12-1-2010

Expression of the ARTEMISIA ANNUA β-CARYOPHYLLENE synthase in Synechocystis sp. strain PCC6803 results in cyanobacterial sesquiterpene accumulation

Robert Eric Reinsvold
University of Northern Colorado

University of Northern Colorado

Greeley, Colorado

The Graduate School

EXPRESSION OF THE *ARTEMISIA ANNUA* B-CARYOPHYLLENE SYNTHASE IN *SYNECHOCYSTIS SP*. STRAIN PCC6803 RESULTS IN CYANOBACTERIAL SESQUITERPENE ACCUMULATION

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Robert Eric Reinsvold

College of Natural and Health Science

School of Biological Sciences

Biology: Thesis

December, 2010

This Thesis by: Robert Eric Reinsvold

Entitled: *Expression of the Artemisia annua β-caryophyllene synthase in Synechocystis sp. strain PCC6803 results in cyanobacterial sesquiterpene accumulation*

has been approved as meeting the requirement for the Degree of Master of Science, from the College of Natural and Health Sciences, in the School of Biological Sciences.

Accepted by this Thesis Committee:

_____

Dr. Chhandak Basu, Chair

_____

Dr. Susan Keenan, Committee Member

_____

Dr. Richard M. Hyslop, Committee Member

Accepted by the Graduate School

_____

Robbyn R. Wacker, Ph. D.
Assistant Vice President for Research
Dean of the Graduate School & International Admissions

ABSTRACT

Reinsvold, Robert Eric. *Expression of the Artemisia annua β-Caryophyllene Synthase in, Synechocystis sp.* strain PCC6803 *Results in Cyanobacterial Sesquiterpene Accumulation*, Published Masters of Science thesis, University of Northern Colorado 2010.

The sesquiterpene β-caryophyllene is a ubiquitous component of many plant resins and has been used in the cosmetic industry to provide a woody, spicy aroma. Recent clinical studies have shown its potential effectiveness as an antibiotic, anesthetic, and anti-inflammatory agent. Additionally, there is significant interest in sesquiterpene accumulation in phototrophic microorganisms for the production of biofuels. Currently, the isolation of β-caryophyllene relies on purification methods from oleoresins extracted from large amounts of plant material. The design of a cyanobacterium platform that produces β-caryophyllene may provide a more sustainable and controllable means of production, allowing for an increase in yield without the potential deleterious effects to natural sources. To this end, the β-caryophyllene synthase gene (*QHS1*) from *Artemisia annua* (Chinese wormwood) was stably inserted via double homologous recombination into the genome of the cyanobacterium *Synechocystis sp.* strain PCC6803, under the control of the strong endogenous *psbAII* promoter. Gene insertion into *Synechocystis* was confirmed through polymerase chain reaction assays and sequencing reactions. Transcription and expression of *QHS1* was confirmed using reverse transcription PCR assays and production of β-caryophyllene was confirmed in the transgenic strain using gas chromatography with flame ionization detection and mass-spectrometry detection.

ACKNOWLEDGEMENTS

The success of this thesis has been possible because of the direct and indirect help of so many people. I wanted to begin by thanking my wife for her continued and relentless support for everything I pursue. Her patience and understanding during this process is even more impressive and endearing, since she has been carrying our first child during the most stressful portion of this thesis project. I want to thank my parents, whose advice continues to guide me as I navigate through new endeavors. I want to thank my father for his editorial help and his life-long effort in preparing me for life in the sciences.

I owe a large debt of gratitude to Dr. Matthew Posewitz and his lab. Without the expert efforts of my collaborators, Robert Jinkerson and Dr. Randor Radakovits, this project would still be floundering with nothing but genetic data. Everyone in the Basu lab deserves my gratitude. Specifically, I want to thank Brenda Thornton for her help and guidance with many of the difficult genetic techniques used in this work. I also would like to thank Sam Zwenger for his enlightening conversations and his continual willingness to act as a sounding board throughout my time at UNC. I want to thank my friend Anthony Endres for his help during the editorial process.

I want to thank my committee members, Dr. Susan Keenan and Dr. Richard M. Hyslop, for their help in crafting this project, keeping me on task, and ensuring that I did quality science. Finally, I want to thank my advisor, Dr. Chhandak Basu, who has provided me with the resources and freedom to design and pursue a project that I was passionate about.

TABLE OF CONTENTS

CHAPTER Page #

I. INTRODUCTION……..……1

II. LITERATURE REVIEW…….……....6

Profile of the cyanobacterium *Synechocystis sp.* strain PCC6803

Profile of *Synechocystis sp.* strain PCC6803 isoprene production

Synechocystis sp. strain PCC6803 transgenic background

Synechocystis sp. strain PCC6803 commercial importance and potential

β-caryophyllene overview

III. MATERIALS AND METHODS……...……..…….……40

Strain and growth conditions

pPSBAIIQHS1 vector design

Escherichia coli transformation

Synechocystis sp. strain PCC6803 double homologous recombination reaction

Reverse transcriptase polymerase chain reaction

Extraction and quantification of β-caryophyllene

IV. RESULTS……..……...…….51

Construction of pPSBAIIQHS1 vector

Transformation and segregation of a transgenic *Synechocystis* strain

Confirmation of QHS1 gene transcription

Production of β-caryophyllene by a transgenic *Synechocystis* strain

Genetic analysis of potential farnesyl diphosphate synthases

V. DISCUSSION AND CONCLUSION……..……66

REFERENCES……..…..……...73

APPENDIX A: EXPERIMENTAL SEQUENCE ANALYSIS……...82

APPENDIX B: *SYNECHOCYSTIS SP*. STRAIN PCC6803
GENE SEQUENCE ANALYSIS……..……..……..……..91

APPENDIX C: *ARTEMISIA ANNUA QHS1* SEQUENCE AND CODON
COMPARISON TO *SYNECHOCYSTIS* SP. STRAIN PCC6803……...96

TABLE OF FIGURES AND TABLES

FIGURE # AND TITLE	PAGE #
Figure 1. The MEP pathway…….	13
Figure 2. The HDR enzymatic mechanism……...	16
Figure 3. Modified MEP Pathway……	18
Figure 4. Proposed mechanism for β-caryophyllene synthesis……	31
Figure 5. Chemical structure of common caryophyllene-like sesquiterpenes…..	32
Figure 6. Restriction digest of the pPSBAIIKS vector…….	42
Figure 7. BG-11-sucrose-agar plate with successful *QHS1* transformant……	46
Figure 8. Graphical representation of the double homologous recombination strategy……..……	47
Figure 9. β-caryophyllene concentration standard curve……..	50
Figure 10. Agarose gel image of the *E. coli* colony PCR screening assays……..	53
Figure 11. Segregation of mutant strains and conformation of *QHS1* insertion…….	55
Figure 12. Reverse transcriptase PCR gel images……	58
Figure 13. Gas chromatograms of the experimental extractions……..	60
Figure 14. Mass spectral comparisons……..	62
Figure 15. Detailed gas chromatogram from the *QHS1*strain……..	63

FIGURE # AND TITLE — Page #

Figure 16. The primary metabolic networks of *Synechocystis* sp. strain PCC 6803…….. 70

Figure 17: Plasmid pPSBAII*QHS1*……..…… 83

TABLE # AND TITLE — PAGE #

Table 1. PCR primer sequences and locations……..…… 43

Table 2. Standard curve data……. 50

CHAPTER I

INTRODUCTION

The broad definition of algae, as generally agreed upon by modern phycologists, is a free-living, aquatic, oxygenic, photo-autrophic bacteria or protozoa (1). This inclusive definition includes disparate, morphologically distinct groups such as the large coenocyctic brown algae with their leaf-like blades, the small unicellular, spherical cyanobacterium *Synechococcus*, and the ornate, armored diatoms. Both macro- and microalgae serve as the primary producers in almost all marine environments; it is estimated that oceanic algae accounts for just under half of the total 'global net primary production,' or the conversion of carbon dioxide into organic compounds (reviewed in 1). The ecological importance of algae in both a modern and natural historic context should not be overlooked. The increase in atmospheric oxygen billions of years ago and the resulting explosion of aerobic eukaryotic diversity is due in large part to the emergence of photosynthetic oxygen-producing cyanobacteria 2.7 billion years ago (1). Eventually the increased levels of atmospheric oxygen resulted in the development of an ozone layer, which allowed life to colonize land without the impediment of UV damage (1).

The size and scope of the algal presence in aquatic environments are often underappreciated by humans until massive algal blooms occur as a result of nutrient rich run-off entering their habitats. These blooms can cause large swaths of 'dead zones' in which most marine life is suffocated due to low dissolved oxygen levels or poisoned from high concentrations of algal toxins (1). The looming specter of climate change makes

understanding algal community dynamics even more important, as they may play a role in increasing carbon sequestration as they have in the past. However, they could have the opposite effect and exacerbate an already precarious situation. The very carbon that is emitted from the tailpipes of cars today may have at one point been the carbon sequestered by algae millions, if not billions, of years ago. The ancient phytoplankton were able to sequester carbon dioxide in resistant structures, such as carbonates or large organic molecules that would have been deposited in ocean sediment upon the organism's death, in a process termed 'biological carbon pumping' (1). After millions of years of environmental fluctuations, this organic sediment would have been converted into what is commonly referred to as 'fossil fuels,' quietly waiting miles below sea level, for humans to realize this hidden energy potential.

Beyond this energetic link between humanity and ancient phytoplankton, humans have used living algal products for as long as they have been settling near large bodies of water. Algae's historical importance for marine societies and its continued use in modern applications make it a transcendent primary producer of a variety of essential products. Countries including England, China, Japan, and Chile have historically utilized algae as a source of food and continue this tradition today. The ever popular nori, used in modern sushi is prepared from dried red algae. In China, the communal cyanobacteria (*Nostoc flagelliforme)* are enjoyed as the delicacy 'Fat Choy.' Red and green algae are used in the traditional Welsh dish 'laver.' In South America, red algae from the genus *Callophyllis* are used as another food source in various coastal communities. Carrageenans, water-soluble sulfated galactans extracted from a variety of algal sources, are used as a thickening agent in a number of food products, including ice cream,

chocolate milk, and jellies (2). Algae produce a variety of carotenoid pigments, which are often used in the food and cosmetic industries to provide a natural color to many products (reviewed in 2). These pigments also provide an additional health benefit, as they have been shown to have significant amount of antioxidant capacity (reviewed in 2).

Algae are cultivated as feed for the farming of fish, crustaceans, and other marine invertebrates. The management of this autotrophic food source is critical to maintaining a viable aquaculture industry. Algae provide more than just the calories needed for growth in aquaculture, but serve as the direct source for many of the desired attributes sought in the final animal product (3). Many pigments derived directly from algae, such as astaxanthin and β-carotene, provide high quality fish with their distinct coloration and increase nutritional value (2). Similarly, the beneficial ω-3-polyunsaturated fatty acids found in fish, such as eicosapentaenoic acid and docosahexaenoic acid, are obtained by the fish from algal sources (2, 3).

Algae have also played an important role in conventional agriculture. Brown algae or seaweed have long been used as a fertilizer by many coastal farmers and nitrogen fixing cyanobacteria are often used in rice fields to provide a steady supply of accessible nitrogen, without the need for fertilizers (4, 5).

More recently, algal products, including agar (sulfated galactans from red algae) and lectins (carbohydrate-binding proteins) have become indispensible tools in modern molecular biology and microbiology labs (2). Agar provides an excellent growth matrix for organisms ranging from single cell prokaryotes to eukaryotic plant seeds. Algae produce a diversity of different lectins that can interact with specific membrane surface

carbohydrates, allowing researchers to use them for immunological and histological experiments and diagnostics tests (2).

The promise of producing commercially important products, ranging from medicines to fuels, using only sunlight and waste water is quite an interesting, yet distant prospect. However, considerable research has been devoted to optimizing and engineering algal systems for such duties. Early genomic work has shed light into the genetic underpinnings of various biochemical pathways in a variety of unique species. Bio-prospecting of ecosystems throughout the world has turned up a number of novel algal species with the endogenous capacity to produce these commercially important compounds *de novo* and without the intervention of modern bioengineering (reviewed in 6).

Currently, there are over a dozen completed algal genome projects with even more in development, providing researchers with a growing body of genomic data for further investigation (reviewed in 7). There are even more species of algae that have been successfully transformed with foreign genes using a variety of transformation techniques ranging from electroporation, sonication, glass bead bombardment, *Escherichia coli* and *Agrobacterium* conjugation, and biolistics (7-10). Various strains of algae have been engineered to successfully produce human proteins and antibodies and some researchers view them as a perfect platform for vaccine production (11-13).

As discussed earlier, algae are often used for its antioxidant pigments and strains have been developed to produce higher concentrations of these endogenous compounds (14, 15). Transgenic algal strains are being used as feed supplements in aquaculture in order to increase the growth rate and quality of fish (16). Researchers are also seeking to

capture the energy potential of an already proven biological system using transgenic algae. Pathways for producing fuel products such as ethanol, lipids, isoprene, and hydrogen gas have all been engineered into various algal systems (reviewed in 7). A growing understanding of the interacting metabolic pathways within a biological system (systems biology) within various algae species is further expanding researchers' tools to produce these high value products (17, 18). There are still a number of hurdles facing commercial-scale algal compound production ranging from high water demands, energy intensive extraction techniques, and the potential release of invasive genetically-modified algae into the natural ecosystem. Despite these challenges, algal transgenic research continues unabated. In this project, a strain of the cyanobacterium, *Synechocystis sp.* strain PCC6803, was developed that produces the medicinally and commercially important sesquiterpene, β-caryophyllene.

CHAPTER II

LITERATURE REVIEW

A. Profile of the cyanobacterium Synechocystis sp. strain PCC6803

The cyanobacterium *Synechocystis* sp. strain PCC6803 is a Gram-negative, unicellular, photoheterotrophic prokaryote that has become a ubiquitous research strain. The suffix 'sp. strain PCC6803' refers to the original isolate, obtained in 1968 from a fresh water source in California by Kunisawa (19). This particular culture has since been maintained by the Pasteur Culture Collection (Paris, France)[1]. Previously used names for this strain include ATCC 27184, IAM M-203, IAM M-208, IPPAS B-288, NIBB 1125, and PPMSU 63, and it has been previously grouped within the genus *Aphanocapsa* (19). The segregation of *Synechocystis* into a distinct genus and the original nomenclature dates back to work done by Sauvageau in 1892 (19).

Synechocystis was originally differentiated from similar cyanobacterial genera based on its unicellular nature, distinct from the genus *Gleocapsa* that often forms colonies (19). Its ability to undergo binary fission in multiple planes at right angles to one another has distinguishes it from *Synechococcous*, which undergoes a radial cleavage separation (19). Within the genera *Synechocystis*, there are two subgroups divided on the basis of the GC content of the genome, either low (35-37% GC) or high (42-48% GC).

[1] http://www.crbip.pasteur.fr/fiches/fichecata.jsp?crbip=PCC%206803, accessed on 5.20.10

Synechocystis sp. strain PCC6803 falls in the latter subgroup with a GC content of 47.7% (20). *Synechocystis* sp. strain PCC6803 is also unique within this higher GC content group because it has been shown to grow heterotrophically on glucose as well as having gliding/twitching motility (19). The unique motility of *Synechocystis* sp. strain seems to be solely dependent on the presence of large, thick pili structures on the polar ends of the cellular surface (21). This is distinct from the more fluid swimming motility seen in *Synechoococcus,* which is a result of smaller translocation proteins that cover the surface of the cell (22). Phycologists were able to distinguish *Synechocystis sp*. strain PCC6803 from other cyanobacterial genera without the aid of phylogenetic analysis using these physiological and morphological distinctions.

In 1996, *Synechocystis* sp. strain PCC 6803 became the first oxygenic photoautotroph to have its genome fully sequenced and published (20). At the time, there had been only two other microbial genomes published and both were relatively small (*Haemophilus influenza* at 1.83Mb and *Mycoplasma gemtalium* 0.58 Mb). Assisted by a previously sequenced 1 Mb region, researchers were able to fully assemble the 3.5 Mb genome using a 'bridging shotgun' approach in which two libraries were created for genome assembly: an element library (average size of 700 bp) and a bridge library (average size of 2.5 Kb). Physical mapping of the contigs and final genome construction were done using overlapping regions assembled from constructed shotgun cosmid and λ libraries. Gap filling and sequence confirmation was done using long PCR reactions and restriction mapping. The overall accuracy was predicted to be 99.99% (20). Open reading frame (ORF) analysis of this first version revealed 3,168 potential protein coding regions, a majority of which had no known function or orthologes at the time (20).

Interestingly, a number of gene characteristics originally thought to only occur in eukaryotes were found in this genome, including genes containing trytophan-aspartic acid-(WD)-repeats, which are commonly found in β-propeller domain motifs that are involved in reversible protein-protein interactions[2]. There is also a high abundance of two different repetitive sequences within the genome; insertion-like elements and highly iterated palindromic (HIPI) sequences (23). In a number of positions, the insertion-like elements appear to be either coding for transposase proteins or disrupting ORFs. There also appear to be a number of homologous recombination events that occurred due to these insertion-like elements, given the conservation of these sequences dispersed throughout the genome (23). The HIPI sequences (GCGATCGC) occur regularly throughout the genome (once per every 1131 bp) and 90% of these sequences appear to fall in coding regions (23). Recently, an updated version of the *Synechocystis sp.* strain PCC6803 genome has been released by the Kazusa DNA Research Institute (24). This updated version increased the genome size to 3.9 Mb and number of genes to 3725, due to the inclusion of 6 unique plasmids, which code for 408 genes (24).

The publication of the *Synechocystis sp.* strain PCC6803 genome and has made this strain a popular organism with researchers investigating photosynthesis, evolution, molecular metabolism, and transgenic expression (reviewed by 1). With Lynn Sagan's revival of the 'Endosymbiotic Theory' in 1967, cyanobacteria have been considered the modern progeny of the proto-oxygenic phototrophic prokaryotes that gave rise to the chloroplast in plants and eukaryotic algae (25). Evidence for this evolutionary relationship was based on the homology between the cyanobacterial cell membranes and

[2] http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=eurekah&part=A76837#top, accessed on 5.21.10

the plastidial membranes, the presence of an indispensable, highly-conserved plastidial plasmid with genetic orthology to cyanobacterial genes, and the overall similarity in both photosynthetic pathways (25, 26). The publication of the *Synechocystis sp.* strain PCC6803 genome further established this relationship. Analysis of the tRNA genes in *Synechocystis sp.* strain PCC6803 showed a high degree of orthology to tRNA genes in plant chloroplast plasmids (23). A long list of genes encoding for photosynthetic-related proteins in *Synechocystis sp.* strain PCC6803 had orthologes in plant and algae genomes, both nuclear and plastidial (23). Beyond photosynthesis, eukaryotic plastidial genes involved in exogenous sulfate incorporation, ribosomal protein construction, and ATPase assembly have been found in the *Synechocystis sp.* strain PCC6803 genome, providing evidence for ancestral relationships (27). However, there were a number of genes crucial for photosynthesis in eukaryotes, including a number of small subunit proteins involved in electron transport in photosystem I (PSI) and photosystem II (PSII), that are absent in the *Synechocystis sp.* strain PCC6803 genome (23). These evolutionary links and gaps provide a rich source for comparative genomic research and the investigation into similar yet contrasting metabolic strategies. The synthesis of terpene precursors in *Synechocystis sp.* strain PCC6803 is one such contrasting, yet similar, metabolic pathway that has generated a considerable amount of interest and research.

Profile of Synechocystis sp. strain PCC6803 *terpenoid production*

Terpenes (or terpenoids), as a group, represent over 40,000 different chemical compounds (28). The group includes a diverse collection of chemical structures using 2-methylbutene subunits (commonly referred to as isoprene units) as the primary building blocks. The structures range from saturated long chain hydrocarbons (C_{40} and greater),

alcohols, ethers, esters, and large aromatic structures (29). The plant kingdom produces the greatest amount and diversity of terpenes. In plants, these metabolites serve in a variety of capacities, such as volatile emissions that attract pollinators, toxic or bitter tasting chemicals that deter herbivores, pigments for light harvesting and UV protection, and internal signal factors that help in growth and reproduction regulation (29, 30). Animals also produce terpenes, either *de novo* or from digested plant terpenes, which are used as pheromones, needed for inter-organismal communication, or as sterols, needed for membrane fluidity and internal cell signaling (29).

Synechocystis sp. strain PCC6803 produces a host of different terpenes and terpene-derived constituents including among others; chlorophyll A, phytol, ubiquinone, β-carotene, zeaxanthin, echinenone, myxo-xanthophyll, and other carotenoid gylcosides (31). Carotenoid glycosides, distinguished by a unique linkage of an isoprene tail to aglucoside, are found exclusively in cyanobacteria (31). These terpenoid constituents serve a variety of functions in *Synechocystis sp.* strain PCC6803, including light absorption, oxidative protection within the cell, and membrane fluidity in the cell membrane, thylakoid, and liposome (31, 32). The long, unsaturated carbon chains of carotenoids contain π-electron conjugation within its double bond network, providing a reservoir for reactive electrons produced in excited chlorophyll that would otherwise produce reactive oxygen species (33). Terpenoids are found throughout the *Synechocystis sp.* strain PCC6803 cell, including the outer cell membrane, the thylakoid, unbound in the cytoplasm, and with or without an associated protein (34, 32). Most of the genes for carotenogenesis have been identified and functionally confirmed within the

Synechocystis sp. strain PCC6803 genome (reviewed in 31). However, the physiological roles of some carotenoids have yet to be determined (31).

There are two distinct metabolic pathways for producing the precursor isoprene units: the mevalonate (MVA) and 2-methyl-D-erythritol-4-phosphate (MEP) pathways. Both pathways ultimately lead to the production of isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), but each route requires a completely different set of enzymes and utilizes different carbon sources from the beginning. The MVA pathway appears to be of archaebacterial origin, based on phylogenetic analysis of archaeabacterial genomes in which only MVA eukaryotic orthologes were found and no MEP genes were present (35). Along with archaebacteria, plants, animals, and fungi also utilize the MVA pathway, where it occurs in the cytosol and peroxisomes of these eukaryotes. However, many important plant terpenes (including pigments and growth factors) are derived using the MEP pathway localized in the plastids of the organism (30). Based on the high level of homology among MEP pathway enzymes found in plants and this cyanobacterium, it appears that *Synechocystis sp.* strain PCC6803 may have evolved from the same line of proto-cyanobacteria that originally developed into the modern plant chloroplast, further increasing the evidence of lateral nuclear gene transfer and endosymbiotic theory (35).

Most eubacteria utilize the MEP pathway exclusively for terpene production and *Synechocystis sp.* strain PCC6803 is no exception. It lacks any MVA orthologous genes and most of the plant genes involved in the MEP pathway have orthologoes in the *Synechocystis sp.* strain PCC6803 genome (36). Metabolic experiments in *Synechocystis sp.* strain PCC6803 have also shown that terpenes are exclusively produced from the

MEP pathway (37, 38). One such experiment used ^{13}C labeled glucose that was then fed to *Synechcocystis sp.* strain PCC6803, which allowed the researchers to track the carbon atoms when incorporated into the terpene phytol (37). When the phytol was extracted and the ^{13}C nuclear magnetic resonance spectrum was analyzed, it was clear that the only route the glucose carbons could have taken to the phytol was through the MEP pathway, based on the position of the carbons within the phytol backbone (37). Despite the strong evidence for the exclusivity of the MEP pathway within *Synechcocystis sp.* strain PCC6803, the specific pathway remains unclear and it appears to deviate from the linear chemical pathway determined within the *E. coli* system (Figure 1).

The difference in the MEP pathway between *Synechcocystis sp.* strain PCC6803 and other eubacteria appears to have arisen relatively recently in an evolutionary sense because, as stated earlier, almost all of the genes for the enzymes involved in the MEP pathway can be found in the genome of *Synechocystis sp.* strain PCC6803. Also, when expressed heterologously in *E. coli*, these genes produce functional proteins (38). There is one notable exception; *Synechocystis sp.* strain PCC6803appears to lack a Type I IPP isomerase. There are no orthologues or open reading frames to any known Type I IPP isomerases found within *Synechocystis sp.* strain PCC6803 genome and metabolic tests have failed to reveal any isomerase activity (39). However, there does appear to be a functional Type II isomerase, based on sequence analysis of the *Synechocystis sp.* strain PCC6803 genome, an ORF, sll1556, was identified that showed slight homology to the Type II isomerase found in *Salmonella enteric* (40). It appears that *Synechcocystis sp.* strain PCC6803 can function without this enzyme, but the enzyme provides an advantage in terpene production (41).

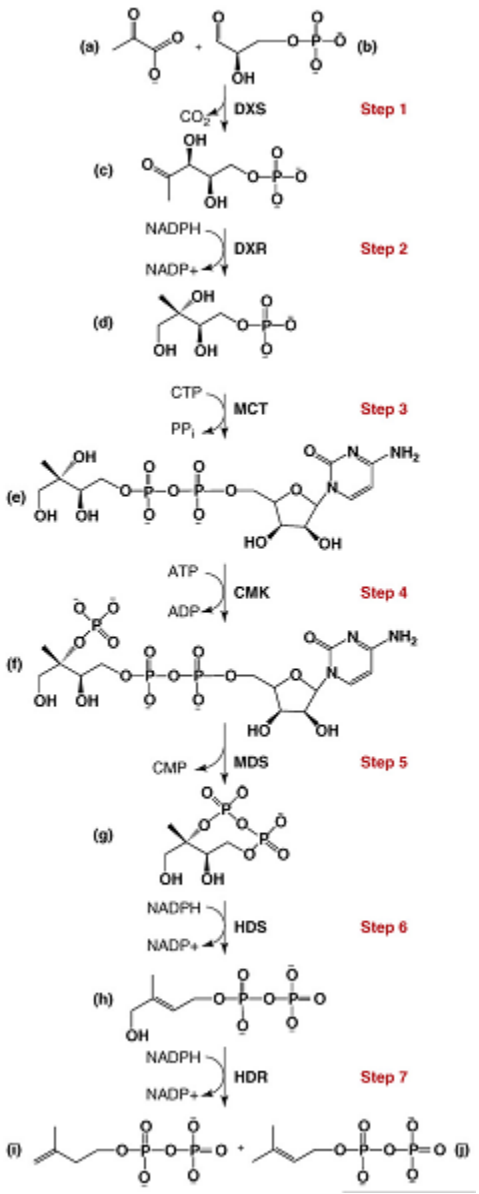

Figure 1. The MEP pathway. The MEP pathway as determined in *E. coli* with a proposed unified nomenclature. The initial condensation of pyruvate (a) and D-glyceraldehyde-3-phosphate (b) catalyzed by 1-deoxy-D-xylulose 5-phosphatesynthase (DXS) to form 1-deoxy-D-xylulose-5-phosphate (c). The reductive isomerization by 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) yields a 2-methyl-D-erythritol-4-phosphate (d). A cytidyl moiety is introduced via a diphospho bridge by 2-methyl-D-erythritol-4-phosphate cytidylyltransferase (MCT) to produce 4-(cytidine-5-diphospho)-2- methyl-D-erythritol (e). This is then phosphorylated (f) by 4-(cytidine-5-diphospho)-2--methyl-D-erythritol kinase (CMK) and then cyclized to yield 2- methyl-D-erythritol-2,4-cyclodiphosphate (g) in a reaction catalyzed by 2-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS). The final two steps catalyzed by (E)-4-hydroxy-3-methylbut-2-enyl diphosphate (h) synthase (HDS) and reductase (HDR) respectively, yield isopentyl diphosphate (i) and dimethylallyl diphosphate (j). Image and legend originally published in (42).

In most cellular systems, DMAPP serves as the priming isoprene unit in downstream biosynthesis reactions but IPP is vital for continual isoprenoid chain elongation (39). Ershov *et al*. also preformed metabolic testing that revealed a clear absence of IPP isomerase activity (39). Soluble cell extracts of *Synechocystis sp.* strain PCC6803 were unable to effectively incorporate isotopically-labeled [1-^{14}C] IPP into allylic diphosphates (39). This result indicates a lack of isomerase activity because there was no labeled allylic diphosphate production, which would require both an IPP and DMAPP to form, suggesting no DMAPP was produced (39). Labeled IPP was only incorporated when the extracts were supplied with exogenous DMAPP, further indicating a lack of IPP isomerase activity (39). Even when a transgenic IPP isomerase is inserted into the genome of *Synechocystis sp.* strain PCC6803, there appears to be no effect on terpene production (14). Researchers inserted the *ipi* gene from *Saccharomyces cerevisiae* into the genome of *Synechocystis sp.* strain PCC6803, using double homologous recombination (14). The resulting mutant strain showed no significant increase in carotenoid production, despite the presence of *ipi* transcripts and the isomerase proteins (14). Researchers in this study did not look for shorter chained terpenoids, so it is conceivable that any increase in DMAPP and IPP relating to the presence of the isomerase could have been shuttled and sequestered in smaller terpenoids (C_{10} or C_{20}) (14). However, similar studies in which this *ipi*^{sc} gene was inserted into *E. coli* showed a significant increase in carotenoid accumulation, indicating that the protein is functional in a prokaryotic system and the system can produce enough precursor to sufficiently supply downstream C_{40} terpene production (43). This evidence indicates the possibility of a separate pathway for DMAPP production in *Synechocystis sp.* strain

PCC6803 and perhaps even other eubacteria, since it has been shown that mutant *E .coli* with *ipi* removed still grew at similar rates as their wild-type counterparts (44).

In 2000, researchers identified a likely candidate gene, *lytB*, that may provide the enzymatic capacity to produce both DMAPP and IPP from the same molecule, hydroxy-2-methyl-2-butenyl-4-diphosphate (HMBPP) (45). The researchers identified *lytB* as a possible MEP pathway gene using comparative genomics across multiple domains, finding homologous copies of *lytB* in both plants and bacteria. They further investigated the enzyme's role in the MEP pathway using mutant *Synechocystis* strains in which the *lytB* promoter was disrupted. This resulted in light sensitive mutants that quickly bleached in low light, but normal physiological growth could occur in these mutants when supplemented with exogenous IPP and DMAPP analogs, indicating that the endogenous IPP/DMAPP production pathway was being inhibited via *lytB* promoter disruption (45). To further establish the role of *lytB* in the MEP pathway, the *Synechocystis sp.* strain PCC6803 gene for *lytB* was inserted into a lycopene producing *E .coli* strain, resulting in a much higher concentration of lycopene compared to the control *E .coli* strain (45). Similar results occurred when the *E .coli* strain was transformed with similar *lytB* genes from plants (45).

It appears that the 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR) enzyme catalyzed the formation of both IPP and DMAPP directly from HMBPP (46). X-ray crystallographic, structural analysis of the IspH protein from *E. col*i, a *lytB* orthologue, interacting with HMBPP revealed an iron/sulfur clustered reactive site, stabilized by a number of hydrogen bonding interactions (47). This permits the allylic

radical HMBPP intermediate to accept an electron at either C1, producing DMAPP, or C3, producing IPP (Figure 2) (47).

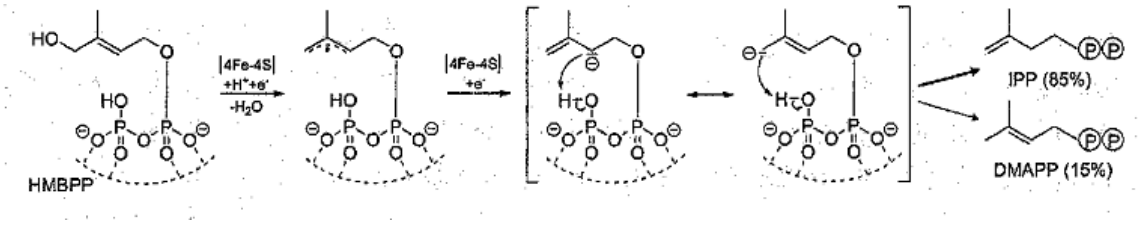

Figure 2. The HDR enzymatic mechanism. The proposed enzymatic mechanism for the conversion of HMBPP into either IPP or DMAPP by HDR. Image from (47)

Another interesting aspect of the *Synechocystis sp.* strain PCC6803 MEP MEP pathway is the identity of the primary carbon sources that feed the pathway. As seen in Figure 1, pyruvate (PYR) and D-glycealdehyde-3-phosphate (GA3P) are traditionally viewed as the starting materials in the MEP pathway for the production of 1-deoxy-D-xylose-5-phosphate (DXP). Researchers investigating a variety of suspected IPP and DMAPP precursors and their respective ability to drive overall isoprenoid production in *Synechocystis sp.* strain PCC6803, utilized [^{14}C] isotopically-labeled precursors to tract isoprenoid assembly (48). The results showed no significant increase in labeled isoprenoid formation when cell-free extracts were incubated with either of the labeled precursors PYR or DXP, however labeled GA3P significantly increased labeled isoprenoid formation (48). In similar *in vitro* experiments, labeled GA3P did significantly increase labeled isoprenoids (48). Other phosphorylated precursors, including a number of intermediates from the pentose phosphate cycle and Calvin-Benson cycle, such as ribose-5-phosphate and xylose-5-phosphate, increased isoprenoid production in *Synechocystis sp.* strain PCC6803 (48). Fosmidomycin, a chemical that inhibits the enzyme 1-deoxy-D-xylulose-5-phosphate reductoiosmerase (DXR)

completely inhibits growth in plants and *E. coli*, demonstrating the indispensability of DXR in terpene production and overall cell viability (49). However, when *Synechocystis sp.* strain PCC6803 was exposed to fosmidomycin, both *in vitro* and *in vivo*, no significant terpene synthesis inhibition was observed (48). All of these data suggest an alternative MEP pathway in *Synechocystis sp.* strain PCC6803, compared to the established pathways in *E. coli* and plant plastids (Figure 3). As suggested by Ershov *et al,* the MEP pathway may be fed with phosphorylated carbon substrates from photosynthetic dark reactions, bypassing the early substrates in the MEP pathway (48).

Along with this suspected interaction with the photosynthetic dark reactions, recent data suggest a potential interplay of the MEP pathway with photosynthetic light reactions, beyond just supplying the reducing capacity of NADPH and energetic potential within the phosphate bonds of ATP and CTP. The enzyme GcpE, found in most cellular systems that utilize an MEP pathway including *Synechocystis sp.* strain PCC6803, catalyzes the conversion of 2-methyl-D-erythritol-2,4-cyclodiphosphate (MEcPP) to 1-hydroxy-2-methyl-2-butenyl-4-diphosphate (HMBPP) through several single electron transfer events. Researchers have shown that the GcpE protein found in the cyanobacterium *Thermosynechococcus elongatus* BP-1 relies on ferrodoxin I, a protein directly involved in electron transfer between Photosystem I and $NADP^+$ reductase, as its electron donor (50). The GcpE protein in *T. elongatus* appears to be homologous to the IspG protein in the *Synechocystis sp*. strain PCC6803 genome, including the enzymatic reaction sites and the ferrodoxin protein interaction sites (50). These interactions of the MEP pathway with photosynthesis seem appropriate and even synergistic, given the reliance of proper photosynthetic function on terpenoid constituent pigments.

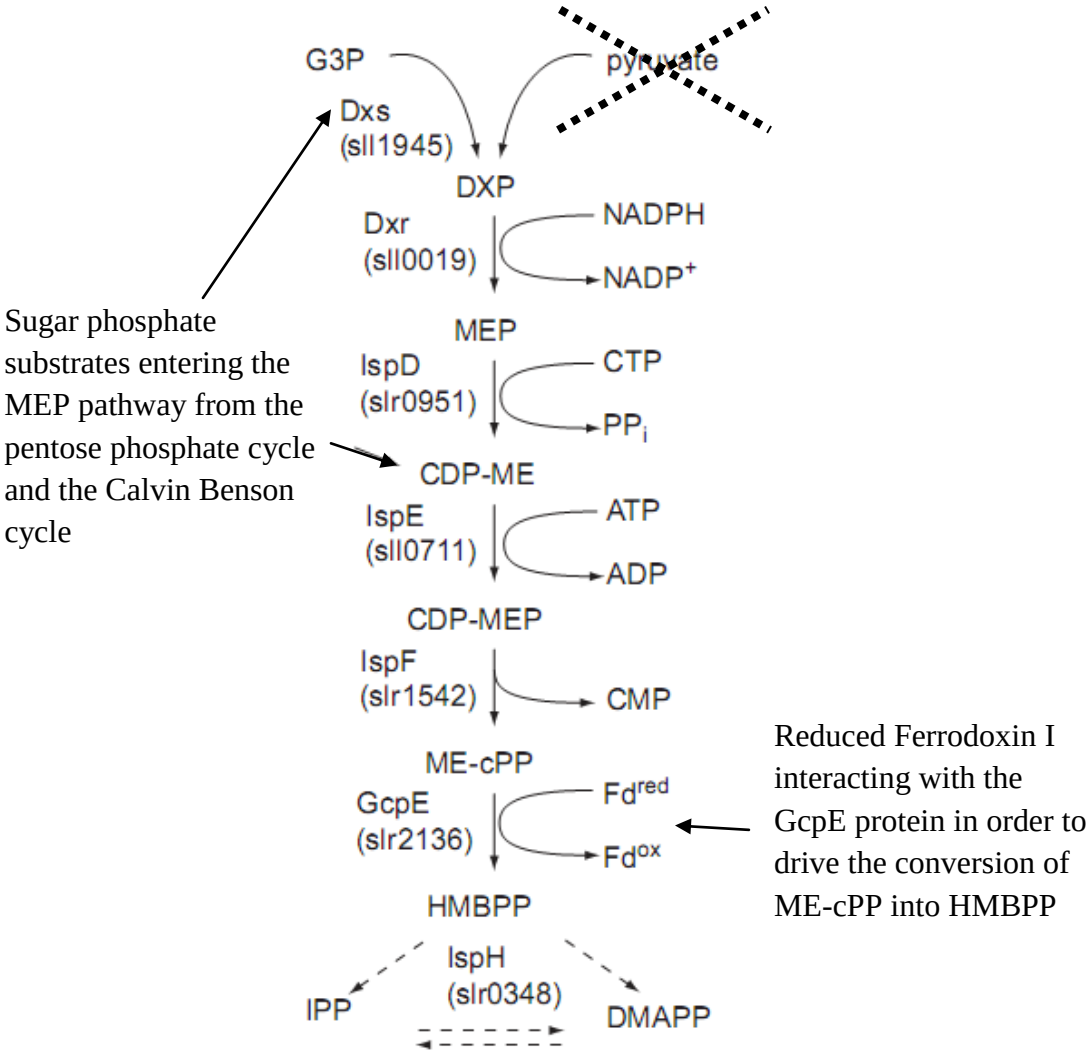

Figure 3. Modified MEP pathway. A proposed modified MEP pathway reflecting the addition of precursors derived from the pentose-phosphate and Calvin-Benson cycle and the interaction of ferrodoxin I with GcpE. Native *Synechocystis sp*. strain PCC6803 metabolites are shown directly below the arrows, the enzymes with their respective gene numbers are listed on the left using annotations (23). Chemical abbreviation names: G3P, glyceraldehydes-3-phosphate; DXP, deoxyxylulose-5-phosphate; MEP, 2-methyl-D-erythritol-4-phosphate; CDP-ME, diphosphocytidylyl methylerythritol; CDP-MEP, CDP-ME 2-phosphate; ME-cPP, methylerythritol 2,4-cyclodiphosphate; HMBPP, hydroxymethylbutenyl-diphosphate; IPP, isopentenyl diphosphate; and DMAPP, dimethylallyl diphosphate. Enzyme abbreviations: Dxs, DXP synthase; Dxr, DXP reductoisomerase, IspD, CDP-ME synthase; IspE, CDP-ME kinase; IspF, ME-cPP synthase; GcpE (IspG) HMBPP synthase; Fd, ferredoxin; IspH, HMBPP reductase. Pyruvate is crossed out indicating it may not play a role in the MEP pathway (48). Image modified from (51).

The redefined MEP pathway within *Synechocystis sp.* strain PCC6803 as shown in Figure 3, with the apparent diversity of routes for IPP and DMAPP formation, reflects an overall flexible metabolism that can be seen in other pathways. Within the *Synechocystis sp.* strain PCC6803 cell, there are two locations where respiratory electron transport (dark production of ATP) occurs: the cytoplasmic-membrane and the thylakoid membrane (34). Both locations appear to use a number of the same protein complexes, though within the thylakoid membrane there appears to be direct protein interaction and electron transfer between the photosynthetic and respiratory electron transport systems (34). There is also evidence that *Synechocystis sp.* strain PCC6803 can survive with either the PSI or PSII systems removed, as long as a fixed carbon source is made available, such as glucose (reviewed in 52). PSII-deficient mutants tend to survive longer and have a faster replication cycle compared to the PSI-deficient strains (reviewed in 52). The PSII-deficient mutants also display only minor physiological differences to the wild-type strains, compared to the distinct morphological differences observed in the PSI-deficient mutants (reviewed in 52). These data suggest that cyclical electron flow in PSI is a more important biochemical process for producing ATP then the production of free electrons from the splitting of water in PSII (52, 53). The cyclical electron flow occurs by the reduction of $NADP^+$ into NADPH by PSI, which is then re-oxidized, with the resulting electrons being transferred back to PSI by cytocrome b_6f (52, 53). Based on observations in PSI-deficient mutants, it is believed that electrons generated in PSII can travel either to PSI or directly to the cytochrome oxidase within the thylakoid respiration pathway (52).

The *Synechocystis sp.* strain PCC6803 PSII reaction center shows significant homology to that of plants. The core heterodimer of the PSII is composed of D1 and D2 proteins, from which other protein bind to form the full reaction complex, with a total size of 250 kDa (34). Within the *Synechocystis sp.* strain PCC6803 genome, there are three different sites that code for the D1 subunit (PsbA) in PSII: *psbAI, psbAII, psbAIII*. The primary genetic difference between these genes is found within the promoter regions, but the overall protein coding regions are all highly conserved with only some genetic variations (54). Both *psbAII* and *III* are up-regulated in high light environments, yet evidence for an environmental response in *psbAI* transcription is lacking (54). The p*sbAII* transcript is the prominent *psbA* transcript in *Synechocystis sp.* strain PCC6803 under normal photosynthetic growth conditions, indicating a stronger endogenous promoter, which has made this site a target of homologous combination experiments (14). *Synechocystis psbAII* knockout mutants remain viable, via an increase in *psbAIII* transcription, and show little negative physiological effects under normal growth conditions (54).

Synechocystis sp. strain PCC6803 *transgenic background*

Synechocystis sp. strain PCC6803 has long been a critical organism in transgenic research. In 1982, Grigorieva and Shestakov were the first researchers to transform wild type *Synechocystis sp*. strain PCC6803 (55). The researchers conferred resistance to the antibiotics erythromycin and ethionine, by incubating exogenous DNA isolated from Ery^R/Eth^R *Synechocystis* spontaneous mutants with wild type *Synechocystis sp.* strain PCC6803 cells on selective media. They were also successful in transferring DNA from two other sub-strains (6702 and 6714) into 6803. However, they were unable to save the

wild type 6803 strain using either Ery^R/Eth^R resistant cassettes from *Anacystis nidulans* 602 or *Synechocystis sp*. strain PCC6701 mutants. This inability to transfer resistance between similar species of cyanobacteria highlights the difference between the sub-groups of *Synechcocystis*, one with 42-48% GC content (6803, 6702, 6714) and the other with 35-37% (6701) (55). The researchers suggested that *Synechocystis sp.* strain PCC6803 may serve as an appropriate organism to test phylogenetic relationships but hesitated to venture any more significance into this initial work (55).

Synechocystis sp. strain PCC6803 has become a popular strain for transgenic experiments, because of its utility and ease of use in this type of research. Vermass highlights a number of these advantages in a 2004 review (56). As stated earlier, it can grow photo-heterotrophically, which allows for different selection/marker techniques and interesting avenues of research dealing with deficiency in photosynthesis. There is not as significant of a codon bias in *Synechocystis sp.* strain PCC6803 as there are in other commonly used algal strains, such as *Chlamydomonas reinhardtii* (57). Severe codon bias is an important factor that can seriously affect translation of an exogenous gene. The underlying factor that can affect transgene expression is the maintenance of an efficient ribozyme complex, which is dependent upon an ample supply of amino acyl-tRNAs. Different tRNAs for separate codons occur in various concentration in the cell owing to the semi-redundant nature of the genetic code (58). There is a degree of 'wobble' within the codon and anti-codon pairing between mRNA and tRNA inside the ribozyme complex, allowing for some imprecise matching at the third base of the codon (58). However, there are certain cases where a 'rare' codon is presented within the mRNA in which there are little to no complementary charged tRNAs present. This can

result in a stalled ribozyme complex and even termination of translation. This occurs less often in prokaryotes than eukaryotes, but it can still result in stalled and terminated translation of transgene sequences that contain 'exotic' codons (58). Within the *Synechocystis sp.* strain PCC6803 systems, all codons appear to be recognized and translated, though codon optimization does increase heterologous gene expression (51).

Synechocystis sp. strain PCC6803 is one of a few cyanobacteria that are naturally and spontaneously transformable, which is a result of a robust homologous recombination system. With the sequencing of the *Synechocystis sp.* strain PCC6803 genome, researchers can effectively design homologous recombination vectors that can readily insert foreign genes or knockout native genes at site-specific locations. These vectors can be designed and synthesized easily in *E. coli* allowing for efficient vector maintenance. *E. coli* is also an effective transconjugation partner for *Synechocystis* and transgenic vectors can pass easily between these two systems (59). This relationship may be a result of a porous membrane in *Synechocysits* under non-saline mating incubation conditions, as suggested by Sode, or perhaps it is the presence of large pili structures that appear to be indispensable for both motility and natural transformation (21, 59).

Early work investigating the nature of homologous recombination within *Synechocystis sp.* strain PCC6803 was conducted with antibiotic-resistant cassettes ligated with small pieces of a digested *Synechocystis sp.* strain PCC6803 genome serving as the source for the selectable markers in early transformation reactions (60). These experiments showed that the marker had to be flanked by digested genomic pieces, otherwise transformation would not occur, and that the flanking regions of this marker were retained within the *Synechocystis sp.* strain PCC6803 chromosome (60). The

researchers did suggest that homology to the genome to the flanking regions may be necessary, but admitted that alternatively, the flanking regions may be protecting the marker from digestion (60). At the time, the actual mechanism for recombination was not fully understood; however based on the research by LaBarre *et al*., it was concluded that a substitution mechanism was the most plausible explanation, in which the exogenous gene was replacing a region of DNA between flanking sites, rather than a single crossing over event (60). These researchers also argued that through homologous recombination it would be easier to develop homozygous recessive mutants after multiple generations based on random segregation during cellular replication (60). Previous attempts to produce homozygous recessive mutants using chemical and UV mutagenesis was comparatively more difficult and imprecise, given the high number of genome copies (up to 12) in *Synechocystis sp.* strain PCC6803 (60).

In 2007, thorough transformation optimization experiments were conducted in which various transformation techniques and vector designs were tested for stable genomic insertion of a kamamycin-resistance cassette into *Synechocystis sp.* strain PCC6803 (10). The researchers found that a "natural transformation" approach, in which concentrated cells were incubated under normal growth conditions with a plasmid for several hours, produced a significantly greater number of resistant transformants (efficiency of 1.4×10^{-5}) as compared to both electroporation and ultrasonication (10). Similar to LeBarre's findings in 1989, the researchers also reported that plasmids designed with larger homologous fragments (up to 1.7 kb) flanking the exogenous gene produced more successful transformants than plasmids with smaller to no flanking sequence (10, 60). This paper also presented the importance of the plasmid:cell

incubation time (optimal at 5 hours), cell culture growth phase (optimal at mid-log growth or early latent stage), and a recovery period in which tranformants were grown briefly in non-selective media before being transferred to kanamycin-selective plates (10). This relatively simple approach to produce mutant and/or transgenic strains of *Synechocystis sp.* strain PCC6803, in which high rates of transformation can occur with the use of homologous genomic sequences and without the need of expensive modern devices, highlights the reason for the overall popularity of this cyanobacterial system. The utility of site-directed transformation is evident, especially when compared to the imprecise nature of common eukaryotic transgenic practices.

Synechocystis sp. strain PCC6803 *commercial importance and potential*.

The versatility of *Synechocystis sp.* strain PCC6803 has made it an effective strain for the production of commercially relevant compounds. *Synechocystis sp.* strain PCC6803 produces a number of endogenous anti-oxidative compounds such as β-carotene, lycopene, and zeaxanthin, all of which have commercial and medicinal uses. Genetic engineering of *Synechocystis sp.* strain PCC6803 has resulted in strains that produce higher concentrations of these anti-oxidants, compared to the wild type (14). LaGarde et al. developed a double homologous recombination system targeting the *psbAII* site, as described earlier, to create several mutants with the goal of increasing zeaxanthin, β-carotene, myxoxanthophyll, and echinenone (14). The researchers produced four different mutant strains of *Synechocystis* by either inserting the IPP isomerase gene (*ipi*) from *Saccharomyces cerevisiae*, the β-carotene hydroxylase gene from *Synechocystis* (*crtR*), or the phytoene synthase and desaturase (*crtB* and *crtP*) from *Synechocystis*. Along with the insertion of these genes, the β-carotene ketolase gene that

produces an enzyme involved in the conversion of β-carotene to echinenone was disrupted. This resulted in a strain that produced a nonfunctional ketolase enzyme, in an attempt to increase zeaxanthin and β-carotene production (14). The role of echinenone in the cell is unclear and deficient mutants have been shown to have no obvious physiological problems (14). The transgenic *ipi* mutant did not appear to have any greater concentrations of the targeted carotenoids compared to the wild type *Synechocystis* (14). This lack of response to the *ipi* gene in *Synechocystis* is part of the evidence supporting an alternative pathway for the terminal steps in the MEP isoprene production, as discussed earlier. The authors also suggested that either the *S. cerevisiae* protein may be too complex to remain active in this prokaryotic system or the IPP isomerization is not a rate-limiting step in this pathway (14). The mutant strain with *crtR*, a gene that codes for the β-carotene hydoxylase which is involved in the conversion of β-carotene into zeaxanthin, produced a significantly higher concentration of zeaxanthin than the wild type strain (0.98 vs. 0.39 μg $(mL)^{-1}$ $(OD_{730}\ unit)^{-1}$) (14). Interestingly, the strain to yield the highest overall carotenoid concentration, 2.49 μg $(mL)^{-1}$ $(OD_{730}\ unit)^{-1}$, compared to 1.62 μg $(mL)^{-1}$ $(OD_{730}\ unit)^{-1}$ unit in the wild type, was the crtPB2ΔcrtO strain (14). This mutant had extra copies of the genes involved in the production of phytoene and ζ-carotene (*crtP and crtB* respectively), which are two of the more abundant terpenoids found in this *Synechocystis* strain. This strain was also designed with a silenced gene (*ΔcrtO*) that would otherwise be involved in echinenone production. The increase in carotenoid levels within this crtPB2ΔcrtO mutant strain is related to a large spike in myxoxanthophyll, though both zeaxanthin and β-carotene were increased significantly (14). The major technical significance of the LeGarde *et al.*

research is the development of an efficient double homologous recombination system that has been used by a number of other researchers. Their work also highlights effective strategies for increasing terpene production through knockouts of seemingly unnecessary genes and insertion of endogenous genes at stronger promoter sites in order to increase overall transcript levels.

A *Synechocystis sp.* strain PCC6803 strain was developed that produced a growth hormone from the flounder fish *Paralichthys olivaceus* (61). The researchers chose *Synechocystis sp.* strain PCC6803 for this experiment because of its ability to be directly consumed by the fish, allowing the cyanobacterium and the transgenic protein to be easily incorporated into the fish's diet without increased processing costs. The researchers obtained the cDNA sequence for the growth hormone gene (fGH) from *P. olivaceus* and inserted it into the vector pZPGK. The fGH gene was inserted in front of an *isiA* promoter on the plasmid, which is activated at low iron levels. The pZPGK plasmids also contained portions of the *groESL* untranslated regions flanking the open reading frame containing the *isiA* promoter, fGH gene, and a kanamycin-resistance cassette. The addition of the *groESL* flanking sequences in pZPGK allowed for homologous recombination at the *groESL* site inside the *Synechocystis sp.* strain PCC6803 genome, which would otherwise be involved in producing heat-shock related proteins that are not required for normal physiological function of *Synechocystis sp.* strain PCC6803. This strategy effectively inserted the fGH gene into the *sp.* strain PCC6803 genome and induction of transcription was easily controlled through adjusting iron levels in the liquid media. The fGH gene was stably inserted into the genome and despite the retention of the kanamycin-resistance cassette, fGH stability and retention did

not seem to require the continual maintenance of kanamycin in the liquid culture. The use of the iron-sensitive *isiA* promoter provides a unique strategy to control transgenic expression through chemical addition/withdrawal, contrasting to early transgenic experiments in which the transgene was constitutively active due to habitually strong promoters. Preliminary experiments in 2007 found the transgenic fGH strain to effectively increase *P. olivaceus* growth rates (61).

Further research investigating the transgenic fGH *Synechocystis* as a dietary supplement for *P. olivaceus* showed similar positive results (16). In a seven week feeding trial, fish were fed diets supplemented with either a 0%, 0.5%, or 2% concentration of lyophilized fGH *Synechocystis* cell concentrate. Fish fed with transgenic *Synechocystis* showed a significant increase in both the specific growth rate (0.58 %/day in the control group vs. 0.74 %/day in the 2% group) and the final overall weight (2% fGH *Synechocystis* fed fish were approximately 49% heavier than the control group) (16). Further histological examination of the fish fed with either the control feed or the transgenic algal feed showed no significant difference in the "stomach, intestine, liver, spleen and kidney" histology (16). This research demonstrates the efficacy and potential of using transgenic *Synechocystis* in aquaculture.

Synechocystis sp. strain PCC6803 strains have also been developed that produce mammalian proteins. The gene for the human liver protein metallothionein was transferred into *Synechocystis sp.* strain PCC6803, in order to increase the heavy metal sequestration capacity of the cyanobacteria for environmental remediation purposes (11). The researchers used vector cloning and insertion with plasmids designed with both light-activated and metal-activated promoters in front of the metallothionein cDNA. This

effectively developed strains that were resistant to 70 µmol of cadmium $(L)^{-1}$ in the growth media (11). The researchers also inserted the metallothionein cDNA, along with an ampicillin-resistance cassette into the *psbB* genomic site using homologous recombination, but were only able to reach cadmium tolerance of 60 µmol $(L)^{-1}$ (11). With these examples of fish hormone production and human protein production, *Synechocystis sp.* strain PCC6803 may be an effective platform for producing human pharmaceuticals.

More recently, researchers have been engineering novel energy-producing pathways into the genome of *Synechocystis sp.* strain PCC6803. In 2009, researchers inserted the *Zymononas moblius'* pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adh*) genes into *Synechocystis sp.* strain PCC6803 in order to create a pathway for ethanol production using carbon dioxide as the lone carbon source (62). There had been earlier attempts to introduce ethanol production pathways into the cyanobacterium *Synechococcus* PCC 7942, using the same *Z. moblius'* genes listed above (63). These researchers designed plasmids with the *pdc* and *adh* genes inserted downstream of an endogenous promoter for ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcLS*) (63). Interestingly, the researchers in this early work replaced the ribosomal recognition binding sequences of the *Z. moblius'* genes with the endogenous ribosomal recognition sequences of *Synechococcus' rbcLS*. This *pdc/adh* plasmid was not designed for homologous recombination, but rather it was maintained in the cell as a replicating plasmid (63). The plasmid also contained an ampicillin-resistance gene that allowed for continual selective retention through antibiotic selection (63). This design resulted in a yield of 54 nmol of ethanol $(OD730\ unit)^{-1}$ $(liter)^{-1}$ $(day)^{-1}$

(63). These researchers went on further to optimize this system and received a patent for a strain of *Synechococcus* that yields about 1.7 µmol of ethanol (mg of chlorophyll)$^{-1}$ (hour)$^{-1}$ in a transgenic strain of *Synechococcus* PCC 7942 (64). The Dexter group was able to significantly increase the yield of ethanol in *Synechocystis sp.* strain PCC6803 using a double homologous recombination system, with the *pdc/adh* cassette driven by the *psbaII* promoter, similar to the method developed by LaGarde (14, 62). These researchers also developed an efficient bench-top autotrophic-fermentative photobioreaction that resulted in a final yield of 5.2 mmol of ethanol(OD730 unit)$^{-1}$ (liter)$^{-1}$ (day)$^{-1}$ (62). The increase in yield of ethanol, without the need of antibiotics, demonstrates the utility of the *psbAII* promoter site and double homologous recombination, as opposed to using replicating exogenous plasmids. However, the novel design of the photobioreactor may have improved the yield of ethanol more significantly then the transgenic strategies.

Volatile isoprene emission has recently been achieved in *Synechocystis sp.* strain PCC6803 with the introduction of the isoprene synthase gene (*ispS*) based on the sequence from the plant *Pueraria montana* (51). Linderg's group, again, used a similar strategy to those of LaGarde and Dexter, targeting a double homologous recombination at the *psbAII* site (51). However, the researchers optimized the codon usage of the *P. montana* IspS sequence using the Kazusa Codon Usage Database, with a 12% codon frequency cutoff. Based on western blot analysis, the codon optimization appeared to increase the yield of isoprene synthase accumulation by about ten-fold, when compared to the expression of the native *P. montana* sequence in *Synechocystis* (51). Gas samples of the volatile emissions in the transgenic strain yielded about 50 µg of isoprene (gram of

dried algal weight)$^{-1}$ (day)$^{-1}$ (51). Lindberg's work shows the importance of codon selection and introduces an interesting idea on metabolite harvesting from liquid algae cultures. Designing a system in which the desired metabolite is volatile and can pass through the cell membrane allows for extraction of the chemical without destroying the cell culture. This approach may not be practical for many desired metabolites, such as oil and various long chained pigments, but it does offer a novel approach to large scale production.

β-caryophyllene overview

The sesquiterpene (E)-β-caryophyllene, more commonly referred to as β-caryophyllene, is a ubiquitous molecule that is produced and emitted by many members of the plant kingdom. The chemical synthesis of (E)-β-caryophyllene involves the interaction of a metal cation with a *trans*-farnesyl diphosphate (FPP), resulting in the formation of a large cyclic carbocation intermediate that quickly stabilizes into a cyclobutane ring attached to a *trans* nine-carbon ring (Figure 4) (65). Plants that produce (E)-β-caryophyllene also often produce both α-humulene (previously α-caryophyllene), a chemical with a similar *trans* cyclic structure that lacks the butane ring, and (Z)-β-caryophyllene, the *cis* isomer of (E)-β-caryophyllene (Figure 5) (66). In plants, α-humulene and (E)-β-caryophyllene are typically the dominant sesquiterpenoids produced (66).

It was originally proposed that FPP was synthesized in plants solely from MVA-derived precursors because farnesyl diphosphoste synthases (FPPS) were primarily expressed in the cytoplasm; however, labeling experiments have revealed a cross flow of MEP-derived GPP reacting with cytosolic IPP and DMAPP (67). There are FPPS in

organisms, such as green algae and cyanobacteria, which lack a MVA pathway (NCBI database, 10.7.10). However, there is no confirmed FPPS identified in the *Synechocystis sp.* strain PCC6803 genome, though sequence analysis suggests some possible homologues (see Chapter IV Results). There are no reported β-caryophyllene synthases found outside of the plant kingdom (NCBI database 10.7.10). This synthase has been characterized in plant species including *Zea mays* (68), *Oryza sativa* (69), *Cucumis sativus* (70), *Salvia officinalis* (71) , *Makania micrantha* (72), and *Artemesia annua* (65)

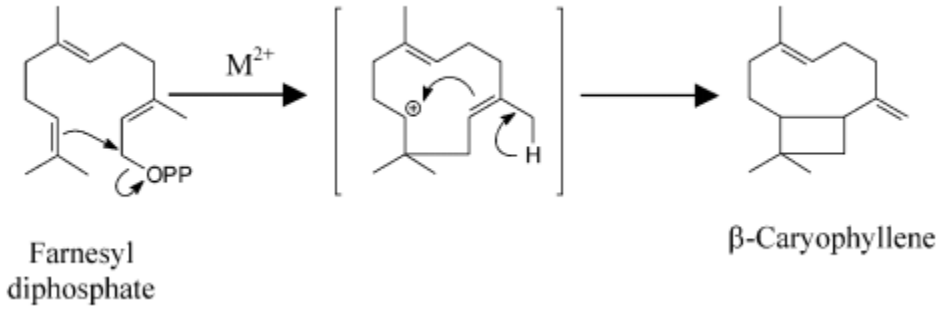

Figure 4. Proposed mechanism for β-caryophyllene synthesis. Cyclization of *trans*-farnesyl diphosphate into β-caryophyllene. The metal cation cofactor is often either a Mg^{2+} or Mn^{2+} ion and helps stabilize the carbocation intermediate. Image from (65).

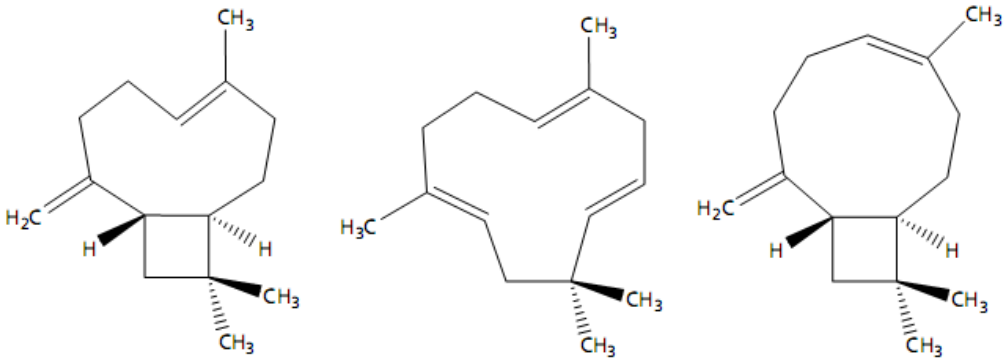

Figure 5. Chemical structure of common caryophyllene-like sesquiterpenes. (E)-β-caryophyllene (left), α-humulene (center), and (Z)-β-caryophyllene (right). Image from (73)

The β-caryophyllene synthase used in this thesis was originally isolated from a cDNA library generated from the plant *A. annua* in 2002 (65). Researchers used primers with conserved bases from other identified β-caryophyllene synthases to probe the *A. annua* cDNA library for any orthologous sequences (65). They found a 1902 bp clone that showed a high degree of orthology with other synthases and further analysis revealed a 1644 bp open reading frame that would produce a 60.3 kD protein. The sequence lacked an N-terminal organelle targeting sequence, suggesting the enzyme and the reaction are located in the cytosol (65). The cDNA sequence was cloned into an *E .coli* expression vector and then expressed in this prokaryotic system, resulting in a soluble protein fraction that had β-caryophyllene synthase activity based on a chemical assay in which radio-labeled farnesyl diphosphate was converted into β-caryophyllene (65). *In vitro*, this synthase had optimal activity at pH 7.7 with a Mg^2 cofactor (65). The protein contains a highly conserved GVYXEP motif similar to other sesquiterpene enzymes as well as a DDXXD metal cation-substrate binding motif which is also highly conserved in angiosperm terpene synthases (65).

Research suggests that β-caryophyllene is involved in a number of defense mechanisms deployed by plants. Mechanical stress and exposure to insect saliva increases plant β-caryophyllene emissions and overall plant-wide production of this sesquiterpene (65, 74, 75). There has been some interesting research conducted on the emissions of β-caryophyllene in response of arthropod herbivory as both a means to repel herbivores or attract natural predators to these herbivores. Researchers investigating the temporal emission of volatile terpenes induced by insect herbivory observed specific

nocturnal emissions in *Nicotiana tabacum* that seemed to deter moth oviposition (76). β-caryophyllene was not emitted exclusively at night, but it was emitted at higher concentrations at night, indicating it may be involved in repelling nocturnally active moths (76). These researchers did not conclude that these emissions were a specific defense mechanism deployed by the plants to prevent herbivory, but instead proposed this behavior was a response by the maternal moth to find undamaged plants to lay their eggs, in order to provide a low competition nursery for her developing progeny (76). However, the researchers did suggest a potential benefit enjoyed by uninjured plants surrounding the infested repellent-emitting plant that may provide a fitness advantage to these healthy plants (76).

Upon herbivore infestation, plants often emit volatile organics, both above and below ground, which can attract natural predators of theses noxious herbivores. Researchers observed wild varieties of *Zea mays* emitting volatile organic chemicals when the roots were infested with Western corn rootworm (*Diabrotica virgifera)* (77). These emissions were shown to attract the nematode *Heterorhabditis megidis,* an effective predator to the Western corn rootworm (77). The researchers further isolated the chemical species being emitted in the soil by the wounded plant and found that β-caryophyllene was the prominent constituent (77). They established β-caryophyllene's role in predator attraction by adding exogenous β-caryophyllene into the soil of healthy *Z. mays* plants. The nematode showed a similar attractive response in this case as compared to the wounded root emissions (77). Interestingly, these researchers observed several North American commercial cultivars of *Z. mays* that seem to have lost their ability to emit significant levels of β-caryophyllene and found them to be far more susceptible to

rootworm infestation, even with the predatory nematodes in the soil (77). They also showed the ability of β-caryophyllene to quickly diffuse through sand and soil, further displaying its integral role in attracting predators from a relatively long distance (77). The work by Rasmann et al. was one of the first research papers to identify a below-ground plant emission constituent, β-caryophyllene, specifically responsible for attracting predators to the host emitter (77). The ability of the plant to emit a soil-soluble, predator-attracting chemical in response to herbivory is quite an advantageous survival technique that opens the door to a number of interesting questions relating to plant-insect coevolution.

β-caryophyllene may also be involved in an allelopathic response to prevent competitive seed germination (74). In 2009, the β-caryophyllene synthase gene was isolated from the invasive weed *Mikania micrantha* (74). Researchers determined that leaves of this plant emitted high concentrations of β-caryophyllene upon mechanical damage (74). These researchers also determined that high concentrations of β-caryophyllene (3 mg $(liter)^{-1}$) inhibited the growth and germination of competitive plant species to *Mikania micrantha*, though small concentrations of the chemical seemed to stimulate germination of these competitive seed (74). It remains to be determined whether living plants are actively emitting β-caryophyllene to prevent seed germination; though decaying plant debris appears to release enough of this sesquiterpene to inhibit seed development, which may be providing a competitive advantage for subsequent generations (74). β-caryophyllene may also be involved in the plant's response to fungal infection. Researchers showed an increase in β-caryophyllene synthase transcripts when fungal *Verticillium dahlia* extracts were applied to 8-week old *Artemisia annua* saplings

(65). However, other studies have shown β-caryophyllene to be ineffective against inhibiting the growth of the fungus *Fusarium oxysporum*, a plant pathogen that causes 'rhizome rot' in ginger (78).

β-caryophyllene, with its woody and spicy aroma, has been used in the fragrance and cosmetic industry since the 1930s (79). The U.S. Food and Drug Administration and the Flavor and Extract Manufacturers Association view this chemical as GRAS ('generally regarded as safe') because it causes very little irritation when subjects are exposed to high concentrations, both on the skin and by ingestion (reviewed by 80). The LD_{50} of β-caryophyllene in rabbit and rat models exceeded 5 grams (kilogram of body mass)$^{-1}$(reviewed by 79, 80). Concentrations as high as 300 mg (kg)$^{-1}$ have been administered orally without any negative physiological effect in mice models (reviewed by 79, 80). Researchers have also shown that high levels of β-caryophyllene (2g (kg)$^{-1}$) administered *in vivo* in mice models did not significantly increase the occurrence of micronucleated polychromatic erythrocytes, which are commonly found in cells exposed to genotoxic chemicals and are a sign of genomic damage (81). It appears β-caryophyllene may even prevent the emergence of micronuclei, because it has been reported that human lymphocyte cells exposed to the genotoxic chemical ethyl methanesulfonate showed a significant reduction of micronuclei development when β-caryophyllene was added either before or during toxicant exposure (82). However, β-caryophyllene did not reverse the damage of ethyl methanesulfonate, because when β-caryophyllene was delivered after toxicant exposure, micronuclei emergence was equivalent to toxicant-exposed cells with no β-caryophyllene treatment (82). This

suggests β-caryophyllene is neutralizing the toxicant rather than eliciting a cellular repair/defense response (82).

β-caryophyllene is found at high concentrations in a number of traditionally used natural remedies such as balsam (83, 84), ginger (78), and clove oils (85). Researchers investigating the bioactivity of the Asteraceae family tested a number of oil extractions from various *Helichrysym* species (86). Results from 5-lipoxygenase assays, which gauge anti-inflammatory capacity by measuring the activity of the human inflammation modulating enzyme 5-lipoxygenase (5-LOX), revealed that extracts from *Helichrysym* species with the highest concentration of β-caryophyllene had the highest level of 5-LOX inhibition (86). These results are consistent with work done by Baylac and Racine, who showed a high level of 5- LOX inhibition in commonly used aromatherapy oils that have high levels of β-caryophyllene among their volatile constituents (87).

β-caryophyllene alone has been shown to reduce colitis-related inflammation in a mouse model (80). Researchers induced intestinal swelling by feeding mice a 5% dextran-sulfate-sodium-water solution, at which point the mice were separated into several treatment groups and received varying concentrations of orally administered β-caryophyllene (80). At a treatment level of 300 mg of β-caryophyllene per kg of body weight (the highest concentration tested), researchers observed a significant reduction of colon shortening, weight loss, and mucosal tissue inflammation (80). They also observed a reduced activity of the inflammation-modulating enzyme myeloperoxidase and a lower level of the inflammatory cytokine interleukin-6 in these colitis mice at the 300 mg $(kg)^{-1}$ treatment level (80).

β-caryophyllene has also been shown to interact with G-protein-coupled Type-2 cannabinoid receptors (CR_2), binding in the same hydrophobic pocket in which endogenous arachidonic acid-derived endocannabinoids bind, and the psyochoactive cannabinoids Δ9-tetrahydrocannabinol (THC), Δ8-tetrahydrocannabinol, and cannabinol interact (66). Similar to other CR_2 agonists, β-caryophyllene showed a capacity to inhibit inflammatory cytokine expression in both human cell cultures and mice models, and the researchers suggested this down regulation was a result of an inhibition of the Erk1/2 and JNK1/2 kinase signaling pathways (66). This interaction may explain the anti-inflammatory properties of β-caryophyllene because the CR_2 receptor group is often located in peripheral tissue and its activation has been shown to help modulate inflammation response in a number of regions (66). This may also explain local anesthetic properties observed in cells treated with β-caryophyllene. Electrically evoked contractions of the hemidiaphragm *in vitro* rat models and the reduction of a conjunctival reflex response *in vivo* rabbit models were observed when these systems were treated with 1 μg $(mL)^{-1}$ of β-caryophyllene (85). However, β-caryophyllene's anesthetic response was short lived, as the electric sensitivity returned after only 15 minutes (85).

β-caryophyllene has also been shown to be a possible inhibitory compound in the growth of several human pathogens. The essential oils from the β-caryophyllene rich-rhizomes (42% composition of the extracted essential oils) of the ginger plant *Zingiber nimmonii* were extracted and screened against a number of human pathogens and their activity was compared against the antibacterial agent streptomycin and the anti-fungal compound fluconazole (78). These β-caryophyllene-rich extracts showed inhibition against the bacterial strains *Bacillus subtilis*, *Salmonella typhi*, and *Pseudomonas*

fluorescens and a significant inhibition against the fungal strains *Aspergillus niger*, *Candida glabrata*, and *Candida albicans* that was almost identical to the treatment with fluconazole (78). Essential oils from several species of *Copaifera* were screened against the leishmaniasis-causing protozoa *Leishmania amazonensis* to determine any inhibitory capacity on promastigote growth (88). The two species that had the highest level of inhibition against *L. amazonensis, C. reticulate* and *C. multijuga* were also the two species to have the highest overall sesquiterpene concentration and the highest level of β-caryophyllene within their respective essential oils (88). Santos *et al.* acknowledge the potential role of β-caryophyllene in inhibiting the proliferation of the *L. amazonensis* promastigote but also suggested that the inhibition may not be due solely to one specific compound because it has been observed that the individual components of the balsam oils never elicit as potent of a response as the oil in its pure state (88).

The synergistic effect of β-caryophyllene with other volatile constituents has also been observed in the treatment of tumor cells. It has been s shown that β-caryophyllene greatly increases the intracellular accumulation and cytotoxicity of several other sesquiterpenes and chemotherapy drugs (73). Though these researchers reported that β-caryophyllene alone had little to no cytotoxic effects against MCF-7 tumor cells, treatments with either (Z)-β-caryophyllene, α-humulene, and paclitaxel showed a significant increase in activity against several tumor cell lines when 10 μg $(mL)^{-1}$ of (E)-β-caryophyllene was added along with each individual treatment compound (73).

Current procedures for isolating bioactive β-caryophyllene require a large input of initial plant biomass. A hexane-based hydrodistillation or ethanol Soxhlet procedure can effectively extract anywhere from 3.9% essential oils when preprocessed, dried plant

powder is used or as low as 0.04% essential oils when fresh plant material is used (78, 89). In both of these examples, the small essential oil volumes would need to be further processed to obtain a pure β-caryophyllene fraction. Further processing can expose β-caryophyllene to high temperatures and pressures, leading to oxidation and a subsequent decrease in the chemical's bioactivity (90). Recent research in isolating and maintaining chemically active β-caryophyllene has been focused around supercritical CO_2 extraction or steam distillation (90, 91). Supercritical CO_2 extraction appears to produce the most pure and active β-caryophyllene fraction, but yields are significantly lower than hydrodistillation protocols (90, 91). Much more research remains to be completed in finding effective methods of isolating β-caryophyllene for use in medicinal applications, especially when considering the relatively high concentrations needed for effective clinical treatments.

CHAPTER III

MATERIALS AND METHODS

Strain and growth conditions

Wild type *Synechocystis sp.* strain PCC803 was obtained from the lab of Dr. Louis Sherman at Purdue University. The strain was maintained on BG-11 media (92) with 1.5% agar supplemented with 1 mM sodium thiosulfate; subcultures were supplemented with 50 µg $(mL)^{-1}$ kanamycin and/or 5% sucrose as needed. Standard liquid cultures were grown in BG-11 liquid media buffered to pH 8.0 and supplemented with 50 µg $(mL)^{-1}$ kanamycin and/or 5% sucrose as needed. Both agar plates and liquid cultures were grown at 30 °C under constant illumination of 1000 lux (12.5 µE $(m^2s)^{-1}$). Liquid cultures were aerated using a shaker table set at 130 RPMs. Bioline ElectroShox (Tauton, MA) electropotant *E. coli* cells were grown at 37 °C on LB plates with 1.5% agar after electroporation and screened with 75 mg $(L)^{-1}$ of ampicilin as needed. Successful *E. coli* transformants were maintained both at 4 °C on LB agar plates and in 70% glycerol at -80°C. *E. coli* was grown at 37°C in liquid LB media, pH 7.0, supplemented with 75 mg $(L)^{-1}$ ampicilin for selection.

pSBAIIQHS1 vector design

The plasmid PSBAII-KS (14) used during the first round of homologous transformation and contains the selective cassette *sacB/aphX,* flanked between the

homologous sequences to the flanking regions of the *Synechocystis sp.* strain PCC6803 *psbAII* gene. This plasmid was obtained from Dr. Wim Vermass at Arizona State University. The *A. annua* β-caryophyllene synthase gene, housed within plant expression plasmid pET*QHS1* (65), was obtained from Dr. Rodney Corteau at Washington State University. Cloning primers were designed to amplify the entire coding region of *QHS1* along with introducing the restriction sites AseI at the 5' site and a BamHI at the 3' site (Table 1). Using these designed primers, the full coding sequence of *QHS1* was amplified from pET*QHS1*. The PCR assay was run at annealing temperature of 60°C for 35 cycles. This PCR and subsequent ones were run using the Dream Taq Master Mix solution from Fermentas (Glenburie, MD, USA) on a Bio-Rad MJ Mini Personal Thermal Cycler (Hercules, CA, USA). The resulting PCR product was electrophoresed on a 1% agarose gel, stained with 0.1% SYBR Safe from Invitrogen (Carlsbad, CA, USA,) to confirm a single band product and correct size. The appropriate *QHS1* band was excised from the gel and the DNA was purified using a GeneJET ^{TM} clean up kit obtained from Fermentas (Glenburie, MD, USA). To increase the concentration of the *QHS1* insert, the PCR product, verified to contain only the insert, was directly cleaned using GeneJET^{TM}. The resulting DNA was digested with the restriction enzymes AseI and BamHI obtained from New England Biolabs (Ipswich, MA, USA) following the manufacturer's protocol. The digested *QHS1* gene was further cleaned using GeneJET ^{TM}. Simultaneously, the vector pPSBAII-KS was digested with NdeI and BamHI, from New England Biolabs (Ipswich, MA, USA). The digested vector was run on a 0.8% agarose gel, which produced two bands; the larger of which (~4.2 kB) is the pPSBAII backbone (Figure 6), was excised from the gel and purified following the prescribed protocol of the GeneJET

™ plasmid purification kit. The purified *QHS1* and pPSBAII fragments were subsequently ligated using Novegen's T4-Ligation kit (Darmstadt, Germany). The ligation product was clean and eluted with nuclease-free water using the GeneJET ™ kit. Henceforth, this constructed plasmid will be referred to as pPSBAII*QHS1*.

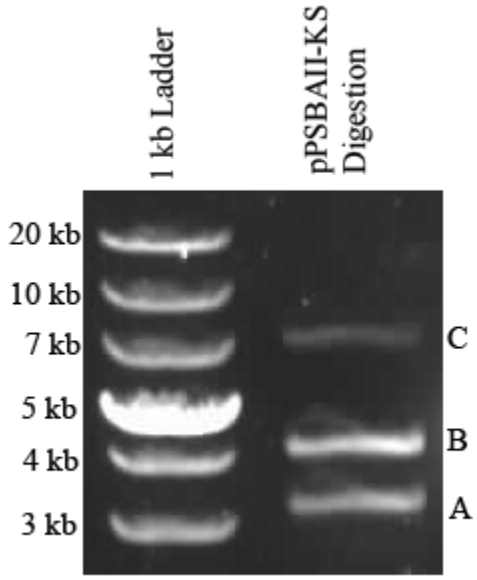

Figure 6. Results of the restriction digest of the pPSBAII-KS vector. The agarose gel image of vector pPSBAIIKS digested with the restriction enzymes NdeI and BamHI. (A) Represents the excised *sacB/aphX* cassette, ~3240 bp (theoretically 3233 bp), (B) represents the remaining pPSBAII vector backbone, ~4300 bp (theoretically 4298 bp), (C) undigested pPSBAII-KS vector. Band (B) was cut from the gel and subsequently used in ligation reaction with the *QHS1* insert.

Table 1. Primer sequences and putative genomic positions.

Primer	Sequence	Position
5' pET*QHS1*	5'- aga***attaat*** ATGTCTGTTAAAGAAGAGAAAGTAATTCG-3'	24-53[a]
3' pET*QHS1*	5'- aacaga***ggatcc***TTATATAGGTATAGGATGAACGAGCAAAG-3'	1642-1671[a]
5' $QHS1_{RT}$	5'- ATGTCCCAAACAATCTAC-3'	805-823[a]
3' $QHS1_{RT}$	3'- GTTCCATAAGCATCATAAG-3'	948-967 [a]
5' $rnPB_{RT}$	5'- AGAGTTAGGGAGGGAGTT -3'	153167- 153184[b]
3' $rnPB_{RT}$	5'- CTGTTTACTGGTTGCTGTTT-3'	153308-153327[b]
5' ^{UP}pPSBAII$QHS1_{sq}$	5'- TACACAGCCCAGAACTAT -3'	6980-6997[b]
3' ^{UP}pPSBAII$QHS1_{sq}$	5'- CTTGCTCAATCTCCTCTT-3'	285-268[a]
5'^{DOWN}pPSBAII$QHS1_{sq}$	5'- GTCAGTTGCGTTTGTAAGC-3'	207-1226[a]
3' ^{DOWN}pPSBAII$QHS1_{sq}$	5'- CCCCCATAATTCCTCTATCG-3'	8479-8498[b]

The "RT" subscript indicates primers used for reverse transcriptase PCR and qPCR. The "sq" subscript indicates primers used for the sequencing of both the pP*QHS1* plasmid and the *QHS1* mutant *Synechocystis* strain. The 'UP' and 'DOWN' superscripts refer to the relative position targeted for sequencing. The resulting sequences are available within the Appendix. The lowercase lettering within the sequences indicate modifications introduced to assist with cloning. The bold face lettering shows the restriction sites inserted for the cloning and ligation reaction (AseI for 5' pET*QHS1* and BamHI for 3' pET*QHS1*). The underlined 'ATG' is the start codon for the coding sequence of *QHS1*. ([a]) Position numbering based on the published sequence of *A. annua*'s β-caryophyllene synthase *QHS1* mRNA (NCBI Accession-AF472361). ([b]) Position numbering based on the published sequence of *Synechocystis sp.* strain PCC803 (NCBI Accession-BA000022.2).

Escherichia coli transformation

For the *E .coli* transformation, 50 µL of thawed Bioline ElectroShox electropotant *E. coli* cells were mixed with 5 µL of the cleaned pPSBAII*QHS1* ligation mixture, inside a 1 mm electroporation cuvette. The mixture was then electroporated at 1800V for 5 ms, with a capacitance of 25 µF and a resistance of 200 Ω, using a Bio-Rad GenePusle Xcell Electroporator (Hercules, CA, USA). The cells were immediately diluted with 1 mL of super optimal broth media and incubated at 37 °C for 1 hour. Aliquots of the electroporation mixture, measured at 200 µL, were spread on ampilicin selection plates, as describes above, and incubated overnight at 37 °C. Colonies that grew on the ampicilin plates were screened using a colony PCR assay with the cloning primers pET*QHS1*, at an annealing temperature of 60°C for 35 cycles (Table 1). *E. coli* colonies that tested positive for *QHS1* presence were inoculated into liquid LB/ampicilin media and grown overnight. The pPSBAII*QHS1* plasmid was subsequently extracted from these *E. coli* cells using a Qiagen Miniprep Plasmid Extraction kit (Valencia, CA, USA) following the manufacturer's protocol. The colony that did show positive results was screened further with a sequencing reaction to determine *QHS1* retention and orientation. As described above, the plasmid was directly extracted from the *E. coli* cells. This cleaned plasmid DNA, along with the primers $5'^{UP}$ pPSBAII$QHS1_{sq}$ and $5'^{DOWN}$ pPSBAII$QHS1_{sq}$, were sent to MCLAB DNA Sequencing Services (San Francisco, CA, USA) for the sequencing reaction (Table 1).

Synechocystis sp. Strain PCC6803 *double homologous recombination reaction*

Wild type *Synechocystis sp*. strain PCC803 was grown until mid-log phase, at which point aliquots of the liquid suspension were centrifuged, washed, and resuspended

with BG-11 media. The pPSBAII-KS plasmid, at 1 μg, was added to the concentrated *Synechocystis* suspension and incubated at 30 °C for 6 hours. Small 100 μL aliquots of the cell:plasmid mixture were inoculated onto 80 mm filter paper, placed in non-selective BG-11 agar plates and allowed to incubate for 24 hours at 30°C. The filter paper containing the cells was then transferred onto selective plates with 50 μg $(mL)^{-1}$ of kanamycin and grown for 4 days before further screening. Successful transformants were transferred to separate kanamycin-selective plates. Individual isolated colonies were grown in liquid BG-11 cultures and further screened for the presence of the *sacB/aphX* cassette and full segregation of the mutant by testing sucrose sensitivity on 5% sucrose BG-11 plates and running a PCR using the $5'^{UP}$ pPSBAII$QHS1_{sq}$ and $3'^{DOWN}$ pPSBAII$QHS1_{sq}$ primers, at an annealing temperature of 60° C for 35 cycles (Table 1). The *sacB* gene codes for a levansucrase protein, that causes a lethal phenotype in gram-negative bacteria cells (93).

After establishing segregation, the pPSBAII-KS mutant strain was grown to mid-log phase in a liquid culture, at which point the 15 mLs of the culture was concentrated down to 1 mL using a centrifuge run at 2000xG. Then three μg of the pPSBAII*QHS1* vector were introduced. The cell:plasmid mixture was incubated at 6 hours before being diluted 20:1 with non-selective BG-11 liquid media. This mix was incubated for 48 hours at normal growth conditions before a sterile-filtered 50% sucrose solution was added, bringing the final concentration of sucrose in the BG-11 liquid media to 5%. The sucrose-selective liquid culture incubated for 4 days before small aliquots were streaked on selective 5% sucrose BG-11 agar plates (Figure 7). After 1 week, individual colonies that grew on the sucrose were screened for the presence of the *QHS1* using PCR assay

with either the pET*QHS1* primers or the the 5'^{UP} pPSBAII*QHS1*$_{sq}$ and 3'^{DOWN} pPSBAII*QHS1*$_{sq}$ primers. Several screened colonies showed some presence of the *QHS1* gene during early screening, but only one mutant colony was further sub-cultured and used in the proceeding RNA and terpene extraction (Colony 4, Quadrant 2, Figure 7). Genomic DNA from this mutant colony, along with the primers 5'^{UP} pPSBAII*QHS1*$_{sq}$ and 3'^{DOWN} pPSBAII*QHS1*$_{sq}$ primers, were sent to MCLAB DNA Sequencing Services. Sequencing reactions were used to confirm orientation and position of the *QHS1* gene inside the *Synechocystis* genome. The overall double-homologous recombination transformation strategy is graphically detailed in Figure 8.

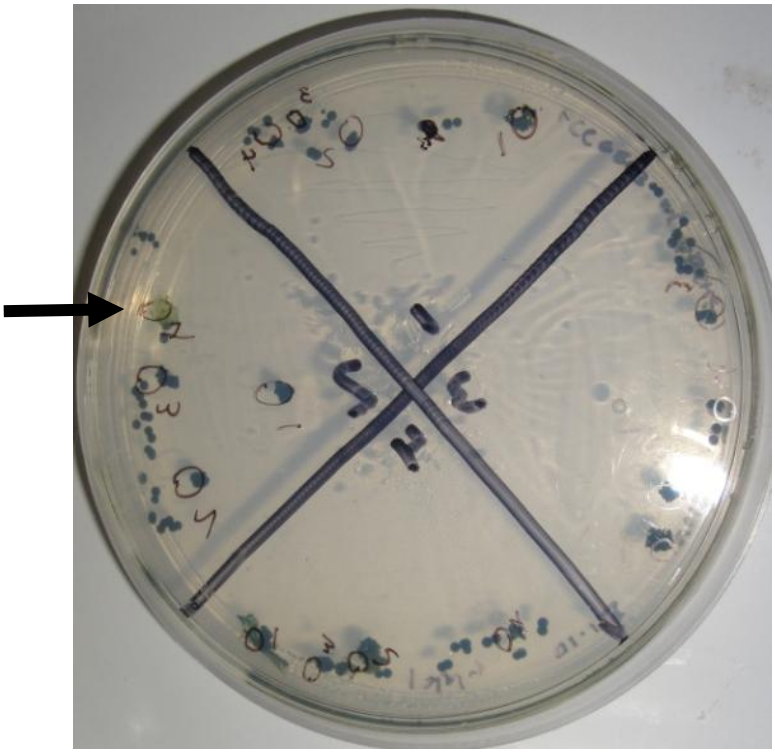

Figure 7. BG-11-sucrose-agar plate with successful *QHS1* transformant. This image displays the modified streaking method used during the second homologous recombination reaction. The arrow indicates the Colony 4, Quadrant 2, the transformant that showed consistent positive colony PCR results and that was selected for further sequencing and β-caryophyllene production assays.

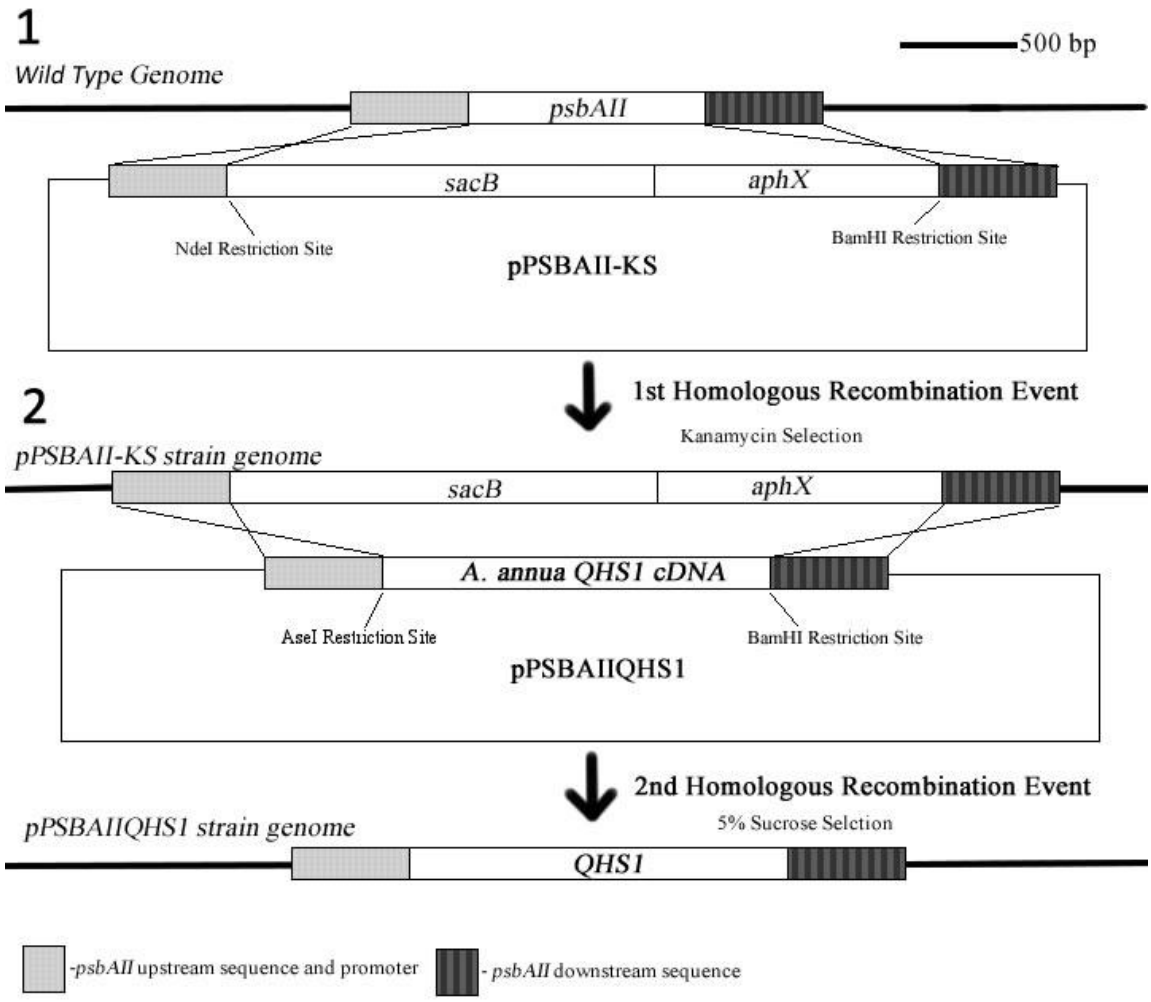

Figure 8. Graphical representation of the double homologous recombination strategy. *QHS1* insertion into the *Synechocystis sp.* strain PCC6803 genome, using *A. annua's* native cDNA sequence inserted into the previously designed pPSBAII vector. (1) The first crossing over event in which the endogenous gene *psbAII* is replaced by the *sacB/aphX* cassette, causing stable insertion of the selection cassette into the *Synechocystis* genome (14). (2) The final crossing over reaction in which the genomic *sacB/aphX* cassette is replaced with the *A. annua* cDNA sequence for β-caryophyllene synthase, creating a mutant strain of *Synechocystis* with the *psbAII* promoter driving *QHS1* transcription.

Reverse transcriptase polymerase chain reaction

DNA and RNA were extracted from mid-log phase liquid cultures of wild type *Synechocystis sp.* strain PCC6803, the *Synechocystis* pPSBAII-KS strain, and the *Synechocystis* pPSBAII*QHS1* strain. For each strain used in this experiment, 14 mL of the liquid culture were centrifuged and the supernatant was removed. From this pellet, RNA and DNA were extracted using a modified Chang extraction protocol (94). An aliquot of the final extraction was retained and used as a DNA template for sequencing and PCR confirmation. The remainder of the extracted solution was treated with Fermentas' DNaseI (Glenburie, MD, USA), following the manufacturer's protocol. The RNA was reverse transcribed to produce cDNA using iScript cDNA synthesis kit from Bio-Rad (Hercules, CA, USA), following the manufacturer's protocol. The cDNA from each strain was used as a template in reverse transcriptase polymerase chain reaction (RT-PCR) assays. The primers $QHS1_{RT}$ were used to detect the presence of any *QHS1* RNA that was transcribed and the RT-PCR was run at an annealing temperature of 56 °C for 35 cycles. The $rnPB_{RT}$ primers were used to detect RNA transcribed from the house-keeping gene rnPB, that codes for subunit B of RNase P that has been shown to be constitutively active in *Synechocystis sp.* strain PCC6803 (95). The housekeeping RT-PCR was run at an annealing temperature of 57 °C for 35 cycles. Resulting cDNA bands were visualized on an ethidium bromide stained 2% agarose gels, run at 80V for 60 minutes.

Extraction and quantification of β caryophyllene

Wild type *Synechocystis* and the *QHS1* transgenic *Synechocystis* liquid cultures were grown for one week under standard growing conditions, to an OD_{730} of

approximately 3.5. Lipids and terpenes were extracted according to Bligh and Dyer (96). Cells were pelleted followed by the addition of chloroform and methanol to obtain a monophasic solution of the ratio 1:2:0.8 (chloroform:methanol:water/biomass) by volume. Samples were vortexed and allowed to stand for 15 minutes. Chloroform and water were added (1:1 v/v) to form a biphasic system (final solvent ratio: 1:1:0.9 chloroform:methanol:water v/v/v). The samples were vortexed and the lipids and terpenes were allowed to partition into the chloroform layer. Chloroform extracts were washed with deionized water and analyzed directly by gas chromatography-mass spectrometry (GC-MS) and gas chromatography-flame ionization detection (GC-FID). β-caryophyllene was identified by comparison of GC retention times to standards (Sigma, St. Louis, MO) and by comparison of mass spectra to the National Institute of Standards and Technology (NIST) Mass Spectral Library (V 2.0d). Mass spectral analysis was conducted on a Varian 3800 GC and Varian 1200 quadrupole MS/MS operated with the following conditions: Rxi-5 ms column (30 m x 0.25 mm; 0.25 µm film thickness); temperature program of 60-240 °C (at 3 °C $(min)^{-1}$); He carrier gas: 1.5 ml $(min)^{-1}$; mass spectra, 70eV, in EI mode; Ion source temperature 250 °C; scan mass range, 25-600 (97). β-caryophyllene levels were quantified via a GC-FID (Agilent 7890A) equipped with a DB5-ms column (Agilent Technologies, Santa Clara, CA; 30m x 0.25mm; 0.25 µm film thickness) operated under the following conditions: temperature program: 60-240 °C (at 3 °C/min); He carrier gas: 1.5ml/min (98). Concentrations of β-caryophyllene in experimental samples were determined using a standard curve. The standard curve was generated by coupling the gas chromatograms of various known concentrations of the β-caryophyllene with the respective areas of these β-caryophyllene peaks (Table 2 and

Figure 9). From this data and the resulting equation of the standard curve, the concentrations of experimental samples could be determine based on the area of the peak from the experimental gas chromatograms.

Table 2. Standard curve data.

Concentration of β-caryophyllene (ng/µL)	Average area of the peak with error limits
28.125	496.95 ± 20.15254
56.25	992.25± 13.36432
112.5	1955.15± 1.343503

Compiled data used for the generation of the standard curve.

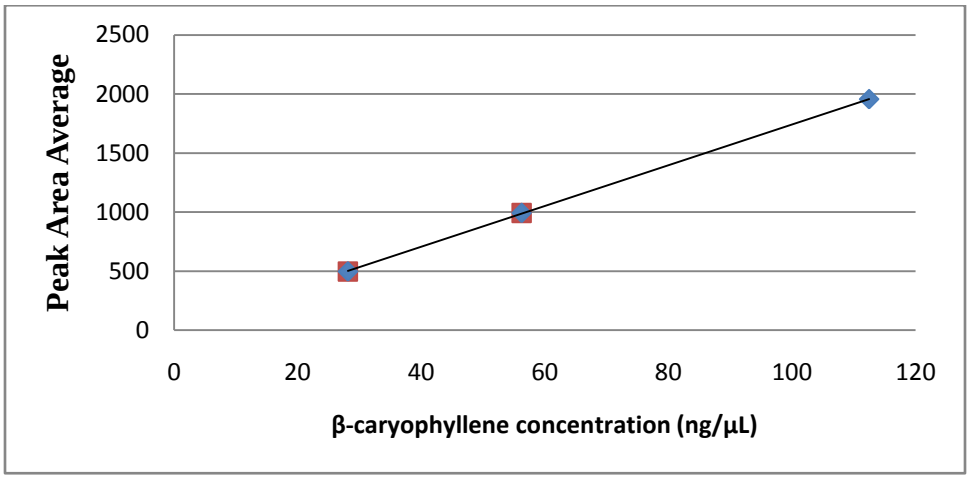

Figure 9. β-caryophyllene concentration standard curve. The standard curve equation was $y = 16.324x + 73.874$, with an R^2 value of 0.9999. The equation was generated from the data in Table 2 was used to calculate the concentration of experimental samples analyzed.

CHAPTER IV

RESULTS

Construction of the pPSBAIIQHS1 vector

The pET*QHS1* vector, originally developed by Cai *et al*. in 2002, contains a 1902 bp sequence originally isolated from the *A. annua* cDNA library (65). The researchers used degenerative primers based on known, conserved sequences to identify candidate sesquiterpene synthase cDNA clones (65). A cDNA clone was isolated that showed a high level of homology to the δ-cadinene synthase from *Gossypium arboreum* and the researchers determined this cDNA clone contained a 1644 bp open reading frame that would produce a 60 kD protein. This clone was subsequently ligated into the *E. coli* expression vector pET28b, which was then expressed in *E. coli*. The transgenic *E. coli* produced a soluble protein fraction that showed a high level of sesquiterpene activity based on positive assay results using [1-^{3}H] farnesyl diphosphate to determine *in vitro* radioactive incorporation of this sesquiterpene precursor (65). The ability of the isolated cDNA to produce a functional protein in a prokaryotic system made it an appropriate candidate for expression in *Synechocystis sp.* strain PCC6803 because of the apparent absence of any requisite post-translational modification.

Therefore, cloning primers were designed (Table 1) to amplify the *QHS1* 1644 bp open reading frame from the pETQHS1 and insert the sequence into the pPSABII *Synechocystis sp.* strain PCC6803 homologous recombination vector. In order to insert

the *QHS1* sequence into the multiple cloning site on the pPSBAII backbone, restriction sites were designed into the *QHS1* insert at the 5' (AseI) and 3'(BamHI) ends, immediately down/upstream to the start and stop codons, respectively. These restriction sites, when digested, produced sticky ends that were complementary to the existing 5' NdeI and 3' BamHI sites located between the upstream and downstream *psbAII* un-translated sequence within the pPSBAII vector. The restriction enzymes NdeI and AseI produce complimentary sticky ends, but the *QHS1* sequence has several NdeI recognition sites within the coding sequence, therefore AseI was used for the restriction digest. The fragmented pPSBAII-KS vector (Figure 6) was used as the source for the digested pPSBAII backbone (~4300 bp band) which was ligated with the digested *QHS1* amplicon. The resulting plasmid was transformed into *E. coli* and the colonies that grew were screened for *QHS1* retention (Figure 10 and Appendix A for plasmid image) (99). The pPSBAII$QHS1_{sq}$ sequencing primers were used to create two different amplicons: one that spanned the *psbAII* promoter region and the 5' region of the *QHS1* sequence and another that spanned the 3' termination region of the *QHS1* sequence and the un-translated downstream region of *psbAII*. The plasmid DNA was sequenced and the resulting sequences confirmed insertion and the correct orientation based on NCBI BLAST analysis that showed a near perfect match to the un-translated *psbAII* regions within the *Synechocystis sp.* strain PCC6803 genome and the *A. annua QHS1* gene (Appendix A).

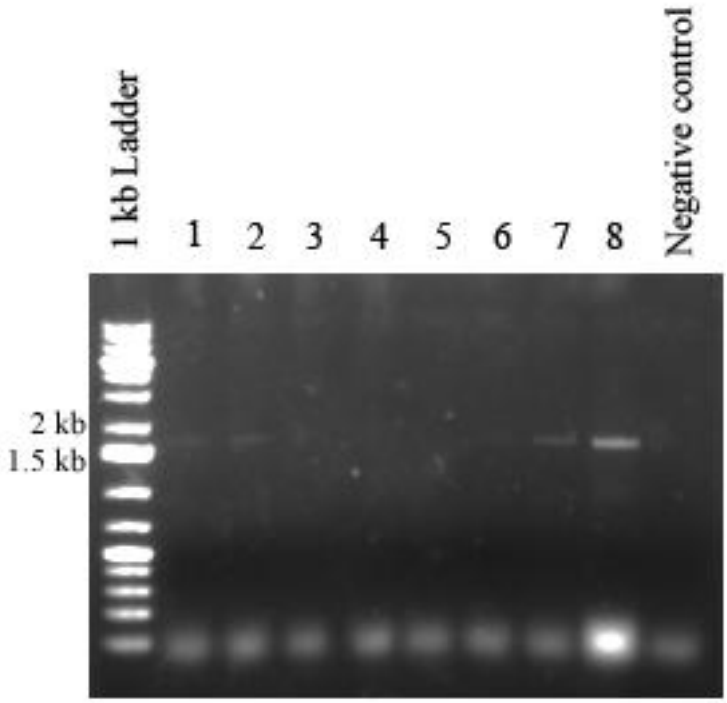

Figure 10. Agarose gel image of the *E. coli* colony PCR screening assays. The *E. coli* colony PCR assay was conducted using the pET*QHS1* primers as probes for the *QHS1* retention. Lanes 1-8 were loaded with PCR products in which DNA from colonies that successfully grew on ampicilin plates served as the template in the PCR assay. The *E. coli* colony that was used as a template for the Lane 8 reaction showed the correct banding size of approximately 1.6 kb and was selected for further plasmid extractions. The plasmid derived from this colony was used for the second homologous recombination reaction and a sequencing reaction.

Transformation and segregation of a transgenic Synechocystis strain

The double homologous recombination strategy for transgene expression in *Synechocystis* sp. strain PCC6803, as employed in this and other research, allows for targeted insertion of an exogenous gene at a specific site without retaining a selectable marker for continued transgene expression (Figure 8) (14, 51, 62). However, retention of the transgene is contingent on full segregation after each transformation reaction. Each

native copy of *psbAII* must be replaced with the *sacB/aphX* cassette during the first round of homologous recombination, in order to prevent repair and reversion back to the wild type genotype. After the first round of homologus recombination, the mutant *Synechocystis* colonies that grew on kanamycin plates were screened using a PCR assay with primers (Table 1) that flanked the *psbAII* insertion site. With this PCR assay, segregation could be determined as depicted in Figure 11.A, in which a fully segregated transformant has only one band at the *psbAII* site with a size corresponding to the inserted cassette. The *Synechocystis* strain transformed with the pPSABII-KS plasmid showed only one band representing the *sacB/aphX* cassette (~3750 bp) after full segregation (Figure 11.A).

The protocol for the first transformation is a slight modification to the optimal transformation protocol developed by Zang *et al.* in 2007 (61). In this experiment less plasmid DNA was used (3 μg of plasmid verse the optimized 15 μg of plasmid), due to low plasmid extraction efficiencies. Also the cell:plasmid reaction was allowed to incubate for 6 hours before proceeding to the kanamycin selection step, as opposed to the optimized 5 hour incubation suggested by Zang *et al.* (61). The pPSBAII*QHS1* vector also had shorter homologous flanking regions (500 bp verse the optimized 1700 bp), but the use and efficacy of this transformation vector's backbone (pPSBAII) was previously demonstrated by LaGarde *et al.*(14). The segregated strain, with the *sacB/aphx* cassette inserted, was maintained in kanamycin-selective liquid culture until the second homologous recombination reaction with the pPSBAII*QHS1* vector.

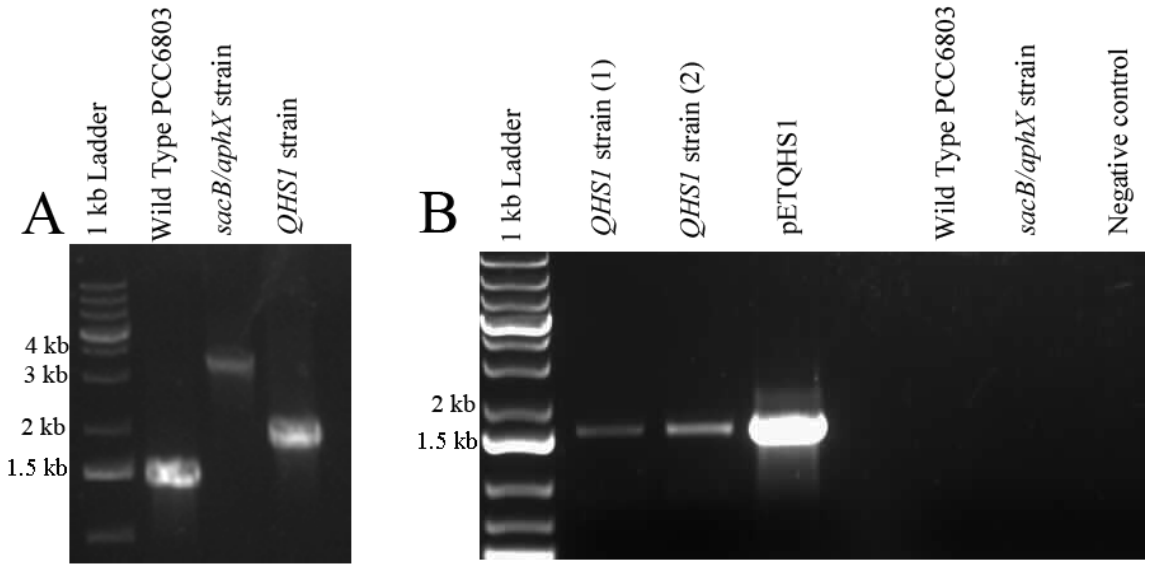

Figure 11. Segregation of mutant strains and conformation *QHS1* insertion. (A) Gel image from the segregation PCR assay using the 5'^{UP} pPSBAII*QHS1*$_{sq}$ and 3' ^{DOWN} pPSBAII*QHS1*$_{sq}$ primers as probes to determine recombinant segregation at the *psbAII* site. All three strains of *Synechocystis* used during this experiment produced only one band, indicating there was only one gene at the *psbAII* insertion site. (B) Gel image from PCR assays using pET*QHS1* primers as probes for retention of *QHS1* in transformed *Synechocystis*. The "*QHS1* strain (1)" lane represents a PCR reaction using DNA isolated from the original secondarily transformed *Synechocystis* colony that showed positive results for *QHS1* presence during the initial screening. The "*QHS1* strain (2)" lane represents a PCR assay with DNA extracted from a *Synechocystis* pPSBAII*QHS1* culture that had been routinely sub-cultured for 3 months after original positive screening. The pETQH1 band is from a PCR reaction using the pET*QHS1* as the DNA template, representing a positive control. The "Wild Type PCC6803" and "*sacB/aphX* strain" lanes were loaded with a PCR reaction using DNA extracted from each of respective strains as the template and the absence of bands indicate there are no endogenous genomic sites that the pET*QHS1* primers would anneal to in the absence of *QHS1* insertion. The "Negative Control" lane used water in lieu of DNA as a template.

During the second round of recombination, the *sacB/aphx* cassette was replaced with the *A. annua's QHS1* gene, in a similar genetic fashion as the first round of transformation. This transformation was driven by the selective pressure generated by the addition of 5 % sucrose to the growth media. Cells that retain the *sacB* gene produce

a secretory protein, levansucrase, when sucrose is added to the growth media. The combination of sucrose and levansucrase causes a lethal phenotype in cyanobacteria, as a result of either membrane receptor disruption or over accumulation of proteins in the cell membrane (93). Therefore, the double transformants were grown on 5% sucrose BG-11 agar plates to select against the retention of *sacB/aphX*. Colonies that grew on the sucrose media were screened using a PCR assay with the original cloning primers, pET*QHS1*, to detect *QHS1* insertion (Figrue 11.B). The colony that showed a consistent putative *QHS1* band was further screened to determine complete genomic segregation, using a similar PCR assay as described above. As seen in Figure 11.A, the double transformant showed only one band (~2100) corresponding to the *QHS1* gene inserted into *psbAII* site. Genomic DNA of the segregated double transformant was also sent out for sequencing to confirm the correct alignment and the resulting sequences confirmed the placement and sequence based on NCBI BLAST results (Appendix A).

The second transformation protocol is significantly different than the first transformation method, for several reasons. Preliminary secondary transformations showed a high number of transformants, beyond expected efficiencies, suggesting the selective pressure of the sucrose was not inducing lethality because of either poor genome incorporation or the lack of sucrose contact. It has been suggested the filter paper was creating a barrier on contact and thus not causing a selective effect (Dr. Wim Vermaas, personal communication, 2.18.10). It may also be a result of the actual material of the filter paper, which was more of a conventional paper media as opposed to the cellulose-acetate membrane filters used in other protocols (61). It was observed that the filter paper greatly inhibited the isolation of an individual colony because a

cyanobacterial lawn would often grow throughout the membrane, which made PCR colony assays very difficult. For these reasons, the concentrated cell:plasmid mixture was incubated in non-selective liquid BG-11, before the addition and subsequent incubation with 5% sucrose for several days. The liquid culture was then streaked on selective sucrose plates, further pressuring segregation of the transformants (see Figure 7 in Materials and Methods). This method has yet to be reported in the literature, but it appears to be sufficiently effective to produce segregated transformants, though the overall efficiency was not determined. Further analysis would have to be conducted to determine this protocol's efficacy for high throughput screening, but it was sufficient for these research purposes.

Confirmation of QHS1 transcription

Previously, researchers have shown the efficacy of the *psbAII* promoter to drive transgene expression (14, 51, 62). To determine if the *A. annua QHS1* sequence was being transcribed, RNA was extracted from the various strains and cDNA was generated for a reverse transcriptase PCR. As detailed in Figure 12, only the *QHS1* mutant showed any sign of the transcript being present when the $QHS1_{RT}$ primers were used (Figure 12.A) and yet all strains showed levels of the housekeeping transcript, rnPB (Figure 12.B). The retention and continual expression of the exogenous gene appears to be sustained, as RT-PCR and chemical analysis has been conducted on mutant strains that have been sub-cultured numerous times other the past 6 months, at this date of reporting (9/8/10).

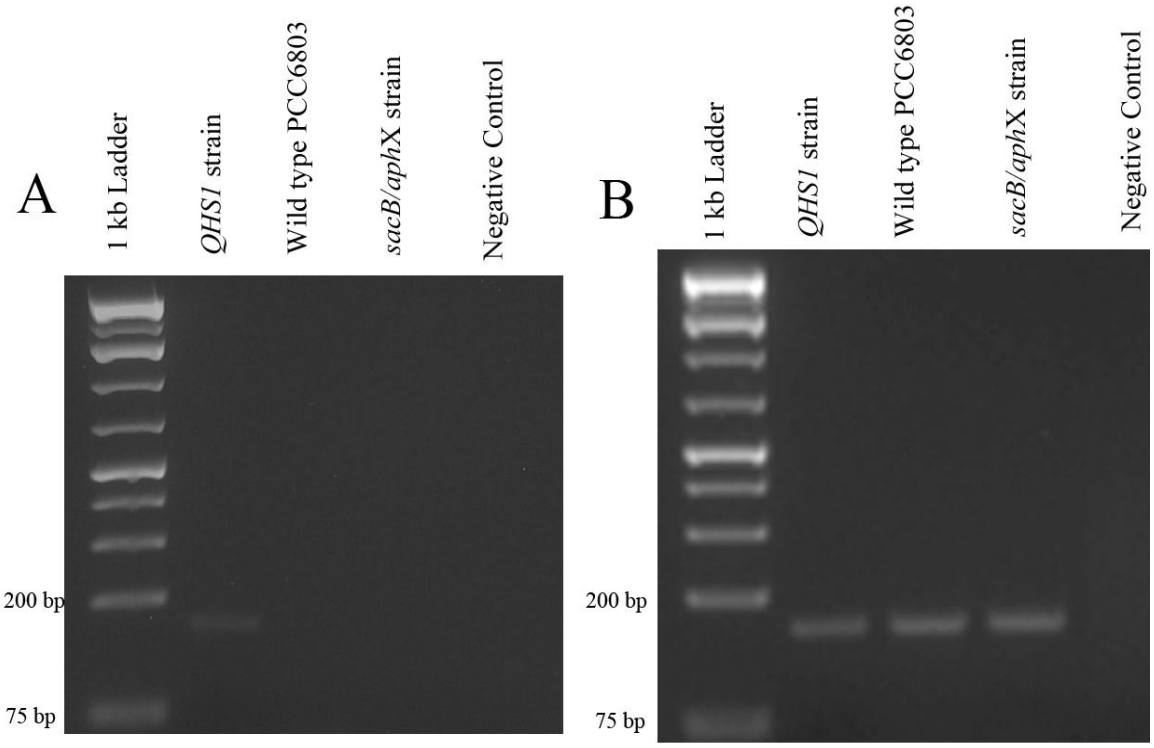

Figure 12. Reverse transcriptase PCR gel images. (A) RT-PCR results using $QHS1_{RT}$ primers as a probe for *QHS1* transcripts. The pPSBAII*QHS1* lane showed the only appropriately sized band, indicating that *QHS1* is only transcribed in the double recombinant *QHS1 Synechocystis* mutant. (B) RT-PCR results using $rnPB_{RT}$ primers as a probe for the constitutively present rnBP transcripts. All RT-PCR reactions using cDNA from each of the three strains showed positive results, indicating the RNA extraction and reverse transcription successfully produced sufficient amounts of cDNA. In both images, the Negative Control lane was loaded with a RT-PCR reaction prepared with water, instead of cDNA.

Production of β-caryophyllene by Mutant Synechocystis Strains

As discussed earlier, *Cai et al*. established that the *A. annua* β-caryophyllene synthase could be produced in a prokaryotic system and that this enzyme was active *in vitro* (65). However, it was not originally reported whether this enzyme was active *in vivo* or if it could even sufficiently compete with endogenous terpene pathways for the farnesyl diphosphate substrate needed to produce β-caryophyllene. In this experiment,

the *in vivo* activity of this enzyme was determined by the extraction of terpenes in both the *QHS1* strain and wild type strains and the analysis of the resulting extracts using GC-FID and GC-MS. In Figure 13, three different gas chromatograms are compiled together in order to highlight the presence of any suspected β-caryophyllene in the *QHS1* trangenic strain (13.C) and the absence of a β-caryophyllene peak in the wild type PCC6803 strain (13.B). The chromatogram in Figure 13.A represents the β-caryophyllene standard, which displays only a single peak with a retention time of 24.5 minutes. This peak was slightly asymmetric, which may have been a result of the GC column being overloaded with the standard. However, the resulting mass spectrum of the "Peak 1" in Figure 13.A has the same profile to the β-caryophyllene spectrum found in the NIST Mass Spectral Library (Figure 14.A). Therefore, despite the asymmetry of this peak, its retention time was used to determine the presence of any putative β-caryophyllene peak in either the wild type or *QHS1* extraction. There was no detectable peak present at 24.5 minutes on the wild type lipid extraction chromatogram (Figure 13.B), suggesting that no β-caryophyllene is produced in wild type *Synechocystis sp*. strain PCC6803. However there was a distinguishable peak at 24.5 minutes in the *QHS1* strain chromatogram (Peak 2 in Figure 13.C). The contrasting chromatograms provide strong evidence for the production of β-caryophyllene in *Synechocystis* as a result of the insertion of the *A. annua QHS1* gene. The identity of the putative β-caryophyllene peak (Peak 2) in Figure 13.C was further investigated by comparing its mass spectrum to the β-caryophyllene mass spectrum found in the NIST Mass Spectral Library (Figure 14.B). All of these data suggests that Peak 2 in Figure 14.C is β-caryophyllene, which provides further evidence for the presence of an active β-caryophyllene synthase *in vivo* in the

QHS1 trangenic strain. Under the normal growth conditions described previously and using a crude extraction method, a concentration of 46.4 +/- 2.9 ng of β-caryophyllene (mL of culture)$^{-1}$ (week)$^{-1}$ was obtained, which corresponds to approximately 3.7 µg (g of dry cell weight)$^{-1}$ (week)$^{-1}$. The concentrations were determined by comparing the area of the putative β-caryophyllene peak to a standard curve generated using known concentrations of the β-caryophyllene standard (see Materials and Methods).

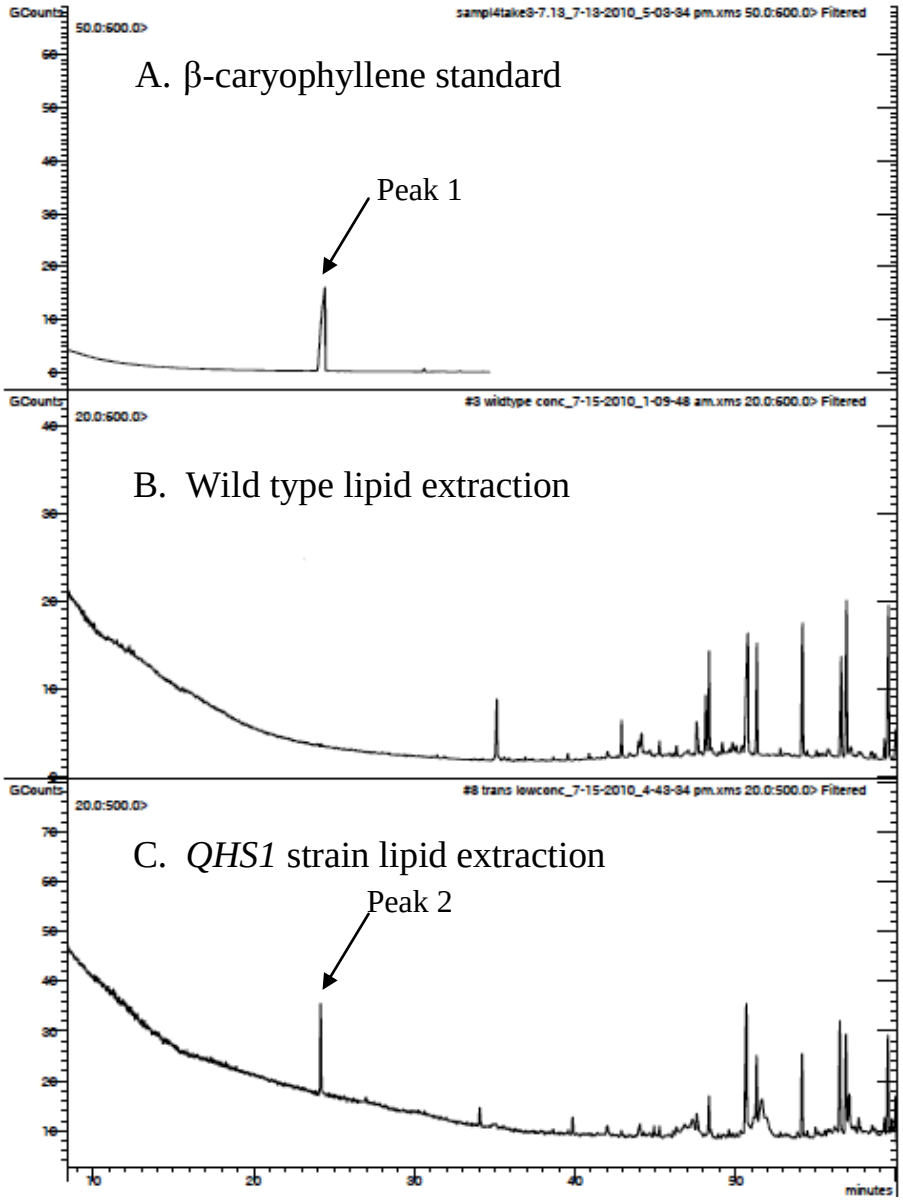

Figure 13. Gas chromatograms of the experimental extractions. (A) β-caryophyllene standard, (B) wild type lipid extraction, (C) QHS1 strain lipid extraction. The arrows represent putative β-caryophyllene peaks, with a retention time of 24.5 minutes. Note the absence of any peak at 24.5 minutes in the wild type extraction. The resulting mass spectral analysis of these peaks (1 and 2) is shown in Figure 14.

Cai *et al.* originally reported that the β-caryophyllene synthase, as expressed in *E. coli*, also produced minor amounts of α-humlene, another common sesquiterpene, but this not detected in this current experiment (Figure 15), (65). There were some unexpected differences that were observed between the wild type and *QHS1* strain lipid extraction. A comparison between the gas chromatograms in Figure 13.B and 13.C shows several wide peaks at 46 minutes and 53 minutes that are present in the *QHS1* strain but absent in the wild type. A more detailed image of the *QHS1* strain gas chromatogram revealed that these peaks were the fatty acids palmitic acid (46 min) and oleic acid (53 min) based on mass spectral analysis of these peaks and comparisons to the NIST database (solid arrows in Figure 15). This seems odd because theoretically, the *QHS1* gene should not have any impact on fatty acid metabolism. Further investigation would need to be conducted to determine if *QHS1* is having any impact on fatty acid metabolism or if this difference arose from a technical sampling error.

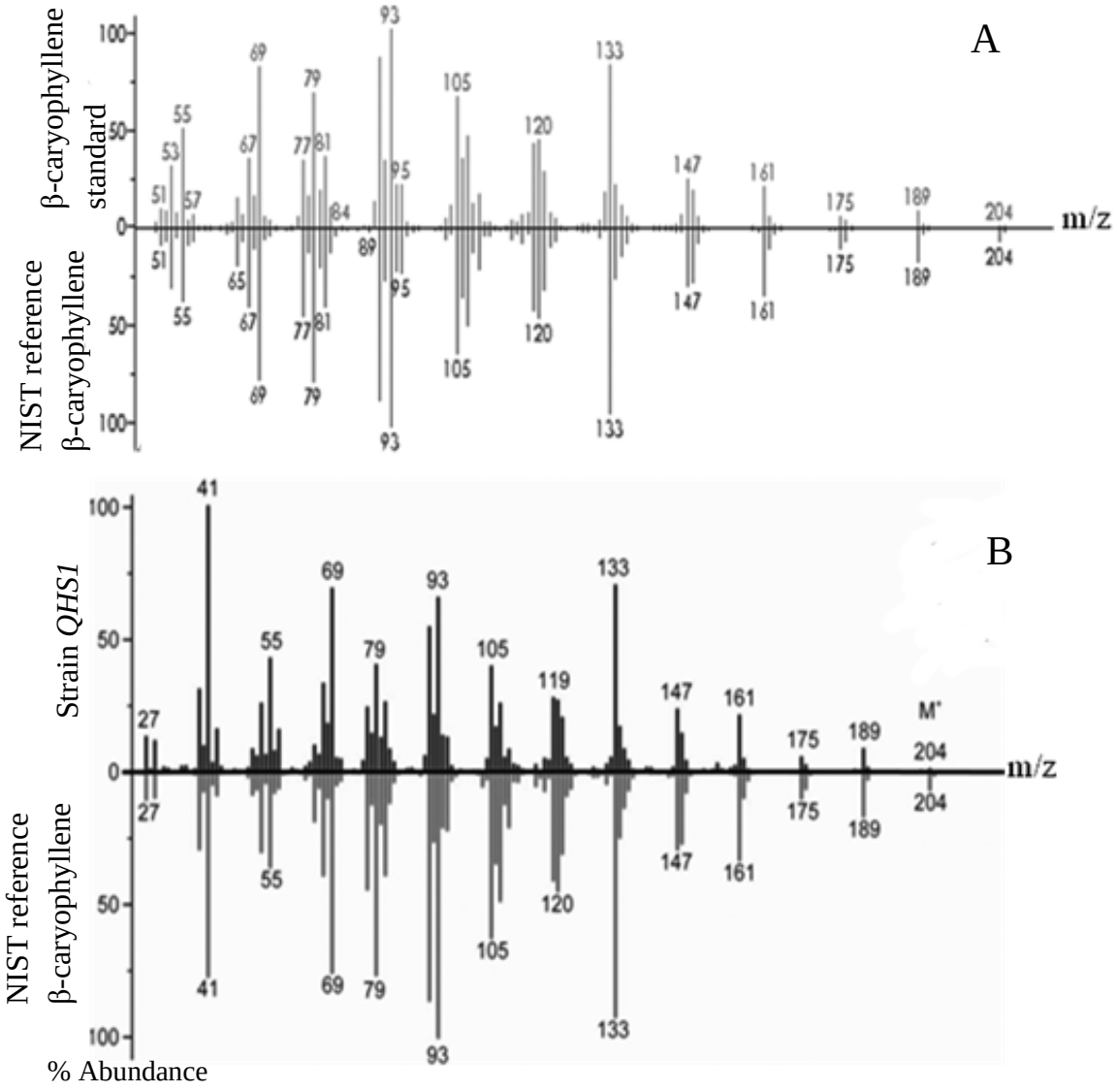

Figure 14. Mass spectrum comparisons. (A) The mass spectrum of the putative β-caryophyllene peak (Peak 1 Figure 13.A) from the β-caryophyllene standard aligned with and compared against the β-caryophyllene spectrum found in the NIST Mass Spectral Library. (B) The mass spectrum of the putative β-caryophyllene peak (Peak 2 Figure 13.B) from transgenic *QHS1* strain aligned with and compared against the reference mass spectrum of β-caryophyllene found in the NIST Mass Spectral Library. Note that the respective experimental spectra are not in the same horizontal (m/z) scale. However, both match almost perfectly with their scaled NIST standard comparison.

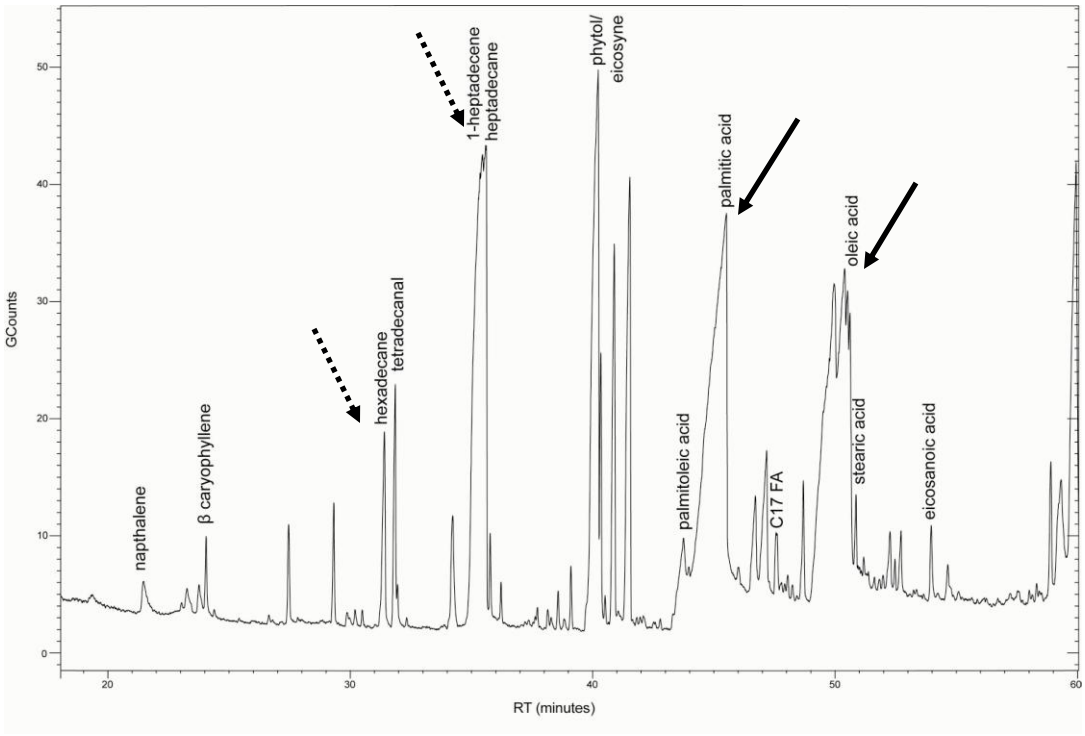

Figure 15. Detailed gas chromatogram from the *QHS1*strain. All identified molecules listed above prominent peaks were determined based on further mass spectral analysis and comparison to the NIST Mass Spectral Library. The dashed arrows indicate putative alkanes and alkenes. The solid arrows indicate the fatty acid peaks that are more conspicuous in the *QHS1* chromatogram than the wild type chromatogram.

Genetic analysis of potential farnesyl diphosphate synthases

As discussed in Chapter II, there is no reported farnesyl-diphosphate synthase (FPPS) found in the genome of *Synechocystis sp*. stain PCC6803 (31). However, based on the results of this research, there is strong evidence to suggest that *Synechocystis* does indeed contain an unidentified FPPS gene. In order for the transgenic strain to produce any amount of β-caryophyllene, which was reported here, there would need to be some underlying concentration of farnesyl diphosphate in order to drive a reaction through the β-caryophyllene synthase. Cai *et al*. reported that the *A. annua* β-caryophyllene synthase reacted poorly with geranyl diphosphate as a substrate *in vitro* with aK_M > 25 μM compared to a K_M of 1.8 μM with farnesyl diphosphate, and though it reacted with

geranyl diphosphate, no β-caryophyllene was produced (65). Therefore, there is indirect evidence of FPPS activity based on the production of β-caryophyllene with only the insertion of an exogenous β-caryophyllene synthase with no other metabolic manipulations.

Based on BLAST analysis and sequence alignments, there are two *Synechocystis* genes that have a significant degree of orthology to other FPPS genes. These gene are annotated as "solanesyl diphosphate synthase" (NCBI Accession BAA16579) and "geranylgeranyl pyrophosphate synthase" (GGPS) (NCBI Accession BAA16690.1). They have a significant amount orthology to other cyanobacterial FPPSs, as determined using the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST), which compares query sequences to the entire NCBI genomic database in order to find conserved regions (see Appendix B) (100). Both of these genes appear to categorized within the larger enzymatic family cd00685, the "*trans*-Isoprenyl Diphosphate Synthases," base on conserved Mg^{2+} binding domains, substrate binding pocket domain, active site lid domain, and the cataylitic site domain (100). This specific group of enzymes facilitates the 1'-4' condensation reaction of IPP with geranyl (10 carbons), farnesyl (15 carbons), and geranylgeranyl diphosphate (20 carbons) (101). The active site is distinct in this group with the presence of two aspartate-rich regions, located on anti-parallel alpha-helices that associate with the substrates allylic and homolylic components (101). Upon association of a prenyl pyrophosphate and the Mg^{2+}, the active site undergoes a conformational change, closing off the active site to any solvents and allowing the formation to occur (101). The *Synechocystis sp.* strain PCC6803 GGPS proteins appears to be the stronger candidate for the suspected farnesyl diphosphate

synthase, because it shares a stronger orthology to other Type II eubacterial FPPS, specifically at the aspartate-rich motif (FARM) in the Mg^{2+}binding domain (100). Bacterial FPPS are subcategorized as Type II FPPS based on the unique conservation amino acids before and within FARM, SLIHDDLPXMDX, that is absent in Eukaryotic and Archaeal FPPS (102) (100). This degree of sequence conservation is not present in the *Synechocystis sp.* strain PCC6803 solanesyl synthase, despite other site conservation among various active site domains. A NCBI BLAST search, using the conserved FARM domain of the Type II FPPS as the query search within the *Synechocystis* sp. strain PCC6803 genome, revealed only one protein, GGPS, to have the full conserved FARM domain (Appendix B). Clearly more research would need to be conducted to verify the identity of the active FPPS in *Synechocystis sp.* strain PCC6803, but GGPS seems like a good candidate to begin investigating based on this preliminary bioinformatic analysis (Appendix B).

Finally, there is growing interest in using microbial systems to produce fuel like alkanes and alkenes (103). In Figure 16, there are two relatively large peaks (dashed arrows) that correspond to a hexadecane and heptadecene, as determined from their respective mass spectra and comparison to the NIST database. This work supports the recent evidence presented by Schirmer *et al.*, who were the first to confirm alkane/alkene production in *Synechocystis sp.* strain PCC6803, as well as identify the gene responsible for alkane/alkene production in *Synechocystis sp.* strain PCC6803 (103). The presence of these large hydrocarbon peaks offers another intriguing avenue of research.

CHAPTER V

DISCUSSION AND CONCLUSIONS

The ability to transform *Synechocystis sp.* strain PCC6803 for the production of β-caryophyllene, as detailed in this paper, is a promising step towards developing strains of autotrophic microorganisms that can produce important terpenoids on a commercial scale. This is of relevance to both the biofuel and pharmaceutical communities. In 2006, Chang and Keasling highlighted the potential of *E. coli* and *Saccharomyces cerevisiae* as microbial platforms for pharmaceutical isoprenoid production (104). The authors make the case that a designed microbial system could avoid many of the short falls of synthetic chemical production or natural source extraction (104). More recently, Radakovits et al. have reviewed the potential of engineering phototrophic microalgae for the accumulation of isoprenoids, particularly C10 and C15 products, as these hydrocarbons are excellent diesel fuel surrogates (7). My current work combines these preceding approaches, as it incorporates the production of important terpenoids into an autotrophic microorganism. To my knowledge, this is the first reported strain of *Synechocystis sp.* strain PCC6803 to produce any sesquiterpene.

However, the β-caryophyllene production levels reported here are quite low and this designed system would need to be further optimized to justify continued use of this platform. There are a number of potential areas in which optimization could be pursued. Simple modifications to the growth system may provide significant gains. Terpenes and

lipids were extracted from the transgenic *Synechocystis* liquid cultures growing in optimal conditions for genetic transformations and may not necessarily be optimal for terpene production. Controlled experiments testing various pHs, temperatures, and light conditions may reveal a significant increase in β-caryophyllene. It has been reported that in dark-adapted, glucose fed *Synechocystis sp.* strain PCC6803 cells, carotenogensis is increased as a result of high cytoplasmic pH levels (105). A fluctuating pH strategy may be able to stimulate terpene synthesis continuously in large outdoor production facilities, even under low/no light growth conditions. Further, using a shaker table does not provide optimal conditions for the absorptions of metabolically important gases (O_2 and CO_2), therefore using a direct injected gas bubbler or a continuous culture reactor could increase growth rates and even terpene concentrations through increased gas:cell contact. Under nitrogen starvation conditions, the accessory light-harvesting protein structures, phycobilisomes, are often degraded to release more usable nitrogen into the cytoplasm and the cell responds with an increase in carotenoid accumulation in order to maintain sufficient photosynthetic capacity (106). It has been reported that higher light intensities stimulate carotenogenesis, suggesting an increase of carbon flux through the MEP pathway, perhaps leading to an increase in the β-caryophyllene precursor farnesyl diphosphate (31). The current system implemented in this work is designed around the light activated *psbAII* promoter. In previous experiments, this promoter has been shown to increase transgene expression as light intensities rise (51), therefore growing this transgenic strain in higher light intensities could increase the number of *QHS1* transcripts.

Modifying the coding sequence of the native *A. annua* gene may also significantly increase the final synthase concentration. As discussed earlier, *Synechocystis sp.* strain PCC6803 does not display as severe of a codon bias as other photosynthetic model organism, such as *Chlamydomonas reinhardtii* (57, 107). However, recent transgenic research has shown a significant increase in protein levels when exogenous genes are modified to better resemble the native codon usage of *Synechocystis sp.* strain PCC6803 (51). Lindberg et al. synthetically modified the native *IspS* coding sequence from *Pueraria montana,* prior to transformation, in order eliminate all codons that occurred less than 12% in the *Synechocystis sp.* strain PCC6803 tRNA pool (51). This resulted in a ten-fold increase in modified *IspS* translation, compared to the native *P. montana* sequence translation, based on quantified western blot analysis (51). This method also resulted in a significant increase in overall isoprene emission within this modified transgenic strain (51). There are complementary tRNAs within the *Synechocystis sp.* strain PCC6803 genome for all codons presented in the *A. annua QHS1* coding sequence (Appendix C) (108). However, if a 12% codon occurrence cutoff were used, as established by Lindberg et al., the native *A. annua* sequence would need to be significantly modified because there are a number of codons that fall below the 12% cutoff, with some occurring as low as 2% (Appendix C) (51). The coupled approach of increasing the chemical precursors through optimized terpene production and the optimization of exogenous gene transcription/translation, may be able to significantly increase the yield of β-caryophyllene.

The previous discussion focused on suggesting direct methods for optimizing the current system, but these are only modest. They do not address some fundamental

hurdles that face a microbial platform producing a secondary metabolite, such as β-caryophyllene. As Linderberg *et al.* discussed, the carbon flux in *Synechocystis sp.* strain PCC6803 during photosynthesis is highly favored toward carbohydrate accumulation (80-85%), and then lipid accumulation (~10%), and then finally terpenoid synthesis (3-5%) (51). Therefore any platform for the production of terpenes at titers high enough for commercial production needs to be designed in such a way as to increase the carbon flux through the isoprenoid pathways. Key to this approach is first understanding the interconnected metabolic pathways and the associated signaling network that influences metabolism. From this, strategies can be developed to reroute endogenous pathways for producing desired exogenous metabolites. Genomic investigation into glycogen/starch catabolism and the resulting carbon flux has been reported in several research papers, yet very little has been reported on any resulting terpene changes (109, 110). In 2010, a comprehensive photosynthetic metabolic profile was developed for *Synechocystis sp.* strain PCC 6808 (18). The researchers complied decades worth of work in fields ranging from genomics to biochemistry, providing a thorough flux-balance analytical model to explore the primary metabolism of this strain (Figure 16), (18). The term 'flux-balance analysis' (and more generally 'systems biology') are relatively new concepts within molecular biology and bioengineering, but the ultimate goal is to provide a characterization of large metabolic networks under steady-state conditions in order to maximize the production of a particular metabolite (18). Despite this recent, thorough profiling of the *Synechocystis sp.* strain PCC6803 metabolic network, terpenoid synthesis is largely ignored (Figure 16), (18). This is disappointing, given the evidence of the interconnected nature of the MEP pathway, photosynthesis, and reliance of efficient

photosystems on resulting terpenoid products (see Chapter II). However, this gap in understanding does provide an excellent area for continued research since there are already some results that can be incorporated into this existing model.

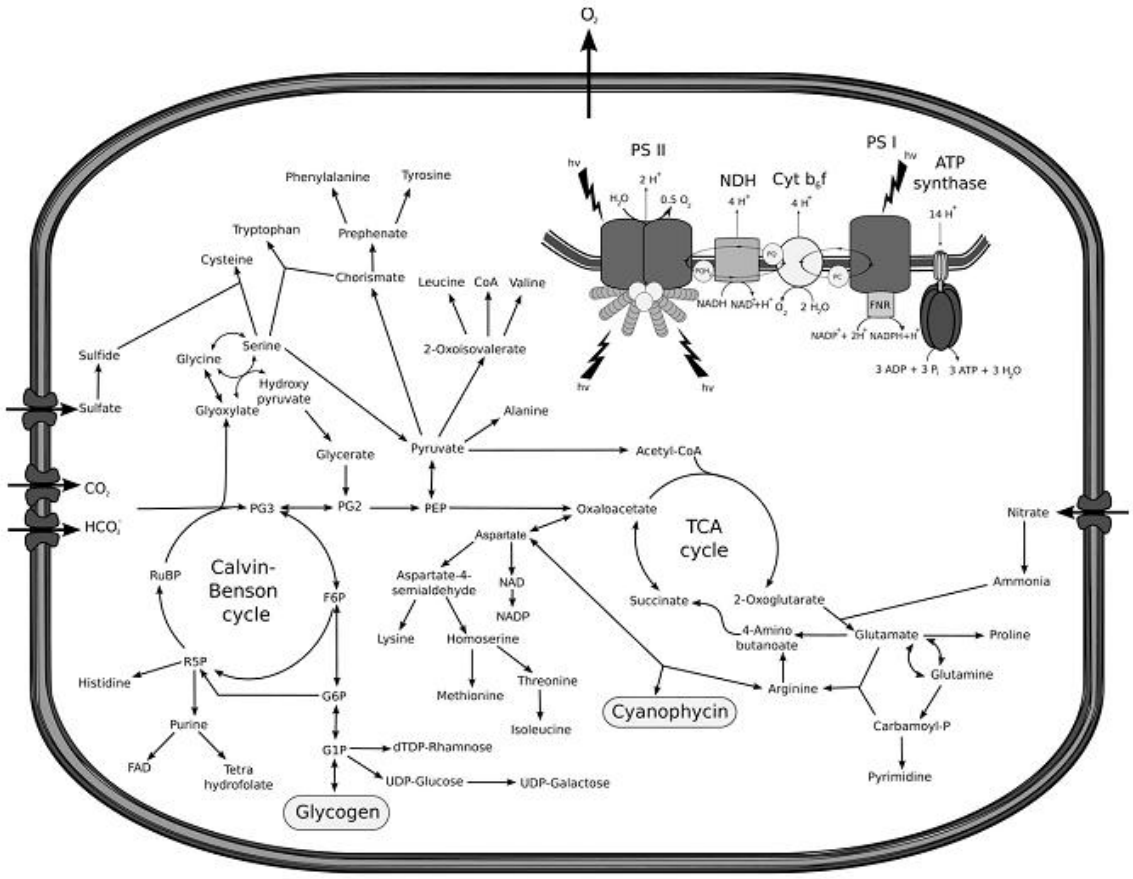

Figure 16. The primary metabolic networks of *Synechocystis* sp. strain PCC 6803 A graphic representation of the photosynthetic pathway of *Synechocystis sp*. strain PCC6803. Image is modified from Knoop et al. (18)

Another interesting area that has shown to effectively reroute terpenoid precursors into other products is the use of mutant knockouts. In 2000, LeGarde *et al*. produced a *Synechocystis* knockout strain lacking the gene *crtO*, which codes for an enzyme responsible for the production of the carotenoid echinenone. This mutant strain produced no echinenone, but produced higher overall levels of total carotenoids and specifically

myxoxanthophyll, without any perceived physiological deficiencies (14). Using a systems biology approach, endogenous terpenoids with no reported physiological role could be identified and if their synthase genes were deleted, an increased flux through the β-caryophyllene synthase pathway may occur. However, there is a risk of shunting too many terpenoid precursors toward a non-essential product, such as β-caryophyllene. The longer-chained carotenoids play a crucial role in photosynthesis and oxidative protection in *Synechocystis sp.* strain PCC6803, therefore any increased production of smaller sesquiterpenes may limit the pool of available precursors needed in important downstream terpenoid synthesis, leading to physiologically impaired strains (41). However, if the cell responded to a genetic barrier for carotenoid accumulation (via knockouts) through an increase in phycobilisome assembly, potential physiological deficiencies may be avoided, as the light harvesting apparatus could be maintained through the diverse phycobilisome pigments absorbing any excess light otherwise handled by carotenoids. This could become a precarious balancing act, since chlorophyll A, the primary photoreceptor in the photosystems, is assembled in part from terpene precursors. Therefore any inhibition of chlorophyll A production could be even more deleterious to the cell. Also, a reduction of carotenoids could increase oxidative damage to the cell. Just as the production of the terpenoid precursors seems to be interconnected and reliant upon a number of sources, the absence or rerouting of terpenoid precursors could have an effect on various systems. An increased understanding of the role of terpenoids in the cell and any resulting cellular response to lower levels is needed to better develop strategies for the production of any exogenous terpenoid.

The work begun in this thesis and the resulting discussion displays the immensely complicated nature of a relatively simple project. Fundamentally, this was an exercise of inserting one gene into a genome with the goal of producing one product from an already existent precursor. However, nothing occurs in isolation inside a cell, therefore the resulting upstream and downstream effects of any changes to the native system, no matter its perceived simplicity, must be investigated to fully appreciate the overall impacting consequences.

References

1. Grahm, L., J. Grahm, & L. Wilcox. 2009. *Algae*. Pearson Edducation, San Francisco, CA

2. Cardozo, K., T. Guaratini, M. Barros, V. Falcão, A. Tonon, N. Lopes, S. Campos, M. Torres, A. Souza, P. Colepicolo, & E. Pinto. 2007. Metabolites from algae with economical impact.. Comparative Biochemistry and Physiology 146: 60–78.

3. Reitan, K., J. Rainuzzo, G. Oie, and Y. Olsen. 1997. A review of the nutritional effects of algae in marine fish larvae. Aquaculture 155: 207-221.

4. Khan, Z., Z. Begum, R. Mandall, and M. Hossain. 1994. Cyanobacteria in rice soils. World Journal of Microbiology and Biotechnology 10: 296-298.

5. Lembi, C. and J. Waaland. 1988. Algae and Human Affairs. Cambridge University Press, New York, NY.

6. Leary, D., M. Vierros, G. Hamon, S. Arico, C. Monagle. 2009. Marine genetic resources: A review of scientific and commercial interest. Marine Policy. 33: 183–194.

7. Radakovits, R., R. Jinkerson, A Darzins, and M. Posewitz. 2010. Genetic engineering of algae for enhanced biofuel production. Eukaryotic Cell 9: 486-501.

8. Leon-Banares, R., D. Gonzalez-Ballester, A. Galvan, and E. Fernandez. 2004. Transgenic microalgaeas as green cell-factories. TRENDS in Biotechnology. 22: 45-52.

9. Coll, J. M. 2006. Review: Methodologies for transferring DNA into eukaryotic microalgae. Spanish Journal of Agricultural Research. 4: 316-330.

10. Zang, X., B. Lui, S. Liu, K. Arunakumara, and X. Zhang. 2007. Optimum Conditions for Transformation of *Synechocystis sp*. PCC 6803. Jounral of Microbiology. 45: 241-245.

11. Chen, Z., L. Ren, Q. Shao, D. Shi, B. Ru. 1999. Expression of Mammalian Metallothionein-I Gene in Cyanobacteria to Enhance Heavy Metal Resistance. Marine Pollution Bulletin. 39: 155-158.

12. . Mayfield, S. and S. Franklin. 2005. Expression of human antibodies in eukaryotic micro-algae. Vaccine. 23 : 1828–1832.

13. Griesbeck, C., I. Kobl, and M. Kobl. 2006. *Chlamydomonas reinhardtii*-A Protein Expression System for Pharmaceutical and Biotechnological Proteins. Molecular Biotechnology. 34: 213-223.

14. LaGarde, D., L. Beuf, and W. Vermaas. 2000. Increased Production of Zeaxanthin and Other Pigments by Application of Genetic Engineering Techniques to *Synechocystis sp*. Strain PCC 6803., Applied and Environmental Microbiology. 66: 64–72.

15. Leon, R., I. Couso, and E. Fernandez. 2007. Metabolic engineering of ketocarotenoids biosynthesis in the unicelullar microalga *Chlamydomonas reinhardtii*.. Journal of Biotechnology. 130: 143–152.

16. Liu, S., X. Zhang, X. Zang, B. Liu, K. Arunakumara, D. Xu, and X. Zhang. 2008. Growth, feed efficiency, body muscle composition, and histology of flounder (*Paralichthys olivaceus*) fed GH transgenic *Synechocystis.* Aquaculture. 277: 78–82.

17. . Rupprecht, J. 2009. From systems biology to fuel—*Chlamydomonas reinhardtii* as a model for a systems biology approach to improve biohydrogen production. Journal of Biotechnology. 142: 10-20.

18. Knoop, H., Y. Zilliges, W. Lockau, and R. Steuer. 2010. The Metabolic Network of *Synechocystis sp.* PCC 6803: Systemic Properties of Autotrophic Growth. Journal of Plant Physiology. 154: 410–422.

19. Rippka, R., J. Deruelles, J. Waterbury, M. Herdman, and R. Stanier. 1979 Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. Journal of General Microbiology. 111: 1-61.

20. Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirosawa, M. Sugiura, S. Sasamoto, T. Kimarura, T. Hosouchi, A. Matsuno, A. Muraki, N. Nakazaki, K. Naruo, S. Okumura, S. Shimpo, C. Takeuchi, T. Wada, A. Watanabe, M. Yamada, M. Yasuda, and S. Tabata. 1996. Sequence Analysis of the Genome of the Unicellular Cyanobacterium *Synechocystis sp*. Strain PCC6803. II. Sequence Determination of the Entire Genome and Assignment of Potential, Protein-coding Regions. DNA Research. 3: 109-136.

21. Yoshihara, S., X. Geng, S. Okamoto, K. Yura, T. Murata, M. Go, M. Ohmori, and M. Ikeuchi. 2001 Mutational Analysis of Genes Involved in Pilus Structure, Motility and Transformation Competency in the Unicellular Motile Cyanobacterium *Synechocystis sp.* PCC 6803. Plant Cell Physiology. 42: 63–73.

22. Brahamsha, B. 1996. An abundant cell-surface polypeptide is required for swimming by the nonflagellated marine cyanobacterium *Synechococcus.* Proceedings of the National Acadamey of Sciences. 93: 6504–6509.

23. Kaneko, T. and S. Tabata. 1997. Complete Genome Structure of Unicellular Cyanobacterium *Synechocystis sp.* PCC6803. Plant Cell Physiology. 11: 1171-1176.

24. Nakao, M., S. Okamoto, M. Kohara, T. Fujishiro, T. Fujisawa, S. Sato, S. Tabata, T. Kaneko, and Y. Nakamura. 2010. CyanoBase: the cyanobacteria genome database update 2010.., Nucleic Acid Research. 38: D379–D381.

25. Sagan, Lynn. 1967. On othe Origin of Mitosing Cells. Journal of Theoretical Biology. 14: 225-274.

26. . Keeling, Patrick J. 2004. Diversity and Evolutionary Historyof Plastids and Their Hosts. American Journal of Botany. 10: 1481–1493.

27. Douglas, S. 1994. Plastid Evolution. p.91-118. *In* D. Bryant. The Molecular Bioology of Cyanobacteria-. Kluwer Academic Publishers, Boston, MA.

28. Roberts, Susan C. 2007. Production and engineering of terpenoids in plant. Nature Chemical Biology. 3: 387-395.

29. Gershenzon, J. and N. Dudareva. 2007. *The* function of terpene natural products in the natural world. Nature Chemical Biology. 3: 408-414.

30. Wanke M., K. Skorupinska-Tudek, and E.Swiezewska. 2001. Isoprenoid biosynthesis via 1-deoxy-D-xylulose5-phosphate/2-C-D-erythritol4-phosphate (DOXP/MEP) pathway. Acta Biochimica Polonica. 48: 663-672.

31. Takaichi, S. and M. Mochimaru. 2007. Carotenoids and cartoengenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. Cellular and Molecular Life Sciences. 64: 2607-2619.

32. Mohamed, H. E., A. M. L. van de Meene, R. W. Roberson, W.F. J. Vermaas. 2005. Myxoxanthophyll Is Required for Normal Cell Wall Structure and Thylakoid Organization in the Cyanobacterium *Synechocystis sp.* Strain PCC 6803. Journal of Bacteriology. 187: 6883–6892.

33. Chemovitz, D., J. Hirschberg and D. Bryant. 1994. Carotenoids in Cyanobaceria. p.559-579. *In* D. Bryant. The Molecular Biology of Cyanobacteria. Kluwer Academic Publishers, Boston, MA.

34. Gantt, E. 1994. Membrane Organization. p. 119-138. *In* D. Bryant. The Molecular Biology of Cyanobacteria. Kluwer Academic Publishers, Boston, MA.

35. Lange, B.M., T. Rujan, W. Martin, and R. Croteau. 2000. Isoprenoid biosynthesis: The evolution of two ancient and distinct pathways across genomes. Proceedings of the National Acadamey of Sciences. 97: 13172–13177.

36. Rohdich, F., K. Kis, A. Bacher, and W. Eisenreich. 2001. The non-mevalonate pathway of isoprenoids: genes, enzymes, and intermediates. Current Opinion in Chemical Biology. 5: 535–540.

37. Proteau, J. 1998. Biosynthesis of Phytol in the Cyanobacterium *Synechocystis sp.* UTEX 2470: Utilization of the Non-Mevalonate Pathway. Journal of Natural Products. 61: 841-843.

38. Rodriguez-Concepcion, M., and A. Boronat. 2002. Elucidation of the Methylerythritol Phosphate Pathway for Isoprenoid Biosynthesis in Bacteria and Plastids. A Metabolic Milestone Achieved through Genomics. Plant Phsyiology. 130: 1079–1089.

39. Ershov, Y., R. R. Gantt, F. X. Cunningham, and E. Gantt. 2000. Isopentenyl diphosphate isomerase deficiency in *Synechocystis sp.* strain PCC6803. FEBS Letters. 473: 337-340.

40. Barkley, S. J., S. B. Desai, and C. D. Poulte. 2004. Type II Isopentenyl Diphosphate Isomerase from *Synechocystis sp.* Strain PCC 6803. Journal of Bacteriology. 186: 8156–8158.

41. Poliquin, K., F. Cunningham Jr., I. MacDonald, R. Gantt, and E. Gantt. 2008. Impaired Isoprenoid Biosynthesis: A Competitive Disadvantage Under Light Stress in *Synechocystis* PCC 6803. P. 763-766. *In*. J., E. Gantt, J. Golbeck, and B. Osmond Allen. Photosynthesis. Energy from the Sun. Springer, Netherlands

42. Phillips, M., P. Leon, A. Boronat, an M. Rodriguez-Concepcion. 2008. The plastidial MEP pathway: unified nomenclature and resources. Trends in Plant Science. 13: 619-623.

43. Kajiwara, S., P. D. Fraser, K. Kondo, and N. Misawa. 1997. Expression of an exogenous isopentenyl diphosphate isomerase gene enhances isoprenoid biosynthesis in *Escherichia coli*.. Journal of Biochemistry. 324: 421-426.

44. Hahn, F.M., Hurlbut, A.P. and Poulter, C.D. 1999. *Escherichia coli* Open Reading Frame 696 Is idi, a Nonessential Gene Encoding Isopentenyl Diphosphate Isomerase. Journal of Bacteriology. 181: 4499–4504.

45. Cunningham, F. X., T. P. Lafond, and E. Gantt. 2000. Evidence of a Role for LytB in the Nonmevalonate Pathway of Isoprenoid Biosynthesis. Journal of Bacteriology. 182:5841–5848.

46. . Altinciceka, B., E. C. Duin, A. Reichenberg, R. Hedderich, A. Kollas, M. Hintz, S. Wagner, J. Wiesner, E. Beck, and H. Jomaa. 2002. LytB protein catalyzes the terminal step of the 2-C-methyl-D-erythritol-4-phosphate pathway of isoprenoid biosynthesis. FEBS Letters. 532: 437-440.

47. Grawert, T., I. Span, W. Eisenreich, F. Rohdich, J. Eppinger, A. Bacher, M. Groll. 2010. Probing the Reaction Mechanism of IspH Protein by X-ray Structure Analysis. Proceedings of the National Acadamey of Sciences. 107: 1077-1081.

48. Ershov, Y. V., R. R. Gantt, F. Cunningham, and E. Gantt. 2002. Isoprenoid Biosynthesis in *Synechocystis sp.* Strain PCC6803 Is Stimulated by Compounds of the Pentose Phosphate Cycle but Not by Pyruvate or Deoxyxylulose-5-Phosphate. Journal of Bacteriology. 184: 5045–5051.

49. Fellermeier, M., K. Kis, S. Sagner, U. Maier, A. Bacher, and M. H. Zenk. 1999. Cell-free conversion of 1-deoxy-D-xylulose 5-phosphate and 2-C-methyl-D-erythritol 4-phosphate

into 13-carotene in higher plants and its inhibition by fosmidomycin. Tetrahedron Letters. 40: 2743-2746.

50. Okada, K. and T. Hase. 2005. Cyanobacterial Non-mevalonate Pathway: (E)-4-Hydorxy-3-Methylbut-2-enyl-diphsosphate Synthase Interacts with Ferredoxin in *Thermosynechoccous elongatuus* BP-1. Jounral of Biological Chemistry. 280: 20672–20679.

51. Lindberg, P., S. Park, and A. Melis. 2010. Engineering a platform for photosynthetic isoprene production in cyanobacteria, using *Synechocystis* as the model organism. Metabolic Engineering. 12: 70–79.

52. Vermaas, W. 1996. Molecular genetics of the cyanobacterium *Synechocystis sp*. PCC 6803:Principles and possible biotechnology applications. Journal of Appliced Phycology. 8: 263-273.

53. Munekage, Y., M. Hashimoto, C. Miyake, K. Tomizawa, T. Endo, M. Tasaka, and T. Shikanai. 2004. Cyclic electron flow around photosystem I is essential for photosynthesis. Nature. 429: 579-582.

54. Mohamed A., J. Eriksson, H. D. Osiewacz, and C. Jansson. 1993. Differential expression of the *psbA* genes in the cyanobacterium *Synechocystis* 6803. Molecular Genetics and Genomics. 238: 161-168.

55. Grigorieva, G., and S. Shestakov. 1982. Transformation in the cyanobacterium *Synechocystis sp.* 6803. FEMS Microbiology Letters. 13: 367-370.

56. Vermaas, W. 2004. Targeted Genetic Modification of Cyanobacteria: New Biological Applications. p.457-470. *In* A. Richmond, Handbook of Microalgal Culture: Biotechnology and Applied Phycology. Blackwell Science, Ames, IA.

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

58. Alberts, B, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. 2008. Molecular Biology of the Cell-Fifth Edition. Garland Science, New York, NY

59. Sode, K., M. Tatara, H. Takeyama, J. G. Burgess, and T. Matsunaga. 1992. Conjugative gene transfer in marine cyanobacteria: *Synechococcus sp., Synechocystis sp. and Pseudanabaena sp.* Applied Microbiology and Biotechnology. 37: 369-373.

60. LaBarre, J., F. Chauvat, and P. Thuriaux. 1989. Insertional Mutagenesis by Random Cloning of Antibiotic Resistance Genes into the Genome of the Cyanobacterium *Synechocystis* Strain PCC 6803., Journal of Bacteriology. 171: 3449-3457.

61. Zang, X., B. Liu, S. Liu, P. Sun, X. Zhang, and X.C. Zhang. 2007. Transformation and expression of Paralichthys olivaceus growth hormone cDNA in *Synechocystis sp.* PCC6803. Aquaculture. 266: 63–69.

62. Dexter, J. and P. Fu. 2009. Metabolic engineering of cyanobacteria for ethanol production. Energy & Environmental Science. 2: 857–864.

63. Deng, M. and J. Coleman. 1999. Ethanol Synthesis by Genetic Engineering in Cyanobacteria. Applied and Environmental Microbioliogy. 65: 523–528.

64. R. P. Woods, J. R. Coleman, and M. de Deng. April, 2002. Genetically modified cyanobacteria for the production of ethanol, the constructs and method thereof. US patent: 6306639

65. Cai, Y., J.W. Jia, J. Crock, Z.X. Lin , X.Y. Chen, and R. Croteau. 2002. A cDNA clone for β-caryophyllene synthase from *Artemisia annua.* Phytochemistry. 61: 523-529.

66. Gertsch, J., M. Leonti, S. Raduner, I. Racz, J. Chen, X. Xie, K. Altmann, M. Karsak, and A. Zimmer. 2008. β-caryophyllene is a dietary cannabinoid., Proceedings of the National Acadamey of Sciences. 105: 9099–9104.

67. Dewick, P. 2002. The biosynthesis of C5–C25 terpenoid compounds., Natural Product Reports Articles. 19: 181–222.

68. Kollner, T., M. Held, C. Lenk, I. Hiltpold, T. Turlings, J. Gershenzon, and J. Degenhardt. 2008. A Maize(E)- β-Caryophyllene Synthase Implicated in Indirect Defense Responses against Herbivores Is Not Expressed in Most American Maize Varieties. The Plant Cell. 20: 482–494.

69. Cheng, A., C. Xiang, J. Li, C. Yang, W. Hu, L. Wang, Y. Lou, and X. Chen. 2007. The rice (E)-β-caryophyllene synthase (OsTPS3) accounts for the major inducible volatile sesquiterpenes. Phytochemistry. 68: 1632–1641.

70. Mercke, P., I. Kappers, F. Verstappen, O. Vorst, M. Dicke, H. Bouwmeester. 2004. Combined Transcript and Metabolite Analysis Reveals Genes Involved in Spider Mite Induced Volatile Formation in Cucumber Plants. Plant Physiology. 135: 2012–2024.

71. Dehal, S. and R. Croteau. 1988. Partial Purification and Characterization of Two Sesquiterpene Cyclases. Archives of Biochemistry and Biophysics. 261: 346-356.

72. Wang, R., S. Peng, R. Zeng, L. Ding, and Z. Xu. 2009. Cloning, expression and wounding induction of β-caryophyllene synthase gene from *Mikania micrantha H.B.K*. and allelopathic potential of β-caryophyllene. Allelopathy Journal. 24: 35-44.

73. Legault, J. and A. Pichette. 2007. Potentiating effect of β-caryophyllene on anticancer activity of a-humulene, isocaryophyllene and paclitaxel. Journal of Pharmacy and Pharmacology. 59: 1643–1647.

74. Ruilong, W., P. Shaolin, Z. Rensen, D. Wen, and X. Zengfu. 2009. Cloning, expression and wounding induction of β-caryophyllene synthase gene from *Mikania micrantha* H.B.K. and allelopathic potential of β-caryophyllene. Allelopathy Journal. 24: 35-44.

75. Rodriquez-Saona, C., S. Crafts-Brandner, L. Williams, and P. Pare. 2002. *Lygus hesperus* feeding and salivary gland extracts induce volatile emissions in plants. Journal of Chemical Ecology. 28: 1733-1747.

76. De Moraes, C., M. Mescher, and J. Tumlinson. 2001. Caterpillar-induced nocturnal plant volatiles repel conspecific females. Nature. 410: 577-580.

77. Rasmann, S., T. Kollner, J. Degenhardt, I. Hiltpold, S. Toepfer, U. Kuhlmann, J. Gershenzon, and T. Turlings. 2005. Recruitment of entomopathogenic nematodes by insect-damaged maize roots. Nature. 434:732-737.

78. Sabulal, B., M. Dan, A.J. J, R. Kurup, N.S. Pradeep, R.K. Valsamma, and V. George. 2006. Caryophyllene-rich rhizome oil of *Zingiber nimmonii* from South India: Chemical characterization and antimicrobial activity. Phytochemistry. 67: 2469–2473.

79. Opdyke, D. 1973. Monographs on fragrance raw materials. Food and Cosmetic Toxicolgy. 11: 1011–108.

80. Cho, J.Y., H. Chang, S. Lee, H. Kim, J. Hwang, and H. S. Chun. 2007. Amelioration of dextran sulfate sodium-induced colitis in mice by oral administration of β-caryophyllene, a sesquiterpene. Life Sciences. 80: 932–939.

81. Molina-Jasso, D., I. Alvarez-Gonzalez, and E. Madrigal-Bujaidar. 2009. Clastogenicity of β-caryophyllenein Mouse. Biological & Pharmaceutical Bulletin. 32: 520-522.

82. Di Sotto, A., G. Mazzanti, F. Carbone, P. Hrelia, and F. Maffei. 2010. Inhibition by β-caryophyllene of ethyl methanesulfonate-induced clastogenicity in cultured human lymphocytes, Mutation Research. 699: 23–28.

83. Cascon, V., and B. Gilbert. 2000. Characterization of the chemical composition of oleoresins of *Copaifera guianensis Desf., Copaifera duckei Dwyer* and *Copaifera multijuga Hayne.* Phytochemistry. 55: 773-778.

84. Coussio, J., G. Cissia, and J. Silvia. 2001. Profisetindin type tannins responsible for anti-oxidant activity in *Copaifera recticulata.* Pharmaize. 53: 573-577.

85. Ghelardini, C., N. Galeotti, L. Di Cesare Mannelli, G. Mazzanti, and A. Bartolini. 2001. Local anaesthetic activity of β-caryophyllene. Il Farmaco. 56: 387–389.

86. Lourens, A., D. Reddy, K. Baser, A. Viljoen, S. Van Vuuren. 2004. In vitro biological activity and essential oil composition of four indigenous South African *Helichrysum* species. Journal of Ethnopharmacology. 95: 253–258.

87. Baylac, S., and P. Racine. 2003. Inhibition of 5-Lipoxygenase by essential oils and other natural fragrant extracts. The International Journal of Aromatherapy. 13: 138-142.

88. Santos, A.O., T. Ueda-Nakamura, B. P. Dias Filho, V. F. Viega, A. C. Pinto, and C. Nakamura. 2008. Effect of Brazilian copaiba oils on *Leishmania amazonensis.* Journal of Ethnopharmacology. 120: 204–20.

89. . Singh, G. and P. Marimuthu. 2006. Antioxidant and Biocidal Activities of *Carum nigrum* (Seed) Essential Oil, Oleoresin, and Their Selected Components. Journal of Agricultural and Food Chemistry. 54: 174-181.

90. Quispe-Condori, S., M. Foglio, P. Rosa, and M. Meireles. 2008. Obtaining β-caryophyllene from *Cordia verbenacea de Candolle* by supercritical fluid extraction.. The Journal of Supercritical Fluids. 46: 27–32.

91. Topal, U., M. Sasaki, M. Goto, and S. Otles. 2008. Chemical compositions and antioxidant properties of essential oils from nine species of Turkish plants obtained by supercritical carbon dioxide extraction and steam distillation. International Journal of Food Sciences and Nutrition. 59: 619 - 634.

92. Allen, M. 1968. Simple conditions for growth of unicellular blue-green algae on plates. Journal of Phycology. 4: 1-4.

93. Cai, Y. and P. Wolk. 1990. Use of a Conditionally Lethal Gene in *Anabaena sp.* Strain PCC 7120 To Select for Double Recombinants and To Entrap Insertion Sequences. Journal of Bacteriology. 172: 3138-3145.

94. Chang S., J. Puryear , and J. Cairney. 1993. A Simple and Efficient Method for Isolating RNA from Pine Trees. Plant Molecular Biology Reporter. 11: 113-116.

95. Vioque, A. 1992. Analysis of the gene encoding the RNA subunit of ribonuclease P from cyanobacteria. Nucleic Acid Research. 20: 6331-6337.

96. Bligh, E. G., and W. J. Dye. 1959, A rapid method of total lipid extraction and purification., Canadian Journal of Biochemistry and Physiology. 37: 911-917.

97. Palmeira, S. F., F. d. S. Moura, V. d. L. Alves, F. M. d. Oliveira, E. S. Bento, L. M. Conserva, and E. H. d. A. Andrade. 2004. Neutral components from hexane extracts of *Croton sellowii.*, Flavour and Fragrance Journal. 19: 69-71.

98. Maia, J. G. S., E. H. A. Andrade, A. C. M. da Silva, J. Oliveira, L. M. M. Carreira, and J. S. Araújo. 2005. Leaf volatile oils from four Brazilian *Xylopia* species. Flavour and Fragrance Journal. 20: 474-477.

99. Dong, X., P. Stothard, I. Forsythe, and D. Wishart. 2004. PlasMapper: a web server for drawing and auto-annotating plasmid maps. Journal of Nucleic Acid Research. 32: 660-664.

100. Marchler-Bauer, A., J. Anderson, F. Chitsaz , M. Derbyshire, C. DeWeese-Scott ,J. Fong, L. Geer , R. Geer , N. Gonzales, M. Gwadz, S. He, D. Hurwitz, J. Jackson, Z. Ke, C. Lanczycki, C. Liebert, C. Liu, F. Lu, S. Lu, G. Marchler, M. Mullokandov, J. Song. 2008.

CDD: specific functional annotation with the Conserved Domain Database. Nucleic Acid Research. 37: D205-10.

101. Liang, P., T. Ko, and A. Wang. 2002. Structure, mechanism and function of prenyltransferases. European Journal of Biochemistry. 269: 3339–3354.

102. Szkopiñska, A., and D. Plochocka. 2005. Farnesyl diphosphate synthase; regulation of product specificity. Acta Biochimica Polinica. 52: 45-55.

103. Schirmer, A., M. Rude, X. Li, E. Popova, and S. del Cardayre. 2010. Microbial Biosynthesis of Alkanes. Science. 329: 559-561.

104. Chang, M. and J. Keasling. 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nature Chemical Biology. 2: 674 - 681.

105. Ryu, J., Y. Song, J. Lee, S. Jeong, W. Chow, S. Choi, B. Pogson, and Y. Park. 2004. Glucose-induced Expression of Carotenoid Biosynthesis Genes in the Dark Is Mediated by Cytosolic pH in the Cyanobacterium *Synechocystis sp*. PCC 6803. Journal of Biological Chemistry. 279: 25320-25325.

106. Zuther, E., H. Schubert, and M. Hageman. 1998. Mutation of a Gene Encoding a Putative Glycoprotease Leads to Reduced Salt Tolerance, Altered Pigmentation, and Cyanophycin Accumulation in the Cyanobacterium *Synechocystis sp*. Strain PCC 6803. Journal of Bacteriology. 180: 1715–1722.

107. Kasai, S., S. Yoshimura, K. Ishikura, Y. Takaoka, K. Kobayahsi, K. Kato, and A. Shinmyo. 2003. Effect of Coding Regions on Chloroplast Gene Expression in *Chlamydomonas reinhardtii.* Journal of Biosciences and Bioengineering. 95: 276-282.

108. Nakamura, Y., T. Gojobori, and T. Ikemura. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. Journal of Nucleic Acid Research. 28: 292.

109. Osanai, T., Y. Kanesaki, T. Nakano, H. Takahashi, M. Asayama, M. Shirai, M. Kanehisa, I Suzuki, N. Murata, and K. Tanaka. 2005. Positive Regulation of Sugar Catabolic Pathways in the Cyanobacterium *Synechocystis sp*. PCC 6803 by the Group 2 sigma Factor SigE. Journal of Biological Chemistry. 280: 30653–30659.

110. . Osanai, T., S. Imamura, M. Asayama, M. Shirai, I. Suzuki, N. Murata, and K. Tanaka. 2006. Nitrogen Induction of Sugar Catabolic Gene Expression in *Synechocystis sp.* PCC 6803. DNA Research, 13: 185-195.

Appendix A

Experimental Sequence Analysis

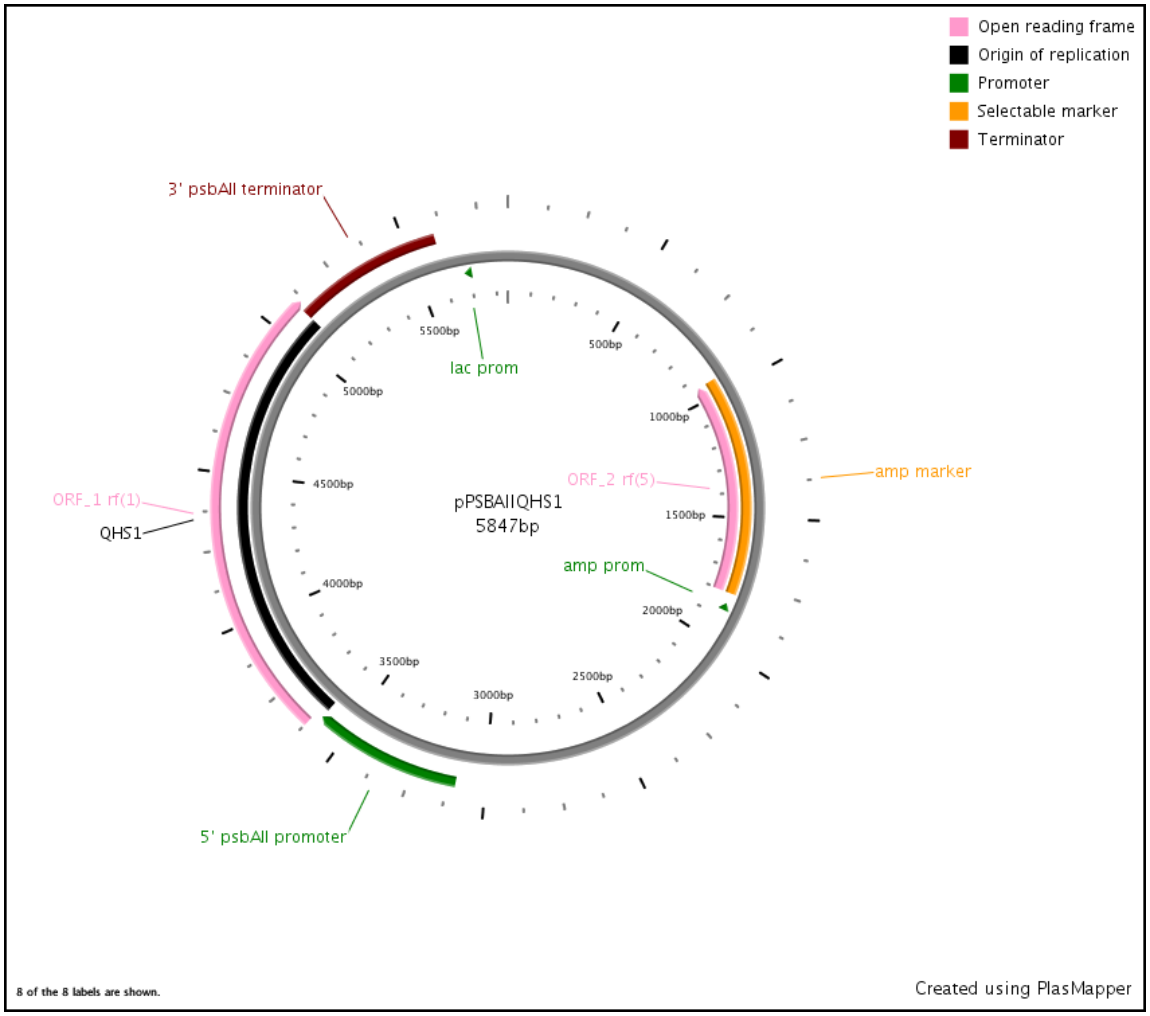

Figure 17: Plasmid pPSBAIIQHS1 image. Images generated by PlasMapper v.2.0 (99). The legend in the upper right corner provides descriptions of respective bands.

Sequencing results: 'Query' sequence represents the sequence returned from MCLABs sequencing service that have been truncated to eliminate non-specific regions at the beginning and end of the raw sequence file. The 'Sbjct' sequence represents the 'Subject' sequence from the NCBI database listed before the alignment results. NCBI data analysis was conducted on the existing NCBI database as of 9.8.10.

DNA Sequence 1: Sequence returned using the pPSBAIIQHS1 vector as a template and the 5' ^{UP} pPSBAII*QHS1* $_{sq}$ primer.

```
LOCUS       dna                         768 bp
BASE COUNT    255 a   144 c   147 g   222 t    0 others
ORIGIN
    1 gccctctgtt tacccatgga aaaaacgaca attacaagaa agtaaaactt atgtcatcta
   61 taagcttcgt gtatattaac ttcctgttac aaagctttac aaaactctca ttaatccttt
  121 agactaagtt tagtcagttc caatctgaac atcgacaaat acataaggaa ttataaccat
  181 aatatgtctg ttaaagaaga gaaagtaatt cggcccattg tccactttcc tcctagcgtt
  241 tgggcagacc agtttctcat cttcgatgac aagcaagcag agcaagccaa tgtagaacaa
  301 gtagtcaacg agttaagaga agatgtgaga aaagatttag tgtcatcttt agatgttcaa
  361 acggagcaca caaatttgct aaaacttatt gatgctatcc aacgtcttgg tatagcgtat
  421 cattttgaag aggagattga gcaagcctta caacatattt atgatacata tggtgatgac
  481 tggaaaggca gaagcccttc cctttggttt cgaatccttc ggcaacaagg tttttatgtt
  541 tcttgtgata tttttaaaaa ctataaaaaa gaggatggtt catttaagga atccctcacc
  601 aacgatgtag aaggcttgct tgagctgtat gaggcgacat atttgagagt gcaaggcgaa
  661 ggggttctag atgatgctct tgttttcaca aggacttgtc ttgagaaaat agcaaaggat
  721 cttgttcaca ccaacccaac actatctacc tacatacaag aagcacta
```

NCBI BLAST results for DNA sequence 1

Subject: dbj|BA000022.2|*Synechocystis* sp. PCC 6803 DNA, complete genome
Length=3573470

Features in this part of subject sequence:
photosystem II D1 protein

Score = 331 bits (179), Expect = 3e-87
Identities = 182/183 (99%), Gaps = 1/183 (0%)
Strand=Plus/Plus

```
Query  1     GCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7049  GCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATCTA 7108

Query  61    TAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTT 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7109  TAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCTTT 7168


Query  121   AGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCAT 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7169  AGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACCA-7227

Query  181   AAT  183
             |||
Sbjct  7228  AAT  7230
```

Subject: gb|AF472361.1| *Artemisia annua* beta-caryophyllene synthase QHS1 (QHS1) mRNA
Score = 1081 bits (585), Expect = 0.0
Identities = 590/592 (99%), Gaps = 1/592 (0%)
Strand=Plus/Plus

```
Query  178  CATA-ATATGTCTGTTAAAGAAGAGAAAGTAATTCGGCCCATTGTCCACTTTCCTCCTAG  236
            |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  17   CATAGACATGTCTGTTAAAGAAGAGAAAGTAATTCGGCCCATTGTCCACTTTCCTCCTAG  76

Query  237  CGTTTGGGCAGACCAGTTTCTCATCTTCGATGACAAGCAAGCAGAGCAAGCCAATGTAGA  296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  77   CGTTTGGGCAGACCAGTTTCTCATCTTCGATGACAAGCAAGCAGAGCAAGCCAATGTAGA  136

Query  297  ACAAGTAGTCAACGAGTTAAGAGAAGATGTGAGAAAAGATTTAGTGTCATCTTTAGATGT  356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  137  ACAAGTAGTCAACGAGTTAAGAGAAGATGTGAGAAAAGATTTAGTGTCATCTTTAGATGT  196

Query  357  TCAAACGGAGCACACAAATTTGCTAAAACTTATTGATGCTATCCAACGTCTTGGTATAGC  416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  197  TCAAACGGAGCACACAAATTTGCTAAAACTTATTGATGCTATCCAACGTCTTGGTATAGC  256

Query  417  GTATCATTTTGAAGAGGAGATTGAGCAAGCCTTACAACATATTTATGATACATATGGTGA  476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  257  GTATCATTTTGAAGAGGAGATTGAGCAAGCCTTACAACATATTTATGATACATATGGTGA  316

Query  477  TGACTGGAAAGGCAGAAGCCCTTCCCTTTGGTTTCGAATCCTTCGGCAACAAGGTTTTTA  536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  317  TGACTGGAAAGGCAGAAGCCCTTCCCTTTGGTTTCGAATCCTTCGGCAACAAGGTTTTTA  376

Query  537  TGTTTCTTGTGATATTTTTAAAAACTATAAAAAAAGAGGATGGTTCATTTAAGGAATCCCT  596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  377  TGTTTCTTGTGATATTTTTAAAAACTATAAAAAAAGAGGATGGTTCATTTAAGGAATCCCT  436

Query  597  CACCAACGATGTAGAAGGCTTGCTTGAGCTGTATGAGGCGACATATTTGAGAGTGCAAGG  656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  437  CACCAACGATGTAGAAGGCTTGCTTGAGCTGTATGAGGCGACATATTTGAGAGTGCAAGG  496

Query  657  CGAAGGGGTTCTAGATGATGCTCTTGTTTTCACAAGGACTTGTCTTGAGAAAATAGCAAA  716
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  497  CGAAGGGGTTCTAGATGATGCTCTTGTTTTCACAAGGACTTGTCTTGAGAAAATAGCAAA  556

Query  717  GGATCTTGTTCACACCAACCCAACACTATCTACCTACATACAAGAAGCACTA  768
            ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  557  GGATCTTGTTCACACCAACCCAACACTATCTACCTACATACAAGAAGCACTA  608
```

DNA Sequence 2: Sequence returned using the pPSBAIIQHS1 vector as a template and the 5'^{DOWN} pPSBAII*QHS1* primer.

```
LOCUS        dna                          587 bp

BASE COUNT    182 a   104 c   132 g   169 t    0 others
ORIGIN
   1 gggtgatatc gttacagatg aagcattcaa atgggctctg acaaaacctc ctatcatcaa
  61 agcttcatgt gcaattgcta gactcatgga tgatattcac tctcaaaagg aggagaaaga
 121 aagaatacat gttgcatcta gtgttgaaag ttacatgaag caatatgatg tgacagagga
 181 gcatgtcctt aaagtattta acaagaaaat cgaggatgcg tggaaagata taacccgaga
 241 atcccttgtg cgtaaagata ttccgatgcc tctgatgatg cgagtgatta acttggcaca
```

```
 301 ggtgatggat gttttatata aacataaaga cggtttcact aatgtgggag aagaactcaa
 361 ggatcatatt aaatctttgc tcgttcatcc tatacctata taaggatcct aattccttgg
 421 tgtaatgcca actgaataat ctgcaaattg cactctcctt caatgggggg tgctttttgc
 481 ttgactgagt aatcttctga ttgctgatct tgattgccat cgatcgccgg ggagtccggg
 541 gcagttacca ttagagagtc tagagaatta atccatcttc gatagaa
```

NCBI BLAST results for DNA sequence 2
Subject: gb|AF472361.1| *Artemisia annua* beta-caryophyllene synthase QHS1 (QHS1) mRNA, Length=1902
Score = 747 bits (404), Expect = 0.0
Identities = 404/404 (100%), Gaps = 0/404 (0%)
Strand=Plus/Plus

```
Query  1     GGGTGATATCGTTACAGATGAAGCATTCAAATGGGCTCTGACAAAACCTCCTATCATCAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1268  GGGTGATATCGTTACAGATGAAGCATTCAAATGGGCTCTGACAAAACCTCCTATCATCAA  1327

Query  61    AGCTTCATGTGCAATTGCTAGACTCATGGATGATATTCACTCTCAAAAGGAGGAGAAAGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1328  AGCTTCATGTGCAATTGCTAGACTCATGGATGATATTCACTCTCAAAAGGAGGAGAAAGA  1387

Query  121   AAGAATACATGTTGCATCTAGTGTTGAAAGTTACATGAAGCAATATGATGTGACAGAGGA  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1388  AAGAATACATGTTGCATCTAGTGTTGAAAGTTACATGAAGCAATATGATGTGACAGAGGA  1447

Query  181   GCATGTCCTTAAAGTATTTAACAAGAAAATCGAGGATGCGTGGAAAGATATAACCCGAGA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1448  GCATGTCCTTAAAGTATTTAACAAGAAAATCGAGGATGCGTGGAAAGATATAACCCGAGA  1507

Query  241   ATCCCTTGTGCGTAAAGATATTCCGATGCCTCTGATGATGCGAGTGATTAACTTGGCACA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1508  ATCCCTTGTGCGTAAAGATATTCCGATGCCTCTGATGATGCGAGTGATTAACTTGGCACA  1567

Query  301   GGTGATGGATGTTTTATATAAACATAAAGACGGTTTCACTAATGTGGGAGAAGAACTCAA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1568  GGTGATGGATGTTTTATATAAACATAAAGACGGTTTCACTAATGTGGGAGAAGAACTCAA  1627

Query  361   GGATCATATTAAATCTTTGCTCGTTCATCCTATACCTATATAAG  404
             |||||||||||||||||||||||||||||||||||||||||||
Sbjct  1628  GGATCATATTAAATCTTTGCTCGTTCATCCTATACCTATATAAG  1671
```

Subject: dbj|BA000022.2| *Synechocystis* sp. PCC 6803 DNA, complete genome Length=3573470
Features in this part of subject sequence:
photosystem II D1 protein
Score = 327 bits (177), Expect = 3e-86
Identities = 177/177 (100%), Gaps = 0/177 (0%)
Strand=Plus/Plus

```
Query  410   TAATTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATGGGGG  469
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8309  TAATTCCTTGGTGTAATGCCAACTGAATAATCTGCAAATTGCACTCTCCTTCAATGGGGG  8368

Query  470   GTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCGCCG  529
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8369  GTGCTTTTTGCTTGACTGAGTAATCTTCTGATTGCTGATCTTGATTGCCATCGATCGCCG  8428
```

```
Query  530   GGGAGTCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGA  586
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8429  GGGAGTCCGGGGCAGTTACCATTAGAGAGTCTAGAGAATTAATCCATCTTCGATAGA  8485
```

DNA Sequence 3: Sequence returned using the genomic DNA from the pPSBAIIQHS1 *Synechocystis* strain as a template and the 5' ^{UP} pPSBAII*QHS1* $_{sq}$ primer.

```
LOCUS       dna                        769 bp
BASE COUNT    255 a   145 c   147 g   222 t    0 others
ORIGIN
  1 acgccctctg tttacccatg gaaaaaacga caattacaag aaagtaaaac ttatgtcatc
 61 tataagcttc gtgtatatta acttcctgtt acaaagcttt acaaaactct cattaatcct
121 ttagactaag tttagtcagt tccaatctga acatcgacaa atacataagg aattataacc
181 ataatatgtc tgttaaagaa gagaaagtaa ttcggcccat tgtccacttt cctcctagcg
241 tttgggcaga ccagtttctc atcttcgatg acaagcaagc agagcaagcc aatgtagaac
301 aagtagtcaa cgagttaaga gaagatgtga gaaaagattt agtgtcatct ttagatgttc
361 aaacggagca cacaaatttg ctaaaactta ttgatgctat ccaacgtctt ggtatagcgt
421 atcattttga agaggagatt gagcaagcct tacaacatat ttatgataca tatggtgatg
481 actggaaagg cagaagccct tccctttggt ttcgaatcct tcggcaacaa ggtttttatg
541 tttcttgtga tatttttaaa aactataaaa aagaggatgg ttcatttaag gaatccctca
601 ccaacgatgt agaaggcttg cttgagctgt atgaggcgac atatttgaga gtgcaaggcg
661 aaggggttct agatgatgct cttgttttca caaggacttg tcttgagaaa atagcaaagg
721 atcttgttca caccaaccca acactatcta cctacataca agaagcact
```

NCBI BLAST results for DNA sequence 3
Subject: dbj|BA000022.2| *Synechocystis* sp. PCC 6803 DNA, complete genome
Length=3573470
Features in this part of subject sequence:
photosystem II D1 protein
Score = 335 bits (181), Expect = 3e-88
Identities = 184/185 (99%), Gaps = 1/185 (0%)
Strand=Plus/Plus

```
Query  36    ACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATC  95
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7047  ACGCCCTCTGTTTACCCATGGAAAAAACGACAATTACAAGAAAGTAAAACTTATGTCATC  7106

Query  96    TATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCT  155
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7107  TATAAGCTTCGTGTATATTAACTTCCTGTTACAAAGCTTTACAAAACTCTCATTAATCCT  7166

Query  156   TTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACC  215
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7167  TTAGACTAAGTTTAGTCAGTTCCAATCTGAACATCGACAAATACATAAGGAATTATAACC  7226

Query  216   ATAAT  220
             | |||
Sbjct  7227  A-AAT  7230
```

Subject: gb|AF472361.1| *Artemisia annua* beta-caryophyllene synthase QHS1 (QHS1) mRNA, Length=1902
Score = 1079 bits (584), Expect = 0.0
Identities = 589/591 (99%), Gaps = 1/591 (0%)
Strand=Plus/Plus

```
Query  215  CATA-ATATGTCTGTTAAAGAAGAGAAAGTAATTCGGCCCATTGTCCACTTTCCTCCTAG  273
            |||| | ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  17   CATAGACATGTCTGTTAAAGAAGAGAAAGTAATTCGGCCCATTGTCCACTTTCCTCCTAG  76

Query  274  CGTTTGGGCAGACCAGTTTCTCATCTTCGATGACAAGCAAGCAGAGCAAGCCAATGTAGA  333
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  77   CGTTTGGGCAGACCAGTTTCTCATCTTCGATGACAAGCAAGCAGAGCAAGCCAATGTAGA  136

Query  334  ACAAGTAGTCAACGAGTTAAGAGAAGATGTGAGAAAAGATTTAGTGTCATCTTTAGATGT  393
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  137  ACAAGTAGTCAACGAGTTAAGAGAAGATGTGAGAAAAGATTTAGTGTCATCTTTAGATGT  196

Query  394  TCAAACGGAGCACACAAATTTGCTAAAACTTATTGATGCTATCCAACGTCTTGGTATAGC  453
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  197  TCAAACGGAGCACACAAATTTGCTAAAACTTATTGATGCTATCCAACGTCTTGGTATAGC  256

Query  454  GTATCATTTTGAAGAGGAGATTGAGCAAGCCTTACAACATATTTATGATACATATGGTGA  513
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  257  GTATCATTTTGAAGAGGAGATTGAGCAAGCCTTACAACATATTTATGATACATATGGTGA  316

Query  514  TGACTGGAAAGGCAGAAGCCCTTCCCTTTGGTTTCGAATCCTTCGGCAACAAGGTTTTTA  573
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  317  TGACTGGAAAGGCAGAAGCCCTTCCCTTTGGTTTCGAATCCTTCGGCAACAAGGTTTTTA  376

Query  574  TGTTTCTTGTGATATTTTTAAAAACTATAAAAAAAGAGGATGGTTCATTTAAGGAATCCCT  633
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  377  TGTTTCTTGTGATATTTTTAAAAACTATAAAAAAAGAGGATGGTTCATTTAAGGAATCCCT  436

Query  634  CACCAACGATGTAGAAGGCTTGCTTGAGCTGTATGAGGCGACATATTTGAGAGTGCAAGG  693
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  437  CACCAACGATGTAGAAGGCTTGCTTGAGCTGTATGAGGCGACATATTTGAGAGTGCAAGG  496

Query  694  CGAAGGGGTTCTAGATGATGCTCTTGTTTTCACAAGGACTTGTCTTGAGAAAATAGCAAA  753
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  497  CGAAGGGGTTCTAGATGATGCTCTTGTTTTCACAAGGACTTGTCTTGAGAAAATAGCAAA  556

Query  754  GGATCTTGTTCACACCAACCCAACACTATCTACCTACATACAAGAAGCACT  804
            |||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  557  GGATCTTGTTCACACCAACCCAACACTATCTACCTACATACAAGAAGCACT  607
```

DNA Sequence 4: Sequence returned using the genomic DNA from the pPSBAIIQHS1 *Synechocystis* strain as a template and the 3' ^{DOWN} pPSBAII*QHS1*$_{sq}$ primer.

```
LOCUS       dna                        815 bp
BASE COUNT    227 a   191 c   132 g   265 t    0 others
ORIGIN
  1 tccccggcga tcgatggcaa tcaagatcag caatcagaag attactcagt caagcaaaaa
 61 gcacccccca ttgaaggaga gtgcaatttg cagattattc agttggcatt acaccaagga
121 attaggatcc ttatataggt ataggatgaa cgagcaaaga tttaatatga tccttgagtt
181 cttctcccac attagtgaaa ccgtctttat gtttatataa aacatccatc acctgtgcca
241 agttaatcac tcgcatcatc agaggcatcg gaatatcttt acgcacaagg gattctcggg
```

```
301 ttatatcttt ccacgcatcc tcgattttct tgttaaatac tttaaggaca tgctcctctg
361 tcacatcata ttgcttcatg taactttcaa cactagatgc aacatgtatt ctttctttct
421 cctccttttg agagtgaata tcatccatga gtctagcaat tgcacatgaa gctttgatga
481 taggaggttt tgtcagagcc catttgaatg cttcatctgt aacgatatca cccatgccaa
541 caaaacatgt tgttgcaagc atgctatacc cgctgcttac aaacgcaact gacatatgct
601 cctctgctgt tggtacatac ccctcatttg cccatttttgc ttccatcata tagctcctaa
661 taaactcttt catagattct ttagcatagc taagatgatg tgcttttccc tcttttccca
721 ttatttcttc catttctatg tatatatcca agactccttg gtataataat agtttttcagg
781 tctggaagca tatctatgca tgtaatcgac caccc
```

NCBI BLAST results (notice subject is in minus strand)

Subject: gb|AF472361.1| *Artemisia annua* beta-caryophyllene synthase QHS1 (QHS1) mRNA, Length=1902

Score = 1253 bits (678), Expect = 0.0

Identities = 683/685 (99%), Gaps = 1/685 (0%)

Strand=Plus/Minus

```
Query  130   CTTATATAGGTATAGGATGAACGAGCAAAGATTTAATATGATCCTTGAGTTCTTCTCCCA  189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1671  CTTATATAGGTATAGGATGAACGAGCAAAGATTTAATATGATCCTTGAGTTCTTCTCCCA  1612

Query  190   CATTAGTGAAACCGTCTTTATGTTTATATAAAACATCCATCACCTGTGCCAAGTTAATCA  249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1611  CATTAGTGAAACCGTCTTTATGTTTATATAAAACATCCATCACCTGTGCCAAGTTAATCA  1552

Query  250   CTCGCATCATCAGAGGCATCGGAATATCTTTACGCACAAGGGATTCTCGGGTTATATCTT  309
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1551  CTCGCATCATCAGAGGCATCGGAATATCTTTACGCACAAGGGATTCTCGGGTTATATCTT  1492

Query  310   TCCACGCATCCTCGATTTTCTTGTTAAATACTTTAAGGACATGCTCCTCTGTCACATCAT  369
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1491  TCCACGCATCCTCGATTTTCTTGTTAAATACTTTAAGGACATGCTCCTCTGTCACATCAT  1432

Query  370   ATTGCTTCATGTAACTTTCAACACTAGATGCAACATGTATTCTTTCTTTCTCCTCCTTTT  429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1431  ATTGCTTCATGTAACTTTCAACACTAGATGCAACATGTATTCTTTCTTTCTCCTCCTTTT  1372

Query  430   GAGAGTGAATATCATCCATGAGTCTAGCAATTGCACATGAAGCTTTGATGATAGGAGGTT  489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1371  GAGAGTGAATATCATCCATGAGTCTAGCAATTGCACATGAAGCTTTGATGATAGGAGGTT  1312

Query  490   TTGTCAGAGCCCATTTGAATGCTTCATCTGTAACGATATCACCCATGCCAACAAAACATG  549
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1311  TTGTCAGAGCCCATTTGAATGCTTCATCTGTAACGATATCACCCATGCCAACAAAACATG  1252

Query  550   TTGTTGCAAGCATGCTATACCCGCTGCTTACAAACGCAACTGACATATGCTCCTCTGCTG  609
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1251  TTGTTGCAAGCATGCTATACCCGCTGCTTACAAACGCAACTGACATATGCTCCTCTGCTG  1192

Query  610   TTGGTACATACCCCTCATTTGCCCATTTTGCTTCCATCATATAGCTCCTAATAAACTCTT  669
             |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Sbjct  1191  TTGGTACATACCCCTCATTTGCCCATTTTGCTTCCATCATATAGCTTCTAATAAACTCTT  1132

Query  670   TCATAGATTCTTTAGCATAGCTAAGATGATGTGCTTTTCCCTCTTTTCCCATTATTTCTT  729
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1131  TCATAGATTCTTTAGCATAGCTAAGATGATGTGCTTTTCCCTCTTTTCCCATTATTTCTT  1072

Query  730   CCATTTCTATGTATATATCCAAGACTCCTTGGTATAATAGTTTTCAGGTATTCTGGAAGC  789
             |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
```

```
Sbjct  1071  CCATTTCTATGTATATATCCAAGACTCCTTGGTATAATAGTTT-CAGGTATTCTGGAAGC  1013

Query  790   ATATCTATGCATGTAATCGACCACC  814
              ||||||||||||||||||||||||
Sbjct  1012  ATATCTATGCATGTAATCGACCACC  988
```

Subject: dbj|BA000022.2| *Synechocystis* sp. PCC 6803 DNA, complete genome
Length=3573470
Features in this part of subject sequence:
photosystem II D1 protein
Score = 230 bits (124), Expect = 1e-56
Identities = 124/124 (100%), Gaps = 0/124 (0%)
Strand=Plus/Minus

```
Query  1     TCCCCGGCGATCGATGGCAATCAAGATCAGCAATCAGAAGATTACTCAGTCAAGCAAAAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8432  TCCCCGGCGATCGATGGCAATCAAGATCAGCAATCAGAAGATTACTCAGTCAAGCAAAAA  8373

Query  61    GCACCCCCCATTGAAGGAGAGTGCAATTTGCAGATTATTCAGTTGGCATTACACCAAGGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  8372  GCACCCCCCATTGAAGGAGAGTGCAATTTGCAGATTATTCAGTTGGCATTACACCAAGGA  8313

Query  121   ATTA  124
             ||||
Sbjct  8312  ATTA  8309
```

Appendix B

Synechocystis sp. strain PCC6803 Gene Sequence Analysis

The alignments were conducted exclusively with cyanobacterial Farnesyl Diphosphate Synthase (FPPS) sequences and the results reported are the first three FPPS hits for each respective sequence. The highlighted regions within the sequence represents the conserved sequences within the first aspartic rich motif (FARM) that is used to distinguish Type I FPPS (Eukaryotic and Archaeal) and Type II FPPS (Bacterial) (102). The 'Query' row represents the *Synechocystis* GGPPS sequence and the 'Sbjct' row represents the 'Subject' sequence from the NCBI database listed before the alignment results. NCBI BLASTP analysis was conducted 10.17.2010 (100).

BLASTP analysis and Sequence Alignment of the *Synechocystis sp*. strain PCC 6803 Geranylgeranyl Diphosphate Synthase (GGPPS NCBI Accession BAA16690.1).

Subject: ref|YP_003421200.1| farnesyl-diphosphate synthase [cyanobacterium UCYN-A]
Length= 309
Score = 436 bits (1120), Expect = 2e-120, Method: Compositional matrix adjust.
Identities = 216/300 (72%), Positives = 247/300 (83%), Gaps = 0/300 (0%)

```
Query  2    VAQQTRTDFDLAQYLQVKKGVVEAALDSSLAIARPEKIYEAMRYSLLAGGKRLRPILCIT  61
            V Q   + FDL  YL+ KK +VEAALD+S+ I  PEKIYEAMRYSLLAGGKRLRPILC+T
Sbjct  9    VTQIETSCFDLVSYLKEKKQLVEAALDASIPITCPEKIYEAMRYSLLAGGKRLRPILCLT  68

Query  62   ACELCGGDEALALPTACALEMIHTMSLIHDDLPSMDNDDFRRGKPTNHKVYGEDIAILAG  121
             CEL GG   +ALPTACALEMIHTMSLIHDDLP+MDNDD+RRG+ TNHKVYGEDIAILAG
Sbjct  69   TCELMGGTVEMALPTACALEMIHTMSLIHDDLPAMDNDDYRRGQLTNHKVYGEDIAILAG  128

Query  122  DGLLAYAFEYVVTHTPQADPQALLQVIARLGRTVGAAGLVGGQVLDLESEGRTDITPETL  181
            D LL+YAFE+V T T       L+ VIARLGRTVGAAGLVGGQVLDLESEG+++IT + L
Sbjct  129  DALLSYAFEHVATQTHNVSLTNLVDVIARLGRTVGAAGLVGGQVLDLESEGKSNITSDIL  188

Query  182  TFIHTHKTGALLEASVLTGAILAGATGEQQQRLARYAQNIGLAFQVVDDILDITATQEEL  241
             FIH HKTGALLE SV++GAILAGA  E  +RL+ YAQNIGLAFQ++DDILDIT+TQ EL
Sbjct  189  KFIHVHKTGALLETSVVSGAILAGANLEDIERLSHYAQNIGLAFQIIDDILDITSTQAEL  248

Query  242  GKTAGKDVKAQKATYPSLLGLEASRAQAQSLIDQAIVALEPFGPSAEPLQAIAEYIVARK  301
            GKTAGKD +AQKATYPSL GLE S+ QAQ LI+ AI  L P+G  A PL+AIA +IV+RK
Sbjct  249  GKTAGKDKQAQKATYPSLWGLEESQKQAQILIEDAIAQLAPYGREAYPLEAIARFIVSRK  308
```

Subject: ref|YP_723652.1| farnesyl-diphosphate synthase [*Trichodesmium erythraeum* IMS101]
Length=316
Score = 426 bits (1094), Expect = 3e-117, Method: Compositional matrix adjust.
Identities = 206/296 (70%), Positives = 242/296 (82%), Gaps = 0/296 (0%)

```
Query  5    QTRTDFDLAQYLQVKKGVVEAALDSSLAIARPEKIYEAMRYSLLAGGKRLRPILCITACE  64
            Q  T FDL+ YL  +K +VE ALD+S+ I  PEKIYEAMRYSLLAGGKRLRPILC+  CE
Sbjct  19   QIDTTFDLSNYLVERKALVEKALDNSIKIVYPEKIYEAMRYSLLAGGKRLRPILCLATCE  78
```

```
Query  65   LCGGDEALALPTACALEMIHTMSLIHDDLPSMDNDDFRRGKPTNHKVYGEDIAILAGDGL  124
            L GG+  +A+PTACALEMIHTMSLIHDDLPSMDNDD+RRG+ TNHKVYGEDIAILAGDGL
Sbjct  79   LSGGNIEMAMPTACALEMIHTMSLIHDDLPSMDNDDYRRGQLTNHKVYGEDIAILAGDGL  138

Query  125  LAYAFEYVVTHTPQADPQALLQVIARLGRTVGAAGLVGGQVLDLESEGRTDITPETLTFI  184
            L+YAFE+VVT T    P+ +LQVIARLG+  GAAGLVGGQV+DL SEG + ++ +TL FI
Sbjct  139  LSYAFEFVVTQTKNVLPKQVLQVIARLGKASGAAGLVGGQVVDLASEGISQVSEKTLHFI  198

Query  185  HTHKTGALLEASVLTGAILAGATGEQQQRLARYAQNIGLAFQVVDDILDITATQEELGKT  244
            HTHKTGALLEASV+ GAILA  +     RL RYAQNIGLAFQ+VDDILDITAT EELGKT
Sbjct  199  HTHKTGALLEASVVCGAILADVSQANLYRLERYAQNIGLAFQIVDDILDITATTEELGKT  258

Query  245  AGKDVKAQKATYPSLLGLEASRAQAQSLIDQAIVALEPFGPSAEPLQAIAEYIVAR  300
            AGKD++A+KATYP + GLE SR QAQ L+++A   LE FG    PL A+A++I++R
Sbjct  259  AGKDIEAKKATYPVIWGLEKSRFQAQKLVEKAKAELEVFGDKVMPLIALADFIISR  314
```

Subject: ref|ZP_07158172.1| farnesyl-diphosphate synthase [*Arthrospira* sp. PCC 8005]
Length=316
Score = 422 bits (1084), Expect = 3e-116, Method: Compositional matrix adjust.
Identities = 205/293 (70%), Positives = 240/293 (82%), Gaps = 0/293 (0%)

```
Query  10   FDLAQYLQVKKGVVEAALDSSLAIARPEKIYEAMRYSLLAGGKRLRPILCITACELCGGD  69
            FDL+ YL  KK +VEAALD+++A+  PEKIYE+MRYSL+AGGKRLRPILC+  CEL GG
Sbjct  24   FDLSTYLSQKKELVEAALDAAIAVTYPEKIYESMRYSLMAGGKRLRPILCLATCELLGGT  83

Query  70   EALALPTACALEMIHTMSLIHDDLPSMDNDDFRRGKPTNHKVYGEDIAILAGDGLLAYAF  129
              +A+PTACALEMIHTMSLIHDDLP+MDNDD+RRGK TNHKVYGEDIAILAGDGLL+YAF
Sbjct  84   VEMAMPTACALEMIHTMSLIHDDLPAMDNDDYRRGKLTNHKVYGEDIAILAGDGLLSYAF  143

Query  130  EYVVTHTPQADPQALLQVIARLGRTVGAAGLVGGQVLDLESEGRTDITPETLTFIHTHKT  189
            E+V T T    P+ +LQVI RLGR VGA GLVGGQV+DL SEG +DI+ ETLTFIHTHKT
Sbjct  144  EFVATQTQGVPPERVLQVIGRLGRAVGATGLVGGQVVDLASEGLSDISEETLTFIHTHKT  203

Query  190  GALLEASVLTGAILAGATGEQQQRLARYAQNIGLAFQVVDDILDITATQEELGKTAGKDV  249
            GALLEA V+ GAILAGA     +RL+RYA+NIGLAFQ+VDDILDIT    ELGKTAGKD+
Sbjct  204  GALLEACVVCGAILAGANETDLERLSRYAKNIGLAFQIVDDILDITVPSSELGKTAGKDL  263

Query  250  KAQKATYPSLLGLEASRAQAQSLIDQAIVALEPFGPSAEPLQAIAEYIVARKY  302
            +AQKATYPSL GLE SRA+AQ L+++A   L+ FG     L AIA++I  R +
Sbjct  264  EAQKATYPSLWGLEKSRAKAQQLVEEAKAELQAFGNGGCALAAIADFITYRSH  316
```

BLASTP analysis and Sequence Alignment of Solanesyl Diphosphate Synthase (SPPS, NCBI Accession BAA16579). Notice the absence of the conserved Type II FARM domain in the 'Query' rows. The 'Query' row represents the *Synechocystis* SPPS sequence and the 'Sbjct' row represents the 'Subject' sequence from the NCBI database listed before the alignment results. NCBI BLASTP analysis was conducted 10.17.2010.

Subject: dbj|BAA82615.1| farnesyl, geranylgeranyl, geranylfarnesyl, hexaprenyl, heptaprenyl
diphosphate synthase (SelF-HepPS) [Synechococcus elongatus]
Length=324
Score = 114 bits (286), Expect = 3e-25, Method: Compositional matrix adjust.
Identities = 83/222 (38%), Positives = 113/222 (51%), Gaps = 10/222 (4%)

```
Query  31   PILGAAAEHLFEAGGKRVRPAIVLLVSRATLLDQELTARHRRLAEITEMIHTASLVHDDV  90
            PIL     +   +GGKR+RP I LL + A        A      EI   +H  +L+HDD+
Sbjct  24   PILDEPVRYALASGGKRLRPLITLLAAEACGSSPADAAPVAAAVEI---LHNFTLIHDDI  80

Query  91   VDEADLRRNVPTVNSLFDNRVAVLAGDFLFAQSSWYLANLDNLEVVK-LLSEVIRDFAEG  149
            +D + LRR  PTV   +D+  A+LAGD +   +   L+     EV + L++E    F +
Sbjct  81   MDRSPLRRGRPTVYRAWDDATAILAGDTMMGIAYRLLSEHYPAEVCQHLVAECSSAFVDV  140

Query  150  EILQSIN---RFDTDTDLETYLEKSYFKTASLIANSAKAAGVLSDAPRDVCDHLYEYGKH  206
             I Q+ +   R      ++ YL     KTA L A +A A G+++ AP DV   L  YG
Sbjct  141  CIGQAEDLAFRTMPSVTMDDYLRMIGRKTARLFATAAVAGGIVARAPHDVLAALRTYGTA  200

Query  207  LGLAFQIVDDILD-FTSPTEVLGKPAGSDLISGNITAPALFA  247
            LG+AFQI DD+LD F SP    GK  G DL  G   T P L A
Sbjct  201  LGMAFQIRDDVLDLFGSPE--FGKELGQDLREGKKTFPTLSA  240
```

Subject: ref|YP_323212.1| farnesyl-diphosphate synthase [*Anabaena variabilis* ATCC 29413]
Score = 113 bits (282), Expect = 7e-25, Method: Compositional matrix adjust.
Identities = 75/220 (35%), Positives = 120/220 (55%), Gaps = 15/220 (6%)

```
Query  43   AGGKRVRPAIVLLVSRATLLDQELTARHRRLAEITEMIHTASLVHDDV--VDEADLRRNV  100
            AGGKR+RP + L  S       E+       A   EMIHT SL+HDD+  +D  D RR +
Sbjct  55   AGGKRLRPILCLATSEMMGGTIEIA---MPTACAVEMIHTMSLIHDDLPAMDNDDYRRGM  111

Query  101  PTVNSLFDNRVAVLAGDFLFAQSSWYLA-----NLDNLEVVKLLSEVIRDFAE-----GE  150
             T + ++ + +A+LAGD L A +  ++A     ++    V+++++ + R         G+
Sbjct  112  LTNHKVYGDDIAILAGDGLLAYAFEFVAIRTPESVPRDRVLQVVARLGRALGAAGLVGGQ  171

Query  151  ILQSINRFDTDTDLETYLEKSYFKTASLIANSAKAAGVLSDAPRDVCDHLYEYGKHLGLA  210
            ++   +   +D  LET       KTA+L+       G+L+ A  +    L  Y +++GLA
Sbjct  172  VVDLDSEGKSDISLETLNFIHNHKTAALLEACVVCGGILAGASSEDVQRLSRYSQNIGLA  231

Query  211  FQIVDDILDFTSPTEVLGKPAGSDLISGNITAPALFAMEK  250
            FQI+DDILD TS  E LGK AG DL++  +T P+L+ +E+
Sbjct  232  FQIIDDILDITSTQEQLGKTAGKDLLAKKVTYPSLWGIEQ  271
```

Subject: gb|ABB50140.1| farnesyl-diphosphate synthase [*Prochlorococcus marinus* str. MIT 9312]
Length=300
Score = 111 bits (278), Expect = 2e-24, Method: Compositional matrix adjust.
Identities = 80/236 (34%), Positives = 124/236 (53%), Gaps = 14/236 (5%)

```
Query  26   VGARHP-ILGAAAEHLFEAGGKRVRPAIVLLVSRATLLDQELTARHRRLAEITEMIHTAS  84
            +G  +P IL  +  +   AGGKR+RP  +L ++  +L   E +        I EMIHT S
Sbjct  30   LGPENPEILRESMRYSLLAGGKRIRP--ILCLASCSLAGGEPSLAVPTAVAI-EMIHTMS  86

Query  85   LVHDDV--VDEADLRRNVPTVNSLFDNRVAVLAGDFL----FAQSSWYLANLDNLEVVKL  138
            L+HDD+  +D  DLRR  PT + ++ + +A+LAGD L    F   S     +D   ++ +
Sbjct  87   LIHDDLPAMDNDDLRRGRPTNHKVYGDAIAILAGDALLTRAFEMVSLRSPGVDPTRLLNV  146
```

```
Query  139  LSEV-IRDFAEGEILQSINRFDTD---TDLETYLEKSYFKTASLIANSAKAAGVLSDAPR  194
            + E+ +   A G +   +   + +    DLET       KT +L+    +   +++ A
Sbjct  147  IGELSLVAGAPGLVGGQVVDLECEGKEVDLETLEYIHLHKTGALLKACVRTGAMIAGANE  206

Query  195  DVCDHLYEYGKHLGLAFQIVDDILDFTSPTEVLGKPAGSDLISGNITAPALFAMEK  250
            +    L  Y + +GLAFQI+DDILD TS +E LGK AG DL++   T P L  ME+
Sbjct  207  ETLQALTTYAEGIGLAFQIIDDILDLTSSSEKLGKTAGKDLLADKTTYPKLLGMEE  262
```

BLASTP analysis of the Type II FPPS FARM domain ‘SLIHDDLPXMDX’ against the entire Synechocystis sp. strain PCC6803 genome. The report below was the first result and the only ‘hit’ to return full converge of the query (FARM domain).

ref|NP_440010.1| geranylgeranyl pyrophosphate synthase [Synechocystis sp. PCC 6803]
Length=302

GENE ID: 953229 crtE | geranylgeranyl pyrophosphate synthase [Synechocystis sp. PCC 6803] (10 or fewer PubMed links)

Score = 36.7 bits (79), Expect = 1e-04
Identities = 10/11 (91%), Positives = 10/11 (91%), Gaps = 0/11 (0%)

```
Query  1   SLIHDDLPXMD  11
           SLIHDDLP MD
Sbjct  87  SLIHDDLPSMD  97
```

Appendix C

Artemisia annua QHS1 sequence and codon comparison to *Synechocystis* sp. strain PCC 6803

Coding Sequence for *Artemisia Annua* β-caryophyllene synthase gene (NCBI accession number: AF472361). The yellow highlighted region represents the putative start codon and the green highlighted region represents the putative stop codon. The sequence between these two highlighted regions were amplified out of the pETQHS1 vector and inserted into the pPSBAII vector.

```
   1 cataatccaa acttctcata gacatgtctg ttaaagaaga gaaagtaatt cggcccattg
  61 tccactttcc tcctagcgtt tgggcagacc agtttctcat cttcgatgac aagcaagcag
 121 agcaagccaa tgtagaacaa gtagtcaacg agttaagaga agatgtgaga aaagatttag
 181 tgtcatcttt agatgttcaa acggagcaca caaatttgct aaaacttatt gatgctatcc
 241 aacgtcttgg tatagcgtat cattttgaag aggagattga gcaagcctta caacatattt
 301 atgatacata tggtgatgac tggaaaggca gaagcccttc cctttggttt cgaatccttc
 361 ggcaacaagg tttttatgtt tcttgtgata tttttaaaaa ctataaaaaa gaggatggtt
 421 catttaagga atccctcacc aacgatgtag aaggcttgct tgagctgtat gaggcgacat
 481 atttgagagt gcaaggcgaa ggggttctag atgatgctct tgttttcaca aggacttgtc
 541 ttgagaaaat agcaaaggat cttgttcaca ccaacccaac actatctacc tacatacaag
 601 aagcactaaa acagccgtta cataaaaggt tgacaagact agaggcattg cgttacattc
 661 ctatgtacga acaacaagca tctcataatg agtctttact caaacttgcg aagctaggtt
 721 tcaacttact tcaatcattg catagaaagg agcttagcga agtttccagg tggtggaaag
 781 gtcttgatgt cccaaacaat ctaccttatg caagagatag aatggttgaa tgttatttt
 841 gggcactagg tgtctatttt gagccaaaat attctcaagc taggatcttt ttagcaaaag
 901 taatttcgct agcaactgtt cttgacgaca cttatgatgc ttatggaacc tatgaagaac
 961 ttaagatctt tactgaagca attcaaaggt ggtcgattac atgcatagat atgcttccag
1021 aatacctgaa actattatac caaggagtct tggatatata catagaaatg gaagaaataa
1081 tgggaaaaga gggaaaagca catcatctta gctatgctaa agaatctatg aaagagttta
1141 ttagaagcta tatgatggaa gcaaaatggg caaatgaggg gtatgtacca acagcagagg
1201 agcatatgtc agttgcgttt gtaagcagcg ggtatagcat gcttgcaaca acatgttttg
1261 ttggcatggg tgatatcgtt acagatgaag cattcaaatg ggctctgaca aaacctccta
1321 tcatcaaagc ttcatgtgca attgctagac tcatggatga tattcactct caaaaggagg
1381 agaaagaaag aatacatgtt gcatctagtg ttgaaagtta catgaagcaa tatgatgtga
1441 cagaggagca tgtccttaaa gtatttaaca agaaaatcga ggatgcgtgg aaagatataa
1501 cccgagaatc ccttgtgcgt aaagatattc cgatgcctct gatgatgcga gtgattaact
1561 tggcacaggt gatggatgtt ttatataaac ataaagacgg tttcactaat gtgggagaag
1621 aactcaagga tcatattaaa tctttgctcg ttcatcctat acctataataa gttccgaaac
1681 ttcatctcgt ttttagcctc cgatacactc aacaactcaa ctttctctca tcttcaaatg
1741 gttggttgtg gcaaatctta agtccttggg tgattgaaat cttcaaatgc gtgtttgtta
1801 ctacatgcat gtgatttttt tccccctctt ctgtgtggca tttgtatttg gatgatttaa
1861 taataaataa aggtgtgttt gtgttaaaaa aaaaaaaaaa aa
```

Codon Plot results

Results of the comparison of the β-Caryophyllene synthase from Artemisia annua odon usage with the codon usage of *Synechocystis sp.* Strain PCC6803. (Nakamura, 2000)
http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=113355

atg, 1 to 3 (Met)-1.00
tct, 4 to 6 (Ser)-0.24
gtt, 7 to 9 (Val)-0.24
aaa, 10 to 12 (Lys)-0.90
gaa, 13 to 15 (Glu)-0.79
gag, 16 to 18 (Glu)-0.21
aaa, 19 to 21 (Lys)-0.90
gta, 22 to 24 (Val)-0.28
att, 25 to 27 (Ile)-0.55
cgg, 28 to 30 (Arg)-0.09
ccc, 31 to 33 (Pro)-0.39
att, 34 to 36 (Ile)-0.55
gtc, 37 to 39 (Val)-0.24
cac, 40 to 42 (His)-0.32
ttt, 43 to 45 (Phe)-0.79
cct, 46 to 48 (Pro)-0.52
cct, 49 to 51 (Pro)-0.52
agc, 52 to 54 (Ser)-0.09
gtt, 55 to 57 (Val)-0.24
tgg, 58 to 60 (Trp)-1.00
gca, 61 to 63 (Ala)-0.20
gac, 64 to 66 (Asp)-0.25
cag, 67 to 69 (Gln)-0.06
ttt, 70 to 72 (Phe)-0.79
ctc, 73 to 75 (Leu)-0.10
atc, 76 to 78 (Ile)-0.38
ttc, 79 to 81 (Phe)-0.21
gat, 82 to 84 (Asp)-0.75
gac, 85 to 87 (Asp)-0.25
aag, 88 to 90 (Lys)-0.10
caa, 91 to 93 (Gln)-0.94
gca, 94 to 96 (Ala)-0.20
gag, 97 to 99 (Glu)-0.21
caa, 100 to 102 (Gln)-0.94
gcc, 103 to 105 (Ala)-0.37
aat, 106 to 108 (Asn)-0.60
gta, 109 to 111 (Val)-0.28
gaa, 112 to 114 (Glu)-0.79
caa, 115 to 117 (Gln)-0.94
gta, 118 to 120 (Val)-0.28
gtc, 121 to 123 (Val)-0.24
aac, 124 to 126 (Asn)-0.40
gag, 127 to 129 (Glu)-0.21
tta, 130 to 132 (Leu)-0.52
aga, 133 to 135 (Arg)-0.36
gaa, 136 to 138 (Glu)-0.79
gat, 139 to 141 (Asp)-0.75
gtg, 142 to 144 (Val)-0.23
aga, 145 to 147 (Arg)-0.36
aaa, 148 to 150 (Lys)-0.90
gat, 151 to 153 (Asp)-0.75
tta, 154 to 156 (Leu)-0.52
gtg, 157 to 159 (Val)-0.23
tca, 160 to 162 (Ser)-0.07
tct, 163 to 165 (Ser)- 0.24
tta, 166 to 168 (Leu)- 0.52
gat, 169 to 171 (Asp)- 0.75
gtt, 172 to 174 (Val)- 0.24
caa, 175 to 177 (Gln)- 0.94
acg, 178 to 180 (Thr)- 0.10
gag, 181 to 183 (Glu)- 0.21
cac, 184 to 186 (His)- 0.32
aca, 187 to 189 (Thr)- 0.16
aat, 190 to 192 (Asn)- 0.60
ttg, 193 to 195 (Leu)- 0.12
cta, 196 to 198 (Leu)- 0.09
aaa, 199 to 201 (Lys)- 0.90
ctt, 202 to 204 (Leu)- 0.11
att, 205 to 207 (Ile)- 0.55
gat, 208 to 210 (Asp)- 0.75
gct, 211 to 213 (Ala)- 0.30
atc, 214 to 216 (Ile)- 0.38
caa, 217 to 219 (Gln)- 0.94
cgt, 220 to 222 (Arg)- 0.19
ctt, 223 to 225 (Leu)- 0.11
ggt, 226 to 228 (Gly)- 0.42
ata, 229 to 231 (Ile)- 0.07
gcg, 232 to 234 (Ala)- 0.13
tat, 235 to 237 (Tyr)- 0.83
cat, 238 to 240 (His)- 0.68
ttt, 241 to 243 (Phe)- 0.79
gaa, 244 to 246 (Glu)- 0.79
gag, 247 to 249 (Glu)- 0.21
gag, 250 to 252 (Glu)- 0.21
att, 253 to 255 (Ile)- 0.55
gag, 256 to 258 (Glu)- 0.21
caa, 259 to 261 (Gln)- 0.94
gcc, 262 to 264 (Ala)- 0.37
tta, 265 to 267 (Leu)- 0.52
caa, 268 to 270 (Gln)- 0.94
cat, 271 to 273 (His)- 0.68
att, 274 to 276 (Ile)- 0.55
tat, 277 to 279 (Tyr)- 0.83
gat, 280 to 282 (Asp)- 0.75
aca, 283 to 285 (Thr)- 0.16
tat, 286 to 288 (Tyr)- 0.83
ggt, 289 to 291 (Gly)- 0.42
gat, 292 to 294 (Asp)- 0.75
gac, 295 to 297 (Asp)- 0.25
tgg, 298 to 300 (Trp)- 1.00

aaa, 301 to 303 (Lys)- 0.90
ggc, 304 to 306 (Gly)- 0.17
aga, 307 to 309 (Arg)- 0.36
agc, 310 to 312 (Ser)- 0.09
cct, 313 to 315 (Pro)- 0.52
tcc, 316 to 318 (Ser)- 0.17
ctt, 319 to 321 (Leu)- 0.11
tgg, 322 to 324 (Trp)- 1.00
ttt, 325 to 327 (Phe)- 0.79
cga, 328 to 330 (Arg)- 0.19
atc, 331 to 333 (Ile)- 0.38
ctt, 334 to 336 (Leu)- 0.11
cgg, 337 to 339 (Arg)- 0.09
caa, 340 to 342 (Gln)- 0.94
caa, 343 to 345 (Gln)- 0.94
ggt, 346 to 348 (Gly)- 0.42
ttt, 349 to 351 (Phe)- 0.79
tat, 352 to 354 (Tyr)- 0.83
gtt, 355 to 357 (Val)- 0.24
tct, 358 to 360 (Ser)- 0.24
tgt, 361 to 363 (Cys)- 0.67
gat, 364 to 366 (Asp)- 0.75
att, 367 to 369 (Ile)- 0.55
ttt, 370 to 372 (Phe)- 0.79
aaa, 373 to 375 (Lys)- 0.90
aac, 376 to 378 (Asn)- 0.40
tat, 379 to 381 (Tyr)- 0.83
aaa, 382 to 384 (Lys)- 0.90
aaa, 385 to 387 (Lys)- 0.90
gag, 388 to 390 (Glu)- 0.21
gat, 391 to 393 (Asp)- 0.75
ggt, 394 to 396 (Gly)- 0.42
tca, 397 to 399 (Ser)-0.07
ttt, 400 to 402 (Phe)- 0.79
aag, 403 to 405 (Lys)- 0.10
gaa, 406 to 408 (Glu)- 0.79
tcc, 409 to 411 (Ser)- 0.17
ctc, 412 to 414 (Leu)-0.10
acc, 415 to 417 (Thr)- 0.36
aac, 418 to 420 (Asn)- 0.40
gat, 421 to 423 (Asp)- 0.75
gta, 424 to 426 (Val)- 0.28
gaa, 427 to 429 (Glu)-
ggc, 430 to 432 (Gly)- 0.17
ttg, 433 to 435 (Leu)-0.12
ctt, 436 to 438 (Leu)-0.11
gag, 439 to 441 (Glu)-0.21
ctg, 442 to 444 (Leu)-0.05
tat, 445 to 447 (Tyr)- 0.83
gag, 448 to 450 (Glu)- 0.21
gcg, 451 to 453 (Ala)- 0.13
aca, 454 to 456 (Thr)- 0.16
tat, 457 to 459 (Tyr)- 0.83
ttg, 460 to 462 (Leu)-0.12
aga, 463 to 465 (Arg)- 0.36
gtg, 466 to 468 (Val)- 0.23
caa, 469 to 471 (Gln)- 0.94
ggc, 472 to 474 (Gly)- 0.17
gaa, 475 to 477 (Glu)- 0.79
ggg, 478 to 480 (Gly)-0.19
gtt, 481 to 483 (Val)- 0.24
cta, 484 to 486 (Leu)-0.09
gat, 487 to 489 (Asp)- 0.75
gat, 490 to 492 (Asp)- 0.75
gct, 493 to 495 (Ala)- 0.30
ctt, 496 to 498 (Leu)-0.11
gtt, 499 to 501 (Val)- 0.24
ttc, 502 to 504 (Phe)-0.21
aca, 505 to 507 (Thr)-0.16
agg, 508 to 510 (Arg)- 0.07
act, 511 to 513 (Thr)- 0.38
tgt, 514 to 516 (Cys)- 0.67
ctt, 517 to 519 (Leu)-0.11
gag, 520 to 522 (Glu)- 0.21
aaa, 523 to 525 (Lys)- 0.90
ata, 526 to 528 (Ile)- 0.07
gca, 529 to 531 (Ala)-0.20
aag, 532 to 534 (Lys)- 0.10
gat, 535 to 537 (Asp)- 0.75
ctt, 538 to 540 (Leu)-0.11
gtt, 541 to 543 (Val)- 0.24
cac, 544 to 546 (His)- 0.32
acc, 547 to 549 (Thr)- 0.36
aac, 550 to 552 (Asn)- 0.40
cca, 553 to 555 (Pro)-0.02
aca, 556 to 558 (Thr)-0.16
cta, 559 to 561 (Leu)- 0.09
tct, 562 to 564 (Ser)- 0.24
acc, 565 to 567 (Thr)- 0.36
tac, 568 to 570 (Tyr)- 0.17
ata, 571 to 573 (Ile)-0.07
caa, 574 to 576 (Gln)- 0.94
gaa, 577 to 579 (Glu)- 0.79
gca, 580 to 582 (Ala)- 0.20
cta, 583 to 585 (Leu)- 0.09
aaa, 586 to 588 (Lys)- 0.90
cag, 589 to 591 (Gln)-0.06
ccg, 592 to 594 (Pro)-0.07
tta, 595 to 597 (Leu)- 0.52
cat, 598 to 600 (His)- 0.68
aaa, 601 to 603 (Lys)- 0.90
agg, 604 to 606 (Arg)- 0.07
ttg, 607 to 609 (Leu)-0.12
aca, 610 to 612 (Thr)-0.16
aga, 613 to 615 (Arg)- 0.36
cta, 616 to 618 (Leu)- 0.09
gag, 619 to 621 (Glu)-0.21
gca, 622 to 624 (Ala)- 0.20
ttg, 625 to 627 (Leu)- 0.12
cgt, 628 to 630 (Arg)- 0.19
tac, 631 to 633 (Tyr)- 0.17
att, 634 to 636 (Ile)- 0.55
cct, 637 to 639 (Pro)- 0.52
atg, 640 to 642 (Met)- 1.00

tac, 643 to 645 (Tyr)- 0.17
gaa, 646 to 648 (Glu)- 0.79
caa, 649 to 651 (Gln)- 0.94
caa, 652 to 654 (Gln)-0.94
gca, 655 to 657 (Ala)- 0.20
tct, 658 to 660 (Ser)- 0.24
cat, 661 to 663 (His)- 0.68
aat, 664 to 666 (Asn)- 0.60
gag, 667 to 669 (Glu)- 0.21
tct, 670 to 672 (Ser)- 0.24
tta, 673 to 675 (Leu)- 0.52
ctc, 676 to 678 (Leu)- 0.10
aaa, 679 to 681 (Lys)- 0.90
ctt, 682 to 684 (Leu)-0.11
gcg, 685 to 687 (Ala)-0.13
aag, 688 to 690 (Lys)-0.10
cta, 691 to 693 (Leu)- 0.09
ggt, 694 to 696 (Gly)- 0.42
ttc, 697 to 699 (Phe)-0.21
aac, 700 to 702 (Asn)- 0.40
tta, 703 to 705 (Leu)- 0.52
ctt, 706 to 708 (Leu)-0.11
caa, 709 to 711 (Gln)- 0.94
tca, 712 to 714 (Ser)-0.07
ttg, 715 to 717 (Leu)-0.12
cat, 718 to 720 (His)- 0.68
aga, 721 to 723 (Arg)- 0.36
aag, 724 to 726 (Lys)-0.10
gag, 727 to 729 (Glu)- 0.21
ctt, 730 to 732 (Leu)- 0.11
agc, 733 to 735 (Ser)- 0.09
gaa, 736 to 738 (Glu)- 0.79
gtt, 739 to 741 (Val)- 0.24
tcc, 742 to 744 (Ser)-0.17
agg, 745 to 747 (Arg)- 0.07
tgg, 748 to 750 (Trp)- 1.00
tgg, 751 to 753 (Trp)- 1.00
aaa, 754 to 756 (Lys)- 0.90
ggt, 757 to 759 (Gly)- 0.42
ctt, 760 to 762 (Leu)-X 0.11
gat, 763 to 765 (Asp)- 0.75
gtc, 766 to 768 (Val) -0.24
cca, 769 to 771 (Pro)- 0.02
aac, 772 to 774 (Asn)- 0.40
aat, 775 to 777 (Asn)- 0.60
cta, 778 to 780 (Leu)-0.09
cct, 781 to 783 (Pro)- 0.52
tat, 784 to 786 (Tyr)- 0.83
gca, 787 to 789 (Ala)-0.20
aga, 790 to 792 (Arg)- 0.36
gat, 793 to 795 (Asp)- 0.75
aga, 796 to 798 (Arg)- 0.36
atg, 799 to 801 (Met)- 1.00
gtt, 802 to 804 (Val)- 0.24
gaa, 805 to 807 (Glu)- 0.79
tgt, 808 to 810 (Cys)- 0.67
tat, 811 to 813 (Tyr)- 0.83
ttt, 814 to 816 (Phe)- 0.79
tgg, 817 to 819 (Trp)- 1.00
gca, 820 to 822 (Ala)-0.20
cta, 823 to 825 (Leu)- 0.09
ggt, 826 to 828 (Gly)- 0.42
gtc, 829 to 831 (Val)- 0.24
tat, 832 to 834 (Tyr)- 0.83
ttt, 835 to 837 (Phe)- 0.79
gag, 838 to 840 (Glu)- 0.21
cca, 841 to 843 (Pro)- 0.02
aaa, 844 to 846 (Lys)- 0.90
tat, 847 to 849 (Tyr)- 0.83
tct, 850 to 852 (Ser)- 0.24
caa, 853 to 855 (Gln)-0.94
gct, 856 to 858 (Ala)- 0.30
agg, 859 to 861 (Arg)- 0.07
atc, 862 to 864 (Ile)- 0.38
ttt, 865 to 867 (Phe)- 0.79
tta, 868 to 870 (Leu)- 0.52
gca, 871 to 873 (Ala)- 0.20
aaa, 874 to 876 (Lys)- 0.90
gta, 877 to 879 (Val)- 0.28
att, 880 to 882 (Ile)- 0.55
tcg, 883 to 885 (Ser)-0.04
cta, 886 to 888 (Leu)- 0.09
gca, 889 to 891 (Ala)- 0.20
act, 892 to 894 (Thr)- 0.38
gtt, 895 to 897 (Val)- 0.24
ctt, 898 to 900 (Leu)- 0.11
gac, 901 to 903 (Asp)- 0.25
gac, 904 to 906 (Asp)- 0.25
act, 907 to 909 (Thr)- 0.38
tat, 910 to 912 (Tyr)- 0.83
gat, 913 to 915 (Asp)- 0.75
gct, 916 to 918 (Ala)- 0.30
tat, 919 to 921 (Tyr)- 0.83
gga, 922 to 924 (Gly)- 0.22
acc, 925 to 927 (Thr)- 0.36
tat, 928 to 930 (Tyr)- 0.83
gaa, 931 to 933 (Glu)- 0.79
gaa, 934 to 936 (Glu)- 0.79
ctt, 937 to 939 (Leu)- 0.11
aag, 940 to 942 (Lys)- 0.10
atc, 943 to 945 (Ile)- 0.38
ttt, 946 to 948 (Phe)- 0.79
act, 949 to 951 (Thr)- 0.38
gaa, 952 to 954 (Glu)- 0.79
gca, 955 to 957 (Ala)- 0.20
att, 958 to 960 (Ile)- 0.55
caa, 961 to 963 (Gln)- 0.94
agg, 964 to 966 (Arg)-0.07
tgg, 967 to 969 (Trp)- 1.00
tcg, 970 to 972 (Ser)-0.04
att, 973 to 975 (Ile)- 0.55
aca, 976 to 978 (Thr)- 0.16
tgc, 979 to 981 (Cys)- 0.33
ata, 982 to 984 (Ile)- 0.07

gat, 985 to 987 (Asp)- 0.75
atg, 988 to 990 (Met)- 1.00
ctt, 991 to 993 (Leu)-0.11
cca, 994 to 996 (Pro)-0.02
gaa, 997 to 999 (Glu)- 0.79
tac, 1000 to 1002 (Tyr)-0.17
ctg, 1003 to 1005 (Leu)- 0.05
aaa, 1006 to 1008 (Lys)- 0.90
cta, 1009 to 1011 (Leu)-0.09
tta, 1012 to 1014 (Leu)- 0.52
tac, 1015 to 1017 (Tyr)-0.17
caa, 1018 to 1020 (Gln)- 0.94
gga, 1021 to 1023 (Gly)-0.22
gtc, 1024 to 1026 (Val)- 0.24
ttg, 1027 to 1029 (Leu)-0.12
gat, 1030 to 1032 (Asp)- 0.75
ata, 1033 to 1035 (Ile)- 0.07
tac, 1036 to 1038 (Tyr)- 0.17
ata, 1039 to 1041 (Ile)- 0.07
gaa, 1042 to 1044 (Glu)- 0.79
atg, 1045 to 1047 (Met)- 1.00
gaa, 1048 to 1050 (Glu)- 0.79
gaa, 1051 to 1053 (Glu)- 0.79
ata, 1054 to 1056 (Ile)- 0.07
atg, 1057 to 1059 (Met)- 1.00
gga, 1060 to 1062 (Gly)-0.22
aaa, 1063 to 1065 (Lys)- 0.90
gag, 1066 to 1068 (Glu)-0.21
gga, 1069 to 1071 (Gly)- 0.22
aaa, 1072 to 1074 (Lys)- 0.90
gca, 1075 to 1077 (Ala)- 0.20
cat, 1078 to 1080 (His)- 0.68
cat, 1081 to 1083 (His)- 0.68
ctt, 1084 to 1086 (Leu)- 0.11
agc, 1087 to 1089 (Ser)- 0.09
tat, 1090 to 1092 (Tyr)- 0.83
gct, 1093 to 1095 (Ala)- 0.30
aaa, 1096 to 1098 (Lys)- 0.90
gaa, 1099 to 1101 (Glu)- 0.79
tct, 1102 to 1104 (Ser)- 0.24
atg, 1105 to 1107 (Met)- 1.00
aaa, 1108 to 1110 (Lys)- 0.90
gag, 1111 to 1113 (Glu)-0.21
ttt, 1114 to 1116 (Phe)- 0.79
att, 1117 to 1119 (Ile)- 0.55
aga, 1120 to 1122 (Arg)- 0.36
agc, 1123 to 1125 (Ser)-0.09
tat, 1126 to 1128 (Tyr)- 0.83
atg, 1129 to 1131 (Met)- 1.00
atg, 1132 to 1134 (Met)- 1.00
gaa, 1135 to 1137 (Glu)- 0.79
gca, 1138 to 1140 (Ala)- 0.20
aaa, 1141 to 1143 (Lys)- 0.90
tgg, 1144 to 1146 (Trp)- 1.00
gca, 1147 to 1149 (Ala-0.20
aat, 1150 to 1152 (Asn)- 0.60
gag, 1153 to 1155 (Glu)-0.21
ggg, 1156 to 1158 (Gly)- 0.19
tat, 1159 to 1161 (Tyr)- 0.83
gta, 1162 to 1164 (Val)- 0.28
cca, 1165 to 1167 (Pro)-0.02
aca, 1168 to 1170 (Thr)-0.16
gca, 1171 to 1173 (Ala)- 0.20
gag, 1174 to 1176 (Glu)- 0.21
gag, 1177 to 1179 (Glu)- 0.21
cat, 1180 to 1182 (His)- 0.68
atg, 1183 to 1185 (Met)- 1.00
tca, 1186 to 1188 (Ser)- 0.07
gtt, 1189 to 1191 (Val)- 0.24
gcg, 1192 to 1194 (Ala)-0.13
ttt, 1195 to 1197 (Phe)- 0.79
gta, 1198 to 1200 (Val)- 0.28
agc, 1201 to 1203 (Ser)- 0.09
agc, 1204 to 1206 (Ser)- 0.09
ggg, 1207 to 1209 (Gly)- 0.19
tat, 1210 to 1212 (Tyr)- 0.83
agc, 1213 to 1215 (Ser)- 0.09
atg, 1216 to 1218 (Met)- 1.00
ctt, 1219 to 1221 (Leu)- 0.11
gca, 1222 to 1224 (Ala)-0.20
aca, 1225 to 1227 (Thr)- 0.16
aca, 1228 to 1230 (Thr)- 0.16
tgt, 1231 to 1233 (Cys)- 0.67
ttt, 1234 to 1236 (Phe)- 0.79
gtt, 1237 to 1239 (Val)- 0.24
ggc, 1240 to 1242 (Gly)-0.17
atg, 1243 to 1245 (Met)- 1.00
ggt, 1246 to 1248 (Gly)- 0.42
gat, 1249 to 1251 (Asp)- 0.75
atc, 1252 to 1254 (Ile)- 0.38
gtt, 1255 to 1257 (Val)- 0.24
aca, 1258 to 1260 (Thr)-0.16
gat, 1261 to 1263 (Asp)- 0.75
gaa, 1264 to 1266 (Glu)- 0.79
gca, 1267 to 1269 (Ala)-0.20
ttc, 1270 to 1272 (Phe)- 0.21
aaa, 1273 to 1275 (Lys)- 0.90
tgg, 1276 to 1278 (Trp)- 1.00
gct, 1279 to 1281 (Ala)- 0.30
ctg, 1282 to 1284 (Leu)-0.05
aca, 1285 to 1287 (Thr)-0.16
aaa, 1288 to 1290 (Lys)- 0.90
cct, 1291 to 1293 (Pro)- 0.52
cct, 1294 to 1296 (Pro)- 0.52
atc, 1297 to 1299 (Ile)-
0.38
aaa, 1303 to 1305 (Lys)- 0.90
gct, 1306 to 1308 (Ala)- 0.30
tca, 1309 to 1311 (Ser)- 0.07
tgt, 1312 to 1314 (Cys)- 0.67
gca, 1315 to 1317 (Ala)-0.20
att, 1318 to 1320 (Ile)- 0.55
gct, 1321 to 1323 (Ala)- 0.30
aga, 1324 to 1326 (Arg)- 0.36

ctc, 1327 to 1329 (Leu)- 0.10
atg, 1330 to 1332 (Met)- 1.00
gat, 1333 to 1335 (Asp)- 0.75
gat, 1336 to 1338 (Asp)- 0.75
att, 1339 to 1341 (Ile)- 0.55
cac, 1342 to 1344 (His)- 0.32
tct, 1345 to 1347 (Ser)- 0.24
caa, 1348 to 1350 (Gln)- 0.94
aag, 1351 to 1353 (Lys)- 0.10
gag, 1354 to 1356 (Glu)-0.21
gag, 1357 to 1359 (Glu)-0.21
aaa, 1360 to 1362 (Lys)- 0.90
gaa, 1363 to 1365 (Glu)- 0.79
aga, 1366 to 1368 (Arg)- 0.36
ata, 1369 to 1371 (Ile)-0.07
cat, 1372 to 1374 (His)- 0.68
gtt, 1375 to 1377 (Val)- 0.24
gca, 1378 to 1380 (Ala)-0.20
tct, 1381 to 1383 (Ser)- 0.24
agt, 1384 to 1386 (Ser)- 0.39
gtt, 1387 to 1389 (Val)- 0.24
gaa, 1390 to 1392 (Glu)- 0.79
agt, 1393 to 1395 (Ser)- 0.39
tac, 1396 to 1398 (Tyr)- 0.17
atg, 1399 to 1401 (Met)- 1.00
aag, 1402 to 1404 (Lys)- 0.10
caa, 1405 to 1407 (Gln)- 0.94
tat, 1408 to 1410 (Tyr)- 0.83
gat, 1411 to 1413 (Asp)- 0.75
gtg, 1414 to 1416 (Val)-0.23
aca, 1417 to 1419 (Thr)- 0.16
gag, 1420 to 1422 (Glu)- 0.21
gag, 1423 to 1425 (Glu)-0.21
cat, 1426 to 1428 (His)- 0.68
gtc, 1429 to 1431 (Val)- 0.24
ctt, 1432 to 1434 (Leu)- 0.11
aaa, 1435 to 1437 (Lys)- 0.90
gta, 1438 to 1440 (Val)- 0.28
ttt, 1441 to 1443 (Phe)- 0.79
aac, 1444 to 1446 (Asn)- 0.40
aag, 1447 to 1449 (Lys)-0.10
aaa, 1450 to 1452 (Lys)- 0.90
atc, 1453 to 1455 (Ile)- 0.38
gag, 1456 to 1458 (Glu)- 0.21
gat, 1459 to 1461 (Asp)- 0.75
gcg, 1462 to 1464 (Ala)-0.13
tgg, 1465 to 1467 (Trp)- 1.00
aaa, 1468 to 1470 (Lys)- 0.90
gat, 1471 to 1473 (Asp)- 0.75
ata, 1474 to 1476 (Ile)-0.07
acc, 1477 to 1479 (Thr)- 0.36
cga, 1480 to 1482 (Arg)- 0.19
gaa, 1483 to 1485 (Glu)- 0.79
tcc, 1486 to 1488 (Ser)- 0.17
ctt, 1489 to 1491 (Leu)-0.11
gtg, 1492 to 1494 (Val)- 0.23
cgt, 1495 to 1497 (Arg)-0.19
aaa, 1498 to 1500 (Lys)- 0.90
gat, 1501 to 1503 (Asp)- 0.75
att, 1504 to 1506 (Ile)- 0.55
ccg, 1507 to 1509 (Pro)- 0.07
atg, 1510 to 1512 (Met)- 1.00
cct, 1513 to 1515 (Pro)- 0.52
ctg, 1516 to 1518 (Leu)-0.05
atg, 1519 to 1521 (Met)- 1.00
atg, 1522 to 1524 (Met)- 1.00
cga, 1525 to 1527 (Arg)- 0.19
gtg, 1528 to 1530 (Val)-0.23
att, 1531 to 1533 (Ile)- 0.55
aac, 1534 to 1536 (Asn)- 0.40
ttg, 1537 to 1539 (Leu)-0.12
gca, 1540 to 1542 (Ala)-0.20
cag, 1543 to 1545 (Gln)-0.06
gtg, 1546 to 1548 (Val)-0.23
atg, 1549 to 1551 (Met)- 1.00
gat, 1552 to 1554 (Asp)- 0.75
gtt, 1555 to 1557 (Val)-0.24
tta, 1558 to 1560 (Leu)- 0.52
tat, 1561 to 1563 (Tyr)- 0.83
aaa, 1564 to 1566 (Lys)- 0.9
cat, 1567 to 1569 (His)- 0.68
aaa, 1570 to 1572 (Lys)- 0.90
gac, 1573 to 1575 (Asp)- 0.25
ggt, 1576 to 1578 (Gly)- 0.42
ttc, 1579 to 1581 (Phe)- 0.21
act, 1582 to 1584 (Thr)- 0.38
aat, 1585 to 1587 (Asn)- 0.60
gtg, 1588 to 1590 (Val)-0.23
gga, 1591 to 1593 (Gly)-0.22
gaa, 1594 to 1596 (Glu)- 0.79
gaa, 1597 to 1599 (Glu)- 0.79
ctc, 1600 to 1602 (Leu)- 0.10
aag, 1603 to 1605 (Lys)-0.10
gat, 1606 to 1608 (Asp)- 0.75
cat, 1609 to 1611 (His)- 0.68
att, 1612 to 1614 (Ile)- 0.55
aaa, 1615 to 1617 (Lys)- 0.90
tct, 1618 to 1620 (Ser)- 0.24
ttg, 1621 to 1623 (Leu)-0.12
ctc, 1624 to 1626 (Leu)- 0.10
gtt, 1627 to 1629 (Val)- 0.24
cat, 1630 to 1632 (His)- 0.68
cct, 1633 to 1635 (Pro)- 0.52
ata, 1636 to 1638 (Ile)- 0.07
cct, 1639 to 1641 (Pro)- 0.52
ata, 1642 to 1644 (Ile)- 0.07
taa, 1645 to 1647 (End)- 0.50