. E. Stevens, Expression of the Escherichia coli lacZ gene on a plasmid vector in a cyanobacterium. *Science*. **230**, 805–807 (1985).