UNIVERSITY OF CALIFORNIA, SAN DIEGO

Structural Variability, Evolutionary History and Biosynthetic Mechanisms of Cyanobacterial Hydrocarbons

A Dissertation submitted in partial satisfaction of the requirement for the degree

Doctor of Philosophy

in

Marine Biology

by

R. Cameron Coates

Committee in Charge:

Professor William H. Gerwick, Chair
Professor Mike Burkart
Professor Lena Gerwick
Professor James W. Golden
Professor Mark Hildebrand
Professor Brian Palenik

2014

The Dissertation of R. Cameron Coates is approved, and is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2014

DEDICATION

I would like to dedicate this thesis to my wonderful parents and my loving wife. They all have continued to supported me through this long journey and I will forever be grateful.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

all = *Anabaena sp*. PCC7120 locus left

alr = *Anabaena sp*. PCC7120 locus right

Sp/Sm = Spectinomycin/Streptomycin

OD600 = Optical Density 600 nm

IPTG = Isopropyl β-D-1-thiogalactopyranoside

PCR = polymerase chain reaction

ACP = acyl carrier protein

KS = ketosynthase

AT = Acyltransferase

KR = ketoreductase

FAAR/ADO = Fatty acyl ACP reductase

OLS = Olefin Synthase

RT-PCR = Real time PCR

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge William Gerwick for his support as my advisor and the chair of my committee. His guidance and friendship have been invaluable and has let me forever changed for the better.

I am so thankful for my wife Simone's patience, love and support throughout my dissertation. I look forward to building our life together after thins long awaited journey!

I would also like to acknowledge my many past and current lab mates of the Gerwick lab, without whom my research would have never been possible. I have made incredible lasting friendships and developed as a scientist because of their support and insights.

I would like to thank my office mates Adam Jones, Jennifer Hull and Nathan Moss for their support and friendships.

My family and friends have constantly supported me and provided me with wonderful opportunities for growth, happiness and connection.

Chapter 1 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

Chapter 2 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

Chapter 3 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

Chapter 4 in full is a manuscript that is in preparation. My co-authors are Arnaud Taton, Jonathan Tram, James Golden, Lena Gerwick, William H. Gerwick.

VITA

Education

2008-2014 **Ph.D.**, Marine Biology
Scripps Institution of Oceanography, UCSD

2008-2012 **M.S.**, Oceanography
Scripps Institution of Oceanography, UCSD

2001-2006 **B.S.**, Biological Science (Emphasis: Marine Biology)
University of California, Davis

Professional Experience

Doctoral Research, Scripps Institution of Oceanography, UCSD
September 2008 - Present
Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. Identification of methyltransferase involved in branched hydrocarbon biosynthesis. Pathway modification for bacterial isooctane production. Cyanobacterial natural product pathway identification
Advisor: William H. Gerwick

Visiting Research Fellow, Synthetic Biology and Biofuel Lab J. Craig Venter Institute
May 2012 – August 2012
Designed and assembled (using Gibson and yeast assembly) constructs for enhanced hydrogenase expression in *Synechocystis* sp. PCC6803 for photosynthetic and heterotrophic H_2 production.
Advisor: Philip Weyman

Staff Research Associate III. Scripps Institution of Oceanography, UCSD.
January 2006 - September 2008
Managed cyanobacterial extraction and cultures. Research expedition and lab diving coordinator. Project manager for DARPA algae jet fuel subcontract with General Atomics.
Advisor: William H. Gerwick

Assistant III. University of California, Davis
June 2004 - December 2005
Tracked invertebrate larval recruitment in Bodega Harbor. Senior thesis: Impact of invasive bryozoan *W. subtorquata.* Field research assistant on impacts of marine diversity projects..
Advisor: Jay Stachowicz

Assistant II. Scripps Institution of Oceanography, UCSD
June 2003 - September 2003
Performed image analysis on underwater bubble trapment data set. Assisted with field deployment on surfzone acoustic arrays.

Advisor: Grant Dean

Assistant II. University of California, Davis
September 2002 - June 2003
Performed bacterial and fungal pathogen screening and culture maintenance. Maintained lettuce strain greenhouse.
Advisor: Oswaldo Ochoa in Michelmore Laboratory

Teaching Experience

AP Environmental Science and Earth Science. University City High School
September 2010 – June 2011
National Science Foundation GK-12 Fellowship. Developed and taught a new curriculum on energy and algae biofuels. Published activities online: http://earthref.org/SCC/lessons/2010/biomass/

Biological Oceanography – BIEB 134 University of California, San Diego
January 2010 – March 2010
Teaching Assistant to Dr. Michael Landry.

Honors and Awards

Claude E. Zobell Fellowship in Marine Microbiology (2014)

Krinsk Research Advancement Initiative (KRAI) Fellowship (2014)

Thomas A. Page Fellowship in Algae Biofuels (2013)

SIO Department Graduate Student Excellence Travel Award (2013)

Marine Biotechnology Training Grant Fellowship (2011-2013)

NSF GK-12 Fellowship - High School Biofuel Curriculum Development (2010)

Sydney Frank Foundation Fellowship (2008-2010)

Korein Foundation Fellowship (2008-2009)

Young Presidents Org. Innovation Challenge Finalist – "Biodegradable Plastics from Algae" (2008)

IGERT Undergraduate Research Fellow (2004)

John H.C.& Mary Louise Riley Fellow (2004)

Eagle Scout (1997)

Publications

Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman D, Allen EE, Gerwick L, Gerwick WH (2013) Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS One.* DOI: 10.137/journal.pone.0085140

Coates RC, Trentacoste E, Gerwick WH 2013. Bioactive and Novel Chemicals from Microalgae. In: Handbook for Microalgal Culture 2nd Ed. Eds. Richmond A, Hu Q. Wiley-Blackwell. DOI: 10.1002/9781118567166.ch26

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu W, Gerwick L, Dorrestein P, Pevzner P, Lasken R, Gerwick W 2011. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS One.* 6(4): e18565

Sellheim K, Stachowicz JJ, Coates RC 2010. Effects of a nonnative habitat-forming species on mobile and sessile epifaunal communities. *Mar. Ecol. Prog. Ser*. 398: 69-80.

Engene N, Coates RC, Gerwick WH 2010. 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J.Phycol.* 46(3): 591-601.

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC, Sorrels CM 2008. Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3(6): 277-284.

Gerwick W, Byrum T, Carland T, Coates RC, Engene N, Esquenazi E, Gerwick L, Grindberg R, Johnson M, Jones A, Malloy K, Nunnery J, Pereira A, Soria I, Sorrels C, Taniguchi M, Tidgewell K, Villa F, Vining O, Dorrestein P, Gu L, Sherman DH 2008. Integrating Chemical and Biochemical Approaches to Natural Products Drug Discovery from Marine Cyanobacteria *Proceedings from the New Developments in Drug Discovery on Natural Products and Traditional Medicines NIPER, India* 33-43.

Esquenazi E, Coates C, Simmons L, Gonzalez D, Gerwick WH, Dorrestein PC. 2008. Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging. *Mol Biosyst.* (6): 562-570.

Simmons TL, Coates RC, Clark BR, Engene N, Gonzalez D, Esquenazi E, Dorrestein PC, Gerwick W. H. 2008. Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *PNAS.* 105(12): 4587-4594.

ABSTRACT OF THE DISSERTATION

Structural Variability, Evolutionary History and Biosynthetic Mechanisms of Cyanobacterial Hydrocarbons

by

R. Cameron Coates

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2014

Professor William H. Gerwick, Chair

Cyanobacteria are a group of photosynthetic bacteria that play major roles in all major ecosystems and biogeochemical cycles. Cyanobacteria have also been shown to exhibit a remarkable morphological and metabolic diversity. All cyanobacteria also seem to exhibit the rare capacity to make hydrocarbons. The physiological or ecological function of these cyanobacterial hydrocarbons is currently unknown. Interest in finding sustainable sources of hydrocarbons could provide opportunities to replace petroleum derivatives like fuels, plastics and chemicals with sustainably derived molecules with identical properties. Cyanobacterial hydrocarbons are biosynthetically derived from fatty acids. There are two pathways that have recently been described to produce

cyanobacterial hydrocarbons. One pathway involves fatty acyl ACP reductase and aldehyde deformylating oxygenase (FAAR/ADO) and is known to produce saturated alkanes one carbon shorter than their fatty acyl ACP substrates. The other is an olefin synthase (OLS) and is a polyketide synthase-type pathway involving elongation/decarboxylation mechanism that is known to produce terminal alkenes that are one carbon longer than their fatty acyl ACP substrate. The evolutionary history of these pathways is unknown. Additionally, the structural variability of the hydrocarbons derived from these pathways is relatively unexplored. This thesis investigates the structural variability, evolutionary history as well as novel biosynthetic pathway variability of cyanobacterial hydrocarbons. Chapter two presents evidence of the involvement of horizontal gene transfer in the evolutionary history of the two hydrocarbon pathways that were identified using a variety of bioinformatic analyses. An investigation of the structural diversity of cyanobacterial hydrocarbons as a consequence of the hydrocarbon pathway from which they are derived is presented in chapter three. Additionally, a variety of biosynthetic mechanisms are considered in the context of newly observed structural features including unique methylation and double bond positions. An effort to identify the methyltransferase responsible for branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120 is explored in chapter four. Finally, chapter five frames the major observations from the previous chapters in a broader context and discusses a wide variety of potential future directions for research on cyanobacterial hydrocarbon biosynthesis.

Chapter 1: Introduction to Cyanobacterial Hydrocarbon Biosynthesis

1.1 Environmental Drivers of Biofuels Research

In recent years, there has been a growing interest in replacing human's extensive use of fossil fuels with more sustainable alternatives. Although there are many factors driving this interest, one of the largest is the projected 56% increased worldwide energy demand over the next 30 years (EIA 2013). More than 85% of the projected increase in global energy is expected to occur in developing nations (EIA 2013). However, advancements in hydraulic fracturing technology is projected to increase the available supply and usage of low cost natural gas worldwide and subsequently decreased concerns about near term energy supplies (EIA 2013). Unfortunately, the sustained emissions of greenhouse gases from combustion of fossil fuels is known to directly increase the impact of climate change and ocean acidification issues (IPCC 2013). Climate change and ocean acidification as well as other environmental impacts of fossil fuel including air pollution and oil spills have created a unique confluence of societal challenges involving energy and the environment (Smith 1993, Peterson et al. 2003, IPCC 2013).

Interest in using agriculturally or other biologically derived fuels (biofuels) has grown in response to increased demand for alternatives to fossil fuels (Hallgren et al. 2013). A variety of biofuels that use established infrastructure and technology, including ethanol and biodiesel, have been increasing in scale of production over the past decade (Searchinger et al. 2008, Hallgren et al. 2013). However, the large-scale development of ethanol and biodiesel fuels that rely on food crops such as corn, soy and palm oil as a feedstock have their own set of unique economic and environmental challenges (Hill et al. 2006, Hallgren et al. 2013). Direct impacts on food price volatility as well as growing

concern over environmental impacts including life cycle greenhouse gas emissions have dimmed the prospects of the future of large scale biofuel development using food crops (Hill et al. 2006, Searchinger et al. 2008).

Interest in finding alternative sources of feedstocks for fuels and other petroleum derived molecules has shifted to a variety of technologies that are the subjects of active research and development (Schirmer et al. 2008, Jones and Mayfield 2012). The more advanced examples of these technologies include cellulosic ethanol and algae biofuels (Schmer et al. 2008, Jones and Mayfield 2012). Cellulosic ethanol presents a major opportunity to enzymatically degrade cellulose, one of the most abundant biomolecules on earth, into fermentable sugars that can be biologically converted into ethanol, a fuel with a widely established infrastructure (Agbor et al. 2011, Schmer et al. 2008). Although this technology continues to rapidly advance, the costs remain high (Hill et al. 2006, Jones and Mayfield 2012, Tao et al. 2014). Additionally, ethanol cannot be used in all fuel infrastructure, including diesel and jet engines (Kwanchareon et al. 2007). Alternatively, algae can be used to produce diesel and jet fuels and exhibit many of same advantages of cellulosic ethanol including high productivity as well as growth on non-arable land (Jones and Mayfield 2012). However, the complex lipids produced by algae require a costly conversion step to produce diesel and jet fuel (Jones and Mayfield 2012). This recognition has driven interest in finding organisms that directly produce hydrocarbons that can be used as drop-in ready fuels without conversion (Schirmer et al. 2010, Peralta et al. 2012). Cyanobacteria are one of only a few types of organisms that are known to directly produce hydrocarbons.

1.2 Cyanobacterial Hydrocarbon Biosynthesis

Cyanobacteria are a diverse group of photosynthetic bacteria that have evolved a remarkable array of adaptive traits including oxygenic photosynthesis, N_2 fixation, a wide morphological diversity, extensive secondary metabolite biosynthetic capacity, and a range of symbiotic relationships with other organisms. Cyanobacteria are estimated to contribute 30% of Earth's annual oxygen production and play a major role in biogeochemical cycles (Sharma et al. 2010). One trait less well characterized and potentially of great societal importance is their universal ability to produce long chain hydrocarbons. First recognition of this latter trait resulted from investigations in the 1960's (Han et al. 1968, 1969, Winters et al. 1969) and was of importance in the context of identifying the origin of hydrocarbons found in sedimentary and oil deposits. Their highest reported native production of alk(a/e)nes is 0.12% of dry biomass (Figure 1.1) (Winters et al. 1969), and the physiological or ecological function of alk(a/e)ne production in cyanobacteria is not yet understood. In this latter regard, various possibilities exist including prevention of grazing from herbivores, intra- or inter- species chemical signaling, prevention of desiccation, enhanced buoyancy, or membrane fluidity/stability. Unfortunately, none of these potential functions have been directly evaluated, and this remains an area in need of further investigation.

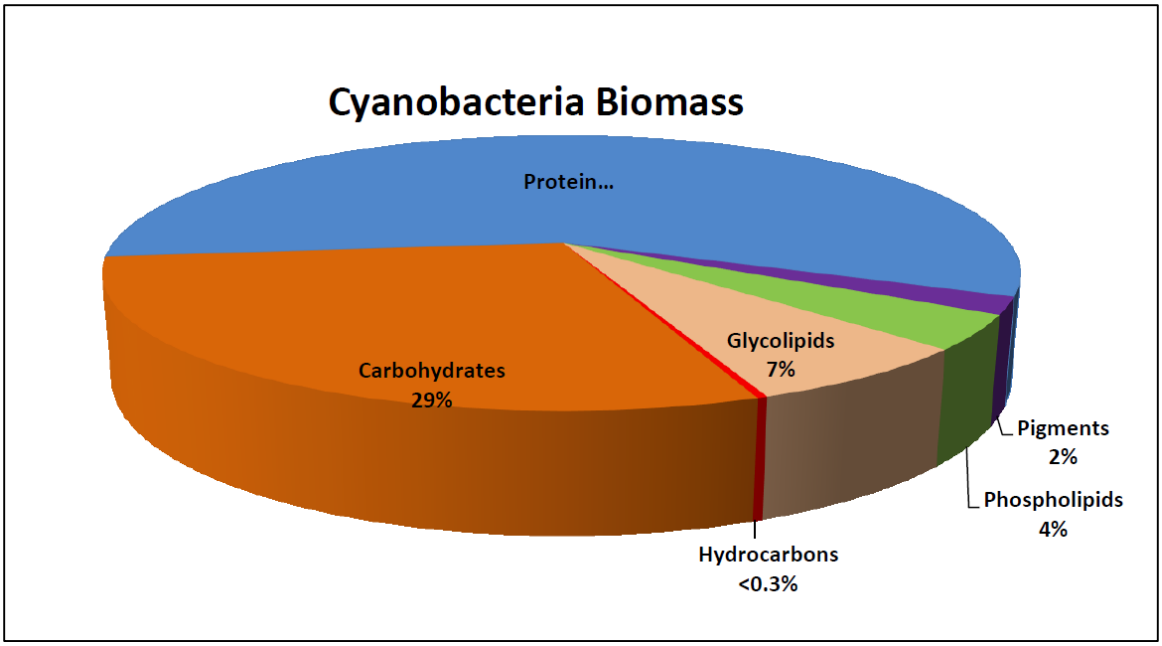


Figure 1.1 Chart depicting the composition of an average cyanobacterial cell. Hydrocarbons make up a relatively small component of the biomass at less than 0.3% (Rodriguez et al. 1989).

Fatty acid-derived hydrocarbons are produced in a variety of life forms. As described more extensively below, cyanobacteria synthesize long chain alk(a/e)nes using two different pathways, one of which involves a deformylation of fatty aldehydes, and the other a decarboxylation of fatty acids (Han et al. 1968, Schirmer et al. 2010, Mendez-Perez et al. 2011). Other prokaryotes such as *Jeotgalicoccus* sp. biosynthesize alk(a/e)nes through a cytochrome P450-catalyzed decarboxylation of fatty acids to form terminal olefins (Rude et al. 2011). Long-chain alkenes in *Micrococcus luteus* ATCC 4698 are formed via head-to-head condensation of fatty acids (Beller et al. 2010). Although some of the final products from these hydrocarbon pathways are identical to the alk(a/e)nes produced by cyanobacteria, the specific biosynthetic steps involved are distinctly different.

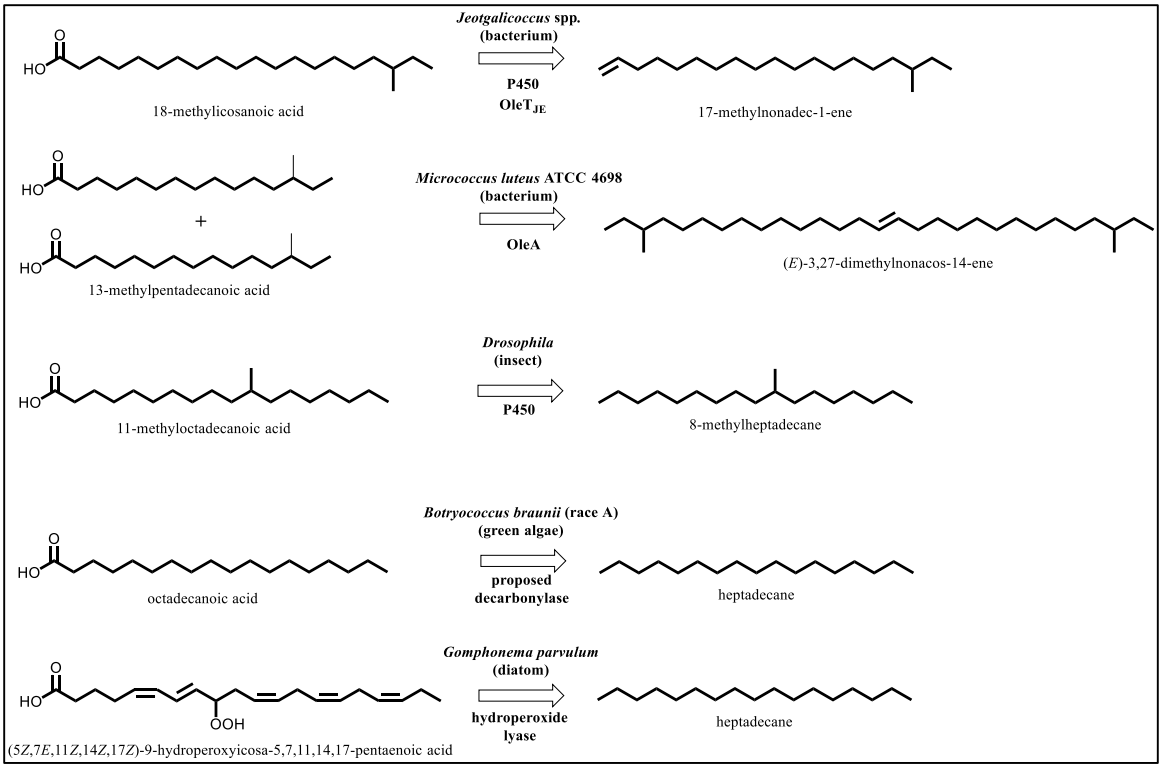


Figure 1.2. Fatty acid derived hydrocarbon pathways that have been observed in organisms other than cyanobacteria.

Hydrocarbons produced via fatty acid pathways have also been described in various eukaryotic organisms including the long-chain fatty acid derived alkenes from *Botryococcus braunii* race A (Dennis and Kolattukudy 1991, Metzger and Largeau 2005), the pheromone attractant hormosirene in the diatom *Gomphonema parvulum* (Pohnert and Boland 2002), and the pheromone 7-methylheptadecane found in lepidopteran insects (Blomquist and Bagneres 2010). However, these eukaryotic hydrocarbons are derived from entirely different biosynthetic pathways than cyanobacterial hydrocarbons. For example, the *Botryococcus braunii* race A alkenes are thought to derive from an aldehyde decarbonylase (Metzger and Largeau 2005), hormosirene biosynthesis in the diatom *Gomphonema parvulum* involves a hydroperoxide lyase (Pohnert and Boland 2002), and a cytochrome P450 has recently

been described as responsible for fatty acid decarbonylation to produce branched alkanes in lepidopteran insects (Qiu *et al*. 2012).

Two separate pathways have recently been described in cyanobacteria that produce hydrocarbons from fatty acid substrates, with each pathway involving a quite different mechanism for removing the terminal oxidized carbon atom (Schirmer et al. 2010, Mendez-Perez et al. 2011). One pathway involves the conversion of fatty acyl-ACPs to fatty acyl-aldehydes by a fatty acyl-ACP reductase (FAAR), followed by a unique oxygen dependent conversion of the aldehyde via an aldehyde-deformylating oxygenase (ADO) to produce an odd-chain length saturated alkane (or alkene if the fatty acid possesses preexisting unsaturation) (Figure 1.3) (Schirmer et al. 2010, Li et al. 2012). FAARs have been found in a wide range of organisms including plants, eukaryotic algae, bacteria and humans (Doan et al. 2009). This reaction is known to require NADPH that likely provides a hydride that attacks the carbonyl of the fatty acid thioester. Nevertheless, the precise mechanism of this FAAR enzyme has not been conclusively verified (Schirmer et al. 2010) and there are no published crystal structures of a FAAR enzyme from any source. The next step involving the ADO enzyme is unique to cyanobacteria (Li et al. 2012). This enzyme was first described by (Schirmer et al. 2010) as an aldehyde decarbonylase after verifying its function through mutagenesis, heterologous expression, and a re-investigation of a previously published crystal structure of the ADO enzyme from *Prochlorococcus marinus* MIT9313. Carbon monoxide was initially considered to be the oxidized byproduct of this reaction (Schirmer et al. 2010, Eser et al. 2011). However, after some debate, it was shown by Li et al. to produce an

alkane and formate from a unique oxygenation reaction, thus justifying the alternative name of 'aldehyde deformylating oxygenase' (Li et al. 2012).

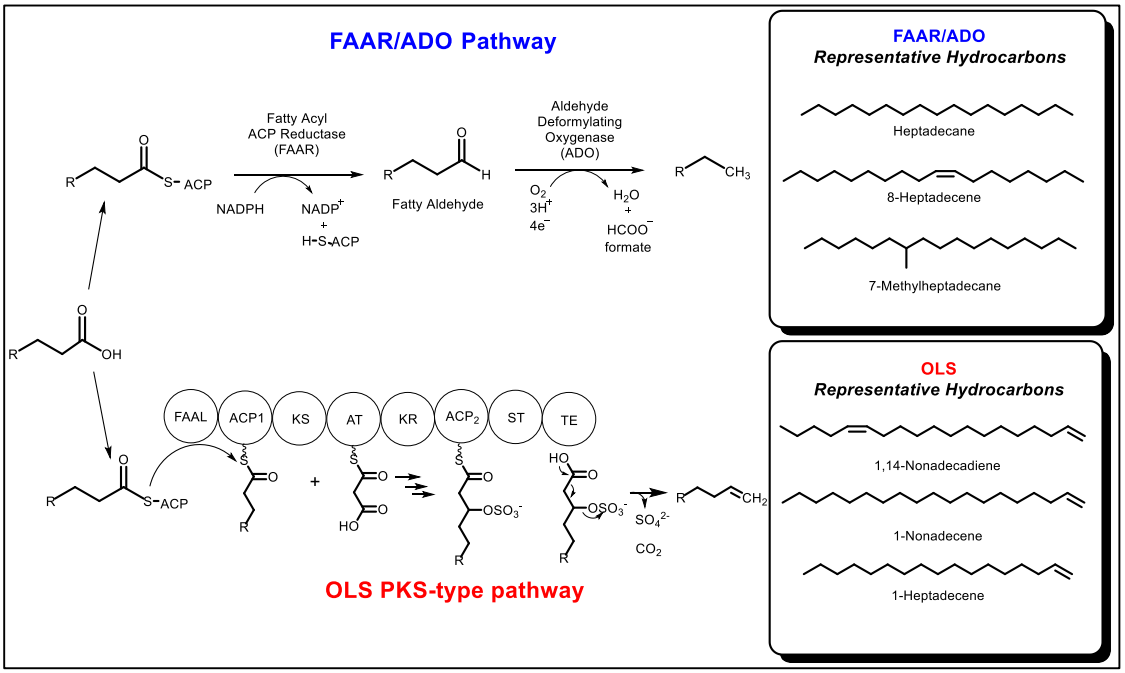


Figure 1.3. Hydrocarbon biosynthetic pathways in cyanobacteria. The Fatty Acyl-ACP Reductase (FAAR)/Aldehyde Deformylating Oxygenase (ADO) involves first a reduction of a fatty acyl substrate to a fatty aldehyde followed by an oxidative conversion to an alkane with the release of formate (Li et al 2012)The OLS (olefin producing) pathway involves a polyketide synthase that elongates a fatty acylACP and subsequently decarboxylates in the TE to form the terminal alkene.

In 2010, Ridley and Reppas and subsequently Mendez-Perez et al. (2011) and Donia et al. (2011) published reports on biosynthetic pathways in *Synechococcus* sp. PCC7002 and *Prochloron didemni* that utilize fatty acid substrates and distinctively produce odd-numbered carbon chains with terminal olefinic bonds. In 2011, Donia et al. and Mendez-Perez et al. identified a large (~10 Kb) multi-domain PKS pathway that proceeds via an elongation/decarboxylation mechanism (Figure 1.3). The pathway was initially located by searching for homologs of the well characterized CurM domain of the

curacin A biosynthetic pathway in *Moorea producens* 3L (formerly *Lyngbya majuscula* 3L) (Mendez-Perez et al. 2011, Donia et al. 2011, Gu et al. 2009, Engene et al. 2012). Curacin A is a secondary metabolite with promising anticancer activity (Chang et al. 2004). CurM is known to produce the terminal alkene functionality in curacin A via an elongation-decarboxylation mechanism and has recently been shown to accept fatty acid substrates to produce terminal alkenes (Gehret-McCarthy et al. 2012, Chang et al. 2004). Mendez-Perez et al. (2011) verified the function of the OLS pathway in *Synechococcus* sp. PCC7002, tested a variety of potential substrates, and enhanced productivity using an alternative promoter.

The elongation/decarboxylation mechanism of this olefin synthase (OLS) pathway is proposed by Mendez-Perez et al. (2011) to include an ATP-consuming fatty acyl ACP ligase (FAAL) domain that transfers a fatty acyl-ACP to the OLS acyl carrier protein (ACP). The ACP-bound substrate subsequently undergoes elongation by an extension module (ketosynthase (KS), acyltransferase (AT), and ketoreductase (KR)) to add two carbons from malonyl-CoA and reduce the β-keto group to a β-hydroxy functionality. Gehret-McCarthy et al. (2012) characterized the final step of the OLS pathway to involve a sulfotransferase (ST) and thioesterase (TE) that transfers sulfonate from the donor 3′-phosphoadenosine-5′-phosphosulfate (PAPS) to the β-hydroxyacyl substrate. This activates the intermediate for subsequent TE-catalyzed hydrolysis from the enzyme surface, decarboxylation and desulfation to form the terminal alkene product **(**Gehret-McCarthy et al. 2012**)**. Gehret-McCarthy et al. also compared the structure and function of the activating sulfotransferases (STs) from the CurM pathway in *M. producens* 3L and the OLS pathway in *Synechococcus* sp. PCC7002, and found that both

STs accepted substrates leading to production of long-chain hydrocarbons (Gehret-McCarthy et al. 2012). The inherent formation of a terminal double bond in the product is a unique structural signature of the OLS pathway compared to the FAAR/ADO pathway (Figure 1.3).

It is important to recognize that the extremely low native yields of hydrocarbons from wild type cyanobacterial strains are insufficient to be cost competitive with petroleum derived fuels, and therefore, some moderation of the expression of cyanobacterial hydrocarbons pathways will be necessary if they are to be used in an industrial or commercial context. Additionally, the unique terminal olefin produced by CurM and the OLS pathway has societal relevance beyond alternative fuels. Terminal olefins are widely used in a variety of plastics, lubricants, surfactants and rubbers (Werpy and Peterson 2004). Thus, a deeper understanding of these pathways is needed in order to rationally engineer and enhance this trait in modified organisms.

Although these two cyanobacterial hydrocarbon biosynthetic pathways (FAAR/ADO and OLS) were only recently identified and characterized, knowledge of the diagnostic structural features of the resulting alk(a/e)nes and their respective gene sequences can be used along with previous reports of cyanobacterial hydrocarbon compositions to better understand the taxonomic distribution of these two pathways and the structural diversity found in cyanobacterial hydrocarbons. A comprehensive investigation of the taxonomic distribution and evolutionary history of the two cyanobacterial hydrocarbon pathways has never been done. Prior to this thesis, it was unknown if these two pathways occur together in the same organism or if their distinct products can be used as diagnostic features to suggest which pathway the organism

possesses. Additionally, it has not been established that all cyanobacteria possess only one of the two pathways. If it can be established that all cyanobacteria possess only one of the two pathways then new insight will be gained into the fundamental role of hydrocarbons in cyanobacteria.

Chapter two presents an investigation of the taxonomic distribution and evolutionary history of the two cyanobacterial pathways. To map the pathway distribution to cyanobacterial phylogeny, a 16S rRNA phylogenetic tree was constructed for a broad phylogenetic distribution of cyanobacteria including all publically available genome sequence strains. The distribution of the two cyanobacterial hydrocarbon pathways revealed a distribution of the OLS pathway that was inconsistent with vertical transmission of the pathway. Evidence of disruption of gene synteny, altered GC content and the presence of transposases is presented as evidence of the involvement of horizontal gene transfer in the evolutionary history of the OLS pathway in cyanobacteria. Opportunities for further investigation of the evolutionary history of the OLS as well as the FAAR/ADO pathway were identified.

Previous characterizations of the structural variability of the hydrocarbons produced by these two pathways have focused on a small subset of cyanobacteria, allowing for the possibility that there may be as yet uncharacterized structural diversity. A comprehensive investigation into the structural variability of hydrocarbons produced by the FAAR/ADO and OLS pathways in a wide range of cyanobacteria would establish the basis for broad perspectives on the role of hydrocarbons in cyanobacteria. Additionally, characterization of new structural variability produced by these pathways may provide insights into novel biochemical steps involved in their production. Chapter three presents

an investigation of the hydrocarbon composition for 32 phylogenetically diverse cyanobacteria. This investigation revealed new double bond (2- and 3-heptadecene) and methyl group positions (3-, 4- and 5-methylheptadecane) for a variety of strains. These observations revealed new insights into variability in the cyanobacterial hydrocarbon pathways and provide the foundation for future investigations.

Observations of a narrow distribution of branched hydrocarbon production as well as prior investigations indicated that a unique methyltransferase was involved in branched hydrocarbon biosynthesis in some cyanobacteria. Chapter four presents an investigation into the methyltransferase responsible for branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120. Three methyltransferases were bioinformatically identified as targets for investigation. A variety of *in vivo, in vitro* as well as mutagenesis experiments were used to narrow down the target methyltransferases. A variety of opportunities for future investigations were identified through this research.

Through a coupling of the phylogenetic distribution and evolutionary history of cyanobacterial hydrocarbon pathways with the detailed compositional analysis of the resulting structural diversity of hydrocarbons from each pathway, this thesis reveals new insights concerning the fundamental role of hydrocarbons in cyanobacteria as well as novel biosynthetic pathway variability that may serve as targets for future investigation or improvements for sustainable fuel and chemical development.

1.3 Acknowledgements

Chapter 1 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

1.4 References

Agbor VB, Cicek N, Sparling R, Berlin A, Levin DB, 2011. Biomass pretreatment: Fundementals toward application. Biotech. Adv. 29(6). 675-685.

Beller HR, Goh EB, Keasling JD, 2010 Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. Appl Environ Microbiol 76. 1212-1223.

Blomquist GJ, Bagneres AG (eds) 2010. Insect hydrocarbons: biology, biochemistry, and chemical ecology. New York: Cambridge University Press.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH, 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. J Nat Prod 67: 1356-1367.

Dennis MW, Kolattukudy PE, 1991 Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus braunii*. Arch Biochem Biophys 287. 268-275.

Doan TTP, Carlsson AS, Hamberg M, Bülow L, Stymne S, Olsson P, 2009. Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli.* J Plant Physiol 166. 787-796.

Donia MS, Fricke WF, Ravel J, Schmidt EW, 2011. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. PLoS ONE 6:e17897. doi:10.1371/journal.pone.0017897

Engene N, Rottacker EC, Kaštovský J, Byrum T, Choi H, Ellisman MH, Komárek J, Gerwick WH, 2012. *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb.

nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Ins J Syst Evol Microbiol 62. 1171-1178.

Eser BE, Das D, Han J, Jones PR, Marsh EN, 2011. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. Biochemistry 50. 10743-10750.

Gehret McCarthy J, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, Sherman DH, Smith JL, 2012. Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. ACS Chem Biol 7. 1994-2003. doi: 10.1021/cb300385m

Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Håkansson K, Smith JL, Sherman DH, 2009. Polyketide decarboxylative chain termination preceded by O-sulfonation in curacin A biosynthesis. J Am Chem Soc. 131. 16033–16035.

Hallgren W, Schlosser CA, Monier E, Kicklighter D, Sokolov A, Melillo J, 2013. Climate impacts of a large-scale biofuels expansion. Geophys. Res. Lett. 8. 1642-1630.

Han J, McCarthy ED, Van Hoeven W, Calvin M, Bradley WH, 1968. Organic geochemical studies, II. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment. PNAS. 59: 29-33.

Han J, Chan HWS, Calvin M, 1969. Biosynthesis of alkanes in *Nostoc muscorum*. J Am Chem Soc 91. 5156-5159.

Hill J, Nelson E, Tilman D, Polasky S, Tiffany D, 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. PNAS. 103(30). 11206-11210.

Intergovernmental Panel on Climate Change 2013. Climate change 2013. The physical science basis. Stocker TF, Qin D, Plattner GK, Tignor MMB, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds) Available: http://www.climatechange2013.org/images/report/WG1AR5_ALL_FINAL.pdf accessed: 1 May 2014.

Jones CS, Mayfield SP, 2012. Algae biofuels: versatility for the future of bioenergy. Curr Opin Biotechnol. 23(3). 346-351.

Kwanchareon P, Luengnaruemitchai A, Jai-In S, 2007. Solubility of a diesel-biodiesel-ethanol blend, its fuel properties, and its emission characteristics from diesel engine. Fuel. 86. 1053-1061.

Li N, Chang W, Warui DM, Booker SJ, Krebs C, Bollinger JM Jr., 2012. Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. Biochemistry 51. 7908-7916.

Lee RF, Loeblich AR, 1971. Distribution of 21:6 hydrocarbon and its relationship to 22:6 fatty acid in algae. Phytochemistry 10. 593-598.

Mendez-Perez D, Begemann MB, Pfleger BF, 2011. Modular synthase-encoding gene involved in α-olefin biosynthesis in *Synechococcus* sp. strain PCC7002 App Envir Micro 77: 4264-4267.

Metzger P, Largeau C, 2005. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. Appl Microbiol Biotechnol 66: 486-496.

Peralta-Yaha PP, Zhang F, del Cardayre SB, Keasling JD, 2012. Microbial engineering for the production of advanced biofuels. Nature 488: 320-328.

Perry GJ, Gillan FT, Johns RB, 1978. Lipid Composition of a prochlorophyte. J Phycol 14: 369-371.

Peterson CH, Rice SD, Short JW, Esler D, Bodkin JL, Ballachey BE, Irons DB, 2003. Long-term ecosystem response to the Exxon Valdez Oil Spill. Science 302(5653). 2082-2086.

Pohnert G, Boland W, 2002. The oxylipin chemistry of attraction and defense in brown algae and diatoms. Nat Prod Rep 19: 108-122.

Rezanka T, Zahradík J, Podojil M, 1982. Hydrocarbons in green and blue-green algae. Folia Microbiol (Praha) 27. 450-454.

Qui Y, Tittiger C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R, 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. PNAS. 109(37). 12858-14863.

Rodriguez H, Rivas J, Guerrero MG, Losada M, 1989. Nitrogen-fixing cyanobacterium with a high phycoerythrin content. AEM. 55(3). 758-760.

Ridley CP, Reppas NP, 2010. Biosynthesis of 1-alkenes in engineered microorganisms. US Patent Appl 12/821,107.

Rude MA, Baron TS, Brubaker S, Alibhai M, del Cardayre SB, Schirmer A, 2010. Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from *Jeotgalicoccus* sp. Appl Environ Microbiol 77. 1718-1727.

Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu TH, 2008. Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319(5867). 1238-1240.

Sharma N, Tiwari S, Tripathi K, Rai A, 2010. Sustainability and cyanobacteria (bluegreen algae): facts and challenges. J Appl Phycol 23.1-23.

Schirmer A, Rude MA, Li X, Popova E, del Cardayre S, 2010. Microbial biosynthesis of alkanes. Science 329. 559-562.

Schmer MR, Vogel KP, Mitchell RB, Perrin RK, 2008. Net energy of cellulosic ethanol from switchgrass. PNAS. 105(2). 464-469.

Smith KR, 1993. Fuel combustion, air exposure, and health: The situation in developing countries. Annu. Rev. Energy Environ. 18. 529-566.

Tao L, Tan ECD, McCormick R, Zhang M, Aden A, He X, Zigler BT, 2014. Techno-economic analysis and life-cycle assessment of cellulosic isobutanol and comparison with cellulosic ethanol and n-butanol. Biofuels, Biprod. Biorefin. 8(1). 30-48.

U.S. Energy Information Administration, International Energy Outlook, http://www.eia.gov/forecasts/ieo/pdf/04842013).pdf, July 25, 2013.

Werpy T, Peterson G, Aden A, Bozell J, Holladay J, White J, Manheim A, 2004. Top value added chemicals from biomass volume 1 – Results of screening for potential candidates from sugar and synthesis gas. US Department of Energy, Office of Scientific and Technical Information. URL: http://www1.eere.energy.gov/bioenergy/pdfs/35523.pdf Accessed Feb. 16, 2014

Winters K, Parker PL, Van Baalen C, 1969. Hydrocarbons of blue-green algae: geochemical significance. Science 163. 467-468.

Chapter 2: Evolutionary history of hydrocarbon biosynthetic pathways in cyanobacteria

2.1 Abstract

Cyanobacteria possess two distinct pathways for hydrocarbon biosynthesis from fatty acid precursors. One pathway comprises a two-step conversion of fatty acids first to fatty aldehydes and then alkanes that involves a fatty acyl ACP reductase (FAAR) and aldehyde deformylating oxygenase (ADO). The second involves a polyketide synthase (PKS) pathway that first elongates the acyl chain followed by decarboxylation to produce a terminal alkene (olefin synthase, OLS). These two pathways were only recently identified in cyanobacteria and have not been extensively investigated. The phylogenetic distribution and evolutionary history of these pathways had not been previously established prior to this investigation. This investigation used a variety of bioinformatic techniques to establish the phylogenetic distribution of the two pathways in cyanobacteria and investigate the involvement of horizontal gene transfer in their evolutionary history. Sixty-one strains possessing the FAAR/ADO pathway and twelve strains possessing the OLS pathway were newly identified through bioinformatic analyses. Strains possessing the OLS pathway formed a cohesive phylogenetic clade with the exception of three *Moorea* strains and *Leptolyngbya* sp. PCC 6406 which may have acquired the OLS pathway via horizontal gene transfer. Hydrocarbon pathways were identified in one-hundred-forty-two strains of cyanobacteria over a broad phylogenetic range and there were no instances where both the FAAR/ADO and the OLS pathways were found

together in the same genome, suggesting an unknown selective pressure maintains one or the other pathway, but not both.

2.2 Introduction

Cyanobacteria are one of the most morphologically diverse prokaryotic phyla and have evolved over the past two billion years wide diversity of metabolic capabilities. For example, cyanobacteria exhibit the unique capacity to produce hydrocarbons from fatty acid precursors. The two biosynthetic pathways that have been shown to produce these hydrocarbons in cyanobacteria are the fatty acyl ACP reductase/aldehyde deformylating oxygenase (FAAR/ADO) pathway and the olefin synthase (OLS) pathway (Figure 1.3). Prior investigations into these pathways have characterized homologs of these pathways in other cyanobacteria, but none of these have undertaken a comprehensive investigation into the distribution of these pathways in cyanobacteria and their evolutionary histories (Schirmer et al. 2010, Mendez-Perez et al. 2011). Previous investigations have thus far identified 58 homologs of the FAAR/ADO pathway (Schirmer et al. 2010) in cyanobacteria and 5 homologs of the OLS pathway in cyanobacteria (Mendez-Perez et al. 2011). Schirmer et al. (2010) identified the FAAR/ADO pathway through a subtractive genome analysis based on the observation that saturated alkanes are not observed in *Synechococcus* sp. PCC7002, and therefore, they made the presumption that *Synechococcus* sp. PCC7002 does not possess the pathway required to make saturated alkanes. Although *Synechococcus* sp. PCC7002 does not produce saturated alkanes it does however produce alkenes. Not until studies by Ridley et al. 2010 and subsequently Donia et al. 2011 and Mendez-Perez *et al*. 2011 was it clear that *Synechococcus* sp.

PCC7002 along with a few other cyanobacterial strains possess an alternative polyketide synthase type pathway that produces the terminal alkenes observed in these strains. The observations that two distinctly different pathways exist that are capable of producing relatively similar molecules raises multiple questions about the prevalence of these two pathways and their evolutionary histories.

Investigations into these questions about the prevalence of the two pathways and their evolutionary histories requires a broad sampling of genome sequence data from a broad range of cyanobacteria. Fortunately 54 phylogenetically diverse cyanobacteria have recently had their genomes sequenced (Shih et al.2013). This broad genomic data set has allowed for a much more comprehensive investigation into the patterns of cyanobacterial hydrocarbon pathway distribution and evolution. Additionally, seven new genome sequences that were not publicly available were made available for this investigation. These newly available cyanobacterial genome sequences coupled with relatively recent insights into the two cyanobacterial hydrocarbon biosynthetic pathways provided a foundation for this investigation into the evolutionary history of cyanobacterial hydrocarbon biosynthesis. This investigation characterized the distribution of the two cyanobacterial hydrocarbon pathways and explored the potential involvement of horizontal gene transfer in their evolutionary history. By establishing the distribution of these pathways in cyanobacteria, this investigation will reveal insights about the role that hydrocarbon production may play in cyanobacteria and the relative importance of this trait during evolutionary history of cyanobacteria.

2.3 Material and Methods

2.3.1 Strain Selection

Strains of cyanobacteria were selected from a wide phylogenetic distribution ensuring that every major clade was sampled (Figure 2.1). Strains of cyanobacteria with available genome sequences and established genetic techniques were given preference. Strains that required extremely low or high temperature or unusual media requirements for growth were excluded due to limited conditions and resources available at the time. *Anabaena (Nostoc)* sp. PCC7120 was provided by James Golden at UCSD. *Synechococcus elongatus* PCC7942 and *Leptolyngbya* BL0902 were provided by Susan Golden at UCSD. *M. producens* 3L, *Moorea bouillonii* PNG5-198 and *Moorea producens* JHB, *Leptolyngbya* sp. PAC10-3, *Moorea* sp. PNG4/22/06-1, *Moorea* sp. PAL8/15/08-1, *Okeania comitata* 3LOSC, *Okeania hirsuta*. PAB10-Feb-10-1, *Okeania hirsuta* PAB21-Jun-10-1, *cf. Phormidium* sp. ISB3/Nov/94-8, *Okeania* sp. PAC18-Feb-10-1.1 were isolated from field collections and are maintained among the Gerwick Laboratory culture collection as described in Engene et al. 2012, Williamson et al. 2002 and Engene et al. 2013. *Planktothrix agardhii* NIVA-CYA 126/8 was provided by Rainer Kurmayer at the Austrian Academy of Sciences Institute for Limnology. *Haplosiphon welwitchii* IC-52-3 and *Westiella intricata* HT-29-1 were provided by Thomas Hemscheidt of the University of Hawaii via Melinda Micallef and Michelle Moffitt (University of Western Sydney). The remaining strains were purchased from the Pasteur Culture Collection of Cyanobacteria (PCC) or the American Type Culture Collection (ATCC) as identified in the text.

2.3.2 Genome Sequencing and Bioinformatic Analysis

The sequenced genomes of *M. bouillonii* PNG5-198*, M. producens* JHB, and *cf. Phormidium* sp. ISB 3/Nov/94-8 were generated at either the Genomic Center at The Scripps Research Institute or at the University of Michigan using Illumina technology (Bennett 2004). Genome sequencing errors were corrected with Quake and assembled using SPAdes 2.4 (Kelley et al. 2010, Bankevich et al. 2012). Contigs were binned by GC content to remove non-cyanobacterial DNA sequences.

Identification of cyanobacterial hydrocarbon pathways was accomplished using blastn searches against newly sequenced genomes and blastp searches using representative genes from each pathway (FAAR/ADO and OLS) against publicly available cyanobacterial genomes (131 total) from GenBank (Benson et al. 2013) and the Joint Genome Institute (JGI) Integrated Microbial Genomes (IMG) database (Version 4.3, (Markowitz et al. 2014). Draft genome sequences for *M. bouillonii* PNG5-198*, M. producens* JHB, *cf. Phormidium* sp. ISB 3/Nov/94-8, *P. agardii* NIVA-CYA 126/8 (provided by Rainer Kurmayer, bio project accession number: PRJNA163669), *H. welwitschii* IC-52-3 and *W. intricata* HT-29-1 (provided by Melinda Micallef and Michelle Moffitt), and *Leptolyngbya* sp. BL0902 (provided by Arnaud Taton, Susan Golden and James Golden) were also used. The sulfotransferase domain in the OLS pathway identified in *Synechococcus* sp. PCC7002 was used for this search because this is a distinctive enzymatic step in terminal alkene biosynthesis (Gehret-McCarthy et al. 2012). The FAAR and ADO enzymes from *S. elongatus* PCC7942 were used as representative sequences in the search for FAAR/ADO pathways. Searches for FAAR/ADO pathway enzymes were also performed using hidden Markov model protein family patterns TIGR04058 (aldehyde-forming long-chain fatty acyl-ACP reductase) and

TIGR04059 (long-chain fatty aldehyde decarbonylase) from the TIGRFAMs database (Haft et al. 2013).

To investigate the role of horizontal gene transfer in the evolutionary history of the FAAR/ADO and OLS pathways JGI's IMG gene neighborhood function was used to compare gene neighborhoods of all FAAR/ADO and OLS containing cyanobacteria (Version 4.3)(Markowitz et al. 2014).

2.3.3 Phylogenetic Analysis

To establish and compare the phylogenetic distribution of hydrocarbon biosynthetic pathways, phylogenies were constructed using the 16S rRNA gene sequences from all cyanobacterial strains for which genome sequence data is publically available (128) as well as all additional cyanobacteria investigated in this study (16). Due to incomplete or partial sequences, *cf. Phormidium* sp. ISB 3/Nov/94-8, *W. intricata* HT-29-1, and *Synechococcus* sp. CB0205 were omitted from these analyses. *Pleurocapsa* sp. PCC7320, *Cyanobacterium* sp. JSC-1 as well as *Synechococcus* sp. CB0101 were not included in the phylogenetic analyses because 16S rRNA sequences are not currently available for these strains. Alignments of the 16S rRNA sequences were completed using MAFFT and trees were generated using both PhyML (GTR+I+G substitution model, 500 bootstrap replicates) and MrBayes (GTR+I+G substitution model). OLS pathway amino acid sequences were recovered as described above and aligned using MAFFT and trees were generated using PhyML (LG substitution model, 500 bootstrap replicates).

2.4 Results

2.4.1 Taxonomic Pathway Distribution

The recent addition of genome sequences from the CyanoGEBA project has more than doubled the amount of genomic information available for cyanobacteria and dramatically improved the distribution of genomic information across the phylogenetic diversity of cyanobacteria (Shih et al. 2013). Prior to the availability of these sequences, only five cyanobacteria were known to possess the OLS pathway (*Synechococcus* sp. PCC7002, *M. producens* 3L, *Cyanothece* sp. PCC7822, *Cyanothece* sp. PCC7424, and uncultured *Prochloron* symbionts (Mendez-Perrez et al. 2011, Donia et al. 2011, Gehret-McCarthy et al. 2012). Using the newly available CyanoGEBA data and newly sequenced genomes of *M. producens* JHB, *M. bouillonii* PNG5-198, *cf. Phormidium* sp. ISB 3/Nov/94-8, *P. agardii* NIVA-CYA 126/8, *H. welwitschii* IC-52-3, *W. intricata* HT-29-1 and *Leptolyngbya* sp. BL0902 (see experimental for sources), twelve additional strains were identified as possessing the OLS pathway, bringing the total to 17 strains (Figure 2.1). Nevertheless, the vast majority of genome-sequenced cyanobacteria possess the FAAR/ADO pathway (122 of 139). Interestingly, some of these latter cyanobacteria possess two homologs of the ADO gene (*Cyanothece* sp. PCC7425, *Cyanobium gracile* PCC6307, *Leptolyngbya* sp. PCC7375, *Synechococcus* sp. CB0101, *Synechococcus* sp. RS9917). Strains that possess the OLS pathway are largely (14 of 17) within a single clade (Clade A), mostly comprised of strains from subdivisions II (formerly Pleurocapsales), subdivision III (formerly Oscillatoriales) and subdivision I (formerly Chroococcales). In addition, two strains in this study that fall into this clade (*D. incrassata* PCC7326 and *Pleurocapsa* sp. PCC7516) are not yet sequenced but produce terminal alkenes, indicating that they likely also possess the OLS pathway. The two clades that are exceptions to this OLS pathway phylogenetic clustering are the clade

containing *M. producens* 3L, *M. producens* JHB and *M. bouillonii* PNG5-198 as well as the more distant clade containing *Leptolyngbya* sp. PCC6406 (Clade B & C, Figure 2.1).

Figure 2.1 Cyanobacterial 16S rRNA phylogeny and hydrocarbon pathway distribution. The 16S rRNA phylogeny of publicly available genome sequenced cyanobacteria (128) and additional strains investigated in this study (14) including *G. violaceus* PCC7421 as the outgroup. Blue strain names indicate strains possessing the FAAR/ADO pathway and red strain names indicate those with the OLS pathway. Purple strain names indicate a strain that does not have a genome sequence and therefore pathway presence cannot yet be verified. Cyanobacterial subdivisions are labeled using colored branches following the key in the upper left: (1) Subdivision 1. Unicellular (Formerly Chroococcales), Subdivision II. Baeocystous (Formerly Pleurocapsales), Subdivision III. Filamentous (Formerly Oscillatoriales), Subdivision IV. Heterocystous (Formerly Nostocales), Subdivision V. Ramified or True Branching (Formerly Stigonematales). The clade indicated as "A" represents the main clade of cyanobacteria with the OLS pathway while clade "B" indicates the *Moorea* strains and clade "C" indicates the clade containing *Leptolyngbya* sp. PCC6404. Baysian posterior probabilities are displayed at nodes (o = posterior support < 0.5).


Contains FAAR/ADO
Contains OLS
Unknown Pathway/No Genome
● = Possible HGT
(I) Unicellular
(II) Baeocystous
(III) Filamentous
(IV) Heterocystous
(V) Ramified
Nostoc sp. PCC 7524
Anabaena sp. PCC 7108
Microchaete sp. PCC 7126
Nostoc azollae 0708
Cylindrospermum stagnale PCC 7417
Nostoc sp. PCC 7107
Anabaena variabilis ATCC 29413
Anabaena Nostoc sp. PCC 7120
Calothrix sp. PCC 7507
Nostoc punctiforme PCC 73102
Anabaena cylindrica PCC 7122
Tolypothrix sp. PCC 9009
Nodularia spumigena CCY9414
Raphidiopsis brookii D9 D9 23
Cylindrospermopsis raciborskii CS-505
Calothrix sp. PCC 6303
Calothrix sp. PCC 7103
Symphyonemopsis sp. VAPOR1
Mastigocladopsis repens MORA PCC 10914
Chlorogloeopsis fritschii PCC 6912
Chlorogloeopsis sp. PCC 7702
Haplosiphon welwitschii IC-52-3
Fischerella sp. PCC 9431
Fischerella sp. PCC 9339
Fischerella sp. PCC 9605
Fischerella sp. JSC-11
Fischerella muscicola PCC 7414
Crinalium epipsammum PCC 9333
Rivularia sp. PCC 7116
Chamaesiphon sp. PCC 6605
Synechocystis sp. PCC 7509
Oscillatoria sp. PCC 10802
Gloeocapsa sp. PCC 7428
Chroococcidiopsis thermalis PCC 7203
Microcoleus sp. PCC 7113
Leptolyngbya sp. PAC-10-03
Moorea producens JHB
Moorea producens 3L
Moorea producens PNG 4 22 06-1
Moorea producens PAL-8-15-08-1
Moorea bouillonii PNG5-198
Planktothrix agardhii NIVA-CYA 126/8
Arthrospira platensis NIES-39
Arthrospira platensis str. Paraca
Arthrospira sp. PCC 8005
Arthrospira platensis C1
Arthrospira maxima CS-328
Lyngbya sp. PCC 8106
Oscillatoria acuminata PCC 6304
Oscillatoria sp. PCC 6506
Oscillatoria sp. PCC 6407
Oscillatoria sp. PCC 7112
Trichodesmium erythraeum IMS101
Okeania sp. PAC-18-Feb-10-1.1
Okeania comitata 3LOSC
Okeania hirsuta PAB-21-Jun-10-1
Okeania hirsuta PAB-10-Feb-10-1
Xenococcus sp. PCC 7305
Chroococcidiopsis sp. PCC 6712
Stanieria cyanosphaera PCC 7437
Dermocarpella incrassata PCC 7326
Pleurocapsa sp. PCC 7319
Pleurocapsa PCC 7516
Gloeocapsa sp. PCC 73106
Prochloron didemni P1 Palau
Spirulina sp. PCC 9445
Cyanobacterium sp. PCC 10605
Geminocystis herdmanii PCC 6308
Cyanobacterium stanieri PCC 7202
Synechocystis sp. PCC 6803
Leptolyngbya sp. PCC 6306
Spirulina major PCC 6313
Synechococcus sp. PCC 7002
Leptolyngbya sp. PCC 7376
Cyanothece sp. PCC 7822
Cyanothece sp. PCC 7424
Pleurocapsa sp. PCC 7327
Microcystis aeruginosa NIES-843
Cyanothece sp. PCC 8802
Cyanothece sp. PCC 8801
Cyanobacterium UCYN-A NC 013771
Crocosphaera watsonii WH 8501
Cyanothece sp. CCY0110
Cyanothece sp. ATCC 51472
Cyanothece sp. ATCC 51142
Halothece sp. PCC 7418
Dactylococcopsis salina PCC 8305
Cyanobacterium sp. ESFC-1
Geitlerinema sp. PCC 7105
Pseudanabaena sp. PCC 7367
Pseudanabaena sp. PCC 7429
Synechococcus sp. PCC 7502
Pseudanabaena sp. PCC 6802
Geitlerinema sp. PCC 7407
Cyanothece sp. PCC 7425
Acaryochloris sp CCMEE 5410
Acaryochloris marina MBIC11017
Synechococcus elongatus PCC 7942
Synechococcus elongatus PCC 6301
Synechococcus sp. WH 5701
Cyanobium gracile PCC 6307
Synechococcus sp. RCC307
Synechococcus sp. RS9917
Synechococcus sp. RS9916
Synechococcus sp. (9)
Prochlorococcus marinus (13)
Prochlorothrix hollandica PCC 9006
Leptolyngbya sp. PCC 7375
Leptolyngbya sp. PCC 6406
Nodosilinea nodulosa PCC 7104
Leptolyngbya sp. BL0902
Synechococcus sp. PCC 7335
Thermosynechococcus elongatus BP-1
Synechococcus sp. PCC 6312
Synechococcus sp. PCC 7336
Synechococcus sp. (6)
Gloeobacter violaceus PCC 7421
0.05
A
B
C

The clade containing the *Moorea* strains cannot be conclusively distinguished from the main OLS containing clade due to low bootstrap and posterior probability support, but *Leptolyngbya* sp. PCC6406 is clearly in a distinct lineage. By 16S rRNA sequence, this freshwater filamentous strain clades with other *Leptolygnbya* strains as well as unicellular marine *Synechococcus* strains. Thus, *Leptolyngbya* sp. PCC6406 is from a distinct evolutionary lineage from other OLS-containing cyanobacteria, yet appears to possess this same metabolic pathway.

2.4.2 Pathway Evolution

An alignment of DNA sequences for all known OLS pathways was constructed with annotations for open reading frames (Figure 2.2). This alignment also includes the CurM domain of the curacin A biosynthetic pathway, and as expected, shows that CurM does not contain the fatty acyl ACP ligase (FAAL) domain or the first acyl carrier protein (ACP) found in the OLS pathway (Figure 2.2) (Gehret-McCarthy et al. 2012).

Figure 2.2. DNA alignment of all 17 known OLS pathways and CurM with annotated ORFs. Four of the strains have the OLS pathway split into two ORFs. In these four strains the fatty acyl ACP ligase is a separate ORF from the PKS portion of the OLS pathway. Panel A shows all 17 aligned sequences with a red square highlighting the break between ORFs for the four pathways with two ORFs. Panel B shows the expanded red highlighted region from panel A. CurM is a part of the curacin A biosynthetic pathway which does not involve a FAAL (Gu et al. 2009)

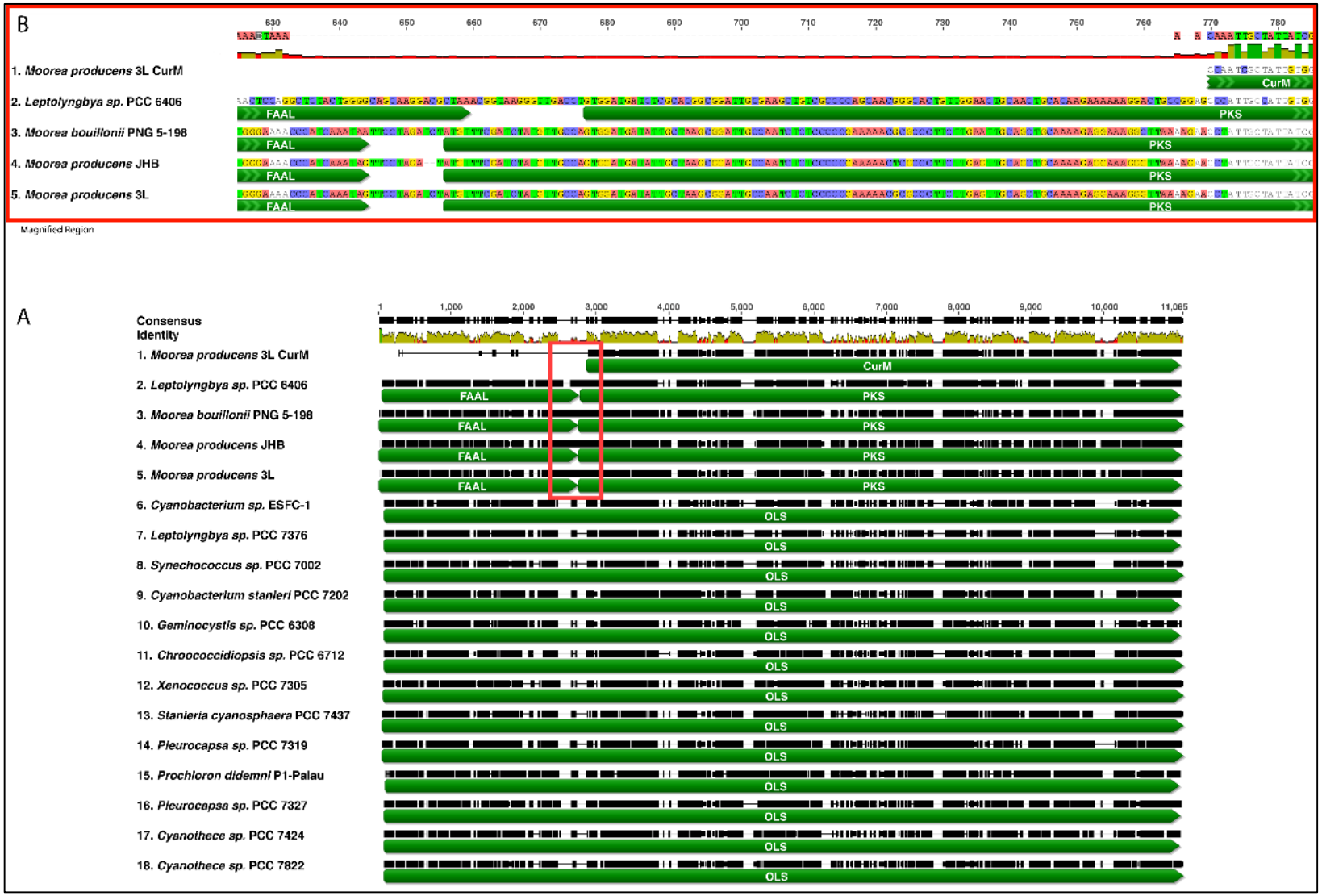
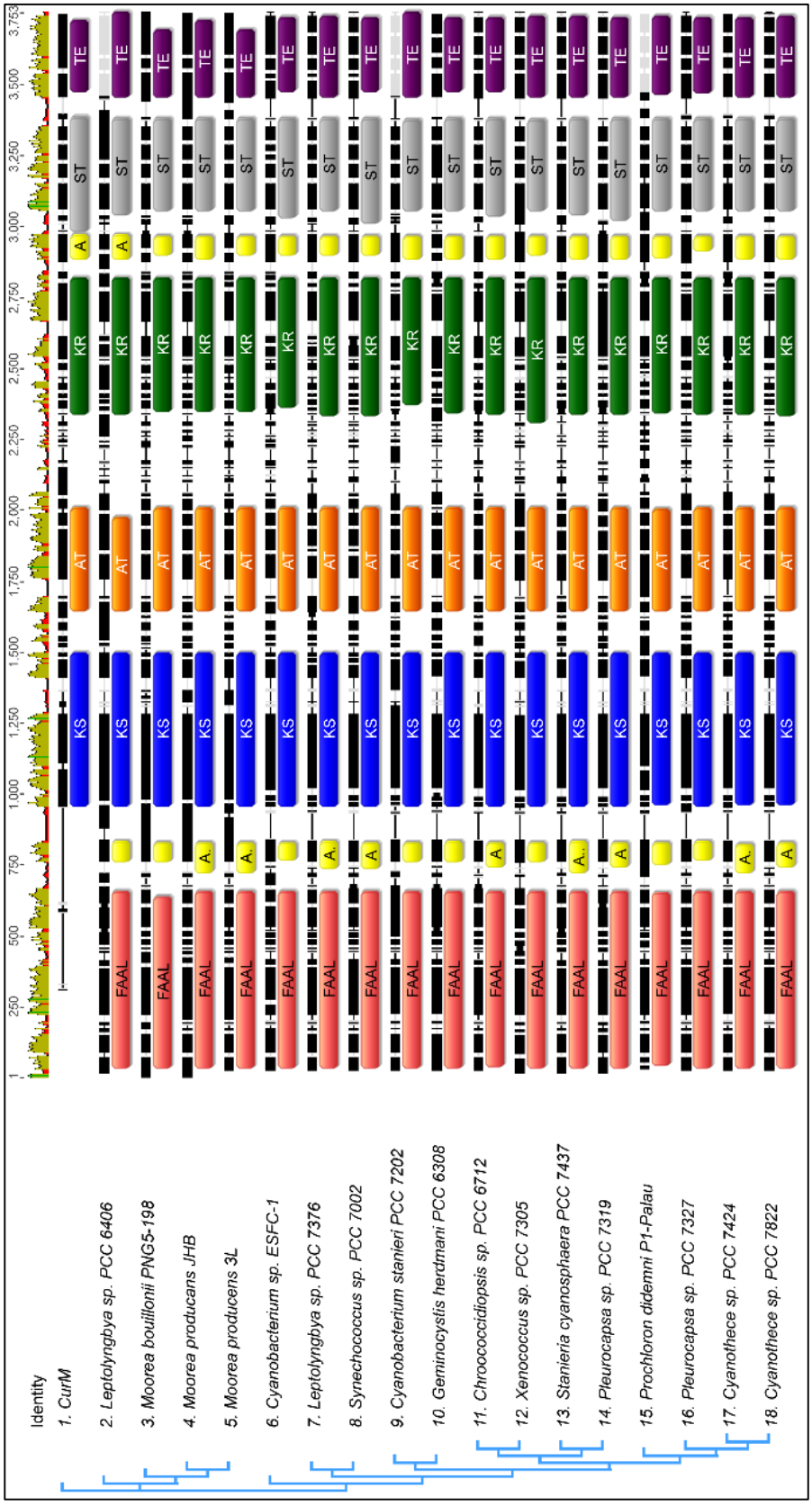

B
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
780
1. Moorea producens 3L CurM
CurM
2. Leptolyngbya sp. PCC 6406
FAAL
PKS
3. Moorea bouillonii PNG 5-198
FAAL
PKS
4. Moorea producens JHB
FAAL
PKS
5. Moorea producens 3L
FAAL
PKS
Magnified Region
A
Consensus
Identity
1
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
11,085
1. Moorea producens 3L CurM
CurM
2. Leptolyngbya sp. PCC 6406
FAAL
PKS
3. Moorea bouillonii PNG 5-198
FAAL
PKS
4. Moorea producens JHB
FAAL
PKS
5. Moorea producens 3L
FAAL
PKS
6. Cyanobacterium sp. ESFC-1
OLS
7. Leptolyngbya sp. PCC 7376
OLS
8. Synechococcus sp. PCC 7002
OLS
9. Cyanobacterium stanieri PCC 7202
OLS
10. Geminocystis sp. PCC 6308
OLS
11. Chroococcidiopsis sp. PCC 6712
OLS
12. Xenococcus sp. PCC 7305
OLS
13. Stanieria cyanosphaera PCC 7437
OLS
14. Pleurocapsa sp. PCC 7319
OLS
15. Prochloron didemni P1-Palau
OLS
16. Pleurocapsa sp. PCC 7327
OLS
17. Cyanothece sp. PCC 7424
OLS
18. Cyanothece sp. PCC 7822
OLS

Figure 2.3 Amino acid alignment and phylogenetic tree of all 17 OLS pathways and the CurM domain. All 17 of the OLS pathways contain the same domain architecture. CurM does not contain the FAAL and ACP1 domains. A maximum likelihood tree is displayed on the left of the alignment to depict the phylogenetic relationships between these pathways.


Identity
1. CurM
2. Leptolyngbya sp. PCC 6406
3. Moorea bouillonii PNG5-198
4. Moorea producans JHB
5. Moorea producens 3L
6. Cyanobacterium sp. ESFC-1
7. Leptolyngbya sp. PCC 7376
8. Synechococcus sp. PCC 7002
9. Cyanobacterium stanieri PCC 7202
10. Geminocystis herdmani PCC 6308
11. Chroococcidiopsis sp. PCC 6712
12. Xenococcus sp. PCC 7305
13. Stanieria cyanosphaera PCC 7437
14. Pleurocapsa sp. PCC 7319
15. Prochloron didemni P1-Palau
16. Pleurocapsa sp. PCC 7327
17. Cyanothece sp. PCC 7424
18. Cyanothece sp. PCC 7822
1
250
500
750
1,000
1,250
1,500
1,750
2,000
2,250
2,500
2,750
3,000
3,250
3,500
3,753
FAAL
A
KS
AT
KR
A
ST
TE

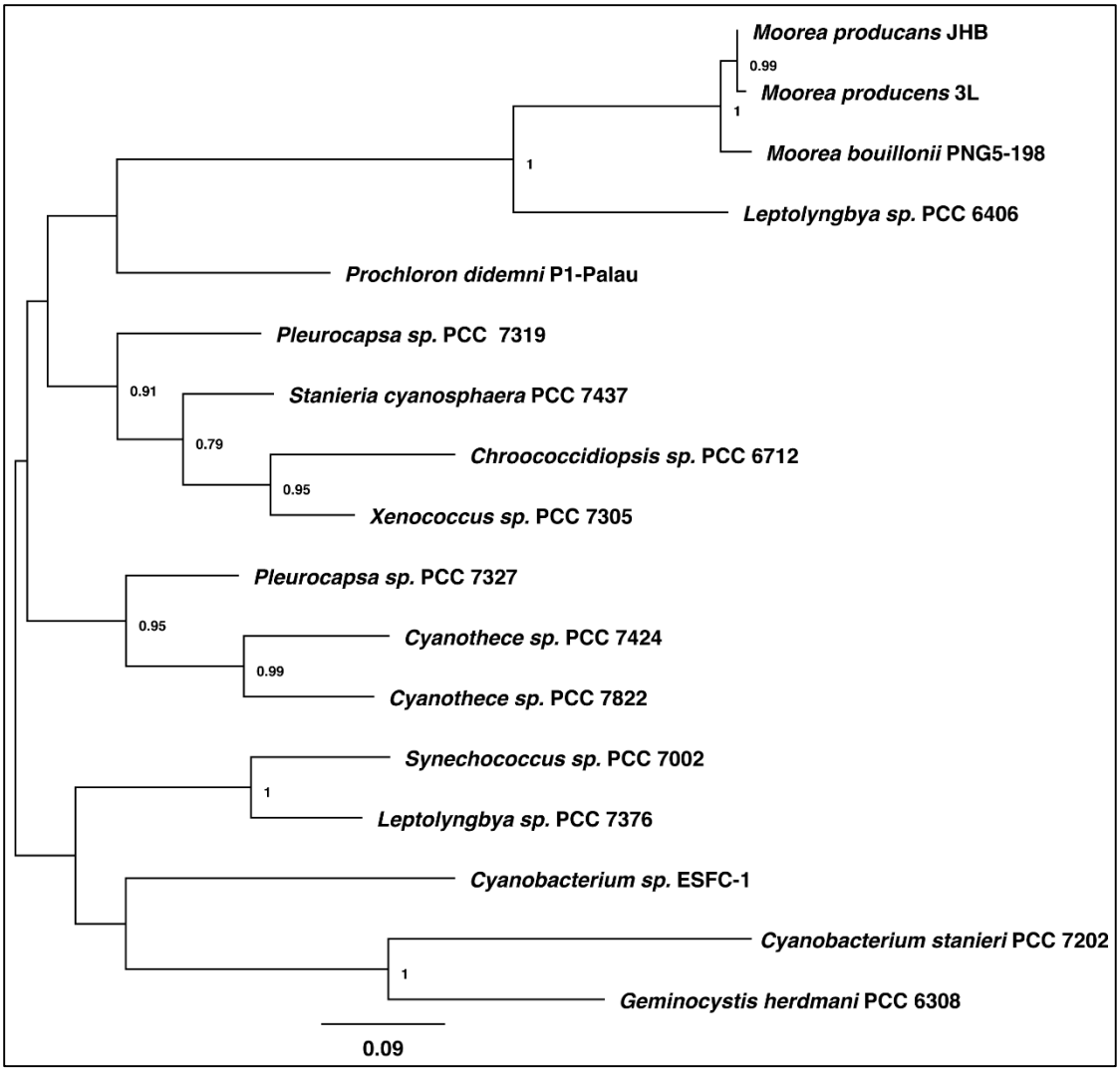


Figure 2.4 Phylogeny of the ketosynthase (KS) domain for 17 cyanobacterial OLS pathways. The KS domains from *Leptolyngbya* sp. PCC6406 and the three *Moorea* strains clade together, suggesting a common evolutionary history.

An alignment of the amino acid sequences of all 17 of the known OLS pathways and CurM shows that all of the OLS pathways contain the same domains and domain architecture (Figure 2.3). A phylogenetic tree of all OLS pathways was created to investigate their evolutionary relationships (Figure 2.4).

This phylogenetic tree (Figure 2.4) shows a similar topology with that produced from the corresponding 16S rRNA sequences (Figure 2.1) with the exception of *Leptolyngbya* sp. PCC6406. In this latter case, the OLS sequence clades with those from the *Moorea* strains and CurM. A phylogenetic tree was generated for all available ADO genes as well; however, evolutionary relationships and topological comparisons to the 16S rRNA phylogenetic tree were prohibited due to poor bootstrap support (data not shown). Significantly, none of the genome sequenced cyanobacteria appear to contain both the FAAR/ADO pathway as well as the OLS pathway. Additionally, no pseudogenes with homology to either of the genes in the FAAR/ADO pathway were detected in OLS-containing cyanobacteria, and similarly, no pseudogenes were detected for the OLS pathway in strains containing the FAAR/ADO pathway.

Table 2.1 Genome and OLS gene percent GC for all 17 OLS pathway containing cyanobacteria. Genome status is indicated (P-Permanent Draft, F-Finished, D-Draft). *Leptolyngbya* sp. PCC6406 exhibits a higher GC content in its OLS pathway compared to its genome, as well as compared to other OLS pathways and their genomes.

Genome	Genome GC%	OlS gene GC%	Status
Leptolyngbya sp. PCC 6406	55	64	P
Synechococcus sp. PCC 7002	49	53	F
Cyanobacterium sp. ESFC-1	47	51	P
Pleurocapsa sp. PCC 7327	45	44	F
Leptolyngbya sp. PCC 7376	44	47	F
Moorea producens 3L	44	47	D
Moorea bouillonii PNG 5-198	44	47	D
Moorea producens JHB	42	47	D
Prochloron didemni P1-Palau	42	44	D
Cyanothece sp. PCC 7822	40	44	F
Xenococcus sp. PCC 7305	40	43	D
Cyanothece sp. PCC 7424	39	42	F
Cyanobacterium stanieri PCC 7202	39	40	F
Pleurocapsa sp. PCC 7319	39	42	P
Stanieria cyanosphaera PCC 7437	36	38	F
Chroococcidiopsis sp. PCC 6712	35	38	F
Geminocystis herdmanii PCC 6308	34	38	P

The GC content for all cyanobacteria that possess the OLS pathway was compared to the GC content of the OLS pathway in those strains (Table 2.1). The GC content of the OLS pathways are all slightly higher than the GC content for all strains that possess the OLS pathway. Additionally, the GC content for *Leptolyngbya* sp. PCC6406 is noticeably higher than all other OLS pathways and significantly higher than its genome GC content.

Figure 2.5 displays the gene neighborhood for six representative cyanobacterial strains that exhibit the conserved gene neighborhood of the FAAR/ADO pathway. Most cyanobacteria with the FAAR/ADO pathway (88 of 128 (69%)) showed the same five

genes downstream of the FAAR/ADO pathway. These five genes are maintained in the same orientation and order in all strains where they are observed together. The remaining 40 cyanobacterial strains that exhibited a disruption in the gene neighborhood around the FAAR/ADO pathway included many strains that are closely related to strains possessing the OLS pathway. The gene neighborhood for a few of the OLS pathways also exhibited signs of gene neighborhood disruption (Figure 2.6). The gene neighborhood for *Moorea producens* 3L, *Moorea bouillonii* PNG5-198, *Cyanobacterium* sp. ESFC-1, and *Leptolyngbya* sp. PCC 6404 each contained some of the genes from the FAAR/ADO gene neighborhood (Figure 2.6). Both *Moorea* strains had all three of the genes from the FAAR/ADO gene neighborhood except the FAAR/ADO genes (Figure 2.6). Transposases as well as phage assembly proteins were observed in close proximity to the OLS pathway suggesting a possible mechanism through which the OLS pathway was acquired.

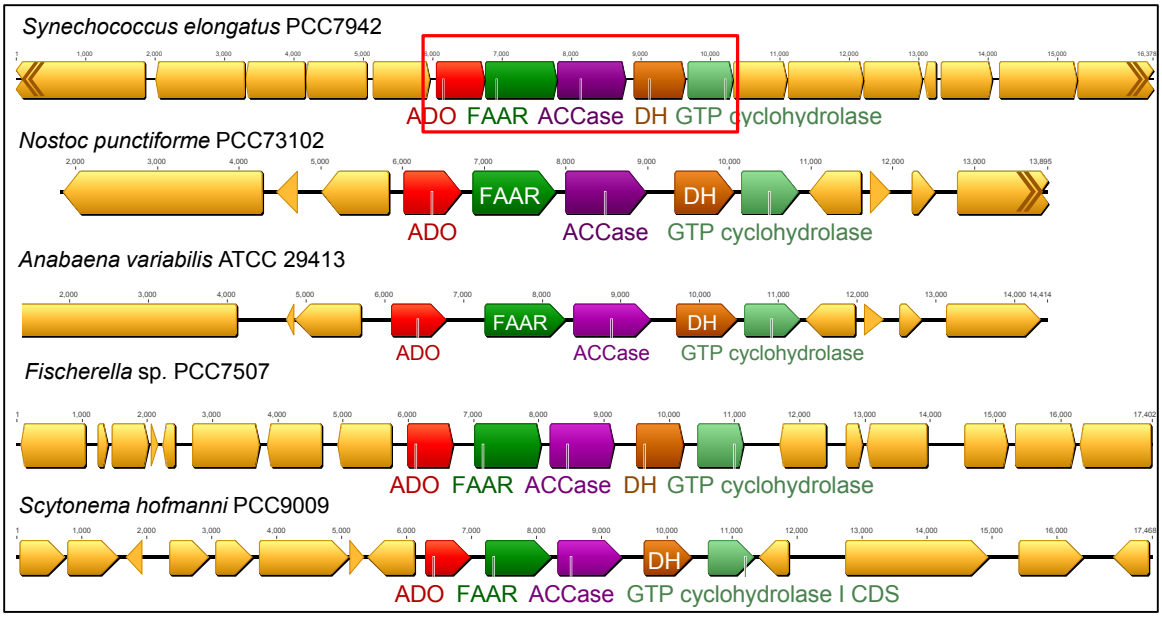


Figure 2.5. Gene neighborhood for six representative cyanobacterial strains that exhibit the conserved gene neighborhood of the FAAR/ADO pathway. The five genes commonly observed in the gene neighborhood with the FAAR/ADO gene are displayed at the top from *S. elongatus* PCC7942. An arrow pointing to each of the five genes and their respective labels is color coded according the color of the ORF in the gene neighborhood. Each of the observed genes is identified by a colored arrow corresponding to the color of the orf.

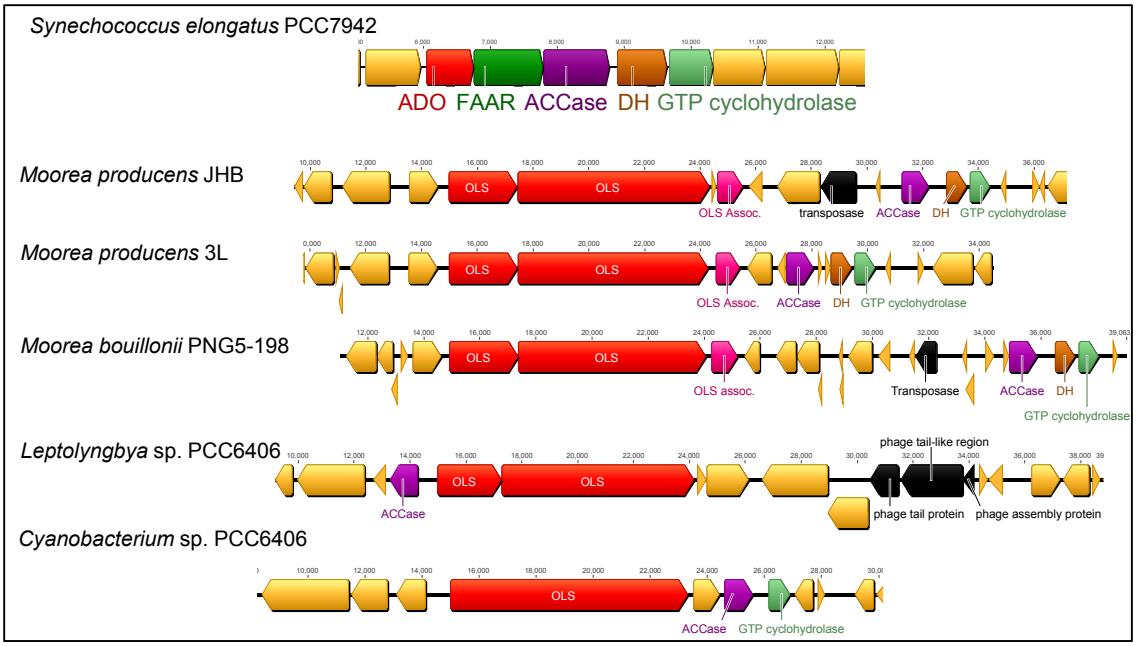


Figure 2.6. Gene neighborhood for four cyanobacteria that exhibit evidence of a disruption of the FAAR/ADO gene neighborhood in close proximity to the OLS pathway. The five genes commonly observed in the gene neighborhood with the FAAR/ADO gene are displayed at the top. An arrow pointing to each of the five genes and their respective labels is color coded according the color of the ORF in the gene neighborhood. Each of the observed genes is identified by a colored arrow corresponding to the color of the orf.

2.5 Discussion

2.5.1 Phylogenetic Distribution

The phylogenetic distribution of the two known hydrocarbon biosynthetic pathways among cyanobacteria is revealing of their evolutionary history. The FAAR/ADO pathway is the most widely distributed pathway taxonomically (122 of 139 strains) and is therefore most likely to be the ancestral hydrocarbon pathway in cyanobacteria (Figure 2.1). Additionally, *Gloeobacter violaceus* possesses the FAAR/ADO pathway and is considered the most primitive cyanobacteria because it lacks thylakoid membranes (Mareš et al. 2013). By comparison, the OLS pathway is only found in a small number of cyanobacteria (17 of 139) and likely evolved later than the FAAR/ADO pathway (Figure

2.1). A striking finding revealed by this genomic investigation is the absence of the alternative hydrocarbon pathway in a strain when either the OLS or the FAAR/ADO pathway is present. More specifically, none of the cyanobacteria with the OLS pathway possess the FAAR/ADO pathway or any pseudogenes derived from it. In reciprocal fashion, none of the organisms that possess the FAAR/ADO pathway also possess the OLS pathway or any derived pseudogenes. Because the OLS pathway contains many of the same enzymatic domains found in PKS and fatty acid biosynthesis pathways (ACP, KS, AT, KR, TE), homologs for each domain or even groupings of domains can be found in all cyanobacteria. Thus, the distinctive domains of this pathway are the fatty acyl ACP ligase (FAAL), sulfotransferase (ST) and thioesterase (TE). These domains impart the specificity of the pathway to first utilize fatty acid substrates, activate the β-hydroxy fatty acid via sulfonation and then catalyze hydrolysis and decarboxylation to produce terminal alkenes. While homologous genes for each of these latter domains can be found throughout cyanobacterial phylogeny, the unique domain architecture that makes up the OLS pathway is found only in OLS-containing cyanobacteria.

2.5.2 Hydrocarbon Pathway Evolution

Most of the strains that possess the OLS pathway are found within a single clade on the 16S rRNA phylogentic tree (Clade A, Figure 2.1). Cyanobacterial phylogenies using multi-locus sequence analyses appear to exhibit the same placement of clade A within the tree topology (Shih et al. 2013). However, this clade also includes many strains that possess the FAAR/ADO pathway, suggesting a unique evolutionary history that may have involved multiple horizontal gene transfer events. Thus, the evolutionary history of the OLS pathway in clade A cannot be definitively attributed to horizontal gene

transfer without additional information including further genome sequences of OLS-containing cyanobacteria and phylogenomic comparisons.

The clade containing the *Moorea* strains (Clade B) as well as the separate clade containing *Leptolyngbya* sp. PCC6406 (Clade C) are phylogenetically outside of this major clade of OLS-containing cyanobacteria. The *Moorea* strains, however, cannot be definitively characterized as separate given the rather low bootstrap/posterior probability support values in the phylogenetic tree in Figure 2.1. Nevertheless, the topological placement of this group of filamentous tropical marine cyanobacteria in Figure 2.1 is consistent with previous phylogenetic analyses (Engene et al. 2012). Additionally, the OLS pathway found in the three genome sequenced strains of *Moorea* and in *Leptolyngbya* sp. PCC6406 is consistently separated into two open reading frames with 9 to 17 bp intervening between the two ORFs (Figure 2.4). This distinctive pathway architecture is not found in any of the other 13 OLS-containing strains, and thus indicates a potentially distinct evolutionary history for these two groups of OLS pathways. Additionally, the complete OLS pathway and the KS alone for all three *Moorea* strains and *Leptolyngbya* sp. PCC6406 clade together suggesting a common evolutionary history (Figure 2.3 and 2.4). Despite the distinct pathway architectures and distinctive evolutionary history, the resulting hydrocarbon products do not appear to be fundamentally different as indicated by the production of 1-heptadecene in *Cyanothece* sp. PCC7822 and the three *Moorea* strains (Figure 3.2). Future investigation into the functional significance of the OLS pathway being split into two ORFs could involve real time PCR (RT-PCR) to investigate if both ORFs are expressed together.

Leptolyngbya sp. PCC6406 contains the OLS pathway, but according to its 16S rRNA phylogeny, is very distantly related to the other strains containing the OLS pathway (Figure 2.1). This latter discrepancy between 16S rRNA and OLS gene tree topologies suggests that the OLS pathway in *Leptolyngbya* sp. PCC6406 may have been obtained by horizontal gene transfer. Supporting this conclusion, the GC content of the *Leptolyngbya* sp. PCC6406 OLS pathway (64%) is higher than that of the entire genome (55%); this contrasts with the OLS pathways in all other cyanobacteria which exhibit similar (though slightly higher) GC contents compared to their respective genomes (Table 2.1).

Additionally, the conserved gene neighborhood around the FAAR/ADO appears to be disrupted in strains that are more closely related to strains possessing the OLS pathway (Figure 2.5 and 2.6). In fact four strains contain many of the genes from this conserved FAAR/ADO gene neighborhood in close proximity to OLS pathway in their genome (Figure 2.6). In these cases the FAAR/ADO pathway may have been displaced by the OLS pathway. Additionally, transposases and phage assembly genes are observed in close proximity to these pathway elements suggesting that the OLS pathway may have been introduced as a transposable element or introduced along with bacteriophage assembly genes during a previous infection (Figure 2.6).

Thus, while horizontal gene transfer appears to have played a role in the evolutionary history of the OLS pathway in *Moorea* and *Leptolyngbya* sp. PCC 6406, the extent to which horizontal gene transfer was involved in the evolutionary history of the other OLS pathway is unclear at this point. The competitive exclusion of the OLS and the FAAR/ADO pathway is intriguing and may suggest the presence of as yet unknown

selective pressures to maintain one or the other of these hydrocarbon biosynthetic pathways, but not both. Further investigation of the physiological and ecological role of cyanobacterial hydrocarbons, as well as further delineation of the phylogenetic distribution of these two pathways, may reveal insights as to the nature of this selection and competitive exclusion pressure.

2.6 Acknowledgements

Chapter 2 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

2.7 References

Bankevich AS, Nurk D, Antipov AA, Gurevich M, Dvorkin AS, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19. 455-477.

Bennett S, 2004. Solexa Ltd. Pharmacogenomics. 5. 433-438.

Benson DA, Cavanaugh M, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL, 2013. GenBank. Nucleic Acids Res 41. D36-42. doi: 10.1093/nar/gks1195

Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17. 540-552.

Donia MS, Fricke WF, Ravel J, Schmidt EW, 2011. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. PLoS ONE 6:e17897. doi:10.1371/journal.pone.0017897

Engene N, Rottacker EC, Kaštovský J, Byrum T, Choi H, Ellisman MH, Komárek J, Gerwick WH, 2012. *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb.

nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Ins J Syst Evol Microbiol 62.1171-1178.

Engene N, Paul VJ, Byrum T, Gerwick WH, Thor A, Ellisman MH, 2013 Five chemically rich species of tropical marine cyanobacteria of the genus Okeania gen. nov. (Oscillatoriales, Cyanoprokaryota). J Phycol doi. 10.1111/jpy.12115

Gehret McCarthy J, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, Sherman DH, Smith JL, 2012. Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. ACS Chem Biol 7. 1994-2003. doi: 10.1021/cb300385m

Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E, 2013 TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 41. D387-95. doi: 10.1093/nar/gks1234.

Kelley DR, Schatz MC, Salzberg SL, 2010 Quake: quality-aware detection and correction of sequencing errors. Genome Biol 11. R116. doi:10.1186/gb-2010-11-11-r116

Mareš, J, Hrouzek, P, Kaňa, R, Ventura, S, Strunecký, O, & Komárek, J, 2013. The primitive thylakoid-less cyanobacterium Gloeobacter is a common rock-dwelling organism. PloS one, 8(6), e66323.

Mendez-Perez D, Begemann MB, Pfleger BF, 2011 Modular synthase-encoding gene involved in α-olefin biosynthesis in *Synechococcus sp*. strain PCC7002 App Envir Micro 77. 4264-4267.

Ridley CP, Reppas NP, 2010. Biosynthesis of 1-alkenes in engineered microorganisms. US Patent Appl 12/821,107.

Shih PM, Wua D, Latifid A, Axena SD, Fewere DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA, 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. PNAS 110. 1053-1058.

Schirmer A, Rude MA, Li X, Popova E, del Cardayre S. 2010. Microbial biosynthesis of alkanes. Science 329. 559-562.

Schirrmeister BE, Antonelli A, Bagheri HC, 2011. The origin of multicellularity in cyanobacteria. BMC Evolutionary Biology 11.45-66. doi:10.1186/1471-2148-11-45

Williamson RT, Boulanger A, Vulpanovici A, Roberts MA, Gerwick WH, 2002. Structure and absolute stereochemistry of phormidolide, a new toxic metabolite from the marine cyanobacterium *Phormidium sp*. J Org Chem 67. 7927-7936.

Chapter 3: Characterization of cyanobacterial hydrocarbon compositions from phylogenetically diverse cyanobacterial strains

3.1 Abstract

Compositional analyses of cyanobacterial hydrocarbons have been previously limited with respect to phylogenetic diversity as well as biosynthetic context. Recently identified cyanobacterial hydrocarbon pathways can now be used to interpret variability in structural features for cyanobacterial hydrocarbon compositions. These biosynthetic interpretations can be used to identify pathway variability and new biochemical functions. They can also be used to predict hydrocarbon compositions in strains that have not yet been characterized. Hydrocarbon compositions of thirty-two phylogenetically diverse strains were characterized to reveal novel structural features and insights into hydrocarbon biosynthesis in cyanobacteria. This investigation revealed new double bond (2- and 3-heptadecene) and methyl group positions (3-, 4- and 5-methylheptadecane) for a variety of strains. Additionally, results from this study and literature reports indicate that hydrocarbon production is a universal phenomenon in cyanobacteria.

3.2 Introduction

Although the FAAR/ADO and the OLS pathway were only recently identified and characterized, knowledge of the diagnostic structural features of the resulting alk(a/e)nes and their respective gene sequences can be used along with previous reports of cyanobacterial hydrocarbon compositions to better understand the taxonomic distribution of these two pathways and their resulting structural diversity

(Table 3.1). However, these previous investigations have focused on a small subset of cyanobacteria, mostly comprised of strains from subdivision I (Unicellular), subdivision III (Filamentous) and subdivision IV (Heterocystous). Only a very limited number of investigations have focused on the hydrocarbons of subdivision II (Baeocystous) and subdivision V (Ramified) have been conducted (Table 3.1) (Castenholz 2001). In this context, the older morphology-based taxonomic perspective of cyanobacteria is severely limited due to the complex evolution of multicellularity in cyanobacteria (Schirrmeister et al. 2011). Indeed, multiple 16S rRNA and multi-locus-based phylogenetic trees of the cyanobacteria reveal have revealed that the cyanobacterial subdivisions are polyphyletic (Dembitsky and Srebnik 2002, Shih et al. 2013) (Figure 2.1). Nonetheless, the limited sampling to date of cyanobacterial diversity for hydrocarbons suggested that there may be as yet uncharacterized structural diversity. Additionally, previous reports of cyanobacterial hydrocarbons have rarely verified chain length, double bond positions or methyl group positions with authentic standards or other rigorous methods of structural analysis, but rather, have relied heavily upon mass spectral library searches (Table 3.1) (Dembitsky and Srebnik 2002, Tan *et al*. 2011). This latter approach introduces ambiguity concerning these key structural features, and therefore, could contribute to misunderstandings concerning the nature and distribution of hydrocarbon biosynthetic pathways and their evolution. For example, a number of reports include even chain-length hydrocarbons that are inconsistent with either of the known mechanisms of hydrocarbon production (Gelpi *et al*. 1970, Paoletti et al. 1976, Dembitsky *et al*. 1999, Liu *et al*. 2013). Both the FAAR/ADO and OLS pathway would be expected to produce only odd-chain alk(a/e)nes given their experimentally verified mechanisms (Schirmer *et*

al. 2010, Mendez-Perez et al. 2011, Gehret-McCarthy *et al*. 2012). This observation could be explained by either the existence of yet another pathway for alk(a/e)ne biosynthesis, or that the reported even chain length alk(a/e)nes have been mis-identified.

This investigation aimed to expand the characterization of structural diversity of cyanobacterial hydrocarbons through rigorous structural identification methods and across a wide range of cyanobacterial diversity. We hypothesized that this expanded structural diversity may reveal new insights concerning cyanobacterial biosynthetic pathways. Additionally, through a comparison of the hydrocarbon pathways found in the currently available sequenced cyanobacterial genomes, we were able to develop a deeper understanding of the relationships and evolution of these biosynthetic pathways in cyanobacteria.

Table 3.1. Summary of literature reports of cyanobacterial hydrocarbon composition. Strains are organized by subdivision on the left-hand side, and a '+' symbol indicates that a particular hydrocarbon was reported for this strain. The relevant reference is given in the right-hand column. Findings that were validated using authentic standards or established using derivatization techniques to verify hydrocarbon structure are denoted with superscript 1.

Strain	C14	meC14	C15	C15:1	C16	meC16	C17	C17:1	C17:2	C17:3	C17:4	meC17	C18	C18:1	meC18	C19	C19:1	C19:2	meC19	References
Subdivision 1 (Unicellular)																				
Synechococcus PCC 7002 (Agmenellum quadruplicum)																	+[1]	+[1]		Winter *et al*. 1969, Mendez-Perez *et al.* 2011
Anacystis cyanea							+[1]					+[1]								Gelpi *et al*. 1970
Anacystis montana			+	+	+		+	+	+											Gelpi *et al.* 1970, Murray & Thomson 1977
Anacystis nidulans			+		+		+[1]	+				+[1]	+							Paoletti *et al.* 1976, Winters *et al.* 1969, McInnes 1980, Han 1968
Chroococcus turgidus					+	+	+					+	+		+	+				Gelpi *et al.* 1970, Murray & Thomson 1977
Coccochloris elabens								+					+				+	+		Winters *et al.* 1969
Cyanothece sp. ATCC 51142			+[1]																	Schirmer *et al*. 2010
Cyanothece sp. PCC 7418							+	+					+	+						Allen *et al.* 2010
Cyanothece sp. PCC 7425							+[1]													Schirmer *et al*. 2010
Gloeocapsa alpicola FACHB 400			+				+													Liu *et al.* 2013
Gloeobacter violaceus PCC 7421							+[1]													Schirmer *et al*. 2010
Microcystis aeruginosa			+		+		+						+			+				Walsh *et al.* 1998
Microcystis sp. PCC 7806							+													Liu *et al.* 2013
Prochlorococcus marinus CCMP 1986			+[1]																	Schirmer *et al.* 2010
Prochloron sp.			+		+		+													Perry et al. 1978
Synechococcus bacillaris			+	+	+	+	+	+[1]	+	+	+									Blumer et al. 1971
Synechococcus elongatus PCC 6301			+[1]				+[1]													Schirmer *et al.* 2010
Synechococcus elongatus PCC 7942			+				+[1]	+												Shirmer *et al*. 2010, Tan *et al.* 2011
Synechocystis sp. PCC 6803							+[1]	+												Schirmer *et al.* 2010, Tan et al. 2011
Synechocystis UTEX 2470							+	+					+							Sakata *et al*. 1997
Subdivision 2 (Baeocystous)																				
Pleurocapsa PCC 7319							+[1]	+[1]												Allen *et al*. 2010
Pleurocapsa PCC 7418																				Allen *et al*. 2010
Pleurocapsa PCC 7516								+[1]	+[1]											Allen *et al*. 2010
Stanieria PCC 7301	+[1]	+[1]	+[1]		+[1]															Allen *et al*. 2010
Subdivision 3 (Filamentous)																				
Limnothrix redekei							+	+	+											Juttner 1991
Lyngbya aestuarii			+		+		+					+	+							Gelpi *et al*. 1970
Lyngbya lagerhaimii			+		+		+	+				+								Winters *et al*. 1969
Microcoleus chthonoplastes PCC 7420							+[1]	+[1]												Allen *et al*. 2010
Microcoleus chthonoplastes			+		+		+						+							Winters *et al*. 1969
Microcoleus vaginatus						+	+	+				+								Dembitsky *et al.* 2000
Oscillatoria f. granulata							+													Tsuchiya & Matsumoto 1999
Oscillatoria williamsii			+		+		+	+					+							Winters *et al*. 1969

Table 3.1 Continued

Strain	C14	meC14	C15	C15:1	C16	meC16	C17	C17:1	C17:2	C17:3	C17:4	meC17	C18	C18:1	meC18	C19	C19:1	C19:2	meC19	References
Osillatoria woronichinii (Sin 24)	+		+		+		+													Lee 1971, Blumer 1971
Phormidium luridum							+					+								Han 1969, Blumer 1971
Phormidium lucidum FACHB 238							+					+								Liu *et al.* 2013
Planktothrix agardhii							+	+				+								Juttner 1991
Planktothrix mougeotii							+	+				+								Juttner 1991
Planktothrix rubescens							+	+				+								Juttner 1991
Planktothrix suspense							+	+				+								Juttner 1991
Plectonema terebrans			+		+		+	+					+							Winter 1969
Spirulina sp.			+		+		+	+				+	+		+	+			+	Paoletti *et al*. 1976
Spirulina platensis			+		+		+	+												Gelpi *et al*. 1970
Spirulina platensis FACHB 419							+													Liu *et al.* 203
Spirulina platensis-Mao	+		+		+		+	+				+	+		+	+			+	Paoletti *et al.* 1976, 1977, 1978, Rezanka *et al.* 1982
Trichodesmium erythraeum			+		+		+	+					+							Winters *et al*. 1969, Capenter *et al.* 1998
Subdivision 4 (Heterocystous)																				
Anabaena (Nostoc) sp. PCC 7120							+[1]					+[1]								Schirmer *et al*. 2010, Tan *et al*. 2011
Anabaena cylindria								+	+				+							Murray & Thomson 1977
Anabaena variablis FACHB 319							+					+								Liu *et al.* 203
Anabaena variabilis							+[1]					+[1]								Han *et al.* 1969, Goodloe 1982, Fehler & Light 1970, Schirmer *et al.* 2010
Aulosira terrastre FACHB 256							+													Liu *et al.* 203
Calothrix scopulorum							+					+								Koster *et al*. 1999
Calothrix sp.			+		+		+	+	+				+		+	+			+	Paoletti *et al*. 1976
Gloeotrichia echinulata FACHB 419							+													Liu *et al.* 203
Nostoc sp.			+		+		+					+								Gelpi *et al*. 1970
Nostoc commune	+		+		+		+					+			+	+			+	Paoletti *et al*. 1976
Nostoc cacicola FACHB 389		+					+													Liu *et al.* 203
Nostoc endophytum							+					+								Murray & Thomson 1977
Nostoc muscorum			+		+		+	+[1]				+	+							Winters *et al*. 1969, Han *et al*. 1968
Nostoc punctiforme PCC 73102							+[1]													Schirmer *et al*. 2010
Nostoc punctiforme FACHB 252		+					+													Liu *et al.* 203
Nostoc spongiaeforme FACHB 130							+					+								Liu *et al.* 203
Nostoc sp.a							+					+	+			+				Dembitsky *et al*. 1999
Richelia sinica FACHB 419							+													Liu *et al.* 203
Scytonema sp.							+	+						+			+			Dembitsky & Srebnik 2002
Scytonema javanicum FACHB 887							+					+								Liu *et al.* 203
Tolypothrix tenuis FACHB							+					+								Liu *et al.* 203
Subdivision 5 (Ramified)																				
Chlorogloea fritschii					+		+[1]					+[1]			+					Han *et al.* 1968

3.3 Material and Methods

3.3.1 Strain Selection

Strains of cyanobacteria were selected from a wide phylogenetic distribution ensuring that every major clade was sampled (Figure 2.1). Strains of cyanobacteria with available genome sequences and established genetic techniques were given preference. Strains that required extremely low or high temperature or unusual media requirements for growth were excluded due to conditions and resources available at the time. *Anabaena (Nostoc)* sp. PCC7120 was provided by James Golden at UCSD. *Synechococcus elongatus* PCC7942 and *Leptolyngbya* BL0902 was provided by Susan Golden at UCSD. *M. producens* 3L, *Moorea bouillonii* PNG5-198 and *Moorea producens* JHB, *Leptolyngbya* sp. PAC10-3, *Moorea sp*. PNG4/22/06-1, *Moorea* sp. PAL8/15/08-1, *Okeania comitata* 3LOSC, *cf. Okeania hirsuta*. PAB 10-Feb-10-1, *Okeania hirsuta* PAB 21-Jun-10-1, *cf. Phormidium sp*. ISB 3/Nov/94-8, *Okeania* sp. PAC-18-Feb-10-1.1 were isolated from field collections and are maintained among the Gerwick Laboratory culture collection as described in (24, 34, 35). *Planktothrix agardhii* NIVA-CYA 126/8 was provided by Rainer Kurmayer at the Austrian Academy of Sciences Institute for Limnology. *Haplosiphon welwitchii* IC-52-3 and *Westiella intricata* HT-29-1 were provided by Thomas Hemscheidt of University of Hawaii via Melinda Micallef and Michelle Moffitt. The remaining strains were purchased from the Pasteur Culture Collection of Cyanobacteria (PCC) or the American Type Culture Collection (ATCC) as identified in the text.

3.3.2 Culture Conditions

Cyanobacterial strains were grown in 2.8 L Fernbach flasks under 16:8 day: night light cycle between 20-60 μE m^{-2} s^{-1} at a constant temperature at 20˚C, 25˚C, or 28˚C in BG-11, SWBG-11, or ASNIII. Cultures were shaken continuously (80 rpm) or grown statically between 14 and 35 days depending upon strain growth rates. Cultures (1L) were harvested via centrifugation at 4000 g for 15 min in 500 mL conical containers and combined (using 0.5 M ammonium formate to remove salts for marine strains and deionized water for freshwater strains) into 50 mL Falcon tubes and subsequently centrifuged again to yield a packed pellet that was frozen and dried for extraction and analysis. For filamentous strains that were not amenable to centrifugation, filaments were removed from the media using forceps (rinsed using 0.5 M ammonium formate to remove salts for marine strains), and frozen for drying, extraction and analysis. All biomass was lyophilized for at least 24 h.

3.3.3 Extraction and Structural Analysis

Dried biomass was ground using a mortar and pestle and weighed. Biomass was extracted using 5 mL of 100% hexanes or 2:1 dichloromethane:methanol (DCM:MeOH) followed by 20 sec of sonication. The extract was filtered using Whatman GF/F and the residual biomass re-extracted two additional times using the same method followed by a 10 mL wash with hexanes or 2:1 DCM:MeOH. Octadecane (Fluka-74691) was added to each extract after initial solvent addition as an internal standard for quantitation. Octadecane was added at approximately 0.1% of the initial dry biomass. DCM:MeOH (2:1) crude extracts were fractionated using a normal phase 500 mg Bonna-Agela Cleanert silica SPE column with collection of the first fraction which eluted with 100% hexanes. Extracts were dried under N_2 gas. A comparison between the hexane and

DCM:MeOH extraction methods for three strains (*Anabaena (Nostoc)* sp. PCC7120, *S. elongatus* PCC7942, *Synechococcus* sp. PCC7002) found that yields and composition of hydrocarbons were not significantly different (data not shown).

3.3.4 FAME Analysis

To characterize potential fatty acid substrates for hydrocarbon biosynthesis, a fatty acid analysis was performed on crude extracts (2:1 DCM:MeOH) of *Anabaena (Nostoc)* sp. PCC7120*, M. producens* 3L, and *Synechococcus* sp. PCC7002. Fatty acid methyl esters (FAMEs) were produced by transesterification by adding 3 mL of 4% H_2SO_4 (in MeOH) to at least 0.1 mg of a crude extract. Samples were then stirred and incubated at 110˚C for 1 h. Four mL of H_2O and 3 mL of hexanes were then added to the sample. After vortexing for 30 sec and allowing time for phase separation (3 min) the hexanes layer (top) was removed and dried in a pre-weighed vial for GC-MS analysis. FAME preparation was followed by dimethyl disulfide (DMDS) derivitization for determination of double bond positions (Vincent et al. 1987).

Each extract was resuspended to a concentration of 100 µg/mL in hexanes and 1 µL was analyzed by gas chromatography mass spectrometry (GC-MS) using a Thermo Trace GC-DSQ instrument equipped with an Agilent DB5-ms column (30 m, ID: 0.25, Film: 0.25µm). Helium (constant flow 1 mL/min) was used as the carrier gas. The inlet temperature was 240˚C and the following temperature program was applied: 40˚C for 1 min with an increase of 4.5˚C/min to 250˚C for 10 min. Data were acquired and processed with the Thermo Xcaliber software. Hydrocarbons were determined using a combination of mass fragmentation patterns, retention time and comparison to authentic standards when available (heptadecane (Fluka-51578), 1-heptadecene (TCI-S0347), 7-

methylheptadecane (kindly provided by Dieter Enders and Wolfgang Bettray, RWTH, Aachen University)), or published mass spectra and the NIST mass spectral library for Xcaliber (2005) when not available. Double bond positions were confirmed for all alkenes and unsaturated fatty acids using the DMDS method (Vincent et al. 1987).

GC-MS detector response factors for heptadecane, 1-heptadecene, and 7-methylheptadecane were determined in comparison with the octadecane standard by creation of standard curves. Standard curves for hydrocarbons with authentic standards were verified using the low mass common to all hydrocarbons analyzed via GC-MS (*m/z* 57). Hydrocarbon concentration was calculated using hydrocarbon peak area compared to the internal standard (octadecane) peak area using *m/z* 57. Percent dry weight was calculated as an average of three biological replicates. Statistical analysis of the hydrocarbon yields between the OLS and FAAR/ADO pathway were completed using a Mann-Whitney-Wilcox Test.

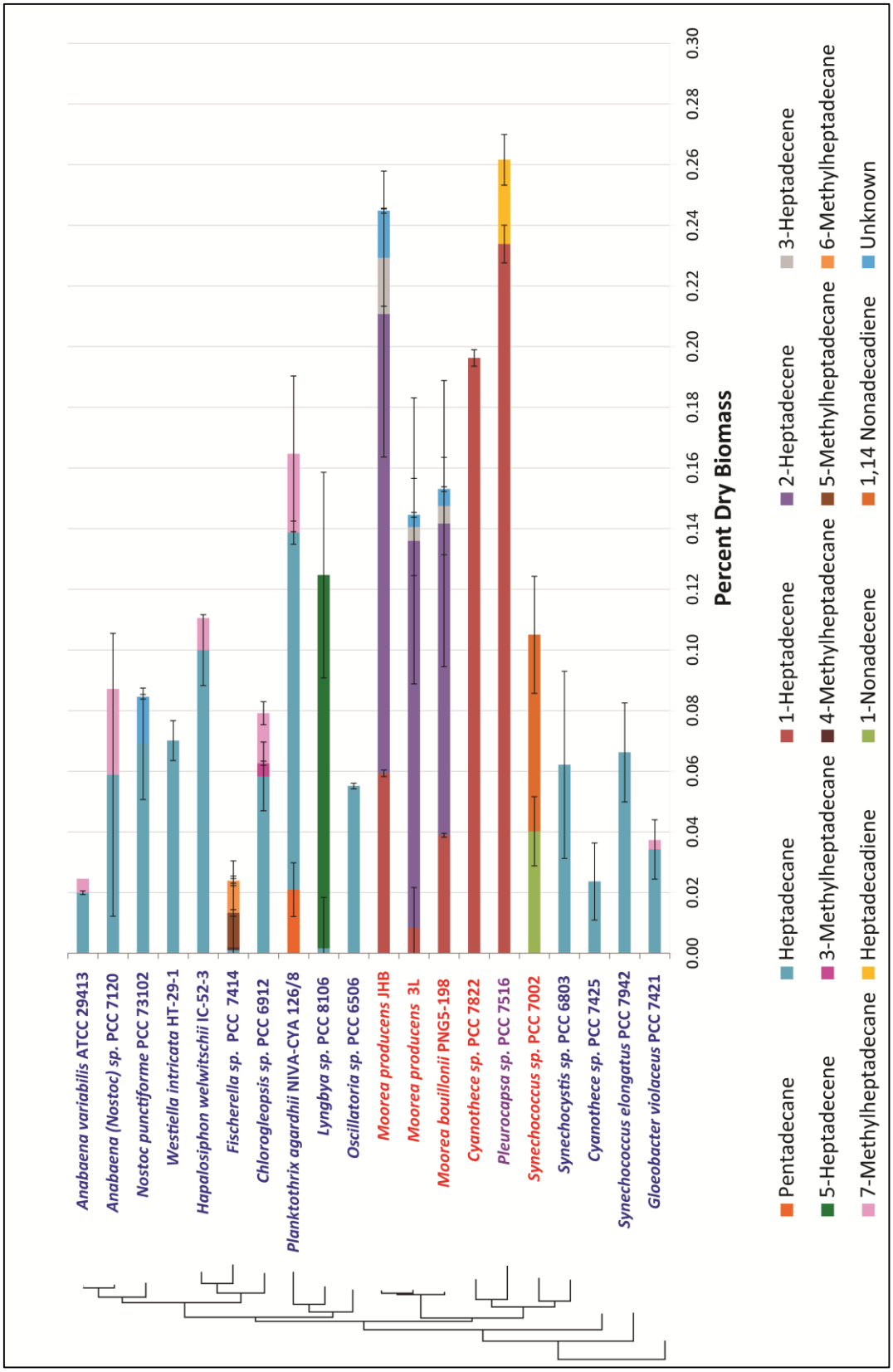
3.4 Results

3.4.1 Hydrocarbon Composition

The hydrocarbon composition of thirty-two cyanobacterial strains were analyzed and compared in the context of their respective biosynthetic pathways. Hydrocarbon yields as a percent of dry biomass ranged from 0.024% ± 0.01% in *Cyanothece* sp. PCC7425 to 0.262% ± 0.01% in *Pleurocapsa* sp. PCC7516 (Figure 3.1 and 3A1), and averaged 0.11% (± 0.015%) across all strains. Strains possessing the OLS pathway appear to have significantly higher levels of hydrocarbons as a percentage of dry biomass (0.173% ±0.032%) than strains with the FAAR/ADO pathway (0.070% ± 0.008%) (p-

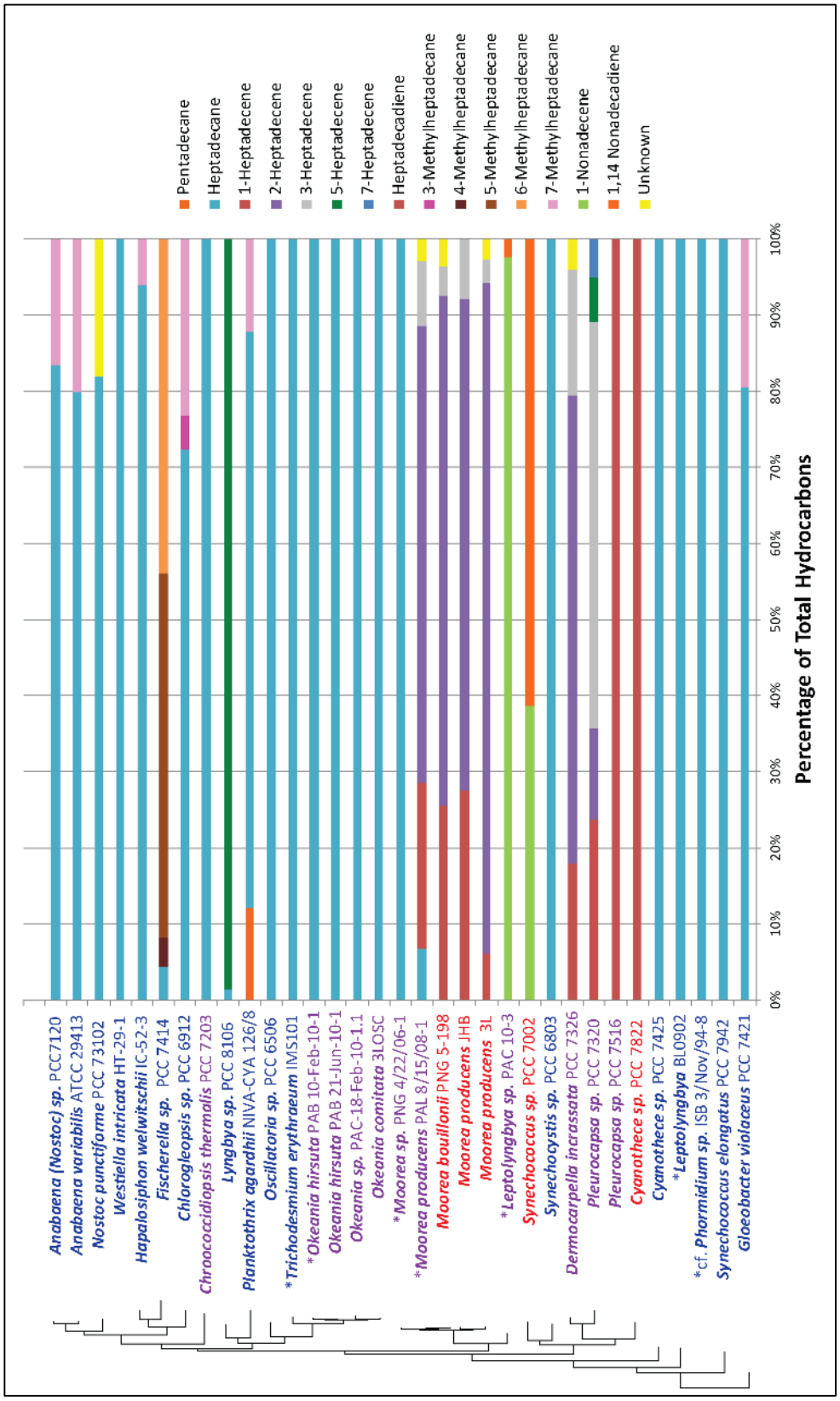
value = 0.0002). Heptadecane was the most commonly observed hydrocarbon (24 out of 32 strains). Alkanes such as heptadecane as well as all branched alkanes (e.g. 7-methylheptadecane), were only observed in strains that contained the FAAR/ADO pathway while strains with the OLS pathway appeared to only produce alkenes. *Lyngbya* sp. PCC8106 is the only strain that produced alkenes and also possessed the FAAR/ADO pathway. However, *Lyngbya* sp. PCC8106 produced 5-heptadecene, an alkene with an internal double bond instead of a terminal alkene consistent with the OLS pathway. Therefore, 5-heptadecene in *Lyngbya* sp. PCC8106 is likely derived from a monounsaturated fatty acid.

Figure 3.1. Quantitative yields of hydrocarbons as percent dry weight of biomass from 20 cyanobacteria displayed by phylogenetic relationship and pathway distribution. Specific hydrocarbons for each strain are color coded and stacked to depict the overall quantitative yield. Standard error bars are given for each hydrocarbon and each strain. Blue strain names indicate strains possessing the FAAR/ADO pathway and red indicates those with the OLS pathway. Purple strain names indicate a strain that does not have a genome sequence and therefore the pathway type is unknown. To the left of the figure, a 16S rRNA phylogenetic tree (Maximum Likelihood, see Figure 3A1 for complete tree) is presented for all 20 displayed strains. Branch tips are aligned to corresponding strain names except for *Westiella intricata* HT-29-1 for which no 16S rRNA sequence is available. Vertical connection lengths were modified to accommodate the location of *W. intricata* HT-29-1 in the table.


Anabaena variabilis ATCC 29413
Anabaena (Nostoc) sp. PCC 7120
Nostoc punctiforme PCC 73102
Westiella intricata HT-29-1
Hapalosiphon welwitschii IC-52-3
Fischerella sp. PCC 7414
Chlorogleopsis sp. PCC 6912
Planktothrix agardhii NIVA-CYA 126/8
Lyngbya sp. PCC 8106
Oscillatoria sp. PCC 6506
Moorea producens JHB
Moorea producens 3L
Moorea bouillonii PNG5-198
Cyanothece sp. PCC 7822
Pleurocapsa sp. PCC 7516
Synechococcus sp. PCC 7002
Synechocystis sp. PCC 6803
Cyanothece sp. PCC 7425
Synechococcus elongatus PCC 7942
Gloeobacter violaceus PCC 7421
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
Percent Dry Biomass
Pentadecane
Heptadecane
1-Heptadecene
2-Heptadecene
3-Heptadecene
5-Heptadecene
3-Methylheptadecane
4-Methylheptadecane
5-Methylheptadecane
6-Methylheptadecane
7-Methylheptadecane
Heptadecadiene
1-Nonadecene
1,14 Nonadecadiene
Unknown

Branched alkanes were observed in many of the strains investigated (7 of 32; Figure 3.3 and 3.4). Most commonly as 7-methylheptadecane, these branched alkanes were observed only in particular clades including heterocystous, ramified, and some filamentous cyanobacteria including *Planktothrix* (Figures 3.2). Additionally, 7-methylheptadecane was observed in this study from *Gloeobacter violaceus* PCC7421, and constitutes a first report of branched hydrocarbons from this organism. *Fischerella* sp. PCC7414 was found to produce several less common methylheptadecanes with the methyl groups assigned to positions C-4, C-5, and C-6 on the basis of fragmentation patterns (Figure 3.2 and 3.3). *Chlorogleopsis* sp. PCC6912 also produces less commonly observed branched alkanes, namely 3- and 4-methylheptadecane (Figure 3.2). However, except for 3- and 7-methylheptadecanes for which authentic standards exist, definitive assignment of the position of methylation in the above alkanes remains tentative.

Figure 3.2. Hydrocarbon composition expressed as a percentage of total hydrocarbons for 32 strains of cyanobacteria displayed by phylogenetic relationship and pathway distribution. Percentages are displayed as mean percentage between three replicates except for those indicated with an asterisk for strains that were characterized using a single sample. Blue strain names indicate strains possessing the FAAR/ADO pathway and red indicates those with the OLS pathway. Purple strain names indicate a strain that does not have a genome sequence and therefore the pathway type is unknown. To the left of the figure, a 16S rRNA phylogenetic tree (Maximum Likelihood, see Figure 3A2 for complete tree) is presented for all 32 displayed strains. Branch tips are aligned to the corresponding strain names. Branches corresponding to *W. intricata* HT-29-1, *cf. Phormidium* sp. ISB 3/Nov/94-8 and *Pleurocapsa sp*. PCC7320 are not shown because 16s rRNA sequences are not available for these strains. Vertical connection lengths were modified to accommodate the location of these three strains in the table.


Pentadecane
Heptadecane
1-Heptadecene
2-Heptadecene
3-Heptadecene
5-Heptadecene
7-Heptadecene
Heptadecadiene
3-Methylheptadecane
4-Methylheptadecane
5-Methylheptadecane
6-Methylheptadecane
7-Methylheptadecane
1-Nonadecene
1,14 Nonadecadiene
Unknown
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Percentage of Total Hydrocarbons
Anabaena (Nostoc) sp. PCC7120
Anabaena variabilis ATCC 29413
Nostoc punctiforme PCC 73102
Westiella intricata HT-29-1
Hapalosiphon welwitschii IC-52-3
Fischerella sp. PCC 7414
Chlorogleopsis sp. PCC 6912
Chroococcidiopsis thermalis PCC 7203
Lyngbya sp. PCC 8106
Planktothrix agardhii NIVA-CYA 126/8
Oscillatoria sp. PCC 6506
*Trichodesmium erythraeum IMS101
*Okeania hirsuta PAB 10-Feb-10-1
Okeania hirsuta PAB 21-Jun-10-1
Okeania sp. PAC-18-Feb-10-1.1
Okeania comitata 3LOSC
*Moorea sp. PNG 4/22/06-1
*Moorea producens PAL 8/15/08-1
Moorea bouillonii PNG 5-198
Moorea producens JHB
Moorea producens 3L
*Leptolyngbya sp. PAC 10-3
Synechococcus sp. PCC 7002
Synechocystis sp. PCC 6803
Dermocarpella incrassata PCC 7326
Pleurocapsa sp. PCC 7320
Pleurocapsa sp. PCC 7516
Cyanothece sp. PCC 7822
Cyanothece sp. PCC 7425
*Leptolyngbya BL0902
*cf. Pharmidium sp. ISB 3/Nov/94-8
Synechococcus elongatus PCC 7942
Gloeobacter violaceus PCC 7421

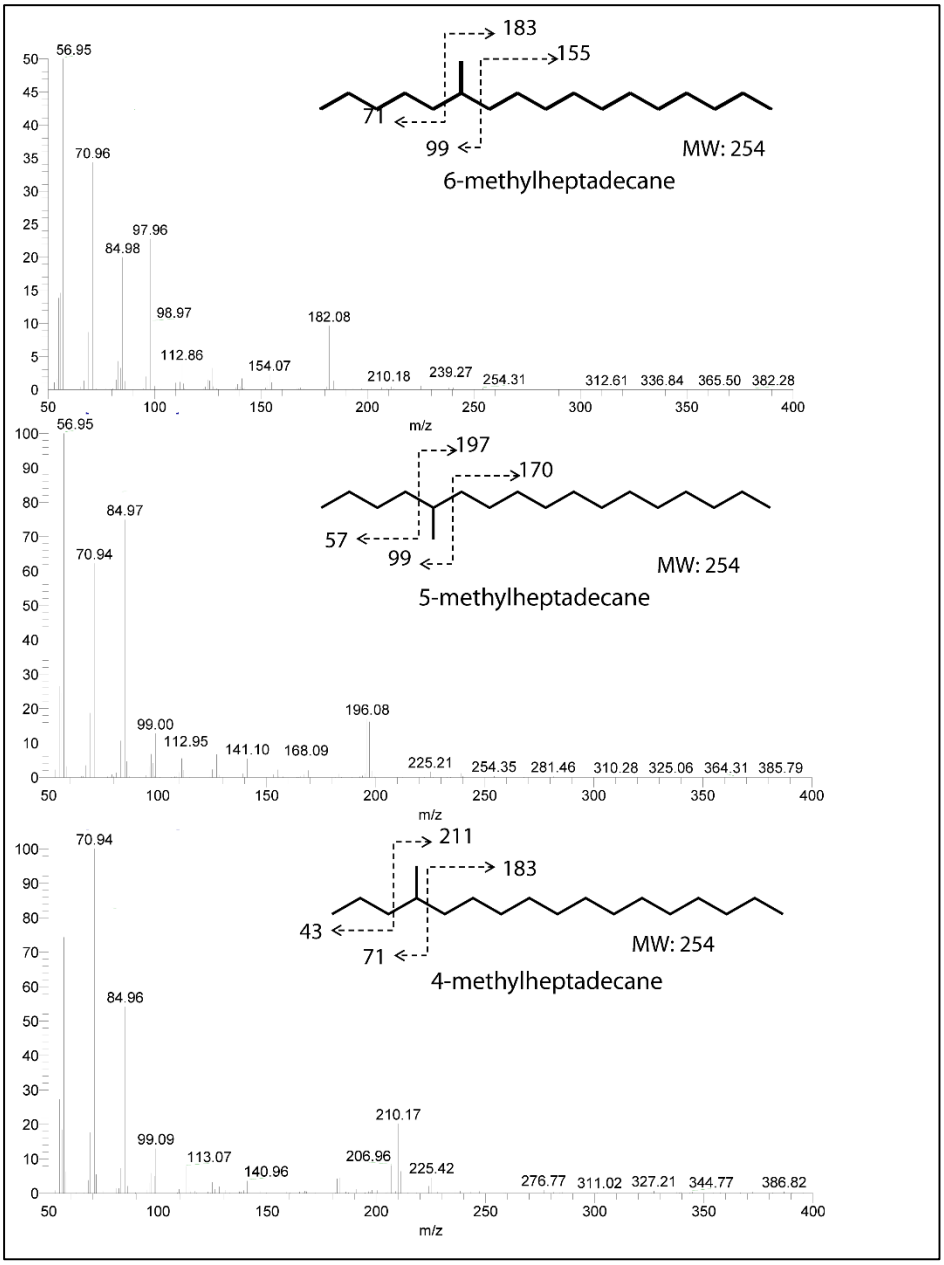


Figure 3.3. Mass spectra of branched hydrocarbons observed in *Fischerella* sp. PCC7414. The fragmentation patterns were used to propose the locations of methyl group substitutions on a heptadecane parent structure. In addition to the fragment losses depicted, neutral ion losses were also present.

Through rigorous identification of double bond positions, this study was able to differentiate which strains produce terminal alkenes and which produce alkenes with double bonds at internal positions. Mass fragmentation of DMDS derivatives revealed

several novel hydrocarbon structures such as 2- and 3-heptadecene (observed in all three *Moorea* strains as well as *Dermocarpella incrassata* PCC7326) as well as the tentative identification of heptadecadiene in *Pleurocapsa sp*. PCC7516.

The fatty acids of *Anabaena (Nostoc)* sp. PCC7120*, M. producens* 3L, and *Synechococcus* sp. PCC7002 were analyzed to evaluate the available substrates for alkane or alkene biosynthesis (Figure 3.4). Hexadecanoic acid was the most abundant fatty acid in all three strains followed by various unsaturated C16 and C18 fatty acids. *Anabaena (Nostoc)* sp. PCC7120 possessed 9-hexadecenoic acid, 9- and 11-octadecenoic acid, and octadecanoic acid in descending order of abundance. *M. producens* 3L also possessed 11-octadecenoic acid but in significantly lower abundance compared to 9- or 11-hexadecenoic acid, the two most abundant fatty acids after hexadecanoic acid. *M. producens* 3L also possessed tetradecanoic acid as well as 9-tetradecenoic acid. *Synechococcus* sp. PCC7002 possessed 9-hexadecenoic acid in addition to 12-octadecenoic acid, 9, 12-octadecadienoic acid, octadecanoic acid and tetradecanoic acid in descending order of abundance. There was no evidence of odd chain length fatty acids in any of the strains investigated.

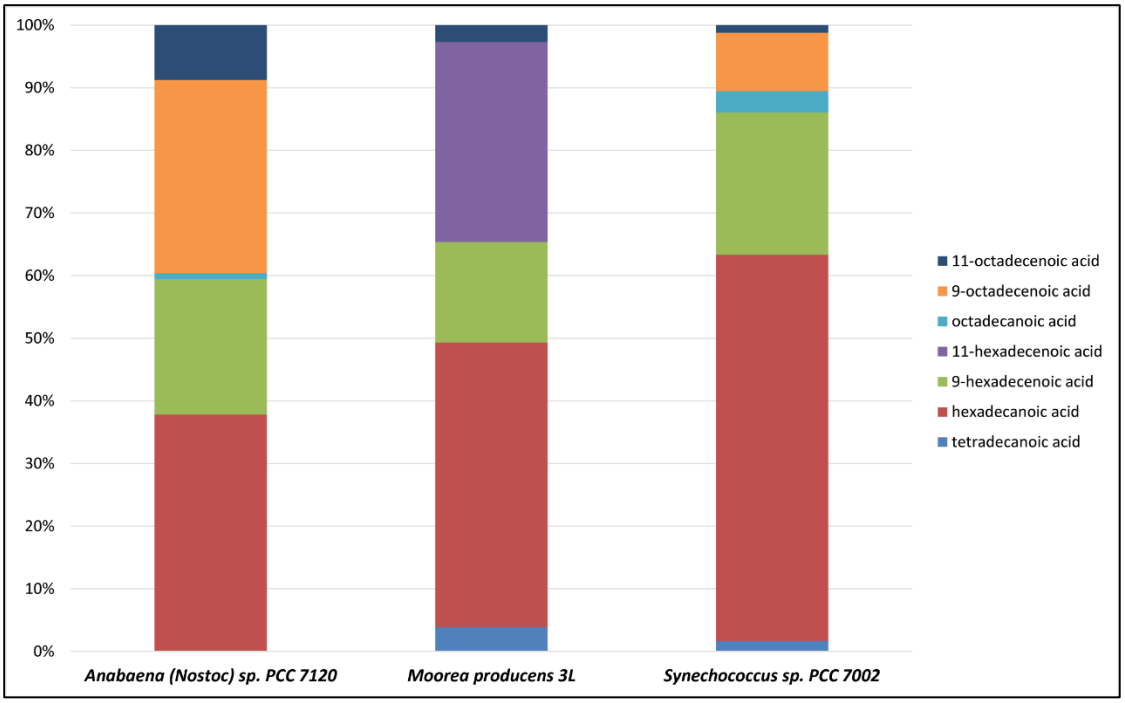


Figure 3.4. Fatty acid analysis of three cyanobacteria. Fatty acid analysis from single samples of A*nabaena (Nostoc)* sp. PCC7120, *M. producens* 3L, and *Synechococcus sp*. PCC7002. All three strains exhibit similar proportions of hexadecanoic acid and 9-hexadecenoic acid; however, A*nabaena (Nostoc)* sp. PCC7120 exhibits a higher proportion of 9-octadecenoic acid and 11-octadecenoic acid and tetradecanoic acid is absent. *M. producens* 3L contains 11-hexadecenoic acid and no octadecanoic acid. *Synechococcus sp*. PCC7002 exhibited a similar composition to A*nabaena (Nostoc)* sp. PCC7120 yet has a higher proportion of hexadecanoic acid.

3.5 Discussion

A striking general conclusion that emerges from review of previous studies along with the results of this study is that hydrocarbon production is a universal phenomenon among cyanobacteria (Figure 2.1, 3.3, and Table 3.1). However, hydrocarbon production in cyanobacteria appears to be derived from at least two very different pathways. Moreover, the specific structural features of the hydrocarbons reflect which pathway is present; saturated alkanes are found in strains with the FAAR/ADO pathway and terminal alkene containing-hydrocarbons are found only in OLS containing strains. As reported previously, heptadecane is the most commonly observed hydrocarbon in cyanobacteria

followed by heptadecene, pentadecane and 7-methylheptadecane (Table 1) (Han et al. 1968, Schirmer et al. 2010, Allen et al. 2010). This observation is consistent with octadecanoic acid (FAAR/ADO) or hexadecanoic acid (OLS) precursor fatty acids, and these along with a variety of unsaturated derivatives are the most common fatty acids in cyanobacteria (Sata and Wada 2010).

3.5.1 Hydrocarbon Composition

Most of the strains investigated in this study had never been previously characterized for their hydrocarbon composition (22 of 32). This investigation found a wide variation in the content of hydrocarbons between these cyanobacterial strains with a range of only 0.024% ± 0.01% in *Cyanothece* sp. PCC7425 to 0.262% ± 0.01% in *Pleurocapsa* sp. PCC7516 (Figure 3.1). These results expand the previously reported range of natural cyanobacterial hydrocarbon yields that was between 0.025 - 0.12% dry weight (Han et al. 1968, Winters *et al*. 1969). Strains possessing the OLS pathway appear to have significantly higher hydrocarbon yields (0.173% ±0.032) than strains with the FAAR/ADO pathway (0.070% ± 0.008) (p-value = 0.0002).

Branched alkanes have been used as a biomarker for cyanobacteria, and consistent with this, they appear to be widely distributed across cyanobacterial phylogeny (Figure 2.1, Table 1) (Han et al.1968, Allen et al. 2010). Branched alkanes have been observed mostly in filamentous cyanobacteria but there are a few reports of branched hydrocarbons in unicellular strains (*Anacystis nidulans, Anacystis cyanea,* and *Chrococoocus turgidus*) (Han et al. 1968, Gelpi et al. 1970, Liu et al. 2013). Expanding this distribution, we show for the first time that the unicellular cyanobacterium *G. violaceus* PCC7421 also produces 7-methylheptadecane. This study also expanded the known structural diversity

of cyanobacterial hydrocarbons to include additional double bond positions (2- and 3-heptadecenes) and methyl group positions (3-, 4- and 5-methylheptadecanes). Overall, branched alkanes were observed in 7 of the 32 strains examined in this investigation. Branched alkane biosynthesis was previously investigated in cyanobacteria and the pendant methyl group found to be derived from S-adenosylmethionine (SAM) through a methyltransferase reaction (Fehler and Light 1970, 1972). However, the methyltransferase involved in this pathway has yet to be identified and characterized. Han et al. (1968) and Fehler & Light (1970) used radiolabeled substrates to verify that vaccenic acid (11-octadecenoic acid) is the likely precursor to 7- or 8-methylheptadecane in *Nostoc muscorum.* We also observed 11-octadecenoic acid in our fatty acid analysis of *Anabaena (Nostoc)* sp. PCC7120, thus confirming the possibility that this fatty acid is the precursor to these branched hydrocarbons in this strain (Figure 3.4). Future investigations into the methyltransferase responsible for branched hydrocarbon biosynthesis may reveal not only the substrate required for this pathway but may also provide insights into the unique methylation patterns observed in this investigation for *Fischerella* sp. PCC7414 and *Chlorogleopsis* sp. PCC6912. Authentic standards are not available for the 4-,5-,6- and 7-methylheptadecanes that are predicted from observed branched hydrocarbons from these strains in this investigation (Figure 3.3). Future investigations into branched hydrocarbon production in these strains would likely require the synthesis of authentic standards to verify methylation positions for these strains.

All three *Moorea* strains as well as *D. incrassata* PCC7326 and *Pleurocapsa* sp. PCC7320 produced the unique alkenes 2- and 3-heptadecene. Two possible pathways may be responsible for the production of these observed alkenes. First, a desaturase could

act upon a fatty acid substrate to produce double bonds at the ω-2 or ω-3 position (Figure 3.5). The unsaturated fatty acid could then undergo elongation and decarboxylation via the OLS pathway followed by a single round of reduction to remove the newly introduced terminal double bond. Alternatively, an isomerase might act on 1-heptadecene to produce these 2- and 3-heptadecenes (Figure 3.5). Fatty acid analysis of *M. producens* 3L did not reveal any unsaturated fatty acids with double bonds at these ω-2 or ω-3 positions, but instead, hexadecanoic acid, 11-hexadecenoic acid, 9-hexadecenoic acid, and 11-octadecenoic were observed in order of decreasing abundance (Figure 3.4). The lack of ω-2 or ω-3 unsaturated fatty acids in this cyanobacterium suggests that an isomerase is likely involved in 2- and 3-heptadecene biosynthesis. Consistent with this hypothesis, the required 16:0 fatty acid is the most abundant fatty acid in *M. producens* 3L.

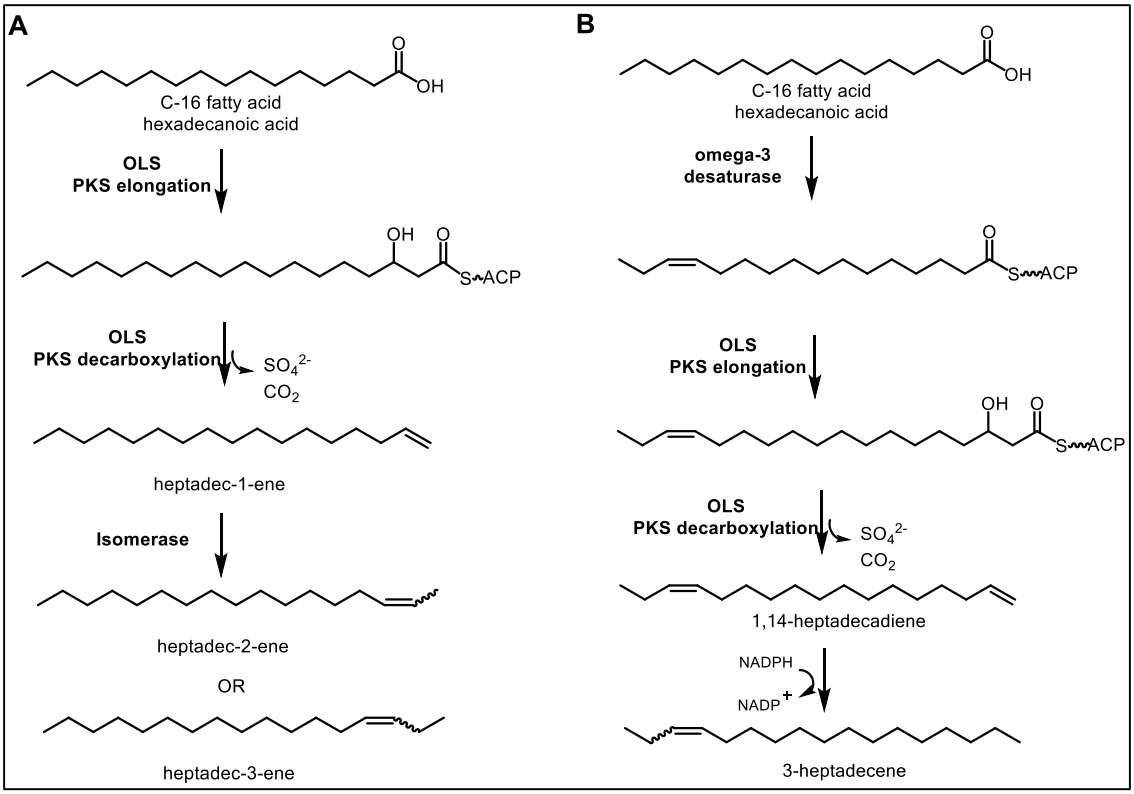


Figure 3.5. Possible pathways for alkene biosynthesis in *Moorea.* (A) A possible pathway involving OLS elongation, desaturation followed by an isomerase. (B) A possible pathway involving a desaturase followed by OLS elongation and desaturation followed by reduction.

D. incrassata PCC7326 and *Pleurocapsa* sp. PCC7320 produce not only 2- and 3-heptadecene as discussed above but also 1-heptadecene thus matching rather closely the composition of hydrocarbons observed in *Moorea* strains. This observation suggests that the OLS pathway found in *Moorea* strains is quite likely to also occur in *D. incrassata* PCC7326 and *Pleurocapsa* sp. PCC7320. Neither *D. incrassata* PCC7326 nor *Pleurocapsa* sp. PCC7320 have had their genomes sequenced (Figure 2.1). *Pleurocapsa* sp. PCC7319, a close relative to these two strains, possess the OLS pathway therefore increasing the likelihood of finding the OLS pathway in these strains.

This investigation established a number of fundamental features of hydrocarbon production among cyanobacteria. Primarily, the observation that hydrocarbon production

is universal among not only previously investigated cyanobacteria as shown in table 3.1 but also in the phylogenetically diverse samples selected in this investigation. This observation suggests that all cyanobacteria produce hydrocarbons and that there is some major selective pressure that has maintained this capacity during the long evolutionary history of cyanobacteria. Future investigations into the evolutionary history and the physiological role of hydrocarbon production may reveal this unknown selective pressure.

This investigation also revealed that strains with the OLS pathway produce more total hydrocarbons as a percentage of dry biomass. However, these results are based upon a final biomass that has accumulated through a growth period of differing lengths for each strain. This did not account for differences in growth rate between strains. Comparisons of total hydrocarbon production levels between strains with different pathways that take growth rate into account may reveal a different pattern. The differences in total hydrocarbon productions levels between the two pathways may reveal new insights into the relative selective advantage between the two pathways and may therefore have implications for understanding the physiological role of hydrocarbons in cyanobacteria as well as the selective advantage of one pathway over the other.

A more extensive investigation of the impacts of each respective pathway upon the cyanobacterium possessing that pathway could reasonably take advantage of the wide variety of genetic tools available for cyanobacteria to rationally manipulate these pathways. Previous investigations that knocked out the OLS of the FAAR/ADO pathway in their host strains have revealed that the resulting mutant strains exhibit a diminished growth rate and potentially decreased pigmentation (personal communication Dr. Brian

Pfleger, Dr. Christian Ridley, Dr. Christer Jansson, Dr. You Chen). These observations could be coupled with additional manipulations to the presence and production level of each respective pathway to reveal the physiological impact of the each hydrocarbon pathway. An extensive characterization of growth rate, photosynthetic rate, hydrocarbon and fatty acid composition production over time as well as cellular ultrastructure in strains with and without either pathway may reveal a physiological connection between pathway and phenotype. Additionally, knockout strains for each pathway could have the hydrocarbon pathway that they do not naturally produce introduced to provide insights into the physiological consequences of each pathway on their host strains.

3.6 Appendix

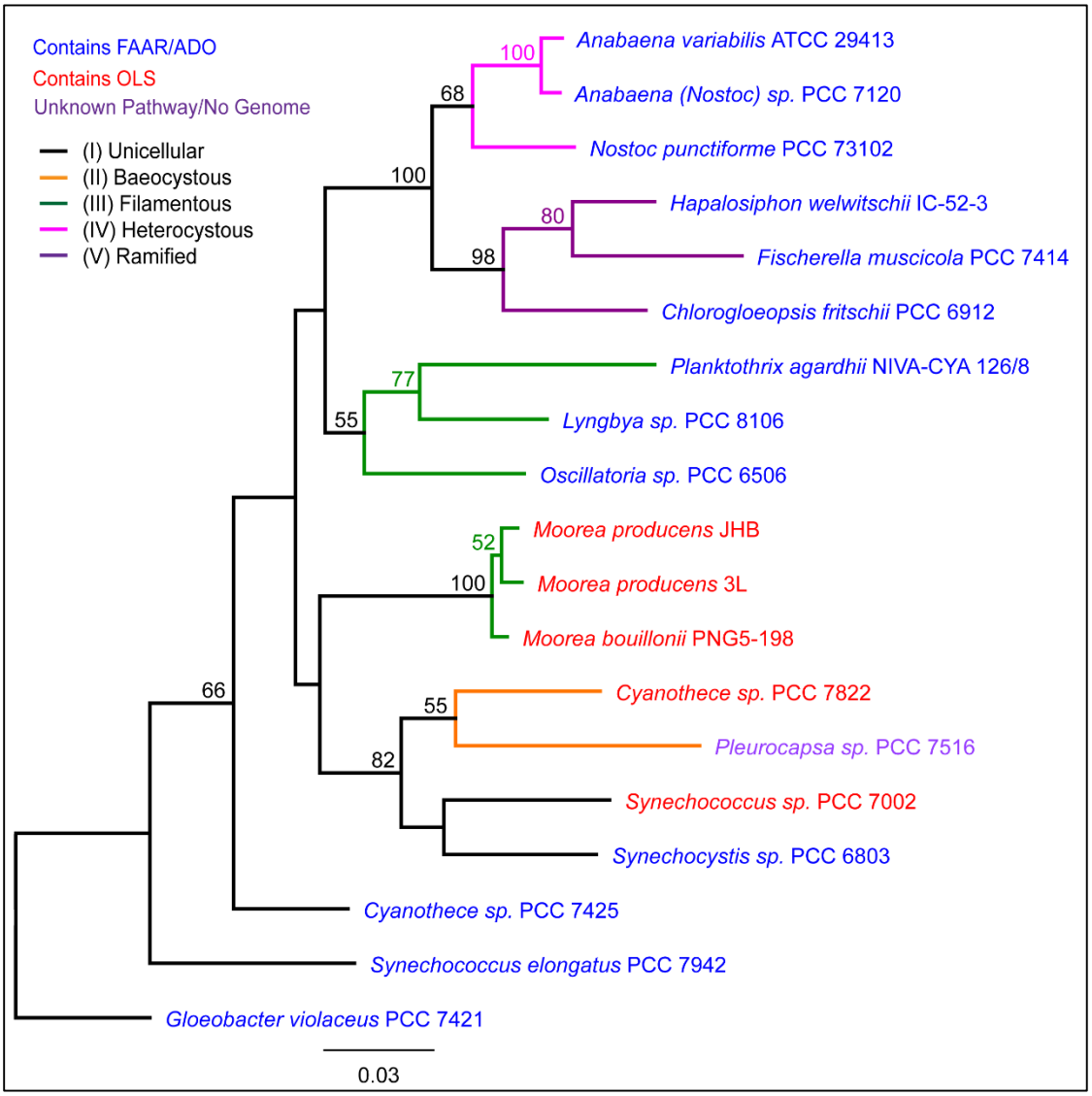


Figure 3A1. Cyanobacterial 16S rRNA phylogeny and hydrocarbon pathway distribution for the compressed tree displayed in Figure 3.1**.** The 16S rRNA phylogeny is displayed for the 20 cyanobacteria quantitatively characterized for their hydrocarbon composition. Blue strain names indicate strains possessing the FAAR/ADO pathway and red strain names indicate those with the OLS pathway. Purple strain names indicate a strain that does not have a genome sequence and therefore the pathway is unknown. Cyanobacterial subdivisions are labeled using colored branches following the key in the upper left: Subdivision I. Unicellular (Formerly Chroococcales), Subdivision II. Baeocystous (Formerly Pleurocapsales), Subdivision III. Filamentous (Formerly Oscillatoriales), Subdivision IV. Heterocystous (Formerly Nostocales), Subdivision V. Ramified or True Branching (Formerly Stigonematales).

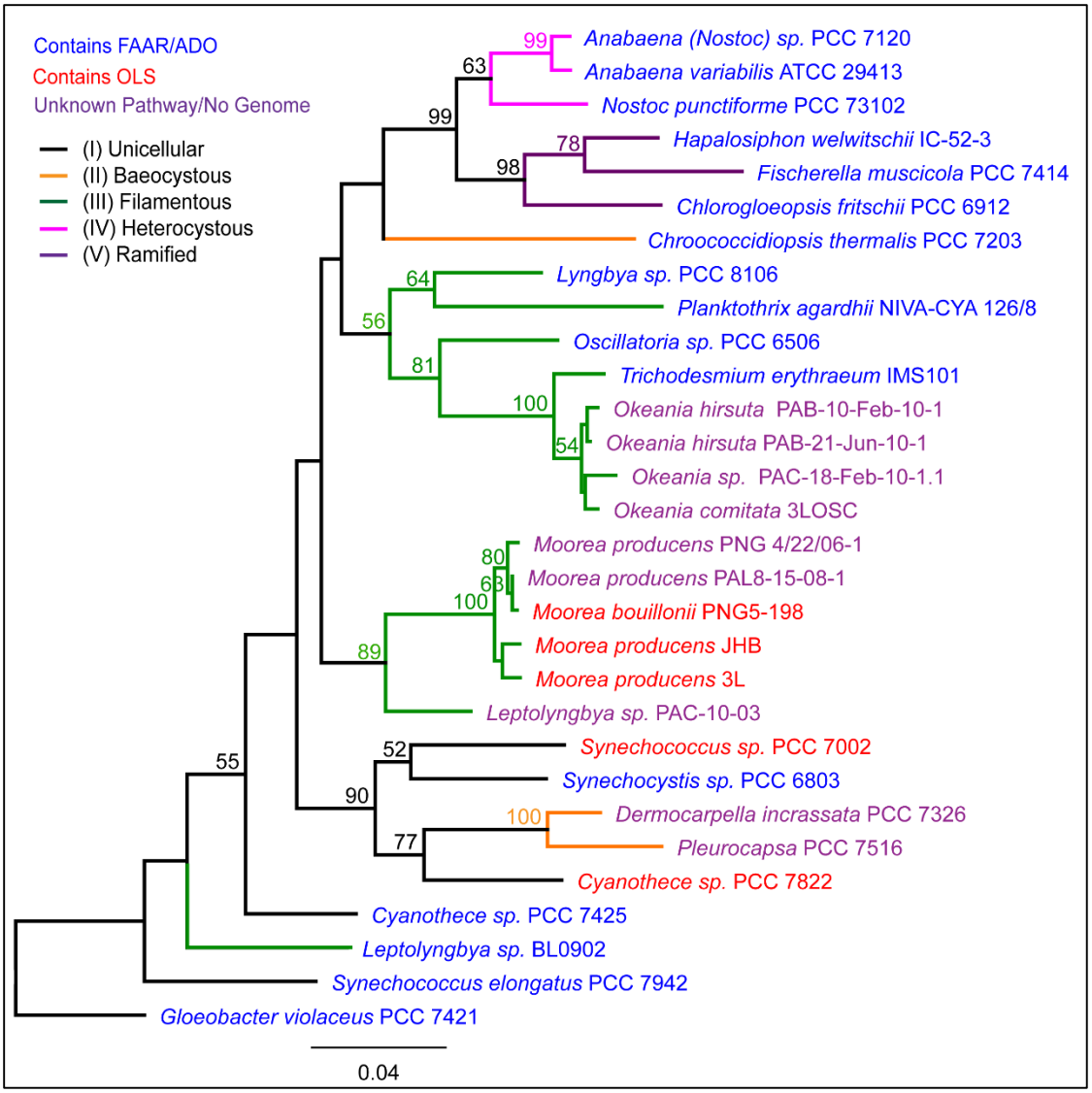


Figure 3A2. Cyanobacterial 16S rRNA phylogeny and hydrocarbon pathway distribution for compressed tree displayed in Figure 3.2. The 16S rRNA phylogeny is displayed for the 32 cyanobacteria characterized for their hydrocarbon composition as an overall percentage. Blue strain names indicate strains possessing the FAAR/ADO pathway and red strain names indicates those with the OLS pathway. Purple strain names indicate a strain that does not have a genome sequence and therefore the pathway is unknown. Cyanobacterial subdivisions are labeled using colored branches following the key in the upper left: Subdivision I. Unicellular (Formerly Chroococcales), Subdivision II. Baeocystous (Formerly Pleurocapsales), Subdivision III. Filamentous (Formerly Oscillatoriales), Subdivision IV. Heterocystous (Formerly Nostocales), Subdivision V. Ramified or True Branching (Formerly Stigonematales).

3.7 Acknowledgements

Chapter 3 includes a part of a manuscript that has been published in the Public Library of Science (PLoS) One in 2014. My co-authors were Sheila Podell, Anton Korobeynikov, Alla Lapidus, Pavel Pevzner, David H. Sherman, Eric E. Allen, Lena Gerwick, William H. Gerwick.

3.8 References

Allen ME, Neilan BA, Burnes BP, Jahnke LL, Summons RE, 2010. Lipid biomarkers in Hamelin Pool microbial mats and stromatolites. Org Geochem 41. 1207-1218.

Balaban AT, 1983. Topological indices based on topological distances in molecular graphs. Pure Appl Chem 55. 199-206.

Bankevich AS, Nurk D, Antipov AA, Gurevich M, Dvorkin AS, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA, 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19. 455-477.

Beller HR, Goh EB, Keasling JD, 2010. Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. Appl Environ Microbiol 76. 1212-1223.

Bennett S, 2004. Solexa Ltd. Pharmacogenomics 5. 433-438.

Benson DA, Cavanaugh M, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL, 2013. GenBank. Nucleic Acids Res 41. D36-42. doi: 10.1093/nar/gks1195

Blomquist GJ, Bagneres AG (eds) 2010. Insect hydrocarbons: biology, biochemistry, and chemical ecology. New York: Cambridge University Press.

Blumer M, Guillard RRL, Chase R, 1971. Hydrocarbons of marine phytoplankton. Mar Biol 8. 183-189.

Carpenter EJ, Harvey HR, Fry B, Capone DG, 1997. Biogeochemical tracers of the marine cyanobacterium *Trichodesmium.* Deep Sea Res Part I. 44. 27-38.

Castresana J, 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17. 540-552.

Castenholz RW, 2001. Phylum BX. Cyanobacteria. Oxygenic Photosynthetic Bacteria. In: Garrity G, Boone DR, Castenholz RW editors. Bergey's Manual of Systematic

Bacteriology. Volume 1: The Archaea and the Deeply Branching and Phototropic Bacteria. Second Edition. New York: Springer-Verlag. 473-599.

Chevenet F, Brun C, Banuls AL, Jacq B, Chisten R, 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics. 10. 439. doi:10.1186/1471-2105-7-439.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH, 2004. Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. J Nat Prod 67: 1356-1367.

Creton B, Dartiguelongue C, de Bruin T, Toulhoat H, 2010. Prediction of the cetane number of diesel compounds using the quantitative structure property relationship. Energy Fuels 24. 5396-5403.

Dembitsky VM, Shkrobb I, Dor I, 1999. Separation and identification of hydrocarbons and other volatile compounds from cultured blue-green alga *Nostoc* sp. by gas chromatography–mass spectrometry using serially coupled capillary columns with consecutive nonpolar and semipolar stationary phases. J Chromatogr A 682. 221-229.

Dembitsky VM, Dor I, Shkrobb I, 2000. Variability of lipid constituents of the soil cyanobacterium *Microcoleus vaginatus* from the dead sea basin and negev desert. Biochemistry (Moscow) 65. 1403-1408.

Dembitsky VM, Srebnik M, 2002. Variability of hydrocarbon and fatty acid components in cultures of the filamentous cyanobacterium *Scytonema* sp. isolated from microbial community "black cover" of limestone walls in Jerusalem. Biochemistry (Moscow) 67. 1276-1282.

Dennis MW, Kolattukudy PE, 1991. Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus braunii*. Arch Biochem Biophys 287. 268-275.

Dereeper A, Audic S, Claverie JM, Blanc G, 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol 10.8. doi:10.1186/1471-2148-10-8

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O, 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36. W465-469. doi. 10.1093/nar/gkn180

Doan TTP, Carlsson AS, Hamberg M, Bülow L, Stymne S, Olsson P, 2009. Functional expression of five *Arabidopsis* fatty acyl-CoA reductase genes in *Escherichia coli.* J Plant Physiol 166. 787-796.

Donia MS, Fricke WF, Ravel J, Schmidt EW, 2011. Variation in tropical reef symbiont metagenomes defined by secondary metabolism. PLoS ONE 6:e17897. doi:10.1371/journal.pone.0017897

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32.1792-1797.

Engene N, Rottacker EC, Kaštovský J, Byrum T, Choi H, Ellisman MH, Komárek J, Gerwick WH, 2012. *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Ins J Syst Evol Microbiol 62. 1171-1178.

Engene N, Paul VJ, Byrum T, Gerwick WH, Thor A, Ellisman MH, 2013. Five chemically rich species of tropical marine cyanobacteria of the genus Okeania gen. nov. (Oscillatoriales, Cyanoprokaryota). J Phycol doi: 10.1111/jpy.12115

Eser BE, Das D, Han J, Jones PR, Marsh EN, 2011. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. Biochemistry 50. 10743-10750.

Fehler WG, Light RJ, 1970. Biosynthesis of hydrocarbons in *Anabaena variabilis*. Incorporation of (methyl- ^{14}C - and (methyl-$^{2}H_3$) methionine into 7- and 8-methylheptadecanes. Biochemistry 9. 418-422.

Fehler WG, Light RJ, 1972. Biosynthesis of methylheptadecanes in *Anabaena variabilis*. In Vitro incorporation of S- (methyl-^{14}C) adenosylmethionine. Biochemistry 11. 2411-2416.

Gehret McCarthy J, Eisman EB, Kulkarni S, Gerwick L, Gerwick WH, Wipf P, Sherman DH, Smith JL, 2012. Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. ACS Chem Biol 7. 1994-2003. doi: 10.1021/cb300385m

Gelpi E**,** Schneider H, Mann J, Oró J, 1970. Hydrocarbons of geochemical significance in microscopic algae. Phytochemistry. 9. 603-612.

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu W, Gerwick L, Dorrestein P, Pevzner P, Lasken R, Gerwick W 2011. Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. PLoS One. 6(4): e18565

Goodloe R, Light RJ, 1982. Biosynthesis of hydrocarbon in *Anabaena variabilis* in vivo incorporation of (18-14C) stearate. Biochem Biophys Acta. 711. 261-265.

Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Håkansson K, Smith JL, Sherman DH, 2009. Polyketide decarboxylative chain termination preceded by O-sulfonation in curacin A biosynthesis. J Am Chem Soc. 131. 16033–16035.

Guindon S, Gascuel O, 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52. 696-704.

Haft DH, Selengut JD, Richter RA, Harkins D, Basu MK, Beck E, 2013. TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 41. D387-95. doi: 10.1093/nar/gks1234.

Han J, McCarthy ED, Van Hoeven W, Calvin M, Bradley WH, 1968. Organic geochemical studies, II. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment. PNAS 59. 29-33.

Han J, Chan HWS, Calvin M, 1969. Biosynthesis of alkanes in *Nostoc muscorum*. J Am Chem Soc 91. 5156-5159.

Juttner F, 1991. Taxonomic characterization of *Limnothrix* and *Planktothrix* using secondary metabolites (hydrocarbons). Algological Studies/Archiv für Hydrobiologie 64. 261-266.

Kelley DR, Schatz MC, Salzberg SL, 2010. Quake: quality-aware detection and correction of sequencing errors. Genome Biol 11: R116. doi:10.1186/gb-2010-11-11-r116

Köester J, Volkman JK, Rullköetter J, Scholz-Böettcher BM, Rethmeier J, 1999. Mono-, di- and trimethyl-branched alkanes in cultures of the filamentous cyanobacterium *Calothrix scopulorum*. Org Geochem 30. 1367-1379

Li N, Chang W, Warui DM, Booker SJ, Krebs C, Bollinger JM Jr., 2012. Evidence for only oxygenative cleavage of aldehydes to alk(a/e)nes and formate by cyanobacterial aldehyde decarbonylases. Biochemistry 51. 7908-7916.
et
Liu A, Zhu T, Lu X, Song L, 2013. Hydrocarbon profiles and phylogenetic analyses of diversified cyanobacterial species. Appl Energy 111. 383-393.

Lee RF, Loeblich AR 1971. Distribution of 21:6 hydrocarbon and its relationship to 22:6 fatty acid in algae. Phytochemistry 10. 593-598.

Markowitz VM, Chen MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Verghese N, Mavromatis K, PAti A, Ivanova N, Krypides NC, 2014. IMG 4 version of the integrated microbial genomes comparative analysis system, Nucl. Acids Res. 42, D560-D567.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM, 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437. 376–380.

Mclnnes AG, Walter JA, Wright JLC, 1980. Biosynthesis of hydrocarbons by algae: Decarboxylation of stearic acid to N-heptadecane in *Anacystis nidulans* determined by 13C - and 2H -labeling and 13C nuclear magnetic resonance. Lipids 15. 609-615

Mendez-Perez D, Begemann MB, Pfleger BF, 2011. Modular synthase-encoding gene involved in α-olefin biosynthesis in *Synechococcus sp*. strain PCC7002 App Envir Micro 77. 4264-4267.

Metzger P, Largeau C, 2005. *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids. Appl Microbiol Biotechnol 66. 486-496.

Murray J, Thomson A, 1977. Hydrocarbon production in *Anacystis montana* and *Botryococcus braunii*. Phytochemistry 16. 465-468.

Paoletti C, Pushparaj B, Florenzano G, Capella P, Lerker G, 1976. Unsaponifiable matter of green and blue-green algal lipids as a factor of biochemical differentiation of their biomasses: I. Total unsapononifiable and hydrocarbon fraction. Lipids 11. 258-265.

Peralta-Yaha PP, Zhang F, del Cardayre SB, Keasling JD, 2012. Microbial engineering for the production of advanced biofuels. Nature 488. 320-328.

Perry GJ, Gillan FT, Johns RB, 1978. Lipid Composition of a prochlorophyte. J Phycol 14. 369-371.

Pohnert G, Boland W, 2002. The oxylipin chemistry of attraction and defense in brown algae and diatoms. Nat Prod Rep 19. 108-122.

Qiu Y, Tittige C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R, 2012. An insect-specific P450 oxidative decarbonylase forcuticular hydrocarbon biosynthesis. PNAS 109:14858–14863.

Rezanka T, Zahradík J, Podojil M, 1982. Hydrocarbons in green and blue-green algae. Folia Microbiol (Praha) 27. 450-454.

Ridley CP, Reppas NP, 2010. Biosynthesis of 1-alkenes in engineered microorganisms. US Patent Appl 12/821,107.

Rude MA, Baron TS, Brubaker S, Alibhai M, del Cardayre SB, Schirmer A, 2010. Terminal olefin (1-alkene) biosynthesis by a novel P450 fatty acid decarboxylase from *Jeotgalicoccus* sp. Appl Environ Microbiol 77. 1718-1727.

Sakata S, Hayes JM, McTaggart AR, Evans RA, Leckrone KJ, Togasaki RK, 1997. Carbon isotopic fractionation associated with lipid biosynthesis by a cyanobacterium: relevance for interpretation of biomarker records. Geochim Cosmochim Acta 61.5379-5389.

Sata N, Wada H, 2010. Lipid biosynthesis and its regulation in cyanobacteria. In: Wada H, Murata N, editors. Lipids in photosynthesis. Advances in photosynthesis and respiration. Dordrect: Springer. 30. pp 157-177.

Shih PM, Wua D, Latifid A, Axena SD, Fewere DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport KW, Han CS, Rubin EM, Eisen JA, Woyke T, Gugger M, Kerfeld CA, 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. PNAS 110. 1053-1058.

Schirmer A, Rude MA, Li X, Popova E, del Cardayre S, 2010. Microbial biosynthesis of alkanes. Science 329. 559-562.

Schirrmeister BE, Antonelli A, Bagheri HC, 2011. The origin of multicellularity in cyanobacteria. BMC Evolutionary Biology 11.45-66. doi:10.1186/1471-2148-11-45

Sharma N, Tiwari S, Tripathi K, Rai A, 2010. Sustainability and cyanobacteria (bluegreen algae): facts and challenges. J Appl Phycol 23.1-23.

Tan XM, Yao L, Gao QQ, Wang WH, Qi FX, Lu X, 2011 Photosynthesis drive conversion of carbon dioxide to fatty alcohols and hydrocarbons in cyanobacteria. Metab Eng 13. 169-176.

Tsuchiya Y, Matsumoto A, 1999. Characterization of *Oscillatoria F. granulata*. producing 2-methyl-isoborneol and geosmin. Water Sci Technol 40. 245–250.

Winters K, Parker PL, Van Baalen C, 1969. Hydrocarbons of blue-green algae: geochemical significance. Science 163. 467-468.

Walsh K, Jones GJ, Hugh R, 1998. Dunstan Effect of high irradiance and iron on volatile odour compounds in the cyanobacterium *Microcystis aeruginosa*. Phytochemistry 49. 1227–1239.

Williamson RT, Boulanger A, Vulpanovici A, Roberts MA, Gerwick WH. 2002. Structure and absolute stereochemistry of phormidolide, a new toxic metabolite from the marine cyanobacterium *Phormidium sp*. J Org Chem 67. 7927-7936.

Vincent M, Guglielmetti G, Cassani G, Tonini C, 1987. Determination of double-bond position in diunsaturated compounds by mass spectrometry of dimethyl disulfide derivatives. Anal Chem 59. 694-699.

Zarycki J, Axen SD, Kinney JN, Kerfeld CA, 2012. Cyanobacterial-based approaches to improving photosynthesis in plants. J Exp Bot 64. 787-798. doi:10.1093/jxb/ers294

Chapter 4: Identification of a Methyltransferase Involved in Branched Alkane Biosynthesis in *Anabaena* sp. PCC7120

4.1 Abstract

Cyanobacterial pathways involved in hydrocarbon biosynthesis from fatty acids have attracted substantial interest as a platform for biofuel and sustainable chemical production. These unique pathways are known to produce saturated alkanes and alkenes, but have also been observed to produce branched alkanes. While branched alkane production in cyanobacteria is found broadly distributed among all major cyanobacterial subdivisions, it is mostly observed in heterocystous cyanobacteria. Investigations of the biosynthesis of branched hydrocarbons in cyanobacteria have been limited and a methyltransferase responsible for production of branched hydrocarbons in cyanobacteria has not yet been identified. This investigation seeks to identify the methyltransferase responsible for branched fatty acid biosynthesis in *Anabaena sp*. PCC7120. Three target methyltransferases (all2121, all3016 and alr3038) were bioinformatically identified as targets for investigation. Heterologous expression in *E. coli* was used to evaluate the specificity of the three target methyltransferases. No branched fatty acids were detected, suggesting that none of the three methyltransferases exhibits this biochemical activity in *E. coli.* Purified proteins derived from all3016 and alr3038 were used for *in vitro* analysis; however branched fatty acids were not produced, indicating that these methyltransferases do not exhibit methyltransferase activity under the conditions tested. A knockout of all3016 was generated by homologous recombination in *Anabaena sp*. PCC7120. However, branched hydrocarbon production was not abolished in the resulting

mutant strain, suggesting that all3016 is not involved in branched hydrocarbon biosynthesis, or that additional methyltransferases contribute or can compensate to maintain branched hydrocarbon production.

4.2 Introduction

Cyanobacteria are a group of photosynthetic bacteria that play major roles in all major ecosystems and biogeochemical cycles. Cyanobacteria have also been shown to exhibit a remarkable morphological and metabolic diversity. This diverse metabolic capacity, rapid auxotrophic growth and a wide variety of established genetic tools make cyanobacteria an attractive source for biofuels, natural products and sustainable chemicals. All cyanobacteria also seem to exhibit the rare capacity to make hydrocarbons that are drop-in ready materials for diesel or jet engines. The relatively rare and societally valuable hydrocarbons found in cyanobacteria are biosynthetically derived from fatty acids (a relatively abundant biomolecule) making this unique pathway attractive as a focus of investigation and development for biofuel applications (Schirmer et al. 2005, Coates et al. 2014). Some cyanobacteria have been found to possess a unique capacity to biosynthesize branched hydrocarbons comprised of mono-methylated heptadecanes (Han et al. 1968, Coates et al. 2014). Production of branched hydrocarbons seems to be most commonly observed in heterocystous (Type IV) cyanobacteria, however, strains from all other major cyanobacterial subdivisions have been reported to produce branched hydrocarbons (Table 4A1) (Coates et al. 2014). The biosynthesis of cyanobacterial branched hydrocarbons has been investigated extensively, however the methyltransferase responsible for incorporation of a methyl group on the resulting hydrocarbon has yet to be identified.

The biosynthesis of cyanobacterial branched hydrocarbons was investigated in early studies that established a few of the key features of the biosynthetic pathway, and these can be used to help guide efforts to identify a methyltransferase involved in the pathway (Han et al. 1968, Han et al. 1969, Fehler and Light 1970, Fehler and Light 1972, Goodloe and Light 1981). In 1968 Han et al. explored the hydrocarbon constituents of four cyanobacteria including *Nostoc muscorum, Anacystisnidulans, Phormidium luridurn,* and *Chlorogloea fritschii*. They established that the branched hydrocarbons produced by *Nostoc muscorum* was a mixture comprised of a 1:1 ratio of 7- and 8-methylheptadecane using NMR analysis of the natural hydrocarbon composition compared to synthetic standards. Later studies established that 7- and 8-methylheptadecane in *Nostoc muscorum* were derived from (*Z*)-11-octadecenoic acid by measuring the amount of palmitic-ω-^{14}C acid, stearic-ω-^{14}C acid, [1-^{14}C] acetate, and tritiated (*Z*) 11-octadecenoic acid that were separately incorporated into the resulting hydrocarbons (Han et al. 1969, Fehler and Light 1970, Goodloe and Light 1981). These studies also established that the methyl group in 7- and 8- methylheptadecane was derived from S-adenosylmethionine (SAM) from results showing separate incorporation of D-,L- and L-methionine-methyl-^{14}C into 7- and 8-methylheptadecane (Han et al. 1968, Fehler and Light 1970, Goodloe and Light 1981). Mass spectrometry of the methylheptadecane mixture formed in cultures grown in the presence of DL-[*methyl*-$^{2}H_3$] methionine showed that the CD_3 group had been incorporated intact, ruling out cyclopropane intermediates in the biosynthetic pathway. Subsequently, Fehler and Light (1972) established that the apparent K_m for SAM was 1.1 X 10^{-4} M and the pH optimum was 7.0, and a partial dependence on NADPH could be demonstrated in a short-term

dialysis experiment. The authors also found that oxygen inhibited the methylation reaction.

The above insights into branched hydrocarbon biosynthesis, taken in conjunction with the recent identification of the fatty acyl ACP reductase/ aldehyde deformylating oxygenase (FAAR/ADO) pathway, can be used to propose a biosynthetic pathway for 7 and 8-methylheptadecane biosynthesis (Figure 4.1) (Coates et al 2014, Schirmer et al 2005). Considering the observation by Han et al. (1969) of specific methylation of the double bond of (*Z*) 11-octadecenoic acid by SAM, the biosynthetic pathway likely involves the formation of (*Z*) 11-octadecenoic acid from a Δ9 desaturase acting upon hexacanoic acid followed by a two carbon elongation, or a Δ11 desaturase acting upon octadecanoic acid. FAME analysis of *Anabaena* sp. PCC7120 indicates that both 9-hexadecenoic acid as well as 11-octadecenoic acid are produced and could potentially be available as a substrate for methylation. Following desaturation, the double bond of (*Z*) 11-octadecenoic acid would be methylated via SAM by a methyltransferase. The reduction of the carbocation intermediate that would be formed after methylation has not been explored in this system, but would be necessary to produce the branched fatty acid, and could explain the partial dependence on NADPH observed by Fehler and Light (1972). The resulting fatty acid would be converted to a fatty acyl ACP given the preference of the FAAR/ADO pathway for this substrate (Schirmer et al. 2005). The branched fatty acyl ACP would then proceed through the FAAR/ADO pathway to produce a branched fatty aldehyde and finally 7- or 8-methylheptadecane (Figure 4.1).

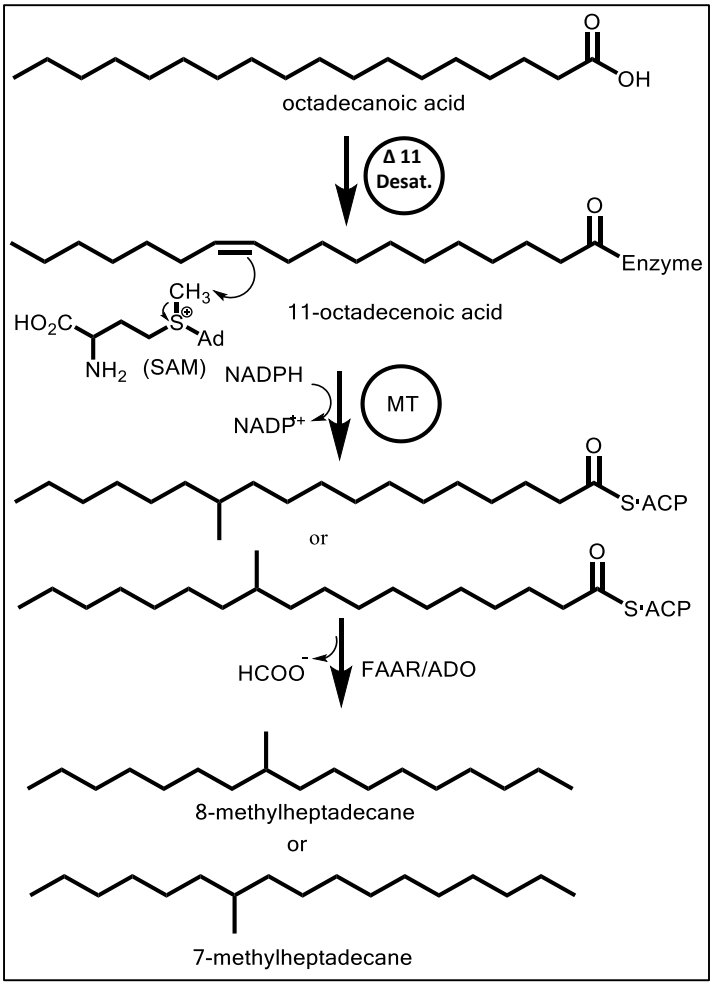


Figure 4.1. Hypothesized biosynthetic pathway for 7-methylheptadecane biosynthesis in *Anabaena* sp. PCC7120.

Despite the work that has been done on their biosynthesis, the physiological or ecological function of branched hydrocarbons in cyanobacteria is still unknown. Fehler and Light (1970) found that the relative amount of the methylheptadecane mixture varies from 22% in young cultures to 84% of total hydrocarbons in cultures grown for 9 days. Additionally they found that the incorporation efficiency of [*methyl*- ^{14}C] methionine into hydrocarbons increased with the age of the culture, matching the increase in production of branched hydrocarbons (Fehler and Light 1970). This suggests that production of branched hydrocarbons may increase over time, however no further studies on this topic

have appeared. Considering other organisms, the branched hydrocarbons 7-methylheptadecane and 7, 11-dimethylheptadecane have been reported as sex pheromones in lepidoptran insects (Bloomquist et al. 2010, Qiu et al. 2012). The physiological role of all hydrocarbons in cyanobacteria is unknown as well and it may be that branched hydrocarbons play the same if not similar role as straight chain or unsaturated hydrocarbons. Until a methyltransferase can be identified and subsequently used to manipulate the presence or abundance of branched hydrocarbons the physiological function of these unique natural products cannot be fully deduced.

The identification of the methyltransferase involved in the biosynthesis of branched hydrocarbons would allow for the investigation of a variety of unexplored features of this pathway, including the substrate specificity of the methyltransferase as well as the enzyme kinetics. Additionally, once a methyltransferase involved in branched hydrocarbon biosynthesis is identified, it can be applied to efforts to produce sustainable biofuels. Highly branched short chain molecules exhibit high octane ratings and can be used to improve fuel quality for gasoline engines (Balaban 1983). This investigation seeks to identify this potentially valuable methyltransferase responsible for branched hydrocarbon production in A*nabaena* sp. PCC7120. Potential methyltransferases were targeted through bioinformatic analysis and were tested through heterologous expression and mutagenesis.

4.3 Materials and Methods

To identify the methyltransferase responsible for branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120 a list of 20 known cyanobacterial

methyltransferases were identified and used to perform a BLAST search against the *Anabaena* sp. PCC7120 genome (Table 4A2). The query methyltransferases were compiled from diverse cyanobacteria and include some methyltransferases that have been biochemically validated as well as others that due to their genetic context in relation to a known biochemical pathway have biochemical functions that are essentially certain. The list of known cyanobacterial methyltransferases was composed of eight C-methyltransferases from polyketide synthases, tocopherol methylation as well as C-methyltransferases that methylate via a radical mechanism, three *N*-methyltransferases, and six *O*-methyltransferases (Table 4A2). The list of query methyltransferases also included one DNA methyltransferase that methylates a carbon of cytosine (Table 4A2). A total of 23 methyltransferases were identified from this investigation of the *Anabaena* sp. PCC7120 genome. These 23 methyltransferases were used as a starting point for investigation into potential methyltransferases involved in branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120.

We proposed to perform heterologous expression and mutagenesis on a subset of these methyltransferases. Bioinformatic comparisons were performed to narrow down this list of methyltransferases to an accessible number of targets for further investigation. This involved a phylogenetic comparison of the 20 known cyanobacterial methyltransferases with the 23 methyltransferases identified from *Anabaena* sp. PCC7120. All of the 43 methyltransferases as well as one human RNA methyltransferase (outgroup) were aligned using MUSCLE (Edgar 2004). A phylogentic tree was constructed using Maximum Likelihood (500 bootstrap replicates).

4.3.1 Bioinformatic Identification of Target Methyltransferases

An alternative approach involving phylogenetic profiling was pursued because the phylogenetic analysis did not effectively narrow the list of target methyltransferases. Phylogenetic profiling is a bioinformatic technique in which gene orthologs and phenotypes are used across multiple species to infer a functional connection with a biosynthetic pathway or gene function (Ternes et al. 2006). Phylogenetic profiling was completed using the online genomic analysis software tool CyanoBike and JGI's IMG phylogenetic profiling function (Elhai et al. 2009, Markowitz et al. 2014). The phylogenetic profiling analysis relied upon the observation that two closely related cyanobacteria (*Anabaena* sp. PCC7120 and *Anabaena variabilis* ATCC 29413) both produced the branched hydrocarbon 7-methylheptadecane while another close relative (*Nostoc punctiforme* ATCC 29133) did not produce branched hydrocarbons. Therefore, it was hypothesized that *Anabaena* sp. PCC7120 and *Anabaena variabilis* ATCC 29413 both possess the methyltransferase responsible for branched hydrocarbon biosynthesis while *Nostoc punctiforme* ATCC 29133 does not. This approach was pursued with the understanding that it would not account for the possibility that *Nostoc punctiforme* ATCC 29133 possesses the methyltransferase responsible for branched hydrocarbon biosynthesis yet does not actively express it under the conditions that this strain was investigated. Orthologous pairwise relationships were computed as bidirectional best BLAST hits between genomes. CyanoBike restricted identified orthologs to hits better than a cutoff E-value of $1E^{-6}$ while JGI IMG's phylogenetic profiler used a cutoff of $1E^{-5}$ and minimum percent identity of 10%.

Additionally, a methyltransferase that performs a similar reaction to the proposed methylation in this investigation was used to bioinformatically identify potential target

genes. A recent study characterized a methyltransferase (*umaA*) responsible for methylation of (Z) 11-octadecenoic acid to produce tuberculosteric acid in *Mycobacterium tuberculosis* (Meena et al. 2013). A BLAST search was performed against the *Anabaena* sp. PCC7120 using *umaA* (E-value cutoff = $1E^{-5}$).

4.3.2 *In vivo* and *in vitro* analysis through heterologous expression in *E. coli*

Each of the three target methyltransferases were cloned into expression vectors and heterologously expressed in *E. coli*. *Anabaena* sp. PCC7120 genomic DNA was extracted using the Promega Wizard Genomic DNA purification kit (A1120) and was used as a template for amplification of all2121, all3016, alr3038. All three methyltransferases were PCR amplified using primers designed for Life Technologies GeneArt® Seamless Cloning and Assembly technique (A14606) (Table 4A3) and cloned into a pET28b vector. Assembly reactions were transformed into *E. coli* DH5α cells. Successful assemblies were assessed using PCR screening, digest analyzed and verified by sequencing (Figure 4A2-4A4).

In vivo testing of all three methyltransferases was completed to screen for fatty acid methylation activity. Sequence verified plasmids were transformed into *E. coli* BL21 cells and upon transformation a single colony was inoculated into 5 ml LB broth with 20 µg/ml Streptomycin/Spectinomycin (Sp/Sm) and incubated overnight at 37°C. One hundred µl of each overnight culture was inoculated into 50 ml LB at 37°C with 20 µg/ml Sp/Sm and the OD600 was monitored. Cultures were induced for expression with 0.1 mM IPTG at approximately OD600 = 0.6 and incubated at 18°C for 12 hrs. Optimal expression conditions for all3016 pET28b were determined experimentally through testing multiple incubation temperatures (37°C and 18°C), IPTG concentrations (0.1 mM

and 0.5 mM), and incubation lengths (2 hrs, 4 hrs, 10 hrs, 12 hrs and 15 hrs). Exogenous fatty acids ((*Z*)-11-octadecenoic acid (10 mg, Sigma V0384), and (9*Z*, 12*Z*) 9-, 12-octadecadienoic acid (10 mg, Sigma L1376)) were added to cultures of *E. coli* expressing all3016 and all2121 in pET28b to test for methyltransferase activity.

Expression was verified through protein purification and SDS-PAGE analysis as well as Western blot analysis. Each culture was split into two 25 ml portions (one for GC-MS analysis and one for protein purification). Each portion was centrifuged at 4000g (4°C), the supernatant was discarded and the remaining pellet was frozen at -80°C. Each pellet for protein purification was thawed at 0°C and resuspended with one ml of native lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM Imidazole, pH8.0). Lysozyme (1 µg/ml) and a protease inhibitor cocktail (Roche: 11873580001) were added and the solution was incubated on ice for 30 min. The solution was sonicated with a probe sonicator with 10 bursts for 10 sec on and 10 sec off at 200W. Lysates were centrifuged for 10 min at 15,000 g and transferred to a new tube. One hundred µl of a 50% slurry of Ni-NTA resin was added to bind His tagged proteins and each tube was incubated for 30 min at 4°C. The pelleted resin was washed twice with 500 µl of wash buffer (15 mM glycerol, 50 µM Tween 20, 1.3 mM β-mercaptoethanol, 40 mM immidazole) and twice with 500 µl of elution buffer (15 mM glycerol, 50µM Tween 20, 1.3 mM β-mercaptoethanol, 250 mM immidazole). Purified proteins were stored at -80°C after purification. Proteins were separated using SDS/PAGE and gels visualized under UV (NuSep). Western blots using Thermo Pierce mouse 6 × His primary anti-His tag antibody and Thermo Pierce goat anti-mouse horseradish peroxidase-conjugate secondary antibody were used to detect recombinant protein.

In vitro analysis was completed using the whole cell lysates for all3016 and alr3038 as well as the purified proteins in two different buffer systems (buffer 1 = elution buffer (15 mM glycerol, 50µM Tween 20, 1.3 mM β-mercaptoethanol, 250 mM immidazole), buffer 2 (25 mM Tris, 0.5 mM EDTA, 1 mM DTT (pH= 7.26). Each reaction included 30 µl of enzyme or whole cell lysate, 22.5 µl (80 µM final) of S-adenosyl-L-methionine (stock solution = 32 mM in 0.005M H_2SO_4 and 10% EtOH), 0.3 µl of NADPH (5 mg/ml), 0.5 µl (10 µg) of (*Z*)-11-octadecenoic acid (stock solution = 20 mg/ml in isopropanol), and 46.7 µl of buffer. *In vitro* activity was tested using the exogenous substrate (*Z*)-11-octadecenoic acid. Each condition was tested using two incubation times (4 hrs and 12 hrs).

Fatty acid methyl esterification (FAME) analysis for each sample was completed on half of the culture biomass (25 ml) that was centrifuged at 4000g (4°C), the supernatant was removed and the remaining pellet was flash-frozen with liquid nitrogen. Samples were then lyophilized overnight prior to extraction. Each pellet was ground with a mortar and pestle followed by the addition of three mL of four percent H_2SO_4 (in MeOH). Samples were then stirred and incubated at 110°C for 1 h. Four mL of H_2O and three mL of hexanes were then added to the sample, vortexed for 20 seconds, separated and extracted with hexane. Samples were dried under N_2 and resuspended to a concentration of 100 mg/mL in hexanes. One µL of each sample was analyzed by gas chromatography mass spectrometry (GC-MS) using a Thermo Trace GC-DSQ instrument equipped with an Agilent DB5-ms column (30 m, ID: 0.25, Film: 0.25 mm). Helium (constant flow one mL/min) was used as the carrier gas. The inlet temperature was 240°C and the following temperature program was applied: 40°C for one min with an increase

of 4.5°C/min to 250°C for 10 min. Data were acquired and processed with the Thermo Xcaliber software. FAME's were determined using a combination of mass fragmentation patterns, retention time and comparison to authentic standards when available (FAME mixture (NuCheck Prep GLC-68A)).

4.3.3 *Anabaena* sp. PCC7120 Extraction and Structural Analysis

Dried biomass was ground using a mortar and pestle and weighed. Biomass was extracted using five mL of 100% hexanes followed by 20 sec of sonication. The extract was filtered using a Whatman GF/F filter and the residual biomass was re-extracted two additional times using the same method followed by a 10 mL wash with hexanes. Extracts were dried under N_2 gas and analyzed via GC-MS analysis as described above. Hydrocarbons were determined using a combination of mass fragmentation patterns, retention time and comparison to authentic standards when available (heptadecane (Fluka-51578), 1-heptadecene (TCIS0347), 7-methylheptadecane (kindly provided by Dieter Enders and Wolfgang Bettray, RWTH, Aachen University)), or published mass spectra and the NIST mass spectral library for Xcaliber 2005) when not available.

4.3.4 *Anabaena* sp. PCC7120 Mutagenesis

In parallel with heterologous expression in *E. coli,* all three methyltransferases (all2121, all3016, alr3038) were targeted for mutagenesis by homologous recombination in *Anabaena* sp. PCC7120. Knockout vectors were constructed as follows. One kilobase regions flanking each side of the methyltransferase were PCR amplified. For each methyltransferase, the PCR products were assembled with a SpSm resistance cassette and a module carrying an *E. coli* origin of replication (pMB1), a bom site (for conjugal transfer) and the *sacB* gene (for positive selection of double recombinant) (Figure 4A1).

Each assembly was completed as a four-piece assembly using the Life Technologies GeneArt® Seamless Cloning and Assembly technique (A14606) using at least 20 nt of overlapping region for each fragment.

Assembly reactions were transformed into *E. coli* DH10B under 20 µg/ml Sp/Sm selection. Successful assemblies were screened by PCR, and restriction digests and verified by sequencing. Assembled vectors were subsequently introduced by electroporation into *E. coli* strain AM1359 that contained the conjugation plasmid pRL443 and the helper plasmid pRL623, which carried a *mob* (ColK) gene and three methylase genes to protect cargo plasmids from *Anabaena* restriction enzymes. Three ml of LB with Sp/Sm/Ap/Cm selection were inoculated with the transformed *E. coli* strain AM1359 and incubated at 37°C overnight. To limit the chance of multiple exconjugants in a single colony, 50 ml of an *Anabaena* sp. PCC7120 WT culture was fragmented by sonication for 100 sec (10x 5 sec pulse on, 5 sec pulse off) at 20% amplitude. Cells were subsequently pelleted by centrifugation (3 min at 400 g) and then transferred to fresh BG-11(+N) for recovery for 24 hrs. Cells from transformed *E. coli* strain AM1359 were pelleted via centrifugation at 4000 g for 3 min and washed twice with antibiotic free LB. Cells were resuspended in 1 ml of LB. Fragmented *Anabaena* sp. PCC7120 cells were centrifuged at 4000 g for 3 min and washed once with BG-11(+N). Cells were resuspended in 10 ml of BG-11(+N). *Anabaena* sp. PCC7120 WT and *E. coli* cultures were mixed together and pelleted via centrifugation at 4000 g for 5 min. Cells were resuspended in 2 ml of BG-11(+N). The mixture was used to plate on agar plates containing BG-11(+N) + 5%LB without antibiotics using glass beads. Two volumes of the mixture were plated (150 µl and 30 µl). Plates were placed at 30°C under low light

(10-20 μE m^{-2} s^{-1}) for 2 days. Plates were underlaid with 2 μg/ml Sp/Sm and transferred to higher light (50 to 100 μE m^{-2} s^{-1}) at 30ºC and high humidity.

Eight individual colonies were observed after 20 days of incubation which were then patched onto agar plates containing BG-11(+N) with 2μg/ml Sp/Sm. After 21 days of incubation patches from each colony were inoculated into 5 ml of BG-11(+N) with 20μg/ml Sp and 5% Sucrose. After 18 days of incubation 100 ul of the liquid culture was then plated onto agar containing BG-11(+N) with 20μg/ml Sp and 5% Sucrose. After 14 days of incubation 5 colonies from each plate that exhibited growth and no contamination were inoculated into 5 ml of BG-11(+N) with 20μg/ml Sp and 5% Sucrose. Each liquid culture that exhibited growth then underwent PCR screening and GC-MS analysis.

PCR screening of potential all3016 KO strains of *Anabaena* sp. PCC7120 was accomplished using primers designed to amplify a 788-bp fragment including the region 48-bp upstream and 78-bp downstream of all3016 in WT *Anabaena* sp. PCC7120 and a 1400-bp fragment in the all3016 KO strains and the KO vector (Figure 4.7). GC-MS analysis was performed on eight ml of potential all3016 KO strains of *Anabaena* sp. PCC7120 and WT cultures. Each culture was pelleted via centrifugation at 4000 g for 15 min and flash frozen with liquid nitrogen. Drying and extraction of the resulting pellet proceeded as described above.

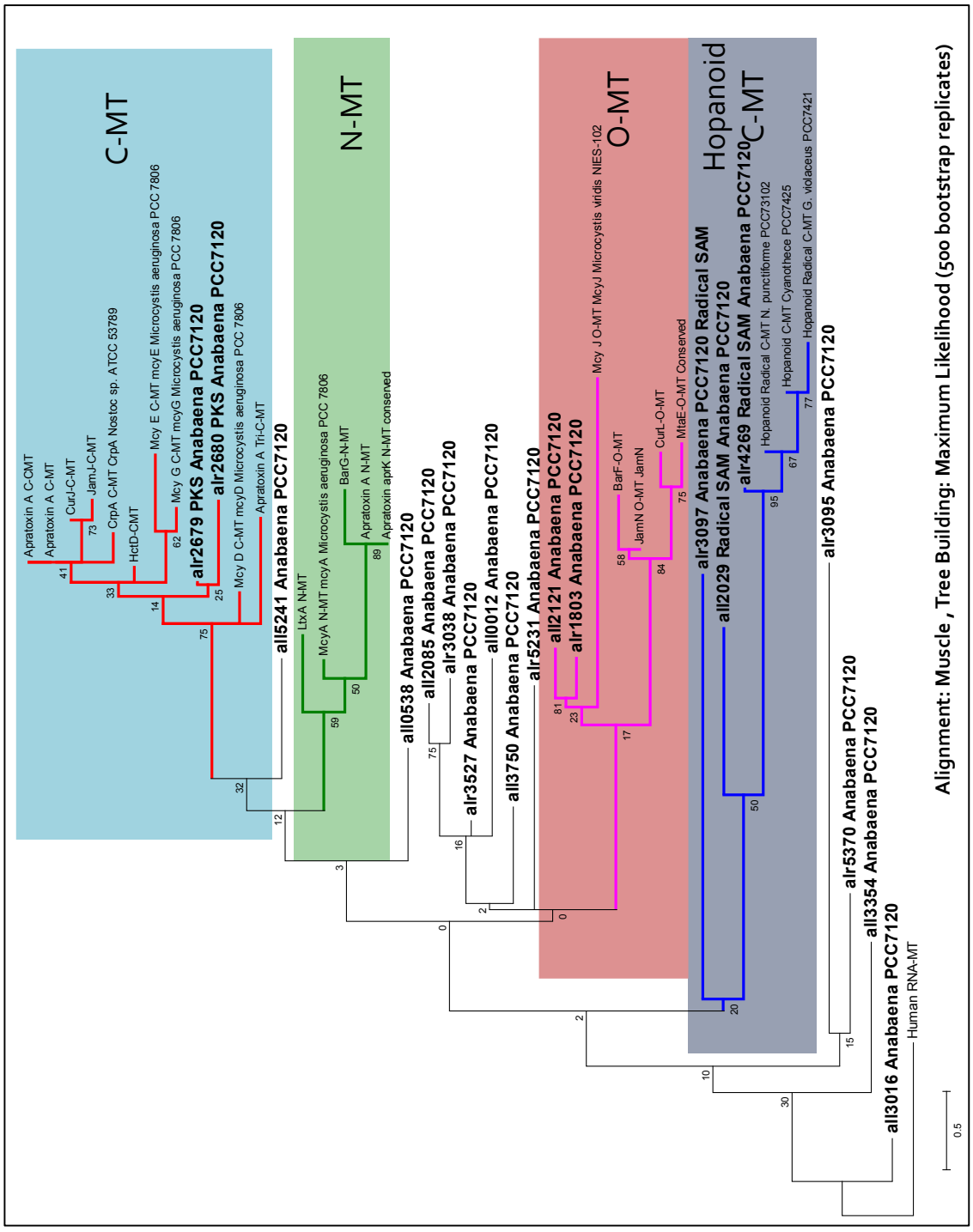
4.4 Results

4.4.1 Identification of Potential Methyltransferases

Bioinformatic identification of potential methyltransferases involved in branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120 identified 23 target genes as

homologs to previously identified cyanobacterial methyltransferases. A phylogenetic analysis of these 23 potential methyltransferases shows relatively poor bootstrap support for the branching that is observed, suggesting a poor conservation between methyltransferases (Figure 4.2). However, clades containing known cyanobacterial methyltransferase were recognizable and are identified in Figure 4.2. Therefore, a phylogenetic analysis of methyltransferases in *Anabaena* sp. PCC7120 is not sufficient to accurately narrow down the list of potential target methyltransferases.

Figure 4.2. Phylogenetic tree of 20 known cyanobacterial methyltransferases with their 23 methyltransferases homologs in *Anabaena* sp. PCC7120. Clades containing known methyltransferases are annotated with colored boxes.


C-MT
N-MT
O-MT
Hopanoid C-MT
Apratoxin A C-CMT
Apratoxin A C-MT
CurJ-C-MT
JamJ-C-MT
CrpA C-MT CrpA Nostoc sp. ATCC 53789
HctD-CMT
Mcy E C-MT mcyE Microcystis aeruginosa PCC 7806
Mcy G C-MT mcyG Microcystis aeruginosa PCC 7806
alr2679 PKS Anabaena PCC7120
alr2680 PKS Anabaena PCC7120
Mcy D C-MT mcyD Microcystis aeruginosa PCC 7806
Apratoxin A Tri-C-MT
all5241 Anabaena PCC7120
LtxA N-MT
McyA N-MT mcyA Microcystis aeruginosa PCC 7806
BarG-N-MT
Apratoxin A N-MT
Apratoxin aprK N-MT conserved
all0538 Anabaena PCC7120
all2085 Anabaena PCC7120
alr3038 Anabaena PCC7120
alr3527 Anabaena PCC7120
all0012 Anabaena PCC7120
all3750 Anabaena PCC7120
alr5231 Anabaena PCC7120
all2121 Anabaena PCC7120
alr1803 Anabaena PCC7120
Mcy J O-MT McyJ Microcystis viridis NIES-102
BarF-O-MT
JamN O-MT JamN
CurL-O-MT
MtaE-O-MT Conserved
alr3097 Anabaena PCC7120 Radical SAM
all2029 Radical SAM Anabaena PCC7120
alr4269 Radical SAM Anabaena PCC7120
Hopanoid Radical C-MT N. punctiforme PCC73102
Hopanoid C-MT Cyanothece PCC7425
Hopanoid Radical C-MT G. violaceus PCC7421
alr3095 Anabaena PCC7120
alr5370 Anabaena PCC7120
all3354 Anabaena PCC7120
all3016 Anabaena PCC7120
Human RNA-MT
0.5
Alignment: Muscle , Tree Building: Maximum Likelihood (500 bootstrap replicates)

Phylogenetic profiling was pursued as an alternative method to narrow down the 23 potential methyltransferases in *Anabaena* sp. PCC7120. CyanoBike and JGI IMG's phylogenetic profiling tools were used to identify orthologs between two strains that are known to produce branched hydrocarbons (*Anabaena* sp. PCC7120 and *Anabaena variabilis* ATCC 29413) and then compared with a closely related strain that does not produce branched hydrocarbons (*Nostoc punctiforme* ATCC 29133). The 23 methyltransferases previously identified in *Anabaena* sp. PCC7120 were compared to the 908 common orthologs identified by CyanoBike between *Anabaena* sp. PCC7120 and *Anabaena variabilis* ATCC 29413 that were not found in *Nostoc punctiforme* ATCC 29133. Only three (gene locus: all2121, all3016, alr3038) of the 23 previously identified methyltransferases were among the 908 common orthologs. Alternatively, IMG's phylogenetic profiler identified only one (gene locus: all3016) of the 23 previously identified methyltransferases among the 727 common orthologs (Figure 3).

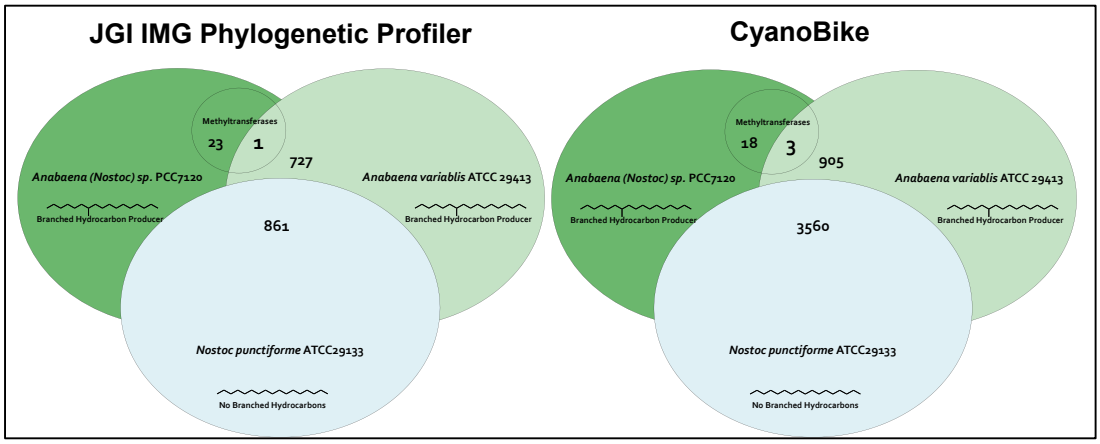


Figure 4.3. Venn diagram representing the two techniques that were used to identify common orthologs identified through phylogenetic profiling between two branched hydrocarbon producers *Anabaena* sp. PCC7120 and *Anabaena variabilis* ATCC 29413 and *Nostoc punctiforme* ATCC 29133, a strain that does not produce branched hydrocarbons.

A third method by which to bioinformatically narrow the list of potential target methyltransferases involved the identification of homologs of *umaA*, a methyltransferase known to perform a similar methylation of an unsaturated fatty acid in *Mycobacterium tuberculosis* (Meena et al. 2013). A recent study characterized the methyltransferase activity of *umaA* establishing its ability to methylate (Z)-11-octadecenoic acid to produce tuberculosteric acid in *Mycobacterium tuberculosis*. A BLAST search using *umaA* against the *Anabaena* sp. PCC7120 genome resulted in a single hit all2121) that was one of the previously identified orthologs identified by CyanoBike.

4.4.2 *In Vivo* and *In Vitro* Analysis

To test the hypothesis that at least one of the targeted methyltransferases (all2121, all3016, and alr3038) encodes a functional methyltransferase that catalyzes the conversion of an unsaturated fatty acid to a branched fatty acid, each of the three genes was cloned, expressed and tested for activity. Expression of all3016 was initially prioritized based upon bioinformatic analysis and the concurrent mutagenesis effort that was on going with all3016 in *Anabaena* sp. PCC7120. Expression of all3016 was optimized using a variety of incubation temperatures, times (2.5 hrs and 12 hrs), and IPTG concentrations (Figure 4.4). Expression was highest for all3016 pET28b in *E. coli* BL21 that was incubated at 18°C for 12 hrs with 0.1 mM IPTG. Expression and activity testing were also completed for incubations of 10, 12 and 15 hrs to explore the possibility that expression or activity differed over this time frame (data not shown). Western blot analysis for all2121, all3016, and alr3038 confirmed that overexpression of all3016

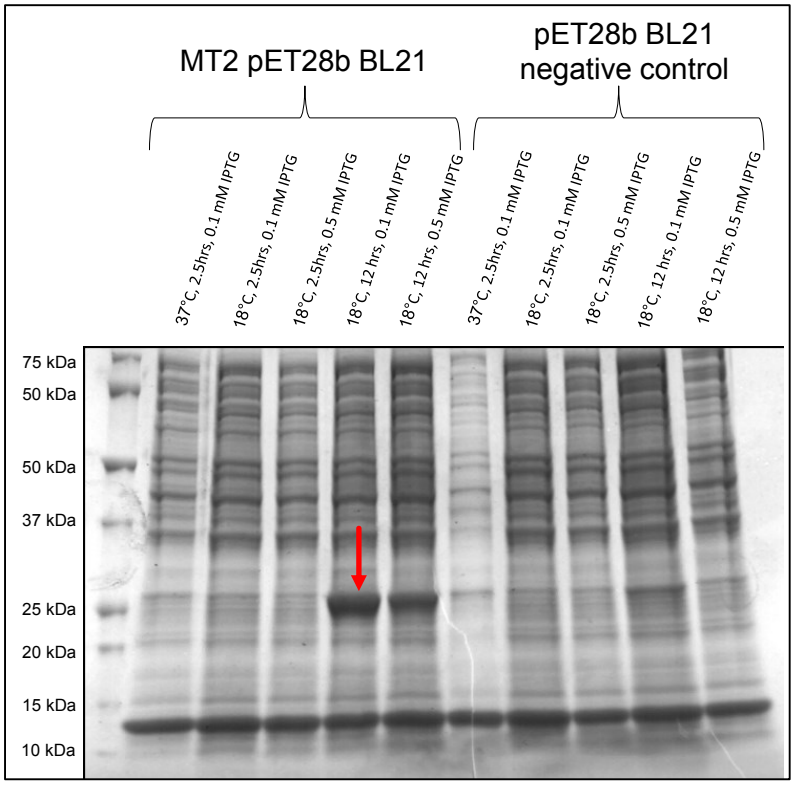


Figure 4.4. A UV visualized Coomassie stained SDS-PAGE gel of whole cell lysates for five conditions tested with all3016 pET28b in *E. coli* BL21 for optimal expression. Expression was highest for all3016 pET28b in *E. coli* BL21 that was incubated at 18°C for 12 hrs with 0.1 mM IPTG (indicated by red arrow).

was successful and that the purified protein exhibited the expected His tag (Figure 4A5). In a subsequent experiment, we successfully expressed alr3038 as well as all3016 under the conditions described above (Figure 4A5). Whole cell lysates and purified protein from this experiment were used for *in vitro* analysis.

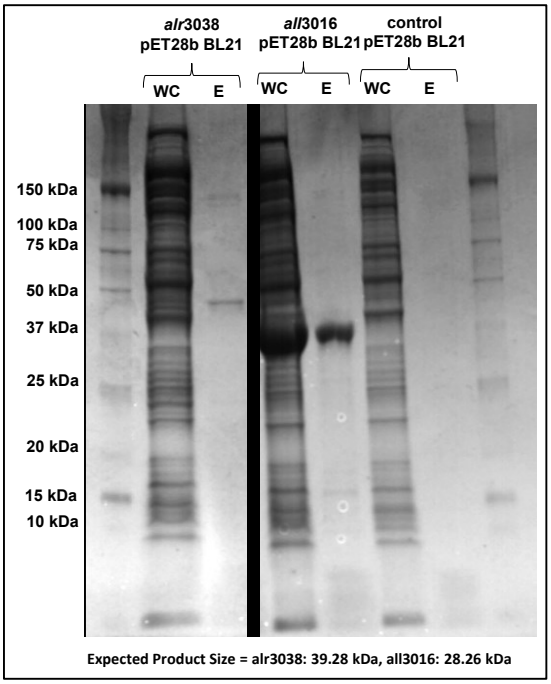


Figure 4.5. SDS-PAGE gel of whole cell lysate (WC) and elution (E) for the purified expressed protein for all3016 pET28b and alr3038 pET28b. Distinct bands are observed in the expected size range for all3016 pET28b (28.26 kDa) and alr3038 pET28b (39.28 kDa). Lanes shown were run together on the same gel. A line indicates where a region of the gel has been removed from the image.

Exogenous fatty acids (Z) 11-octadecenoic acid and (9*Z*,12*Z*) 9-, 12-octadecadienoic acid were tested as substrates for *in vivo* analysis of all3016 pET28b using intact cells. Cultures were incubated for 12 hrs after IPTG induction at 18°C. FAME analysis was completed on the biomass or reactants harvested from each of the *in vivo* and *in vitro* experiments including those involving the addition of exogenous substrates (Figure 4A6). None of the FAME results exhibited any unique branched fatty acid methyl esters that would be consistent with the predicted methyltransferase activity. Chromatograms of the FAME analysis for *in vivo* analyses of all2121 pET28b, all3016 pET28b and all3016 pET28b without exogenous substrates are displayed in Figure 4.6. Extracted ion chromatograms of the same analysis with mass ranges consistent with

branched fatty acids are displayed in Figure 4.7 and Figure 4A8. Chromatograms of FAME analysis for *in vitro* analyses of all2121 pET28b, all3016 pET28b and all3016 pET28b exhibited the same FAME composition as the *in vivo* experiments and are displayed in Figure 4A7.

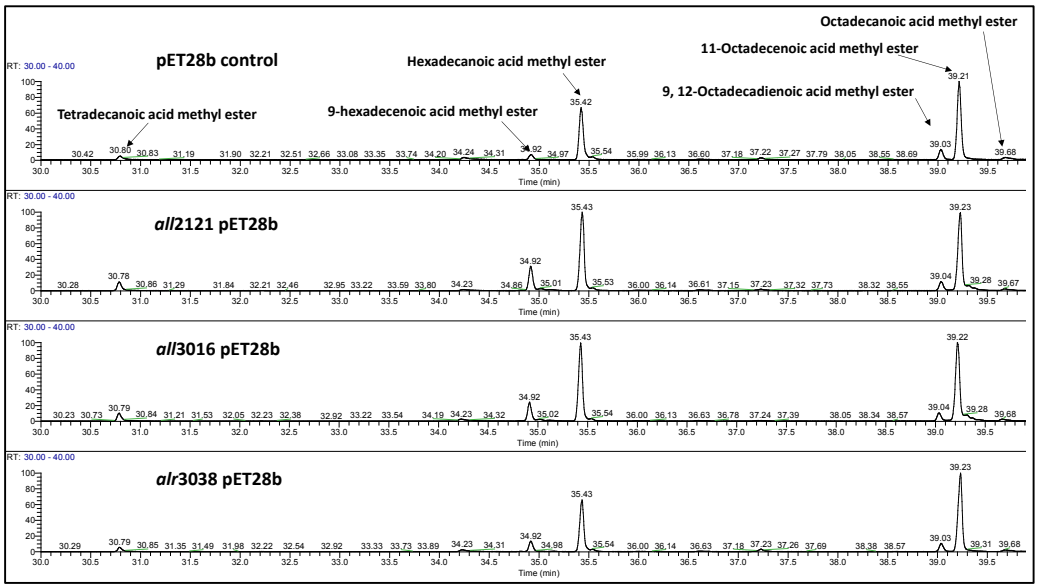


Figure 4.6. Chromatograms of FAMEs from the *in vivo* analyses of all2121 pET28b, all3016 pET28b and all3016 pET28b incubation experiments.

4.4.3 Mutagenesis of all3016 in *Anabaena* sp. PCC7120

To investigate the role of all3016 in the biosynthesis of 7-methylheptadecane biosynthesis in *Anabaena* sp. PCC7120, mutagenesis via homologous recombination was employed as a means to knock out the function of all3016. It was hypothesized that if all3016 was the single methyltransferase responsible for the methylation of unsaturated fatty acids that were subsequently converted to branched alkanes, then knocking out this methyltransferase would abolish the production of branched hydrocarbons in the resulting mutant strain. Assembly of the knockout vector for double recombinant mutagenesis proceeded successfully and following conjugation and selection eight colonies were

observed after 20 days of incubation. Additional selection with increased spectinomycin and counter selection with sucrose yielded five of the initial eight colonies with many subcultures exhibiting sustained growth under selective pressure. PCR verification of five subcultures representing each of the five initial colonies showed 100% success with mutagenesis via homologous recombination for all3016 (Figure 4.7 and Figure 4A10). The expected PCR product was observed for the WT reaction (788-bp), the KO vector control (1400-bp) as well as the five all3016 KO strains (1400-bp). Results from GC-MS analysis of three of the five subcultures indicated that all three

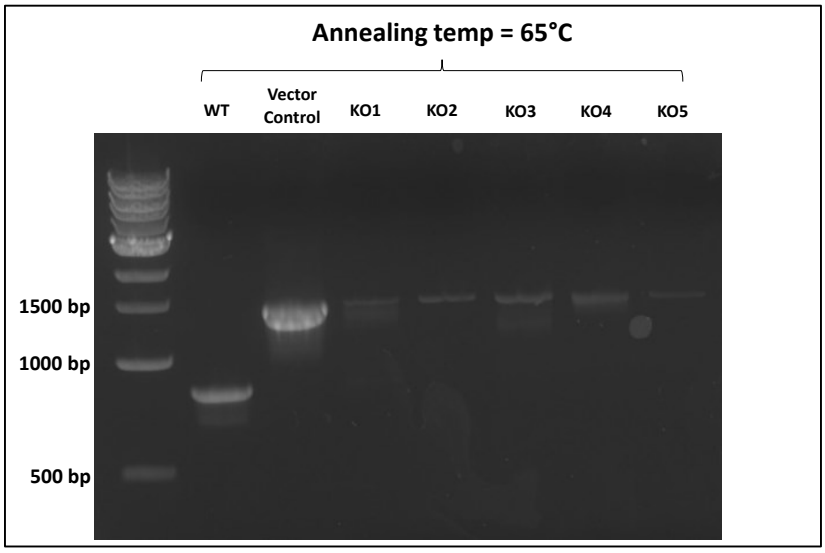


Figure 4.7. PCR screening of all3016 KO mutants using primers designed from left and right regions flanking all3016. The observed PCR products are consistent with successful double recombinant mutants for all five of the all3016 KO mutant strains. The expected PCR product size for each reaction were: WT = 788-bp, vector control = 1400-bp, all3016 double recombinant KO = 1400-bp.

strains still produce 7-methylheptadecane (Figure 4.8). Therefore, either all3016 is not involved in branched hydrocarbon biosynthesis, or alternatively, it is not the only methyltransferase responsible for 7-methylheptadecane biosynthesis in *Anabaena* sp. PCC7120.

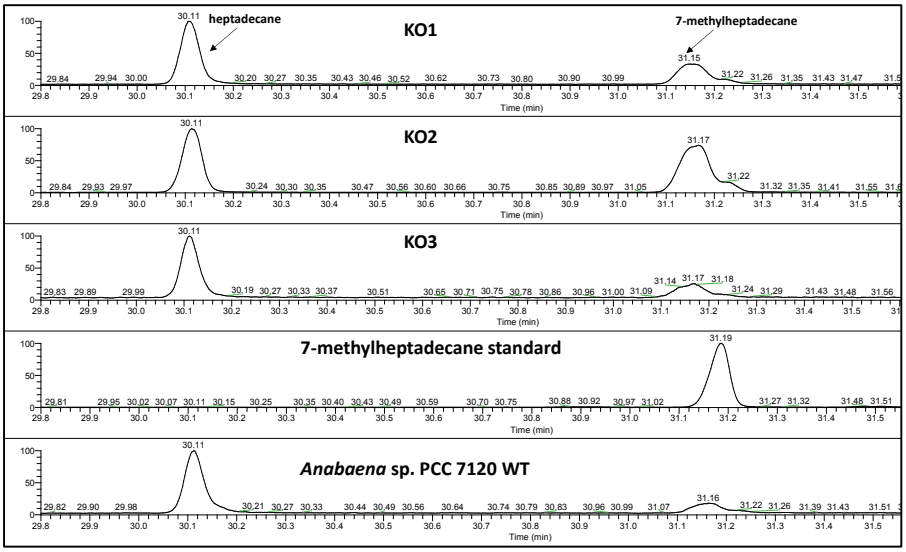


Figure 4.8. GC-MS chromatograms of FAME extracts for three of the all3016 KO mutant strains compared to a standard for 7-methylheptadecane and the hydrocarbon extract of *Anabaena* sp. PCC7120. 7-methylheptadecane production is clearly observed in the three all3016 KO mutant strains.

All three of the target methyltransferases have been tested by at least one method for their involvement in branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120. A summary of the experiments that have been completed to investigate these three methyltransferases is shown in Table 1. The most extensively tested putative methyltransferase was all3016, which has been knocked out in *Anabaena* sp. PCC7120 as well as tested for *in vivo* activity in *E. coli* (with endogenous as well as exogenous susbtrates) as well for *in vitro* activity. Alr3038 has been investigated for *in vivo* activity in *E. coli* (with endogenously and exogenous susbtrates) as well for *in vitro* activity while all2121 has only been tested for *in vivo* activity in *E. coli* once.

Table 4.1. Summary of experiments completed to investigate methyltransferase activity of all2121, all3016, and alr3038.

Protein	*In Vivo* (replicates)	*In Vivo* + FA (replicates)	*In Vitro* (replicates)	*In Vitro* + FA (replicates)	*Anabaena sp.* PCC 7120 KO (replicates)
all2121	No Activity (1)	Untested	Untested	Untested	Untested
all3016	No Activity (5)	No Activity (3)	No Activity (1)	No Activity (1)	No Phenotype (3)
alr3038	No Activity (3)	Untested	No Activity (1)	No Activity (1)	Untested

4.5 Discussion

Cyanobacterial hydrocarbon biosynthesis is a unique phenomenon, and production of branched hydrocarbons is an even more specialized capacity among cyanobacteria. Branched hydrocarbons have been observed from cyanobacterial representatives of every major subdivision (Table 4A1). Among genome sequenced cyanobacteria, branched hydrocarbon production is mostly observed in heterocystous cyanobacteria with the exception of *Planktothrix agardhii* NIVA-CYA 126/8 and *Gloeobacter violaceus* PCC7421 (Coates et al. 2014). An investigation into the evolutionary history of this pathway may reveal new insights into the distribution of branched hydrocarbon biosynthesis after a methyltransferase responsible for branched hydrocarbon production is identified.

Using three separate bioinformatic tools this study identified three (all2121, all3016, alr3038) target methyltransferases as candidates for being involved in branched hydrocarbon biosynthesis in *Anabaena* sp. PCC7120. The CyanoBike analysis returned all three of the target methyltransferases while IMG's phylogenetic profiler returned only all3016 and BLAST searching using the *umaA* methyltransferase involved in tuberculostearic acid methylation returned only all2121. The differences in search

parameters between the three different methods explain these varying results. These three parallel methods for bioinformatically narrowing the list of potential methyltransferases confirmed the findings from each method and allowed for a prioritization of these three methyltransferases for further investigation.

In vivo and *in vitro* characterization of all3016 as one of the three targeted methyltransferases was studied first due to the earlier timing of its successful assembly and expression. Setbacks in assembly steps and expression optimization decreased the number of assays that could be completed using the other two targeted methyltransferases. Successful expression and purification of all three methyltransferases was observed using the pET28b expression vector in *E. coli* BL21 (Figure 4.5 and 4A3). However, GC-MS analysis of FAMEs from the *E. coli* biomass from each experiment did not show production of any of the predicted branched fatty acids. The GC-MS chromatograms showed the presence of the two possible substrates (9-hexadecenoic acid and 11-octadecenoic acid) that we propose could be methylated to produce branched fatty acids, and that would ultimately be converted into 7 or 8-methylheptadecane in *Anabaena* sp. PCC7120. Therefore, if any of the three methyltransferases exhibited activity we would have expected that they would have been able to utilize the unsaturated fatty acids provided endogenously by *E. coli*. Exogenously supplied (*Z*) 11-octadecenoic acid as well as (9*Z*,12*Z*)-9,12-octadecadienoic acid were also supplied to all3016 pET28b for *in vivo* testing; however branched fatty acids were not observed in these experiments. Expression levels for the three methyltransferases exhibited some variability and could be optimized further in future experiments. Expression of each of the methyltransferases using a constitutive promoter system (such as PconII) instead of the inducible T7

promoter may result in different activity. Optimization of expression for all2121 pET28b as well *in vivo* and *in vitro* analysis of activity should be a priority in future investigations of branched hydrocarbon biosynthesis.

In vitro analysis of purified all3016 and alr3038 pET28b involved exogenously supplied (*Z*) 11-octadecenoic acid as well as SAM and NADPH in two different buffer conditions. The conditions that were tested did not yield production of branched fatty acids for all3016 and alr3038 pET28b and did not exhibit any significant difference from the controls (Figure 4A9). Additional exogenous substrates with different chain lengths or double bond positions such as (*Z*) 9-hexadecenoic acid would be reasonable substrates for future investigations of the *in vitro* activity for any of the three methyltransferases. Additional buffer conditions as well as alternative reductants such as NADH or FADH may also provide alternative conditions for testing for methyltransferase activity.

Previous investigations of cyclopropane fatty acids (CFAs) in other bacteria may provide some perspective on the results observed in this investigation, as well as for future investigations. Cyclopropane fatty acids have been investigated extensively in bacteria as diverse as *E. coli* and *Mycobacterium* and their biosynthesis is known to involve a relatively highly conserved group of methyltransferases (Grogan and Cronan 1997). CFAs are produced during the transition between exponential growth and stationary phase. Although the physiological function of CFAs in bacteria has not been definitively established, many researchers hypothesize that CFAs decrease membrane fluidity (Grogan and Cronan 1997). Detection of CFAs are reported to be particularly challenging because the cyclopropane ring can be lost during conversion of the fatty acid moieties to the methyl esters, especially when using acid-catalyzed methylation

conditions. CFAs are often detected as branched or methoxy fatty acids (Grogan and Cronan 1997). The GCMS analysis of *in vivo* and *in vitro* assays completed for this investigation did detect peaks with mass fragmentations patterns consistent with methoxy fatty acids as well as cyclopropane fatty acids. Future investigations of methyltransferase activity using heterologous expression in *E. coli* may benefit from first characterizing production of commonly observed CFAs in *E. coli* BL21, or pursue the use of strains that are deficient in the CFA synthase. It should also be noted that CFA biosynthesis proceeds through a very similar biosynthetic mechanism to what is proposed for cyanobacterial branched hydrocarbon biosynthesis. The methylation of an unsaturated fatty acid for CFA biosynthesis proceeds using the same substrate and co-factor as our proposed branched hydrocarbon pathway in cyanobacteria. However, in CFA biosynthesis the final reduction of the carbocation intermediate comes from abstraction of the methyl proton from the initial carbocation that gives the cyclopropane ring. Variability of active site residues in CFA synthases appears to vary the mechanism of quenching of the carbocation (Grogan and Cronan 1997). BLAST analysis of the CFA synthase found in *E. coli* against the *Anabaena* sp. PCC7120 genome returns a hit with all2121 (e-value = $4E^{-10}$, percent identity = 35%) and all1803 (e-value = $2E^{-8}$, percent identity = 25%). This observation further validates the prioritization of all2121 for future investigations and provides all1803 as an additional target if necessary.

To complement these investigations where the *in vivo* and *in vitro* activity of the three target methyltransferases in *E. coli* was explored, an effort to knock out all3016 in *Anabaena* sp. PCC7120 was also pursued (Figure 4A10). However, the resulting strains did not exhibit a loss of production of branched hydrocarbons, which would have been

expected if this methyltransferase had been involved in branched hydrocarbon biosynthesis. The three strains that were analyzed exhibited some variability in 7-methylheptadecane production, however, it cannot yet be determined if this is due to experimental design or actual changes in production levels between strains. Future experiments to quantify production from all3016 KO strains may clarify this issue. It is possible that more than one methyltransferase is capable of performing the methylation reaction involved in branched hydrocarbon biosynthesis. In this case, concurrently mutated strains would be necessary to observe the abolishment of production of branched hydrocarbons. Additionally, mutagenesis of the remaining two methyltransferases (all2121 and alr3038) is necessary to fully explore their role in branched hydrocarbon biosynthesis.

Identification of the methyltransferase responsible for branched hydrocarbon biosynthesis will be an important finding that will have implications for a variety of different disciplines. Through a deeper understanding of the fundamental biosynthetic process involved in branched hydrocarbon production, a deeper understanding of their functional role should be possible. Additionally, the physiological function of these hydrocarbons could be characterized through heterologous expression in cyanobacterial strains that do not naturally produce branched hydrocarbons. An investigation of the evolutionary history of branched hydrocarbon biosynthesis might provide insights into their distribution as well as their physiological function.

The discovery of a methyltransferase with the ability to produce branched fatty acids from unsaturated fatty acids would have many implications for biofuel research. Highly branched short chain molecules make high quality gasoline fuels. A

methyltransferase from cyanobacterial branched hydrocarbon biosynthesis could be used to increase branched chain fatty acid production or production of other methylated molecules in biofuel or chemical producing organisms, and this would be a highly valuable enzyme. An investigation into the substrate specificity of the methyltransferase would provide a valuable foundation for engineering novel functions for such a methyltransferase enzyme.

Any future investigations of branched hydrocarbon biosynthesis should prioritize mutagenesis of all2121 in *Anabaena* sp. PCC7120 given the multiple lines of bioinformatic evidence suggesting that all2121 exhibits the most homology to the fatty acid methyltransferases *umaA* from tuberculosteric acid biosynthesis as well as *cfa* from cyclopropane fatty acid biosynthesis in *E. coli.* Additionally, further heterologous expression and *in vivo* as well as *in vitro* characterization of methyltransferase activity for all2121 is warranted given the above evidence of and considering that these experiments were not completed in this investigation. Further characterization of the cyclopropane fatty acid composition and level of production by *E. coli* BL21 cells under the conditions used for this investigation may help clarify *in vivo* as well as *in vitro* activity. Future *in vivo* and *in vitro* experiments should also include unsaturated C16 unsaturated fatty acid substrates in addition to the previously tested C18 unsaturated fatty acid substrates. Mutagenesis as well as additional heterologous expression of alr3038 should also be considered. Given the homology between *cfa* from cyclopropane fatty acid biosynthesis in *E. coli* and all1803 an effort to explore methyltransferase activity for all1803 should also be considered.

4.6 Appendix

Table 4A1. Summary of literature reports of cyanobacterial branched hydrocarbon production. Strains are organized by subdivision on the left-hand side, and a '+' symbol indicates that a particular hydrocarbon was reported for this strain. The relevant reference is given in the right-hand column. Findings that were validated using authentic standards or established using derivatization techniques to verify hydrocarbon structure are denoted with superscript 1.

Table A1. Reports of Cyanobacterial Branched Hydrocarbons

Strain	meC14	meC16	meC17	meC18	meC19	References
Subdivision 1 (Unicellular)						
Anacystis cyanea			+[1]			Gelpi *et al*. 1970
Anacystis nidulans			+[1]			Paoletti *et al.* 1976, Winters *et al.* 1969, McInnes 1980, Han 1968
Chroococcus turgidus		+	+	+		Gelpi *et al.* 1970, Murray & Thomson 1977
Synechococcus bacillaris		+				Blumer et al. 1971
Subdivision 2 (Baeocystous)						
Stanieria PCC 7301	+[1]					Allen *et al*. 2010
Subdivision 3 (Filamentous)						
Lyngbya aestuarii			+			Gelpi *et al*. 1970
Lyngbya lagerhaimii			+			Winters *et al*. 1969
Microcoleus vaginatus		+	+			Dembitsky *et al.* 2000
Phormidium luridum			+			Han 1969, Blumer 1971
Phormidium lucidum FACHB 238			+			Liu *et al.* 2013
Planktothrix agardhii			+			Juttner 1991
Planktothrix mougeotii			+			Juttner 1991
Planktothrix rubescens			+			Juttner 1991
Planktothrix suspense			+			Juttner 1991
Spirulina sp.			+	+	+	Paoletti *et al*. 1976
Spirulina platensis-Mao			+	+	+	Paoletti *et al.* 1976, 1977, 1978, Rezanka *et al.* 1982
Subdivision 4 (Heterocystous)						
Anabaena (Nostoc) sp. PCC 7120			+[1]			Schirmer *et al*. 2010, Tan *et al*. 2011
Anabaena variablis FACHB 319			+			Liu *et al.* 203
Anabaena variabilis ATCC 29413			+[1]			Han *et al.* 1969, Goodloe 1982, Fehler & Light 1970, Schirmer *et al.* 2010
Calothrix scopulorum			+			Koster *et al*. 1999
Calothrix sp.				+	+	Paoletti *et al*. 1976
Nostoc sp.			+			Gelpi *et al*. 1970
Nostoc commune			+	+	+	Paoletti *et al*. 1976
Nostoc cacicola FACHB 389	+					Liu *et al.* 203
Nostoc endophytum			+			Murray & Thomson 1977
Nostoc muscorum			+			Winters *et al*. 1969, Han *et al*. 1968
Nostoc punctiforme FACHB 252	+					Liu *et al.* 203
Nostoc spongiaeforme FACHB 130			+			Liu *et al.* 203
Nostoc sp.a			+			Dembitsky *et al*. 1999
Scytonema javanicum FACHB 887			+			Liu *et al.* 203
Tolypothrix tenuis FACHB			+			Liu *et al.* 203
Subdivision 5 (Ramified)						
Chlorogloea fritschii			+[1]	+		Han *et al.* 1968

Table 4A2. Cyanobacterial methyltransferases that were used to perform a BLAST search against the *Anabaena* sp. PCC7120 genome to identify potential methyltransferases.

Methyltransferase	Substrate	Organism	Reference
	Carbon Substrates		
sll0418	Tocopherol Carbon	*Synechocystis PCC 6803*	Shintani et al. 2002
G. viol Hop Radical Mtase	Hopanoid Radical Carbon	*Gloeobacter violeceous*	Welander et al. 2010
N. punct. Hop Radical Mtase	Hopanoid Radical Carbon	*Nostoc punctiforme*	Welander et al. 2010
McyD	Microcystin Carbon	*Microcystis aeruginosa*	Tillet et al. 2001
HctD	Hectachlorin Carbon	*Moorea producens* JHB	Ramaswamy et al. 2007
CurJ	Curacin A Carbon	*Moorea producens* 3L	Chang et al. 2004
CrpA	Cryptophycin Carbon	*Nostoc* sp. ATCC 53789	Magarvey et al. 2006
JamJ	Jamaicamide Carbon	*Moorea producens* JHB	Edwards et al. 2004 (2)
slr0214	DNA Cytosine Carbon	*Synechocystis sp. PCC 6803*	Scharnag et al. 1998
	Nitrogen Substrates		
McyA	Microcystin Nitrogen	*Microcystis aeruginosa*	Tillet et al. 2000
LtxA	Lyngbyatoxin Nitrogen	*Moorea producens*	Edwards et al. 2004
BarF	Barbamide Oxygen	*Moorea producens* JHB	Chang et al. 2002
BarG	Barbamide Nitrogen	*Moorea producens* JHB	Chang et al. 2002
SYNW1914	Betaine Nitrogen	*Synechococcus* sp. WH 8102	Lu et al. 2006
Met	2-methylisoborneol Nitrogen	*Pseudoanabaena limnetica*	Giglio et al. 2010
	Oxygen Substrates		
McyJ	Microcystin Oxygen	*Microcystis aeruginosa*	Tillet et al. 2000
CurL	Curacin A Oxygen	*Moorea producens* 3L	Chang et al. 2004
slr0095	Hydroxycinnamic acids Oxygen	*Synechocystis sp. PCC 6803*	Kopycki et al. 2008
SpoU	RNA Oxygen	*E. coli*	Koonin and Rudd 1993

Table 4A3. Primers used to amplify methyltransferases from *Anabaena* sp. PCC7120 to clone into pET28b and pCONII RiboJ vectors

Primer	Sequence
***E. coli* Expression**	
all2121_pET28bFwd	GTGGACAGCAAATGGGTCGGATGAGTTGGTTGTTTTCTACACTG
all2121_pET28bRev	GAGTGCGGCCGCAAGCTTTATTACTTTTGAGCAACCTTGATC
all3016_pET28bFwd	GTGGACAGCAAATGGGTCGGATGCACAGACAACTAGAACCAGAAGTGATG
all3016_pET28bRev	GCTAAACGCAGTTGGACTAATTAGTAGAGCTTGCGGCCGCACTC
alr3038_pET28bFwd	GTGGACAGCAAATGGGTCGGATGGACAAAAAGAGCAAGCC
alr3038_pET28bRev	GAGTGCGGCCGCAAGCTCTACTAATCTGGTTTCCTCGCAA
***Anabaena* sp. PCC 7120 Double Recombinant Mutagenesis**	
3016_GC_left_R_2	ATCCGCGCGCGCGCGCGCGATTGCTAATACCTCTATCCGTATAAGT
3016_C2G_Right_F_2	ATCCCGCCGCCGCCGCCGGATAAATCTCATACTGTTCGGCTTC
3016_G3C3_Right_R_2	ATCGGGCCCGGGCCCGGGGACCCCCTCAGATAAAGACAGAAC
3016_G5C5_Left_F	CATCGGGGGCCCCCGGGGGGACTGGCTTTAGCATAATCGCCGAT

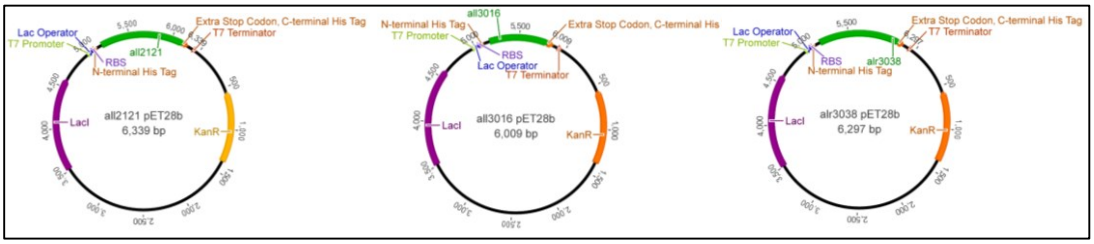


Figure 4A1. Vector maps for the three methyltransferases as they are assembled in the pET28b expression vector.

Figure 4A2. Sequence alignment of successfully assembled all2121 in the pET28b vector. The two sequencing files have been aligned with the vector map. The forward sequence used the T7 promoter as the sequencing primer and the reverse sequence used the T7 terminator as the sequencing primer. The green horizontal bar represents the alignment and any interruption in the green bar represents a less than 100% identity between sequences. The alignment shows that the sequences match the vector map 100%, indicating a successful alignment.

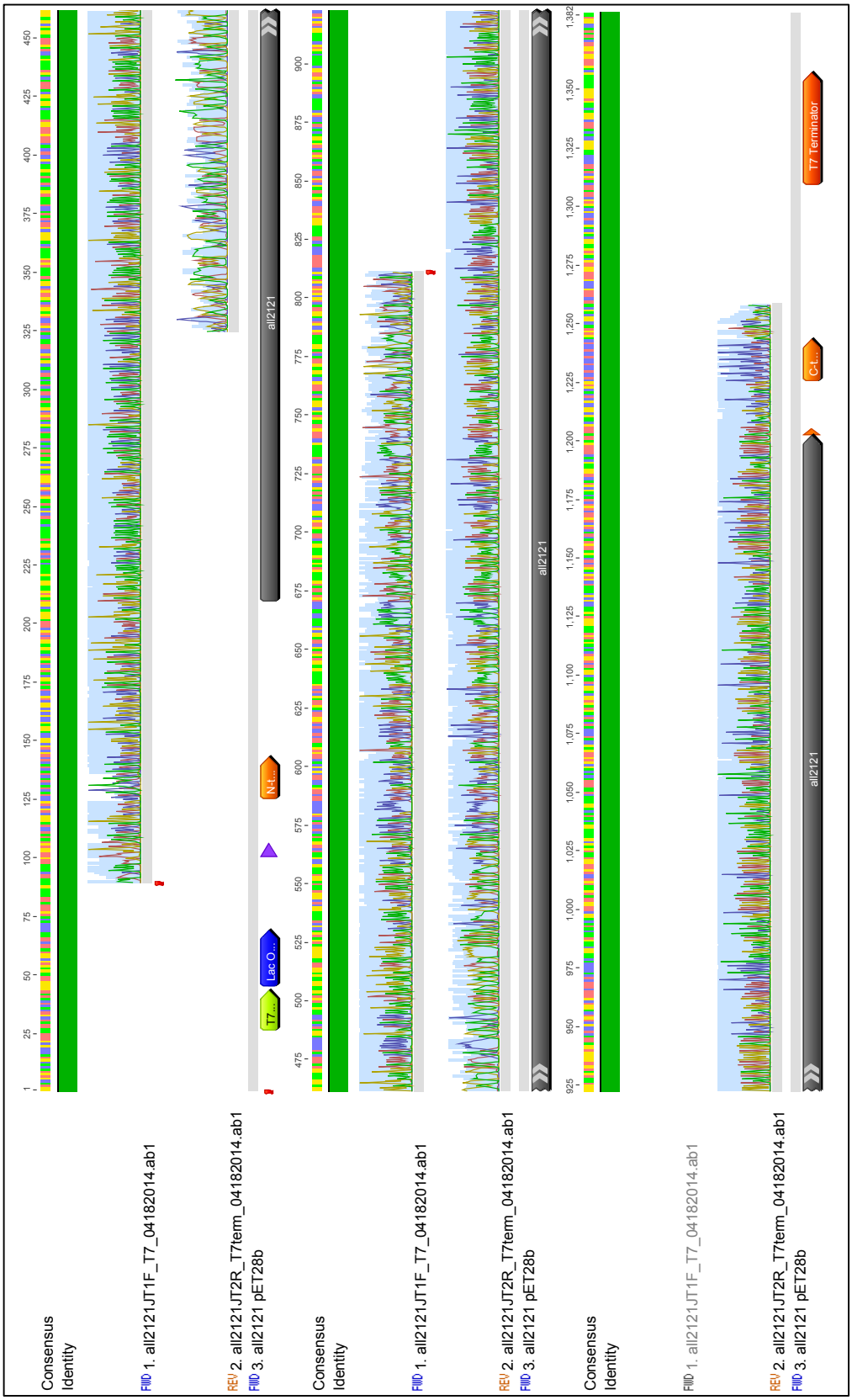

Consensus
Identity
FWD 1. all2121JT1F_T7_04182014.ab1
REV 2. all2121JT2R_T7term_04182014.ab1
FWD 3. all2121 pET28b
T7...
Lac O...
N-t...
all2121
Consensus
Identity
FWD 1. all2121JT1F_T7_04182014.ab1
REV 2. all2121JT2R_T7term_04182014.ab1
FWD 3. all2121 pET28b
all2121
Consensus
Identity
FWD 1. all2121JT1F_T7_04182014.ab1
REV 2. all2121JT2R_T7term_04182014.ab1
FWD 3. all2121 pET28b
all2121
C-t...
T7 Terminator

Figure 4A3. Sequence alignment of successfully assembled all3016 in the pET28b vector. The two sequencing files have been aligned with the vector map. The forward sequence used the T7 promoter as the sequencing primer and the reverse sequence used the T7 terminator as the sequencing primer. The green horizontal bar represents the alignment and any interruption in the green bar represents a less than 100% identity between sequences. The alignment shows that the sequences match the vector map 100%, indicating a successful alignment.

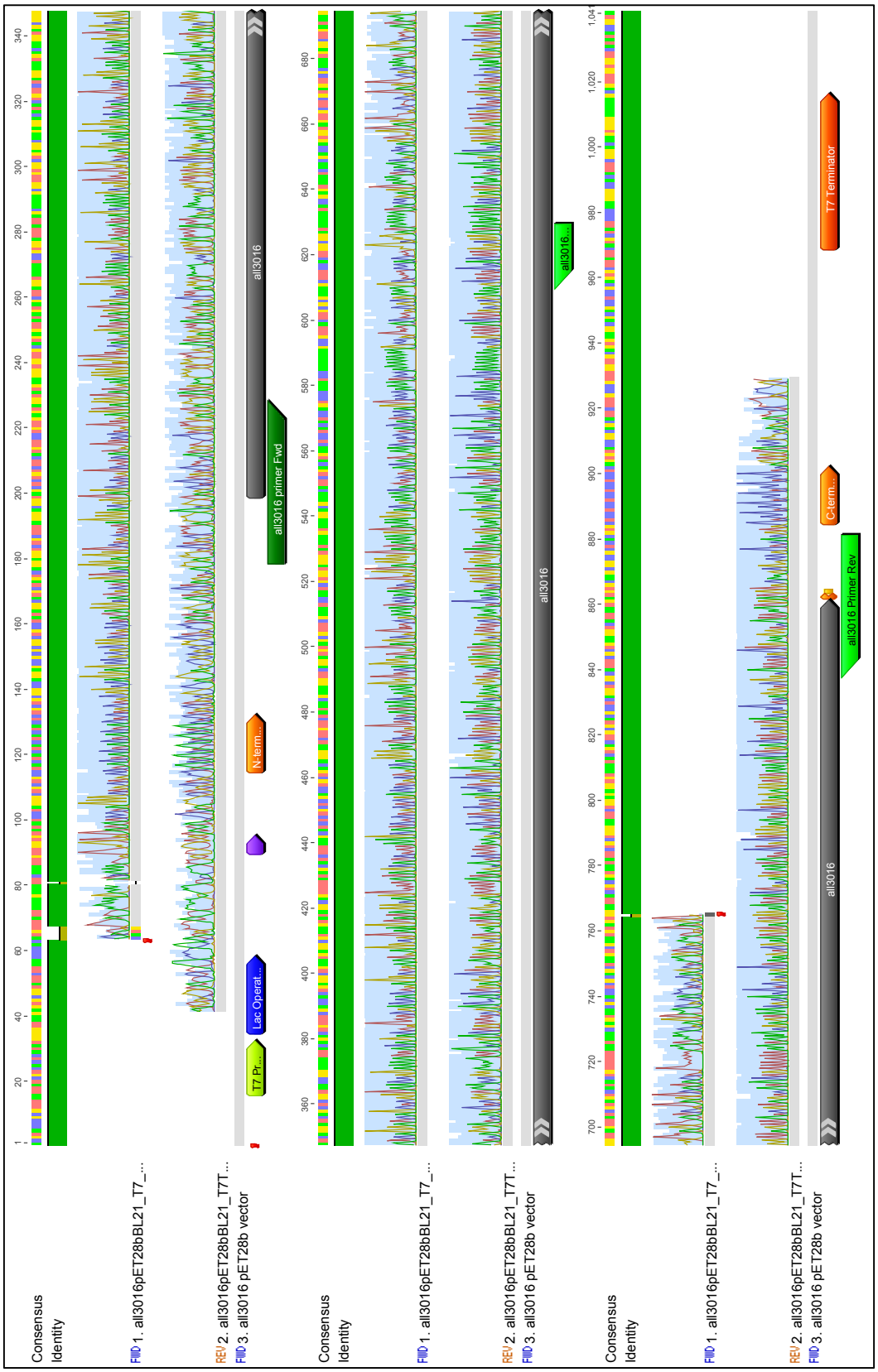
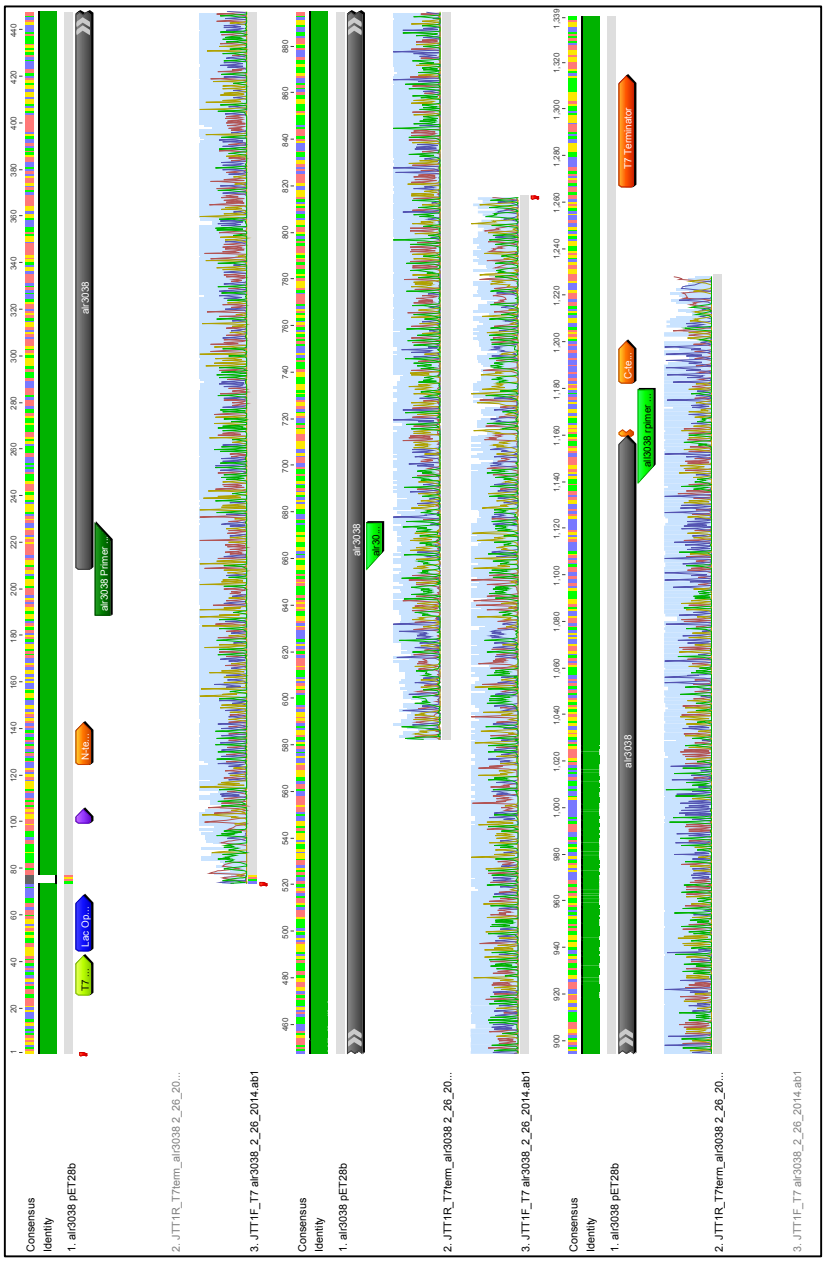

Consensus
Identity
FWD 1. all3016pET28bBL21_T7_...
REV 2. all3016pET28bBL21_T7T...
FWD 3. all3016 pET28b vector
T7 Pr...
Lac Operat...
N-term...
all3016 primer Fwd
all3016
all3016...
C-term...
all3016 Primer Rev
T7 Terminator

Figure 4A4. Sequence alignment of successfully assembled all3016 in the pET28b vector. The two sequencing files have been aligned with the vector map. The forward sequence used the T7 promoter as the sequencing primer and the reverse sequence used the T7 terminator as the sequencing primer. The green horizontal bar represents the alignment and any interruption in the green bar represents a less than 100% identity between sequences. The alignment shows that the sequences match the vector map 100%, indicating a successful alignment.


Consensus
Identity
1. alr3038 pET28b
T7 ...
Lac Op...
N-te...
alr3038
alr3038 Primer ...
2. JTT1R_T7term_alr3038 2_26_20...
3. JTT1F_T7 alr3038_2_26_2014.ab1
Consensus
Identity
1. alr3038 pET28b
alr3038
alr30...
2. JTT1R_T7term_alr3038 2_26_20...
3. JTT1F_T7 alr3038_2_26_2014.ab1
Consensus
Identity
1. alr3038 pET28b
alr3038
C-te...
alr3038 primer ...
T7 terminator
2. JTT1R_T7term_alr3038 2_26_20...
3. JTT1F_T7 alr3038_2_26_2014.ab1

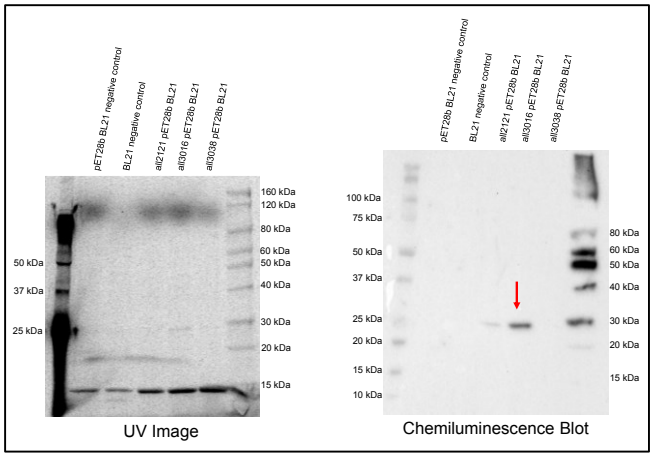


Figure 4A5. Western blot analysis of expression for all2121 pET28b, all3016 pET28b, and alr3038 pET28b confirming that overexpression of all3016 was successful and that the purified protein exhibited the expected His tag. Left image is of the UV image of the SDS-PAGE gel and the right image is of the chemiluminescence observed from the Western blot. The red arrow indicates His-tagged protein for all3016 pET28b.

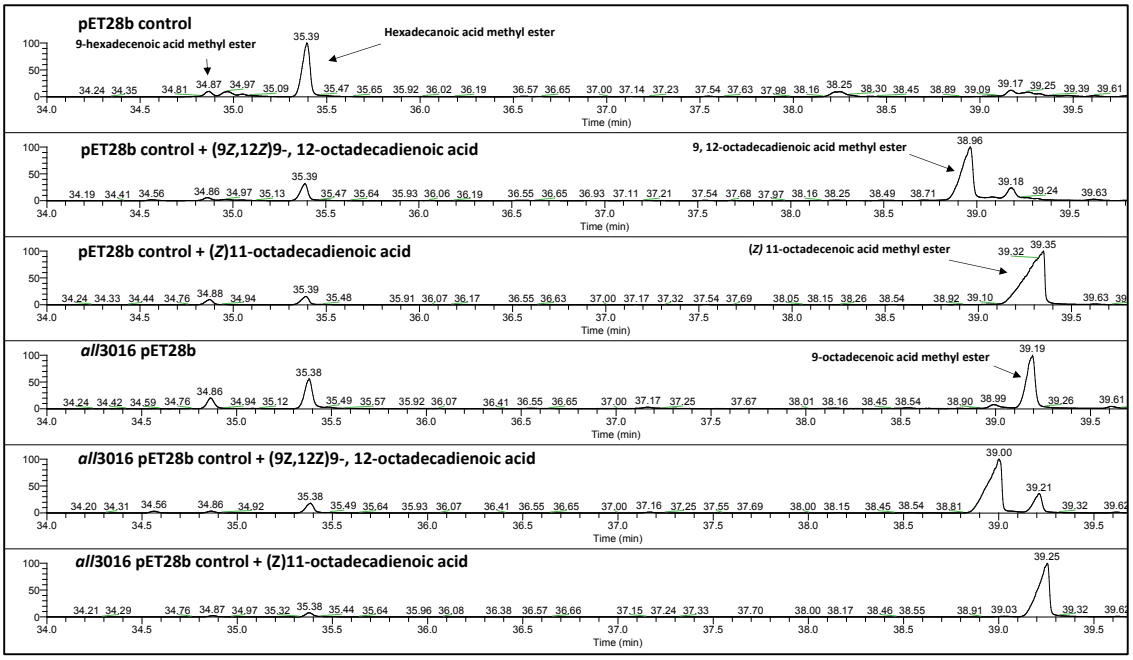


Figure 4A6. Chromatograms of FAME analysis for *in vivo* analyses of all3016 pET28b with exogenously supplied (9*Z*, 12*Z*)9-, 12-octadecadienoic acid and (*Z*)11-octadecadienoic acid.

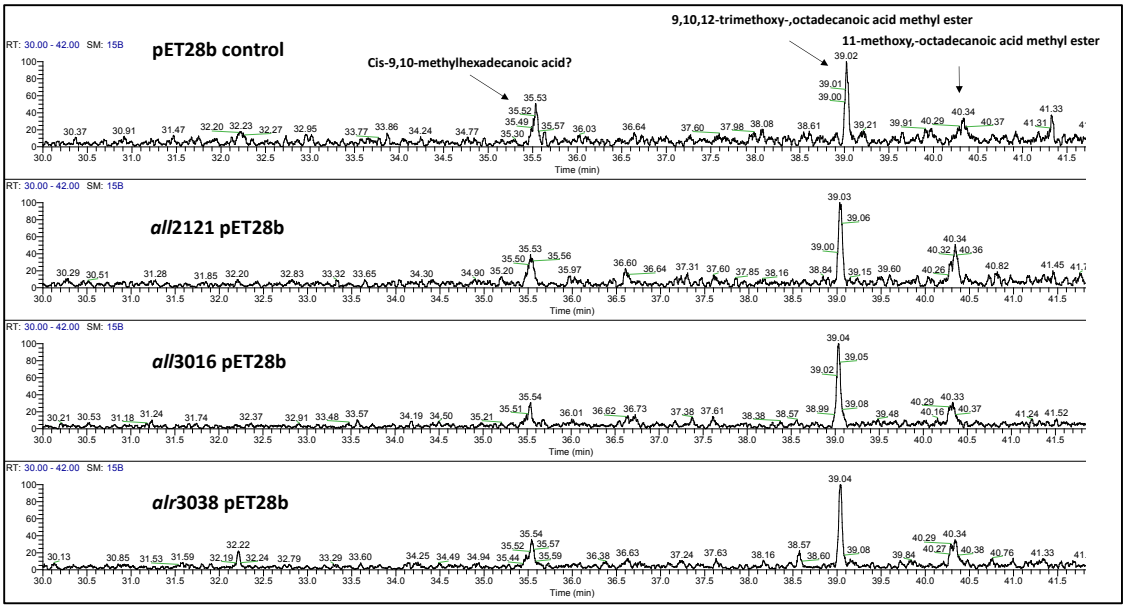


Figure 4A7. Extracted ion chromatograms of FAME from *in vivo* analyses of all2121, all3016, and alr3038 pET28b expressed in *E. coli* BL21. The extracted mass displayed is 283-285 m/z which corresponds to the MW of 9-methylhexadecanoic acid methyl ester, the expected product from methylation of a C16 unsaturated fatty acid.

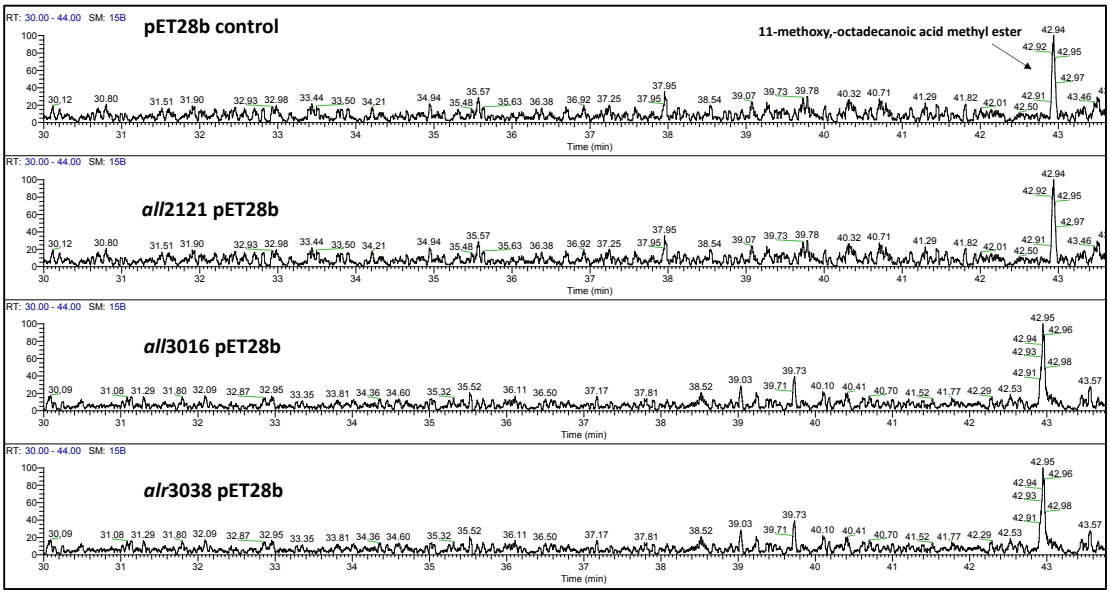


Figure 4A8. Extracted ion chromatograms of FAME from *in vivo* analyses of all2121, all3016, and alr3038 pET28b expressed in *E. coli* BL21. The extracted mass displayed is 311-313 m/z which corresponds to the MW of 9-methyloctadecanoic acid methyl ester, the expected product from methylation of a C18 unsaturated fatty acid.

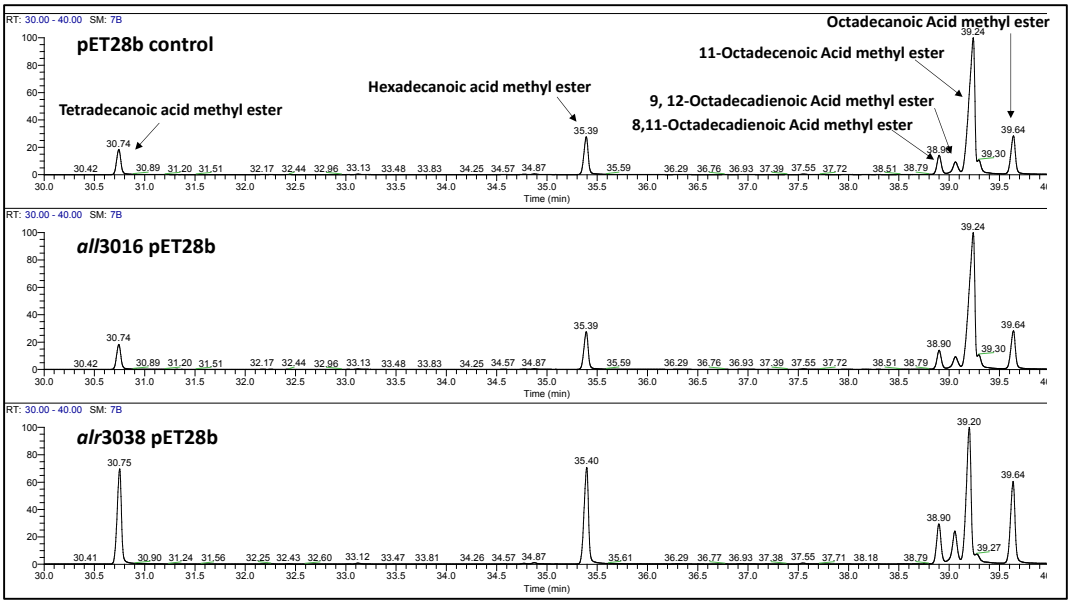


Figure 4A9. Chromatograms of FAME analysis for *in vitro* analyses of all3016 pET28b and all3016 pET28b.

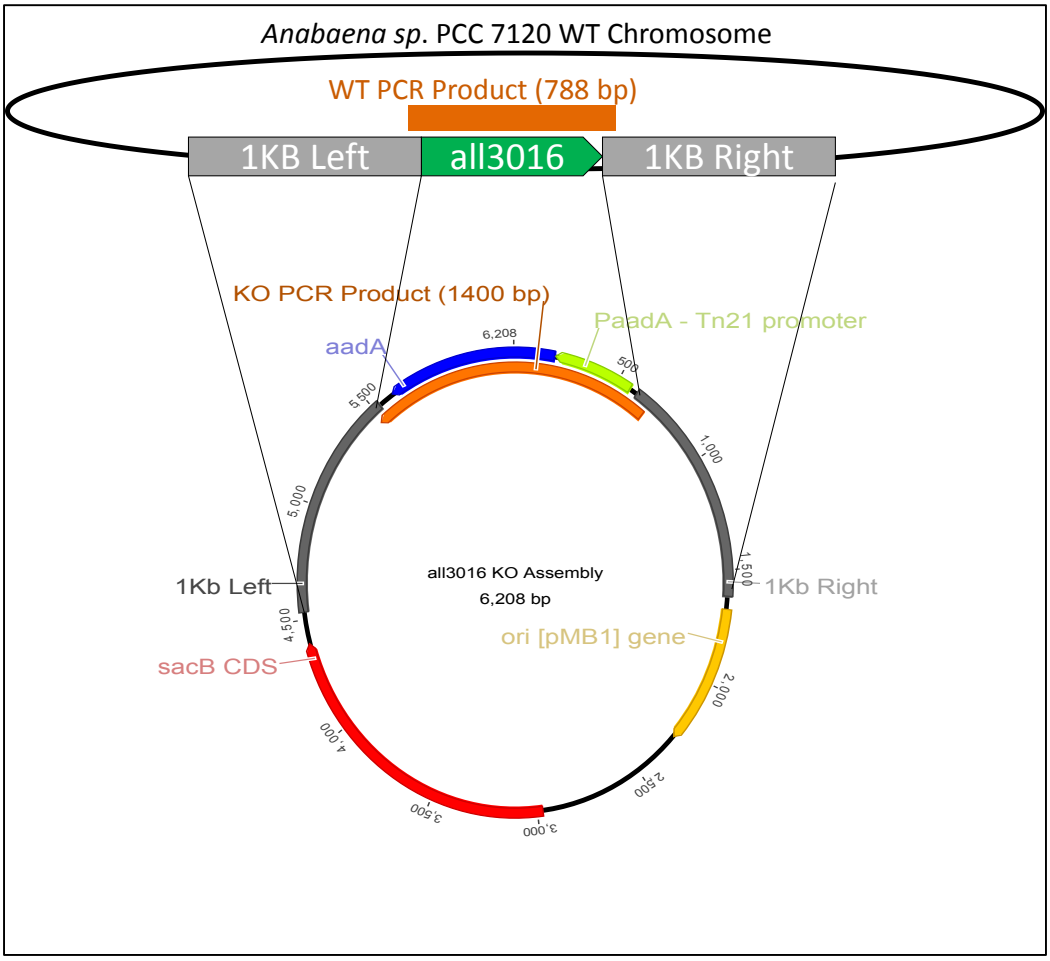


Figure 4A10. Graphical display of vector used to knockout all3016 via homologous recombination. The WT *Anabaena* sp. PCC7120 chromosome is displayed above showing the region 1kb to the left and right of all3016. The orange rectangle above all3016 in the WT region and below the *aadA* (SpSm resistance) gene below represents the PCR products that were used to screen mutant strains.

4.7 Acknowledgements

Chapter 4 in full is a manuscript that is in preparation. My co-authors are Arnaud Taton, Jonathan Tram, James Golden, Lena Gerwick, William H. Gerwick.

4.7 References

Balaban AT, 1983. Topological indices based on topological distances in molecular graphs. Pure Appl Chem 55. 199–206.

Blomquist GJ, Bagneres AG, editors 2010. Insect hydrocarbons: biology, biochemistry, and chemical ecology. New York: Cambridge University Press.

Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH, 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. Gene. 296(1-2). 235-247.

Chang Z, Sitachitta N, Rossi JV, Robers MA, Flatt PM, Jia J, Sherman DH, Gerwick WH, 2004. Biosynthetic pathway and gene cluster analysis of Curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula J. Nat. Prod. 67 (8). 1356–1367

Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman DH, Allen EA, Gerwick L, Gerwick WH, 2014. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. PLoS One 9(1). e85140.

Edgar RC, 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32 (5). 1792–97.

Edwards DJ, Marquez BL, Nogel LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH, 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula. Chem Biol. 11(6). 817-833

Edwards DJ, Gerwick WH, 2004. Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. J Am Chem Soc. 2004 126(37). 11432-11433.

Elhai J, Vepritskiy A, Muro-Pastor AM, Flores E, Wolk CP, 1997. Reduction of conjugal transfer efficiency by three restriction activities of *Anabaena sp*. strain PCC7120. J. Bacteriol. 179. 1998–2005.

Elhai J, Taton A, Massar JP, Myers JK, Travers M, Casey J, Slupesky M, Shrager J, 2009. BioBIKE A web-based, programmable, integrated biological knowledge base. Nucl. Acids Res. 37(2). W28-W32.

Fehler, WG, Light RJ, 1972. Biosynthesis of Methylheptadecanes in *Anabaena variabilis*. *In Vitro* Incorporation of S- (*methyl*-^{14}C) Adenosylmethionine. Biochemistry. 11(13). 2411-2416.

Fehler, WG, Light RJ, 1970. Biosynthesis of Hydrocarbons in *Anabaena variabilis*. Incorporation of (*methyl*- ^{14}C) - and (*methyl*-$^{2}H_3$) Methionine into 7- and 8-Methylheptadecanes. Biochemistry. 9(2). 418-422.

Giglio S, Chou WKW, Ikeda H, Cane DE, Monis PT, 2010. Biosynthesis of 2-Methylisoborneol in cyanobacteria. Environ. Sci. Technol. 45 (3) 992–998.

Goodloe, R.S., R.L. Light, 1982. Biosynthesis of Hydrocarbon in *Anabaena variabilis.* In-vivo incorporation of (18-^{14}C) stearate. Bio. Bioph. Acta. 711, 261-265.

Grogan DW, Cronan Jr. JE, 1997. Cycloproane ring formation in membrane lipids of bacteria. *MMBR* 61(4). 429-441.

Han J, McCarthy ED, Van Hoeven W, Calvin M, Bradley WH, 1968. Organic geochemical studies, II. A preliminary report on the distribution of aliphatic hydrocarbons in algae, in bacteria, and in a recent lake sediment. PNAS 59. 29–33.

Han J, Chan HWS, Calvin M, 1969. Biosynthesis of alkanes in *Nostoc muscorum*. *J.* Amer. Chem. Soc. 91(18). 5156-9.

Kopycki, JG, Stubbs MT, Brandt W, Hagemann M, Porzel A, Schmidt J, Schliemann W, Zenk MH, Vogt T, 2008. Functional and structural characterization of a cation-dependant O-methyltransferase from the cyanobacterium *Synechocystis* sp. PCC6803. J. Biol. Chem. 283 (30) 20888-20896.

Koonin EV, Rudd KE, 1993. SpoU protein of *Escherichia coli* belongs to a new family of putative rRNA methylases. Nucleic Acids Res. 21(23). 5519.

Ladygina, N., E.G. Dedyukhina and M.B. Vainshtein, 2006. A review on microbial synthesis of hydrocarbons. Process Biochemistry. 41(5) 1001-1014.

Liu A, Zhu T, Lu X, Song L, 2013. Hydrocarbon profiles and phylogenetic analyses of diversified cyanobacterial species. Appl Energy 111. 383-393.

Lu WD, Chi ZM, Su CD, 2006. Identification of glycine betaine as compatible solute in *Synechococcus* sp. WH8102 and characterization of its N-methyltransferase genes involved in betaine synthesis. Arch Microbiol 186(6).495-506

Markowitz VM, Chen MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Verghese N, Mavromatis K, PAti A, Ivanova N, Krypides NC, 2014. IMG 4 version of the integrated microbial genomes comparative analysis system, Nucl. Acids Res. 42, D560-D567.

Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK, Abelson D, Moore RE, Sherman DH, 2006 Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. potent anticancer agents from Nostoc cyanobionts. ACS Chem. Biol. 1 (12) 766–779.

Meena LS, Chopra P, Vishwakarma RA, Singh Y, 2013. Biochemical characterization of an S-adenosyl-l-methionine-dependent methyltransferase (Rv0469) of *Mycobacterium tuberculosis.* Biol. Chem 394(7). 871-877

Qiu Y, Tittige C, Wicker-Thomas C, Le Goff G, Young S, Wajnberg E, Fricaux T, Taquet N, Blomquist GJ, Feyereisen R, 2012. An insect-specific P450 oxidative decarbonylase forcuticular hydrocarbon biosynthesis. PNAS 109:14858–14863.

Ramaswamy AV, Sorrels CM, Gerwick WH, 2007. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. J. Nat Prod. 70 (12). 1977–1987.

Schirmer A, Rude MA, Li X, Popova E, del Cardayre S, 2010. Microbial biosynthesis of alkanes. Science 329: 559–562.

Scharnagl M, Richter S, Hagemann M, 1998. The cyanobacterium *Synechocystis* sp. strain PCC 6803 expresses a DNA methyltransferase specific for the recognition sequence of the restriction endonuclease PvuI. J. Bacteriol. 180(16). 4116-4122.

Shintani DK, Cheng Z, DellaPenna D, 2002. The role of 2-methyl-6-phytylbenzoquinone methyltransferase in determining tocopherol composition in *Synechocystis* sp. PCC6803. FEBS Letters 511. 1-5.

Tillet D, Dittmann E, Erhard M, von Döhren, Börner T, Neilan BA, 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system Chem. Biol. 7(10). 753-64.

Ternes P, Sperling P, Albrecht S, Franke S, Cregg JM, Warnecke D, Heinz E, 2006. Identification of fungal sphingolipid C9 –methyltransferase by phylogentic profiling. JBC. 281 (9). 5582-5592.

Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK, 2010 Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. PNAS 107(19). 8537-8542.

Chapter 5: Conclusions from Investigations of Cyanobacterial Hydrocarbon Pathway Mechanisms and Evolutionary History

The fields of natural product chemistry, algae and cyanobacterial biotechnology, and synthetic biology represent a unique confluence of scientific disciplines that have contributed to the development of many societal benefits including new drug therapies, environmental remediation and monitoring technology as well as new opportunities for sustainable production of fuels and chemicals. Cyanobacteria have played a major role in the advancement of all of these fields due to their deep resource of metabolic diversity and genetically tractable model strains. The observation that cyanobacteria produce hydrocarbons naturally and that this phenomenon is ubiquitous among cyanobacteria has implications for all three of these disciplines. Developing a deeper understanding of cyanobacterial hydrocarbon biosynthesis will have implications for understanding novel biosynthetic mechanisms as well as cyanobacterial evolutionary history and physiology.

This research established a number of fundamental insights about cyanobacterial hydrocarbon biosynthesis including a deeper perspective on the evolutionary history of these pathways as well as the characterization of novel hydrocarbon pathway variability. By establishing that hydrocarbon production is a ubiquitous phenomenon among cyanobacteria this work identifies a potentially fundamental insight into the physiology or ecology of cyanobacteria. Through establishing that all genome sequenced cyanobacteria possess one of the two cyanobacterial hydrocarbon biosynthetic pathways, this investigation has leveraged the vast array of genome sequence data available for cyanobacteria to more firmly establish what had been previously a limited observation built upon chemical compositional analyses. Prior to this investigation the hydrocarbon

composition for a wide variety of cyanobacteria had been characterized but the composition of other unexplored cyanobacteria could not logically be inferred. Prior to this thesis insights about newly discovered biosynthetic pathways could not be coupled to compositional analyses and phylogenetic relationships. Observing that the evolutionary history of the narrowly distributed OLS pathway is distinct from that of the more common FAAR/ADO pathway revealed that horizontal gene transfer has been involved in the acquisition of the OLS pathway in some cyanobacteria. However, the evolutionary history of the OLS pathways has not been thoroughly explained. The involvement of horizontal gene transfer in the original acquisition or neofunctionalization of the OLS pathway cannot be fully established without additional sequencing of underrepresented cyanobacterial clades coupled with an extensive phylogenomic comparison of cyanobacterial evolutionary history. Additionally, the observation that gene synteny in and around the FAAR/ADO pathway for a number of cyanobacterial strains has been disrupted by the OLS pathway provides an insight into the possible mechanism through which strains containing the FAAR/ADO pathway might lose this pathway and acquire the OLS pathway. A more extensive investigation of the evolutionary history of these pathways and the role of horizontal gene transfer is warranted.

The observation that hydrocarbon production is ubiquitous among cyanobacteria suggests that there is a selective pressure to maintain the phenotype of hydrocarbon production. Additionally, this unknown selective pressure has somehow maintained only one of the two pathways and never both. An investigation of the physiological function of hydrocarbons in cyanobacteria would provide a fundamental insight into the selective pressure that has maintained these pathways in cyanobacteria

over their long evolutionary history. Considering that photosynthesis is a common trait among cyanobacteria, it may be in some way involved in the selective pressure that is exerted upon cyanobacteria to maintain this metabolic capacity. A variety of possible physiological roles for hydrocarbons could be considered and are discussed more extensively in chapter 1; however, the influence of the presence of hydrocarbons on photosynthesis in cyanobacteria is reasonably testable. This investigation established that the two different hydrocarbon biosynthetic pathways produce different and somewhat predictable hydrocarbon compositions. The physiological impact of these different compositions has yet to be explored. Using strains with their respective hydrocarbon pathways knocked out, the resulting phenotype could be characterized for such parameters as growth rate, photosynthetic rate, as well as cellular ultrastructure using electron microscopy. The two knockout strains could also be used as templates for expression of the alternative hydrocarbon pathway that each respective strain does not naturally possess, thereby providing a means for testing the physiological impact of each respective pathway. These various investigations would establish a much deeper understanding of the physiological role of hydrocarbons in cyanobacteria. Additionally, an investigation of the physical location of hydrocarbons in a cyanobacterial cell may reveal insights about their physiological function. A characterization of the hydrocarbon composition of preparations of different components of the cell derived via differential centrifugation, including thylakoid membranes, may reveal insights about the localization of hydrocarbons and their relationship to physiological function.

Characterizing the hydrocarbon composition of a broad range of cyanobacteria also identified the presence of unique alkenes in marine filamentous cyanobacteria of the

genus *Moorea*. These strains seem to produce alkenes with internal double bonds at the two, three, five and seven positions. Interestingly these double bond positions are not consistent with the known mechanism for the OLS pathway (this is the pathway present in *Moorea* strains by genetic analyses). This finding suggests that an additional modification of the terminal alkene products from the OLS pathway is likely involved in producing the resulting internal double bonds. We propose that either an isomerase or desaturase is responsible for the double positions that are observed. This proposal could be investigated using a ^{13}C-labeled fatty acid at the C1 position as a substrate for a feeding study. If the resulting alkenes can be separated, then the location of the ^{13}C in the alkene would reveal if a desaturase or an isomerase is involved in the biosynthesis of these unique alkenes.

This investigation also established that branched hydrocarbon biosynthesis is a specialized type of hydrocarbon production that is found in a small subset of cyanobacteria, mostly comprised of heterocystous (subdivision IV) cyanobacteria. This observation prompted an investigation into branched hydrocarbon biosynthesis in these organisms. Through bioinformatic analysis, heterologous expression in *E. coli* to perform *in vivo* and *in vitro* analyses as well as mutagenesis of one of three target methyltransferases (all3016), a few promising target methyltransferases remain but further work is required. Additional mutagenesis of two (all2121 and all3018) of the three methyltransferases may reveal if one of these target methyltransferases is solely responsible for the production of branched fatty acids that provide substrates for branched hydrocarbon biosynthesis. Given the homology between all2121 and the

methyltransferases responsible for tuberculosteric acid and cyclopropane fatty acid biosynthesis, additional *in vivo* and *in vitro* characterization of all2121 is also warranted.

Identification of the methyltransferase responsible for branched hydrocarbon biosynthesis may provide a valuable enzyme for engineering highly branched molecules that could be used for a variety of fuel applications or commodity chemicals. Similarly, a thorough investigation of the evolutionary history, physiological function and pathway variability could provide new insights that can be applied to develop higher quality sustainable fuels with decreased processing requirements and cheaper development needs. Scientific advancements, including next generation genome sequencing and editing technology as well as increased sensitivity in mass spectrometry and high powered microscopy will greatly enhance future research into cyanobacterial hydrocarbon biosynthesis. These advancements will significantly increase the chances that insights about cyanobacterial hydrocarbon biosynthesis can be applied to benefit human societies and improve how humans sustainably interact with the environment.