UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Discovery, biosynthesis and evolutionary history of sioxanthin, a novel glycosylated carotenoid from marine bacteria Salinispora /

Permalink

https://escholarship.org/uc/item/7mn416k3

Author

Richter, Taylor Kristen Stratton

Publication Date

2014

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Discovery, biosynthesis and evolutionary history of sioxanthin, a novel glycosylated carotenoid from marine bacteria *Salinispora*

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

in

Marine Biology

by

Taylor Kristen Stratton Richter

Committee in charge:

Professor Bradley Moore, Chair
Professor Eric Allen
Professor Chambers Hughes
Professor Brian Palenik
Professor Kit Pogliano

2014

Copyright

Taylor Kristen Stratton Richter, 2014

All rights reserved.

The Dissertation of Taylor Kristen Stratton Richter is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2014

DEDICATION

For my grandfather, who supported and encouraged all my endeavors

Even when others thought I was crazy

Thanks Papa

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

APCI	atmospheric pressure chemical ionization
ATP	adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
CD	circular dichroism
COSY	correlation spectroscopy
CRISPR	clustered regularly interspaced short palindromic repeats
DCM	dichloromethane
DMAPP	dimethylallyl pyrophosphate
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
ESI	electrospray ionization
FDA	Food and Drug Administration
FPP	farnesyl pyrophosphate
G + C	guanosine and cytosine
GC	gas chromatography
GGPP	geranylgeranyl pyrophosphate
GPP	geranyl pyrophosphate
HGT	horizontal gene transfer
HMBC	heteronuclear multiple-bond correlation spectroscopy
HPLC	high performance liquid chromatography
HR-MS	high resolution-mass spectrometry

HSQC	heteronuclear single-quantum correlation spectroscopy
IPP	isopentyl pyrophosphate
LB	Luria-Bertani
MEGA	Molecular Evolution Genetics Analysis
MEP	2-C-methyl-D-erythritol 4-phosphate
MS	mass spectrometry
MUSCLE	Multiple Sequence Comparison by Log Expectation
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
NOESY	nuclear Overhauser effect
nr	non-redundant
NRPS	non-ribosomal peptide synthase
PCR	polymerase chain reaction
PKS	polyketide synthase
ROESY	rotating frame nuclear Overhauser effect spectroscopy
TOCSY	total correlation spectroscopy
tRNA	transfer ribonucleic acid
UV	ultraviolet
UV/Vis	ultraviolet/visible

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMES

ACKNOWLEDGEMENTS

This work is the result of the guidance of dozens of people who lent their expertise and support over the past six years:

First and foremost, I must acknowledge my advisor, Dr. Brad Moore, whose support was vital throughout my PhD training. Thank you for giving me a place in your lab, for your patience as I explored (and struggled through) new and unfamiliar areas of science, for being available for discussions and advice, and for being an excellent example of how to be a scientist and mentor.

I would like to thank the rest of my committee, Dr. Eric Allen, Dr. Brian Palenik, Dr. Kit Pogliano, and Dr. Chambers Hughes. Thank you for constantly challenging me with your questions and discussions and for supporting my ideas and endeavors. Thank you all for inspiring me with your enthusiasm for science.

I need to acknowledge the rest of the Moore lab who have made coming in to lab everyday both fun and fruitful. The past six years has given me the opportunity to interact with an extraordinary group of scientists in the Moore lab. Too many people to name have given me advice and encouragement, but a few people deserve individual acknowledgement. Dr. Alessandra Eustaquio and Dr. Amy Lane helped get this study of carotenoids going by sharing their expertise on genetic mutation and chemical analysis. Dr. Nadine Ziemert, Dr. Vinayak Agarwal, and Dr. Avena Ross were kind enough to donate their time to editing the content of this dissertation.

I had the wonderful opportunity to work with scientists in other labs. Thanks to the Jensen and Hughes groups for providing me with advice, samples, and lab

space. A special thanks to Dr. Chambers Hughes for his patience with teaching me analytical chemistry. Thank you to Dr. Paul Jensen for all of his guidance in this project and throughout my PhD.

I must acknowledge the SIO administration, who handles the very difficult job of keeping the graduate students sane and functioning with patience and smiles. A big thank you to the rest of the SIO community, which makes Scripps such a special and wonderful place to learn science. I have made such wonderful friends while these past six years who have helped me grow in to the person that I am today. I look forward to seeing what the future holds for all of us. Thank you to my family for supporting my PhD decision, even when they were not sure what I was thinking. Finally, a big thank you to my husband, Danny. I am very grateful for all of the love, support, and laughter that he provides.

VITA

2008-2014	Doctor of Philosophy, University of California, San Diego
2006-2008	Master of Science, University of California, San Diego
2002-2006	Bachelor of Science, University of California, San Diego

PUBLICATIONS

"Ecotype diversity and conversion in *Photobacterium profundum* strains". F.M. Lauro, E. A. Eloe-Fadrosh, **T.K.S. Richter**, S. Ferriera, J. Johnson, D.H. Bartlett. *PLoSOne* 2014 *in press*

"Bioactivity guided genome mining reveals and lomaiviticin biosynthetic gene cluster in *Salinispora tropica*" R.D. Kersten, A.L. Lane, M. Nett, **T.K. Richter**, B.M. Duggan, P.C. Dorrestein, B.S. Moore. *Chembiochem* 2013: 14(8).

"Complete genome sequence of the deep-sea bacterium *Psychromonas* strain CNPT3". F.M. Lauro, **T.K. Stratton**, R.A. Chastain, S. Ferriera, J. Johnson, S.M.D. Goldberg, A.A. Yayanos, D.H. Bartlett. *Genome Announc*. 2013: 1(3).

PRESENTATIONS

2014	**T.K.S. Richter**, B.S. Moore, and C.C. Hughes "Discovery and biosynthesis of sioxanthin, a new glycosylated carotenoid endemic to *Salinispora* marine bacteria" *Gordon Research Conference on Marine Natural Products*, Poster
2011	**T.K. Stratton** and B.S. Moore "Characterization of carotenoid biosynthesis uncovers an unusual genome arrangement in *Salinispora" International Symposium on the Biology of Actinomyces XVI*, Poster.

FELLOWSHIPS

2014	Scripps Institution of Oceanography Graduate Student Excellence Travel Award

2008 UC San Diego Regents Fellowship

2008 NSF Graduate Research Fellowship Honorable Mention

ACADEMIC TEACHING EXPERIENCE

2007 Teaching Assistant in "Marine Biochemistry", Scripps Institution of Oceanography, University of California, San Diego

2007 Teaching Assistant in "Bacteriology", Division of Biology, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Discovery, biosynthesis and evolutionary history of sioxanthin, a novel glycosylated carotenoid from marine bacteria *Salinispora*

by

Taylor Kristen Stratton Richter

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2014

Professor Bradley Moore, Chair

Members of the marine actinomycete genus *Salinispora* constitutively produce an orange pigment during vegetative growth. Investigations into the genome sequences identified putative carotenoid biosynthetic genes in four regions of the genome. Gene inactivation experiments in *S. tropica* CNB-440 confirmed that these four regions, consisting of two gene clusters and two independent genes, contribute to the production of a single carotenoid that is responsible for *Salinispora* pigmentation.

This distributed biosynthetic genome arrangement is unusual among literature about bacterial carotenogenesis and is counter to what is understood about the genome organization of secondary metabolites. Isolation and purification of carotenoids from *S. tropica* enabled the structural elucidation of the compound responsible for orange pigmentation. The compound is a novel carotenoid (2'*S*)-1'-(β-D-glucopyranosyloxy)-3',4'-didehydro-1',2'-dihydro-φ,ψ-caroten-2'-ol, which has been given the trivial name "sioxanthin". Sioxanthin is a C_{40} carotenoid, glycosylated on one end of the molecule and containing an aryl functional group on the other end. Glycosylation is unusual among the actinomycetes and sioxanthin represents a poorly studied group of carotenoids which are polar on one end and non-polar on the other. The addition of a hydroxyl group on the 2'-carbon was not predicted by bioinformatics and has not yet been identified in *Salinispora*. Gene sequence homology predicts that the sioxanthin biosynthetic pathway is present in all of the *Salinispora* as well as other members of the family Micromonosporaceae including the genera *Micromonopsora, Verrucosispora,* and *Actinoplanes.* Additionally, investigations of clustering of carotenoid biosynthetic genes in heterotrophic bacteria showed that a non-clustered genome arrangement is more common than the literature suggests, with nearly half of the investigated genomes showing a non-clustered architecture. The sioxanthin evolutionary history was explored via character trees that predicted the ancestral traits for each region of the pathway. These indicate that genes responsible for cyclization, isomerization, and glycosylation were horizontally acquired more recently in Micromonosporaceae evolution than the rest of the pathway. Comparisons of gene and species trees confirm that each of these acquisitions were from separate bacterial groups. It seems that the novel and unique

carotenoid, sioxanthin, is the result of integration of carotenoid biosynthetic genes from multiple sources.

1 Introduction

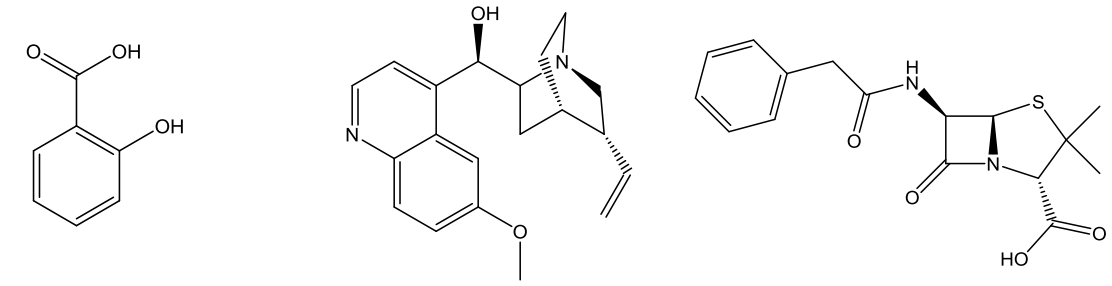
1.1 Background information

1.1.1 Introduction to natural products

Small molecules are the tool set through which living organisms detect, interact with, and adapt to each other and their surrounding environment. These chemical compounds are found all throughout nature and are often the result of an organism's secondary metabolism. Unlike primary metabolism, which includes universal life processes (the buildup and breakdown of cellular machinery, ATP production, respiration, etc.) and regulates basic cell function, secondary metabolism is typically organism-specific and serves ecological or adaptive functions[1]. While the biological function for most such compounds remains a mystery, many have been implicated in roles including feeding deterrents, quorum sensing molecules, defense mechanisms, settling cues, and nutrient scavenging, among others[1,2].

Humans have a long history of exploiting bioactive natural products for medicinal purposes, from the medicinal plants of ancient healers, to current drugs which rely on modern chemical techniques to isolate pure compounds from natural sources[3]. The earliest compounds used in medicine came from plants and included compounds such as salicylic acid **(1)** and quinine **(2)**, both still in use today as a pain reliever and anti-malarial, respectively[3] (Figure 1.1:). Since the discovery of penicillin **(3)**, the scientific world has also been looking to microbes as sources of novel bioactive molecules[3] (Figure 1.1). From 1981-2010, 30% of FDA approved drugs were derived from natural products and target maladies such as infectious disease, parasites, cancer, and inflammation, to name a few[4]. Bacteria, which are responsible for producing

approximately 10% of known bioactive natural products, are a growing source of novel chemistry[3].



salicylic acid (1) **quinine (2)** **penicillin G (3)**

Figure 1.1: Early known natural products with profound biological properties still in use today.

Bacteria are the planet's oldest life forms and inhabit every possible environmental niche[5,6]. This vast range of adaptive capabilities is often reflected in the structural diversity of bacterially-derived natural products. Natural product structural classes from bacteria include ribosomally and non-ribosomally produced peptides, modified amino acids, polyketides, terpenoids, glycosides, alkaloids, and structures that combine multiple classes[1]. Among the most prolific of microbial natural product producers are the actinomycetes, responsible for 45% of microbially-derived bioactive natural products[3].

The bacterial order Actinomycetales (phylum Actinobacteria) is a morphologically diverse group of filamentous, high G+C bacteria with an unprecedented capacity for natural product biosynthesis[7,8]. Of the 110 actinomycete genera, considerable novel chemistry has been discovered in the genus *Streptomyces*[8]. Indeed, a majority of natural product-derived antibiotics used in medicine, in addition to antifungal, anticancer, anti-parasitic, and anti-inflammatory compounds are from *Streptomyces* strains[9]. Though long believed to be only soil-dwelling bacteria, this

genus of over 900 species is now known to inhabit a diverse array of habitats, including deserts, deep sea sediments, and as symbionts of multicellular organisms[9]. Investigations into the chemistry of bacteria in these diverse environments, particularly the marine environment, has increased in recent years as the spread of antibiotic resistance calls for the discovery of novel compound classes[10].

1.1.2 Natural products in the marine environment

The ocean covers greater than 70% of the Earth's surface, making it home to a wide variety of bacteria in underexplored marine habitats that require unique secondary metabolic adaptations. Access to these new habitats has led to the discovery of novel chemistry from a host of marine organisms, including marine actinomycetes[11]. Actinomycetes from the marine environment are the source of a diverse set of compound classes that contain unique structural features and novel bioactivities (Figure 1.2; Table 1.1)[11]. Halogenated molecules, for example, are more common in the organohalogen-rich marine environment than in terrestrial environments and are known to modulate the bioactivity of many compounds[12].

As actinomycetes are thought of as terrestrial bacteria, it was long believed that those found in the marine environments were transients that had been washed in from the shore as dormant spores, without actually residing in or adapted to the ocean conditions[11]. However, studies have since confirmed that marine-adapted actinomycetes not only exist, but represent a significant addition to actinomycete diversity[13]. One such addition is the genus *Salinispora*, the first obligate marine

actinomycete.

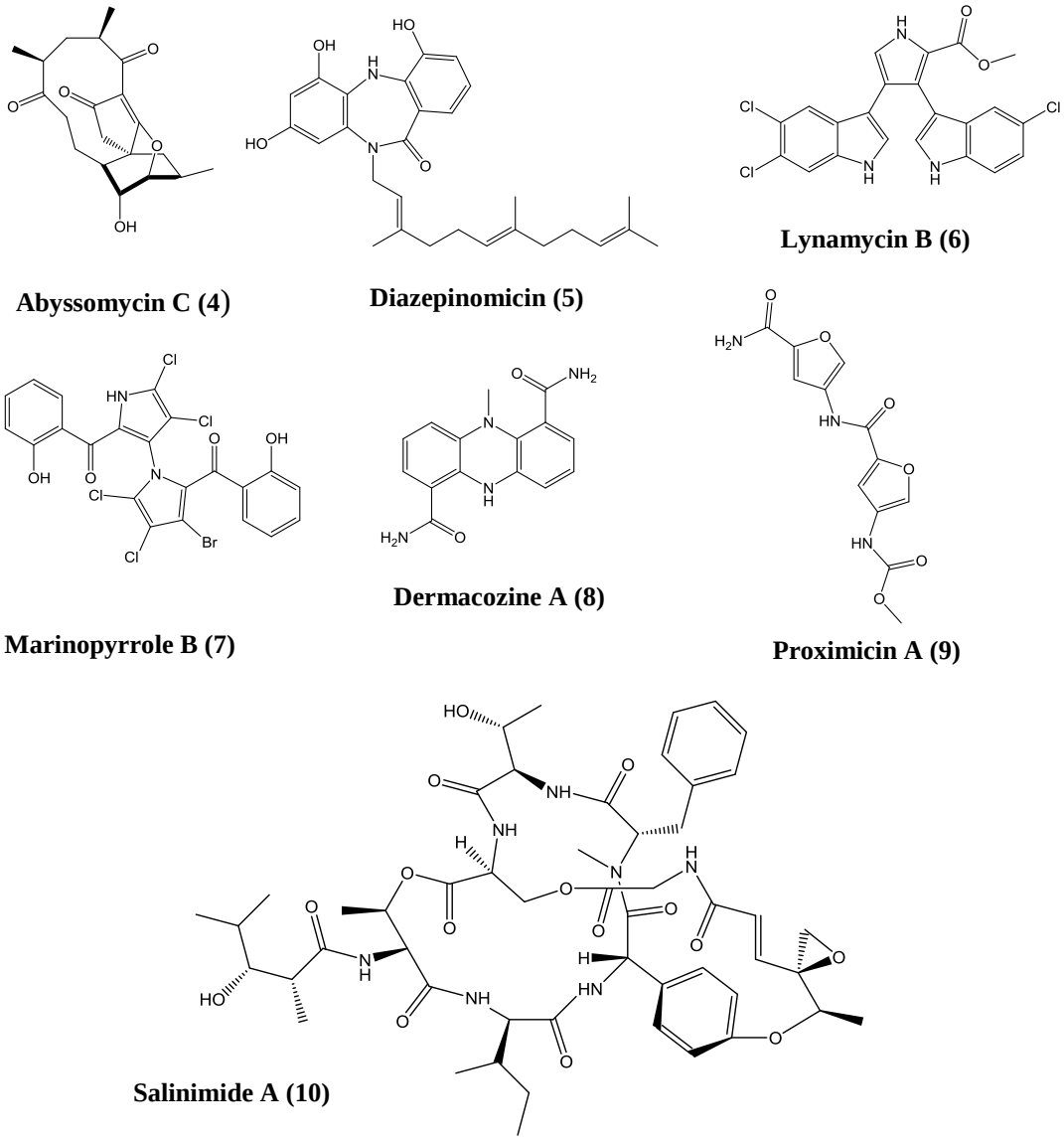


Figure 1.2: Structural diversity of natural products from marine actinomycetes (see Table 1 for information regarding activity, source and structural class)

Table 1.1: Description of the chemical and biological diversity of marine actinomycete natural products. Numbers correspond to structures in Figure 1.2

	Compound	Chemical class	Activity	Isolate genus	Reference
4	Abyssomycin C	Polycyclic polyketide	Antibacterial	*Verrucosispora*	[14]
5	Diazepinomicin	Farnesylated dibenzodiazepinone	Anticancer	*Micromonospora*	[15]
6	Lynamycin B	Bisindole pyrrole	Antibacterial	*Marinispora*	[16]
7	Marinopyrrole B	Bispyrrole	Antibacterial	*Streptomyces*	[17]
8	Dermacozine A	Phenazine	Cytotoxic, radical scavenging	*Dermacoccus*	[18]
9	Proximicin A	Aminofuran	Cytostatic	*Verrucosispora*	[19]
10	Salinamide A	Bicyclic depsipeptide	Antiinflammatory	*Streptomyces*	[20]

* Adapted from Zotchev 2012[11]

1.1.3 Introduction to the *Salinispora* and their secondary metabolism

The *Salinispora* were first identified in sediment samples from the Bahamas in 1991[21], and in the years since have been identified in tropical, sub-tropical, and temperate ocean sediments from around the world[22,23]. The genus contains three formally described species, *S. arenicola*, *S. tropica*, and *S. pacifica*[24,25]. The species are closely related with a 16S rRNA sequence identity greater than 99% and are morphologically indistinguishable with the same orange-pigmented filamentous growth and black spore formation[24,25].

The most distinguishing feature of these closely-related *Salinispora* species is their secondary metabolism. Studies of natural products in *Salinispora* have identified not only a wealth of bioactive compounds, but a species-specific distribution of metabolites (Figure 1.3: Selection of compounds isolated from each species of *Salinispora*. Star indicates that the compound is known only in *Salinispora* [26,28,31-35])[26]. The most promising drug candidate of the *Salinispora* natural products discovered to date, salinosporamide A **(11)**, was found in *S. tropica*, although a structural analog, termed salinosporamide K

(12), has been identified in *S. pacifica*[27,28]. Salinosporamide A is a proteasome inhibitor and has undergone phase II human clinical trials for cancer[29], including multiple myeloma, and is predicted to serve as a feeding deterrent in nature. *S. tropica* is host to seven pathways common throughout the species including the enediyene sporolide A **(12)** and the polyketide salinilactam A **(14)** (Figure 1.3)[26,30,31]. *S. arenicola* has thus far been the most prolific natural product producer among the *Salinispora* species with an average genome size 300kb larger and containing four more operational biosynthetic units (groups of pathways thought to produce related compounds) than the other two species[30]. Like *S. tropica*, *S. arenicola* species host seven pathways found in all members of the species, in addition to several that are unique to *S. arenicola*[26,30,31]. Though, *S. pacifica* has only one species specific pathway that is common throughout the species, it is the most chemically diverse species, with several pathways found only in a few strains, including cyanosporaside **(23)**, salinispyrone **(22)**, and the pacificanones **(24)**[26,30].

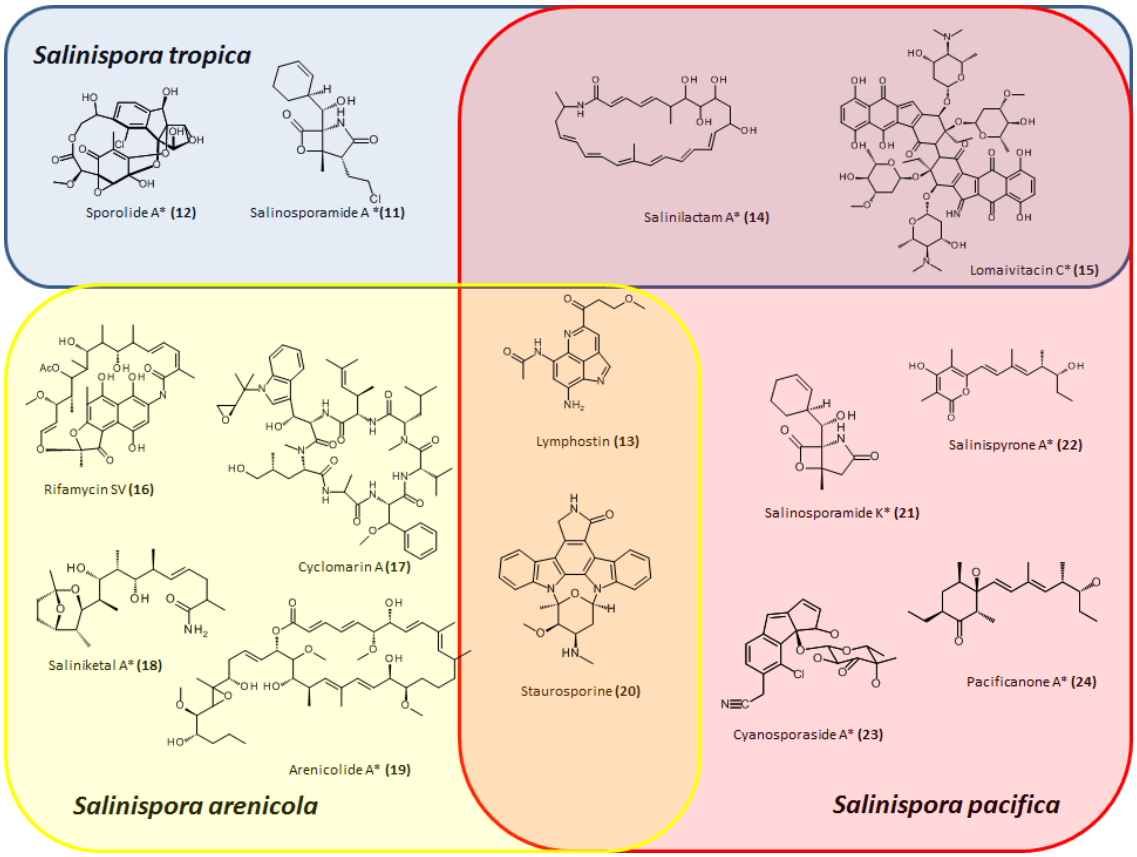


Figure 1.3: Selection of compounds isolated from each species of *Salinispora*. Star indicates that the compound is known only in *Salinispora* [26,28,31-35]

Although pathway diversity more closely follows phylogeny, each species' secondary metabolic capabilities do correlate with their individual biogeographic range, thereby suggesting a role for secondary metabolites in an organism's ability to expand into new habitats or adapt to a variety of niches. *S. arenicola* which produces the greatest number and structural diversity of secondary metabolites also has the most cosmopolitan distribution[23,30]. *S. tropica*, on the other hand, has the smallest chemical arsenal as well as the most limited geographic range, having only been found in the Bahamas and the Sea of Cortez[23,30]. *S. pacifica* has been found globally in sites where *S. tropica* is absent and, like *S. arenicola*, has a large, diverse suite of chemical capabilities[23,30].

1.1.4 Genome sequencing reveals untapped potential and secondary metabolic diversity of the *Salinispora*

Sequenced genomes are available for all three species (two closed and 86 draft genomes), and, along with the biogeography and chemical studies, have provided valuable insight into the diversity of these bacteria. Despite having higher than average sequence identity of individual genes, DNA hybridization across genomes is well below the 70% cutoff used to classify bacteria as a single species[25]. Much of the sequence divergence is localized to genomic islands[36], which are hot spots for horizontal gene transfer and accumulation of secondary metabolic gene clusters. These genome sequences have demonstrated the huge biosynthetic potential that remains untapped in *Salinispora*. This group of bacteria dedicates approximately 10% of their genome to secondary metabolism biosynthetic gene pathways, a higher proportion than even some of the most prolific terrestrial *Streptomyces*[31]. Bioinformatic analyses of genome sequences allows for a genome "mining" approach, in which pathways can be used to predict a compound of interest and targeted for investigation prior to chemical analyses such as bioactivity guided isolation, thus streamlining the hunt for novel chemistry[37]. In *S. tropica*, for example, genome analysis predicted the structures of lymphostin **(13)** and salinilactam A **(14)**, prior to their chemical isolation and identification[31].

The availability of genome sequences has exposed secondary metabolic potential that was overlooked by traditional chemical methods. Bioinformatic analyses of 75 *Salinispora* genome sequences have identified 124 polyketide (PKS) and nonribosomal peptide (NRPS) type pathways, though only nine have been formally linked to a secondary metabolite, showing a huge untapped chemical potential[30].

Furthermore, rarefaction curves of the pathway diversity suggest that the number of potential pathways is currently being greatly underestimated[30]. Combining chemical and molecular analyses with information from the *Salinispora* genomes provides a vital opportunity to link metabolite structures to pathways; a necessary step to understanding enzyme function for pathway engineering.

In addition to providing insight to the undiscovered chemistry of known bacteria, the availability of genome sequences provide insight into the genome arrangement and evolution of secondary metabolism. Microbial secondary metabolite biosynthetic genes are generally organized into gene clusters in which all of the genes of the pathway, including regulatory, resistance and transport elements are co-localized on the same or neighboring operons[1,30]. This conspicuous grouping of associated genes is such a prevalent feature of secondary metabolism that it is thought that there must be a fitness advantage to the evolution and maintenance of this genomic organization. The clustering of genes in a single pathway is proposed to improve coordination during horizontal gene transfer, regulation efficiency, shorter diffusion times for proteins finding their targets and forming complexes, and limits the probability of loss of function due to mutation[38,39]. It is likely a combination of these factors that create the pattern of gene clustering common in bacterial secondary metabolism.

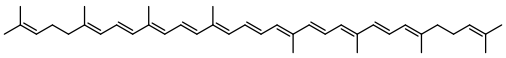
This genome organization is the basis of current genome mining and pathway prediction models (such as antiSMASH)[40]. Bacterial pathways that do not conform to the standard single gene cluster model remain challenging to interpret, making structural predictions for the product compound difficult. Indeed, some pathways are so refractory that they may be overlooked all together.

In *Salinispora*, gene clusters are proposed to have arisen as the result of frequent horizontal gene transfer. A recent study has shown that only five pathways were present in the common ancestor of *Salinispora*, with only two of them shared with the closely related genus *Micromonospora*[30]. This observation suggests that the remaining 96% of the pathways in the genus were acquired via horizontal gene transfer. Genomic islands, regions of the genome rich in secondary metabolic pathways, are home to mobile genetic elements, which may influence the transfer of pathways found in the region and provide further evidence of horizontal gene transfer in the acquisition of secondary metabolite pathways[30,36].

1.1.5 *Salinispora* pigmentation and introduction to carotenoids

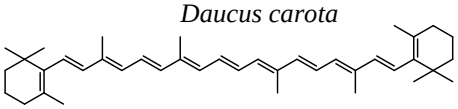
Despite all of the emphasis on *Salinispora* secondary metabolic diversity, there are several pathways that are shared throughout the genus. Surprisingly, some of these ubiquitous molecules still lack chemical characterization, including the compounds responsible for pigmentation. Orange pigmentation in bacteria, such as other actinomycetes, is primarily due to the accumulation of carotenoids. Carotenoids are highly conjugated, linear tetraterpenoids responsible for the majority of yellow, orange, and red pigmentation seen in all three domains of life[41]. Over 750 carotenoid structures have been identified in nature and they share a linear, conjugated backbone and a common biosynthesis [42] (Figure 1.4).

Solanum lycopersicum

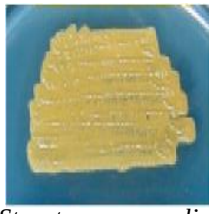


lycopene **(24)**

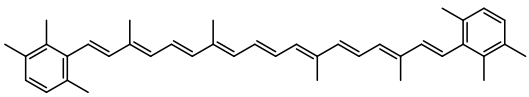
Daucus carota



β-carotene **(25)**

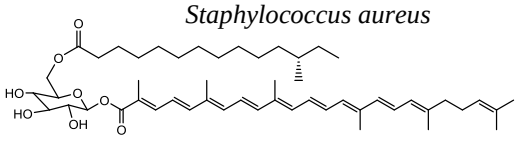


Streptomyces coelicolor



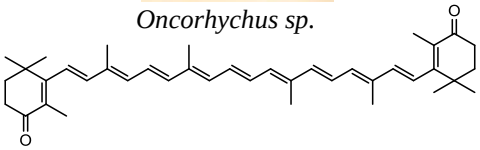
isorenieratene **(26)**

Staphylococcus aureus



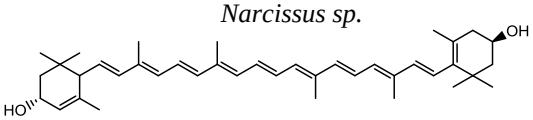
staphyloxanthin **(27)**

Oncorhychus sp.



canthaxanthin **(28)**

Narcissus sp.

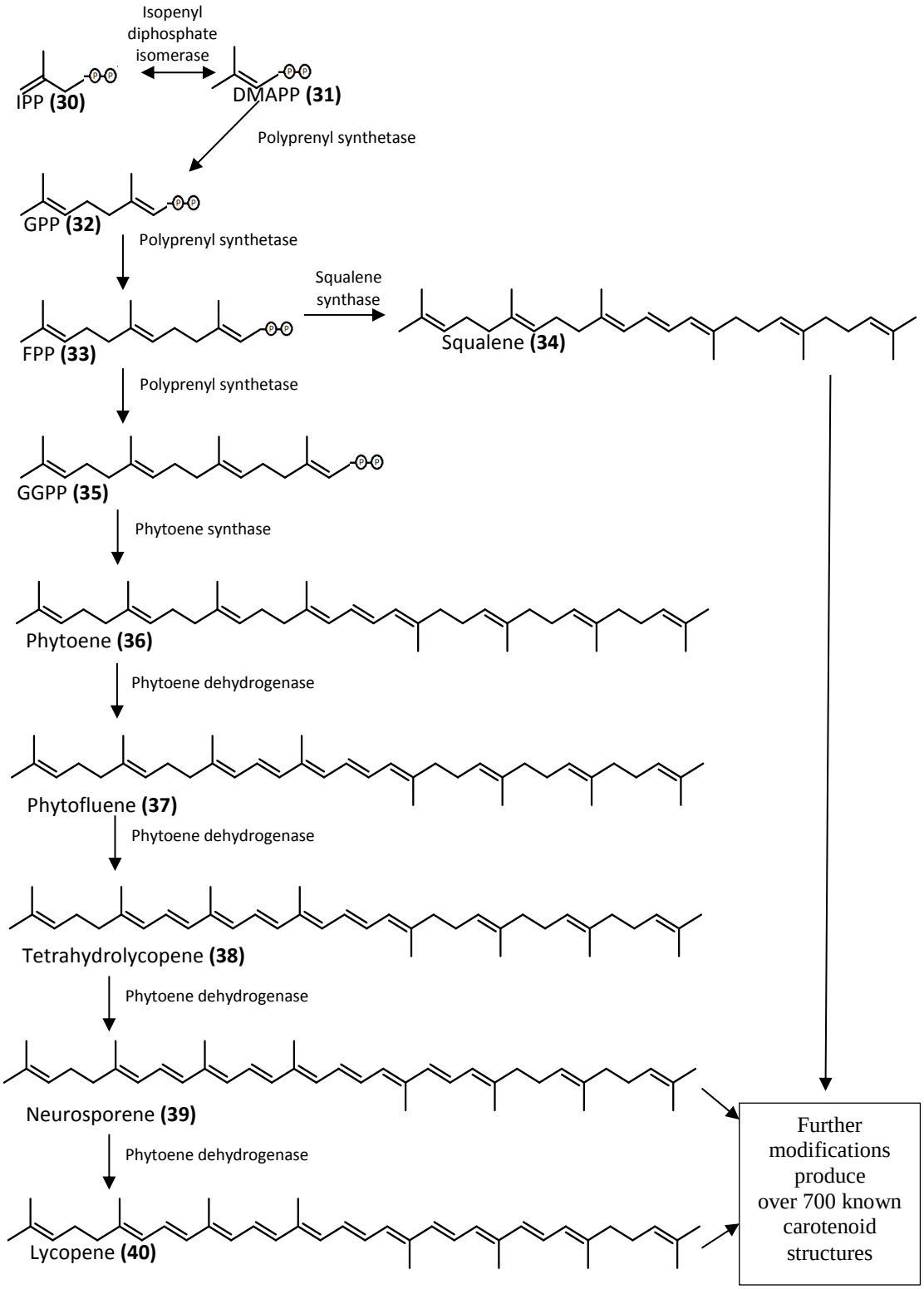


lutein **(29)**

Figure 1.4: Selection of carotenoids in nature and the organisms in which they are found.

The early steps involved in the formation of these molecules are conserved, though evolution has resulted in enzyme sequence divergence as well as diversity in the later steps of carotenoid modification[41]. As shown in Scheme 1.1, carotenoids are built

from the linear condensation of isoprene units, derived from primary metabolism, as is true in other terpene biosynthesis[43,44]. The first dedicated step in carotenoid biosynthesis is the tail-tail condensation of these chains[45]. Most often, it is of two molecules of geranylgeranyl pyrophosphate **(**GGPP **35)**, resulting in the production of phytoene **(36)**. Dehydrogenation of the chain produces the chromophore, characteristic of carotenoids. This long chain of unsaturated carbon bonds results in a cloud of delocalized electrons across the molecule that are easily excited (resulting in pigmentation) or can be donated to quench oxidized molecules (making a carotenoid a good antioxidant)[46]. Further modifications on either end of the molecule give rise to the structural diversity seen in nature. Carotenoids fall into two general categories: carotenes, which are purely hydrocarbon, and xanthophylls, which have oxygen-containing functional groups[47]. Bacterial carotenoid biosynthetic genes were first identified in *Rhodococcus capsulatus* where they form a cluster[47]. Similar arrangements have been found throughout bacteria, where clustering is the common arrangement except in the cyanobacteria[47].



Scheme 1.1: Generalized carotenoid biosynthesis. Isopentyl pyrophosphate(IPP); dimethylallyl pyrophosphate (DMAPP); geranyl pyrophosphate (GPP); farnesyl pyrophosphate (FPP); geranylgeranyl pyrophosphate (GGPP).

Carotenoid biosynthesis occurs in every branch of the tree of life, with the exception of animals where carotenoids are introduced by diet. Perhaps the most well-known role for carotenoids is in photosynthetic organisms. Plants and photosynthetic fungi and bacteria use carotenoids as accessory pigments in their light harvesting centers[48]. Carotenoids, which absorb light in the 420-550 nm range, partially fill in the gap in the visible light range that is missed by chlorophyll, thus expanding the wavelengths of light that photosynthetic organisms can utilize for energy production[48]. Additionally, carotenoids dissipate excess light, thereby protecting the light-harvesting complex from DNA and tissue damage due to UV irradiation[48]. Their biological function in non-phototrophic microorganisms is less obvious, though several functions have been reported for carotenoids in these organisms. The protective functions of carotenoids fall in to two basic categories: oxidative stress relief and membrane stability. As powerful antioxidants, carotenoid biosynthesis has been shown to increase in response to the presence of activated oxygen species directly[49]. In nature, this response has been seen in high copper concentration[50], light[51-53] and biofilm formation[54], each of which acts as a means of reactive oxygen species formation. Carotenoids are localized to cell membranes, where they can increase or decrease membrane fluidity depending on the polarity of their functional groups[55]. This has been shown in nature where pigmented bacteria are better able to withstand temperature extremes of polar and hot spring regions through the membrane modifying effects of carotenoids[56,57]. These molecules are able to exert both oxidative stress and membrane stability functions simultaneously. Expression of carotenoids in *Lactococcus lactis* improved its multi-stress tolerance, increasing its resistance to oxidative stressors

(hydrogen peroxide and low pH) as well as membrane destabilizers (bile and lysozyme)[58].

Carotenoids are also important contributors to industry, human health and nutrition due to their pigmentation and antioxidative properties. As a natural product, carotenoids represented a $1.2 USD billion market in 2010, with isolated compounds being used in everything from food colorants to supplements in the prevention of cancer and macular degeneration[59]. Carotenoids serve as virulence factors in pathogenic microbes such as *Staphylococcus aureus*, where the production of carotenoids makes them less susceptible to neutrophil killing, and have been implicated in suppressing the human immune system[60]. These compounds also enable *Enterococcus* species to persist longer in the environment by protecting against photoinactivation[61]. Better understanding of the biosynthesis and biological functions of carotenoids can have an impact on human health in addition to the food industry.

1.1.5 New insights from this dissertation

In this dissertation, the biosynthesis, structure and evolutionary history of sioxanthin, a novel carotenoid identified in *Salinispora* will be discussed. Using bioinformatics predictions and gene inactivation experiments, a biosynthetic pathway was uncovered in an unusual non-clustered arrangement that involves at least four separate regions of the genome contributing genes to the biosynthesis of a single molecule (see Chapter 2). This molecule is a novel, glycosylated carotenoid, which shares structural elements of other actinomycete carotenoids (see Chapter 3). Finally, by examining the phylogenetics of the biosynthetic pathway and exploring carotenoid biosynthesis in other bacteria, the evolution of this novel compound and the prevalence

of gene clustering in microbial carotenoid biosynthesis can be discussed (see Chapter 4). This work will lay the foundation for exploring the function of carotenoids in *Salinispora* in addition to expanding the understanding of genomic arrangement of secondary metabolite biosynthetic genes.

1.2 References

1 O'Brien, J. & Wright, G. D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **22**, 552-558, (2011).

2 Paul, V. J. & Ritson-Williams, R. Marine chemical ecology. *Nat. Prod. Rep.* **25**, 662-695, (2008).

3 Demain, A. L. & Sanchez, S. Microbial drug discovery: 80 years of progress. *J. Antibiot.* **62**, 5-16, (2009).

4 Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. *J. Nat. Prod.* **75**, 311-335, (2012).

5 Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 6578-6583, (1998).

6 Williams, D. M. & Embley, T. M.. Microbial diversity: Domains and kingdoms Vol. 27 *Annu. Rev. Ecol. Syst.* (ed D. G. Fautin) 569-595, 1996.

7 Goodfellow, M. & Williams, S. T. ECOLOGY OF ACTINOMYCETES. *Annu. Rev. Microbiol.* **37**, 189-216, (1983).

8 Berdy, J. Bioactive microbial metabolites - A personal view. *J. Antibiot.* **58**, 1-26, (2005).

9 Nett, M., Ikeda, H. & Moore, B. S. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat. Prod. Rep.* **26**, 1362-1384, (2009).

10 Donadio, S., Maffioli, S., Monciardini, P., Sosio, M. & Jabes, D. Antibiotic discovery in the twenty-first century: current trends and future perspectives. *J. Antibiot.* **63**, 423-430, (2010).

11 Zotchev, S. B. Marine actinomycetes as an emerging resource for the drug development pipelines. *J. Biotechnol.* **158**, 168-175, (2012).

12 Fenical, W. & Jensen, P. R. Developing a new resource for drug discovery: marine actinomycete bacteria. *Nat. Chem. Biol.* **2**, 666-673, (2006).

13 Mincer, T. J., Jensen, P. R., Kauffman, C. A. & Fenical, W. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Environ. Microbiol.* **68**, 5005-5011, (2002).

14 Bister, B., Bischoff, D., Strobele, M., Riedlinger, J., Reicke, A., Wolter, F., Bull, A. T., Zahner, H., Fiedler, H. P. & Sussmuth, R. D. Abyssomicin C - A

polycyclic antibiotic from a marine *Verrucosispora* strain as an inhibitor of the p-aminobenzoic acid/tetrahydrofolate biosynthesis pathway. *Angew. Chem. Int. Edit.* **43**, 2574-2576, (2004).

15 Charan, R. D., Schlingmann, G., Janso, J., Bernan, V., Feng, X. D. & Carter, G. T. Diazepinomicin, a new antimicrobial alkaloid from a marine *Micromonospora* sp. *J. Nat. Prod.* **67**, 1431-1433, (2004).

16 McArthur, K. A., Mitchell, S. S., Tsueng, G., Rheingold, A., White, D. J., Grodberg, J., Lam, K. S. & Potts, B. C. M. Lynamicins A-E, chlorinated bisindole pyrrole antibiotics from a novel marine actinomycete. *J. Nat. Prod.* **71**, 1732-1737, (2008).

17 Hughes, C. C., Prieto-Davo, A., Jensen, P. R. & Fenical, W. The marinopyrroles, antibiotics of an unprecedented structure class from a marine *Streptomyces* sp. *Org. Lett.* **10**, 629-631, (2008).

18 Abdel-Mageed, W. M., Milne, B. F., Wagner, M., Schumacher, M., Sandor, P., Pathom-aree, W., Goodfellow, M., Bull, A. T., Horikoshi, K., Ebel, R., Diederich, M., Fiedler, H. P. & Jaspars, M. Dermacozines, a new phenazine family from deep-sea dermacocci isolated from a Mariana Trench sediment. *Org. Biomol. Chem.* **8**, 2352-2362, (2010).

19 Fiedler, H. P., Brunter, C., Riedlinger, J., Bull, A. T., Knutsen, G., Goodfellow, M., Jones, A., Maldonado, L., Pathom-aree, W., Beil, W., Schneider, K., Keller, S. & Sussmuth, R. D. Proximicin A, B and C, novel aminofuran antibiotic and anticancer compounds isolated from marine strains of the actinomycete *Verrucosispora*. *J. Antibiot. (Tokyo)* **61**, 158-163, (2008).

20 Moore, B. S., Trischman, J. A., Seng, D., Kho, D., Jensen, P. R. & Fenical, W. Salinamides, antiinflammatory depsipeptides from a marine streptomycete. *J. Org. Chem.* **64**, 1145-1150, (1999).

21 Jensen, P. R., Dwight, R. & Fenical, W. Distribution of actinomycetes in near-shore tropical marine-sediments. *Appl. Environ. Microbiol.* **57**, 1102-1108, (1991).

22 Goo, K.-S., Tsuda, M. & Ulanova, D. *Salinispora arenicola* from temperate marine sediments: new intra-species variations and atypical distribution of secondary metabolic genes. *Anton. Leeuw. Int. J. G.* **105**, 207-219, (2014).

23 Jensen, P. R. & Mafnas, C. Biogeography of the marine actinomycete *Salinispora*. *Environ. Microbiol.* **8**, 1881-1888, (2006).

24 Maldonado, L. A., Fenical, W., Jensen, P. R., Kauffman, C. A., Mincer, T. J., Ward, A. C., Bull, A. T. & Goodfellow, M. *Salinispora arenicola* gen. nov., sp nov and *Salinispora tropica* sp nov., obligate marine actinomycetes belonging to

the family Micromonosporaceae. *Int. J. Syst. Evol. Micr.* **55**, 1759-1766, (2005).

25 Ahmed, L., Jensen, P. R., Freel, K. C., Brown, R., Jones, A. L., Kim, B. Y. & Goodfellow, M. *Salinispora pacifica* sp nov., an actinomycete from marine sediments. *Anton. Leeuw. Int. J. G.* **103**, 1069-1078, (2013).

26 Jensen, P. R., Williams, P. G., Oh, D.-C., Zeigler, L. & Fenical, W. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* **73**, 1146-1152, (2007).

27 Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. & Fenical, W. Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew. Chem. Int. Edit.* **42**, 355-+, (2003).

28 Eustaquio, A. S., Nam, S. J., Penn, K., Lechner, A., Wilson, M. C., Fenical, W., Jensen, P. R. & Moore, B. S. The discovery of salinosporamide K from the marine bacterium "*Salinispora pacifica*" by genome mining gives insight into pathway evolution. *Chembiochem* **12**, 61-64, (2011).

29 Indumathy, S. & Dass, C. R. Finding chemo: the search for marine-based pharmaceutical drugs active against cancer. *J. Pharm. Pharmacol.* **65**, 1280-1301, (2013).

30 Ziemert, N., Lechner, A., Weitz, M., Millan-Aguinaga, N., Chavarria, K. L & Jensen, P. R. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. U. S. A.*, (2014).

31 Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10376-10381, (2007).

32 Freel, K. C., Nam, S.-J., Fenical, W. & Jensen, P. R. Evolution of secondary metabolite genes in three closely related marine actinomycete species. *Appl. Environ. Microbiol.* **77**, 7261-7270, (2011).

33 Kersten, R. D., Lane, A. L., Nett, M., Richter, T. K. S., Duggan, B. M., Dorrestein, P. C & Moore, B. S. Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *Chembiochem* **14**, 955-962, (2013).

34 He, H. Y., Ding, W. D., Bernan, V. S., Richardson, A. D., Ireland, C. M., Greenstein, M., Ellestad, G. A. & Carter, G. T. Lomaiviticins A and B, potent antitumor antibiotics from *Micromonospora lomaivitiensis*. *J. Am. Chem. Soc.* **123**, 5362-5363, (2001).

35 Oh, D.-C., Gontang, E. A., Kauffman, C. A., Jensen, P. R. & Fenical, W. Salinipyrones and pacificanones, mixed-precursor polyketides from the marine actinomycete *Salinispora pacifica*. *J. Nat. Prod.* **71**, 570-575, (2008).

36 Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S. & Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* **3**, 1193-1203, (2009).

37 Corre, C. & Challis, G. L. New natural product biosynthetic chemistry discovered by genome mining. *Nat. Prod. Rep.* **26**, 977-986, (2009).

38 Pal, C. & Hurst, L. D. Evidence against the selfish operon theory. *Trends Genet.* **20**, 232-234, (2004).

39 Ballouz, S., Francis, A. R., Lan, R. & Tanaka, M. M. Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* **6**, (2010).

40 Medema, M. H., Blin, K., Cimermancic, P., de Jager, V., Zakrzewski, P., Fischback, M. A., Weber, T., Takano, E. & Breitling, R. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* **39**, W339-W346, (2011).

41 Armstrong, G. A. in *Annu. Rev. Microbiol.* Vol. 51 *Annual Review of Microbiology* (ed L. N. Ornston) 629-659 (Annual Reviews Inc. {a}, P.O. Box 10139, 4139 El Camino Way, Palo Alto, California 94306, USA, 1997).

42 Britton, G. L.-J. S. P. H. *Carotenoids handbook.* (Birkhäuser Verlag, 2004).

43 Kuzuyama, T., Takahashi, S., Dairi, T. & Seto, H. Detection of the mevalonate pathway in *Streptomyces* species using the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene. *J. Antibiot. (Tokyo)* **55**, 919-923, (2002).

44 Armstrong, G. A. Eubacteria show their true colors - genetics of carotenoid pigment biosynthesis from microbes to plants. *J. Bacteriol.* **176**, 4795-4802, (1994).

45 Sieiro, C., Poza, M., de Miguel, T. & Villa, T. G. Genetic basis of microbial carotenogenesis. *Int. Microbiol.* **6**, 11-16, (2003).

46 Britton, G. Structure and properties of carotenoids in relation to function. *FASEB J.* **9**, 1551-1558, (1995).

47 Armstrong, G. A. & Hearst, J. E. Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. *FASEB J.* **10**, 228-237, (1996).

48 Vershinin, A. Biological functions of carotenoids - diversity and evolution. *Biofactors* **10**, 99-104, (1999).

49 Galbis-Martinez, M., Padmanabhan, S., Murillo, F. J. & Elias-Arnanz, M. CarF mediates signaling by singlet oxygen, generated via photoexcited protoporphyrin IX, in *Myxococcus xanthus* light-induced carotenogenesis. *J. Bacteriol.* **194**, 1427-1436, (2012).

50 Moraleda-Munoz, A., Perez, J., Fontes, M., Murillo, F. J. & Munoz-Dorado, J. Copper induction of carotenoid synthesis in the bacterium *Myxococcus xanthus*. *Mol. Microbiol.* **56**, 1159-1168, (2005).

51 Pezzoni, M., Costa, C. S., Pizarro, R. A. & Oppezzo, O. J. The relationship between carotenoids and sunlight response in members of the family Micrococcaceae. *J. Basic Microbiol.* **51**, 325-329, (2011).

52 Takano, H., Obitsu, S., Beppu, T. & Ueda, K. Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol.* **187**, 1825-1832, (2005).

53 Takano, H., Asker, D., Beppu, T. & Ueda, K. Genetic control for light-induced carotenoid production in non-phototrophic bacteria. *J. Ind. Microbiol. Biotechnol.* **33**, 88-93, (2006).

54 Zheng, Y.-T., Toyofuku, M., Nomura, N. & Shigeto, S. Correlation of carotenoid accumulation with aggregation and biofilm development in *Rhodococcus* sp SD-74. *Anal. Chem.* **85**, 7295-7301, (2013).

55 Gruszecki, W. in *Carotenoids Physical, Chemical, and Biological Functions and Properties* (ed John T. Landrum) Ch. 2, (CRC Press, 2009).

56 Dieser, M., Greenwood, M. & Foreman, C. M. Carotenoid pigmentation in Antarctic heterotrophic bacteria as a strategy to withstand environmental stresses. *Arct. Antarct. Alp. Res.* **42**, 396-405, (2010).

57 Lutnaes, B. F., Strand, A., Petursdottir, S. K. & Liaaen-Jensen, S. Carotenoids of thermophilic bacteria - *Rhodothermus marinus* from submarine Icelandic hot springs. *Biochem. Syst. Ecol.* **32**, 455-468, (2004).

58 Hagi, T., Kobayashi, M., Kawamoto, S., Shima, J. & Nomura, M. Expression of novel carotenoid biosynthesis genes from *Enterococcus gilvus* improves the multistress tolerance of *Lactococcus lactis*. *J. Appl. Microbiol.* **114**, 1763-1771, (2013).

59 Marz, U. The Global Market for Carotenoids. *The Global Market for Carotenoids*, 189 pp, (2011).

60 Liu, G. Y. & Nizet, V. Color me bad: microbial pigments as virulence factors. *Trends Microbiol.* **17**, 406-413, (2009).

61 Maraccini, P. A., Ferguson, D. M. & Boehm, A. B. Diurnal variation in *Enterococcus* species composition in polluted ocean water and a potential role for the enterococcal carotenoid in protection against photoinactivation. *Appl. Environ. Microbiol.* **78**, 305-310, (2012).

2 Identification of carotenoid biosynthetic genes in *Salinispora*

2.1 Introduction

2.1.1 Pigments are a characteristic feature of *Salinispora*

Despite their differences in chemotype, the three species of *Salinispora* are morphologically indistinguishable[1,2]. This is due, in a large part, to pigmentation, which is visibly similar across species. Throughout vegetative growth, *Salinispora* are bright orange in color[1] (Figure 2.1). Then, as the culture ages, a black pigment is formed during sporulation[1] (Figure 2.1). The pigmentation schema is consistent for both solid and liquid cultures of various nutrient regimens, in both light and dark growth conditions. This observation led to the understanding that the orange pigment is not growth condition dependent and is constitutively expressed.

Figure 2.1: Growth of *S. tropica* CNB-440 showing the orange vegetative pigment and the development of black spores over time. Photos show growth at day 6, day 9, and day 18.

A growth stage pigment duality is not uncommon among the actinomycetes. Bacteria of the genus *Streptomyces*, for example, are known producers of a gray spore pigment[3-5]. Though a full structure has never been determined, the PKS-derived biosynthesis has been well studied. The *Salinispora* gene cluster (*strop2486-strop2510*) putatively responsible for spore pigmentation is homologous to that of the *whiE* locus from *Streptomyces coelicolor* A3(2), with additional genes for further modification. Similarly, many other actinomycetes, including *Streptomyces*, produce yellow, orange,

and red pigments during vegetative growth[6-8]. These colors are the result of carotenoid accumulation and are induced by exposure to light[9-11].

Presumably, the distinctive orange pigmentation in *Salinispora* is likewise due, at least in part, to the accumulation of carotenoids. In addition to the production of carotenoids in related bacteria, further support for the hypothesis is provided by identification of carotenoid biosynthesis gene homologs in the *Salinispora* genome[12].

2.1.2 Carotenoid biosynthesis is highly conserved

Carotenoid production is common throughout nature, with all three domains of life possessing versions of the biosynthetic pathway. As described in section 1.1.5, the early steps in carotenoid biosynthesis are highly conserved, with bacteria utilizing isoprene units derived from the non-mevalonate (MEP) pathway[13,14] to form the 40-carbon, desaturated lycopene.

Lycopene, then, becomes the starting material for a series of further enzymatic modifications that result in the great carotenoid diversity known in nature. These further modifications take place on the ends of the linear molecule, preserving the long chromophore backbone, while enabling additional structural and functional diversity.

As is common in other secondary metabolites, carotenoid biosynthetic genes in bacteria are typically clustered together in a single region of the genome. This gene clustering simplifies coordinated regulation of genes as well as complete transfer during horizontal gene transfer[15].

As part of a greater effort to link biosynthetic pathways to their resultant chemical structure, a goal of this project was to identify the biosynthetic genes responsible for orange pigmentation in *Salinispora*. Investigations in to the genome

sequences identified putative carotenoid biosynthetic genes in four regions of the genome. Gene inactivation experiments confirmed that these four genome regions, consisting of two gene clusters and two independent genes, contribute to the production of a single carotenoid compound that is responsible for *Salinispora* pigmentation. This distributed biosynthetic genome arrangement is unusual among bacterial carotenogenesis and is counter to what is understood about secondary metabolic genome organization. Furthermore, the genes identified to contribute to pigmentation are predicted to make a glycosylated compound, making it unique in this family of bacteria.

2.2 Results and Discussion

2.2.1 *Salinispora* carotenoid biosynthesis shared across species

Comparisons of carotenoid extracts from representatives of two *Salinispora* species showed the same HPLC retention times and UV spectra (Figure 2.2). The UV traces have an absorbance between 450-510 nm and a three-pronged peak, indicative of a carotenoid. *Salinispora* bacteria appear to produce two major carotenoids based on HPLC analysis, a polar compound eluting at about 3.5 min and a smaller, less polar molecule at 5 min. The more lipophilic compound is likely a precursor to the more polar compound, evident by the similarity in the UV spectra and the higher proportion in *S. arenicola*, which has a slower growth rate than *S. tropica*. This result suggests that *S. tropica* and *S. arenicola* not only produce carotenoids during vegetative growth, but that they likely produce the same compound. Thus, this observation suggested that genetic and chemical analyses could be focused on the faster growing species, *S. tropica*, and the conclusions applied to the other species.

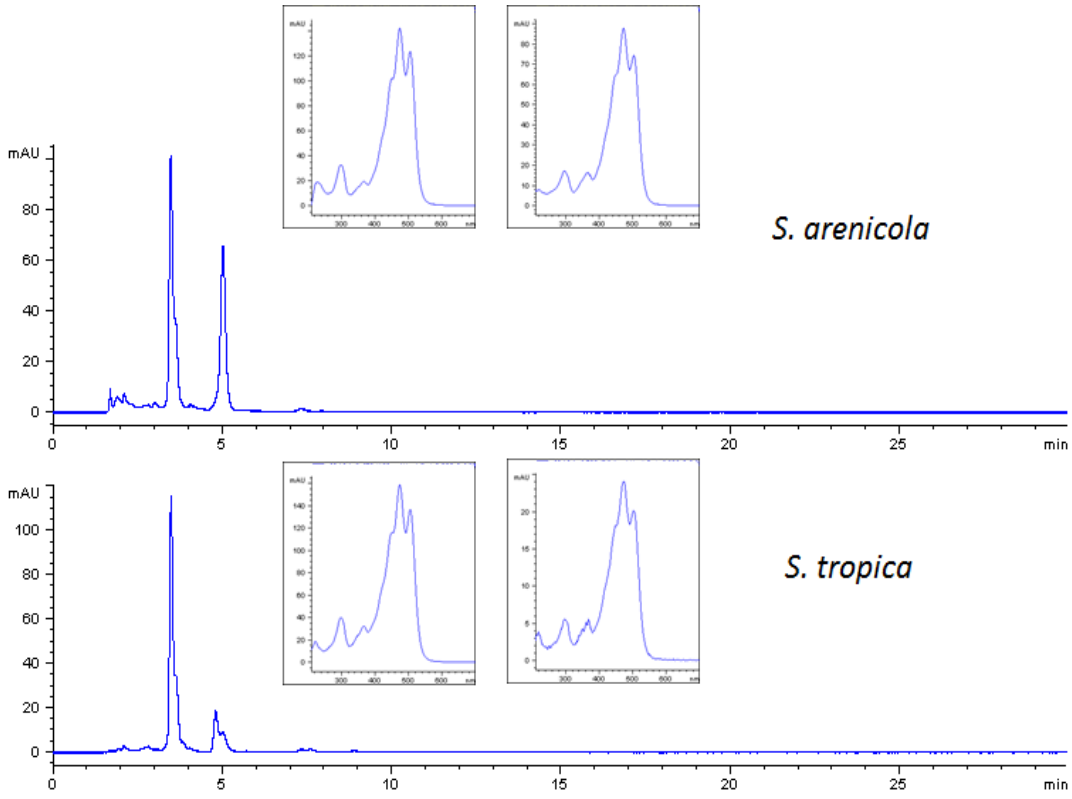


Figure 2.2: HPLC and UV profiles (at 450 nm) of the crude extracts of *Salinispora tropica* CNB-440 and *Salinispora arenicola* CNS-205.

2.2.2 Identification of candidate carotenogenesis genes

To investigate the biosynthetic origin of orange pigmentation believed to derive from carotenoid molecules, gene homology searches of known carotenoid biosynthetic genes in the *S. tropica* CNB-440 genome were performed. Using sequences of biochemically characterized carotenoid biosynthetic enzymes as a search query, putative carotenogenesis genes were identified in four regions of the *S. tropica* genome (Figure 2.3). Two regions had previously been identified as gene clusters *terp1* and *terp2* predicted to function in terpene biosynthesis[16]. The remaining two regions were individual genes located in isolation from other known terpene biosynthesis genes. Initial screening found only carotenogenesis genes homologous to those in

Streptomyces, and suggested the production of the carotene isorenieratene (**26**). HPLC profiles of *S. tropica* extracts differ from isorenieratene extracts from light-induced *Streptomyces coelicolor*, indicating that *S. tropica* does not produce isorenieratene (Figure 2.4). The shorter retention time of the major peak in *S. tropica* suggests a more polar compound, likely due to the presence of oxygen-containing functional groups. A more comprehensive search including sequences from more distantly related bacteria identified genes for additional modifications.

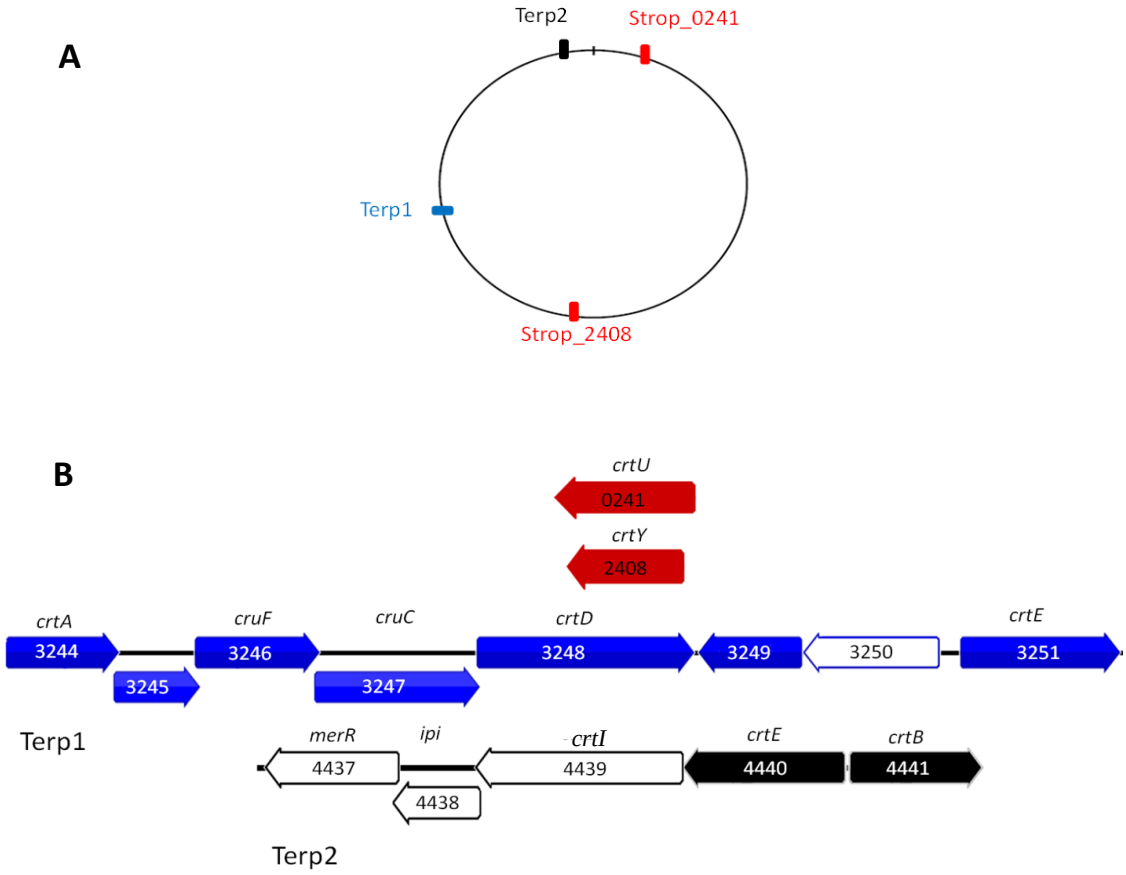


Figure 2.3: A) Map of *S. tropica* showing the relative locations of putative carotenoid biosynthesis genes. B) Expansion of the regions showing gene arrangements including gene locus number. Non-clustered genes are shown in red, *terp1* is in blue, and *terp2* in black. Solid colored genes are those which were successfully inactivated. Genes with letter codes are predicted to be involved in the carotenoid biosynthetic pathway.

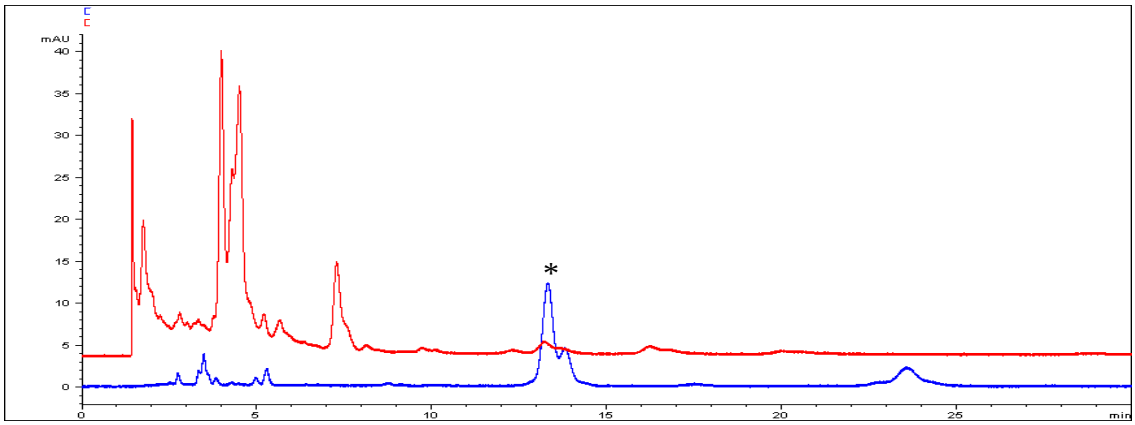


Figure 2.4: HPLC profiles (450 nm) of *S. tropica* crude extract (red) and partially purified isorenieratene (shown with *) from *S. coelicolor* (blue).

Because a non-clustered genomic arrangement of secondary metabolites is unusual in bacteria, we investigated the conservation and arrangement of these genes in other members of the genus. Two closed and 86 draft *Salinispora* genomes are currently available covering all three species. Analysis of the carotenoid genes and surrounding areas indicate that every strain of *Salinispora* had two *terp* clusters with identical gene arrangement in addition to two non-clustered genes in matching gene neighborhoods (Figure 2.5, Figure 2.6, Figure 2.7,

Figure 2.8). The identification of these genes in other species of *Salinispora* demonstrates that the carotenoid biosynthesis genomic pattern, in addition to the compound produced, is conserved throughout the genus.

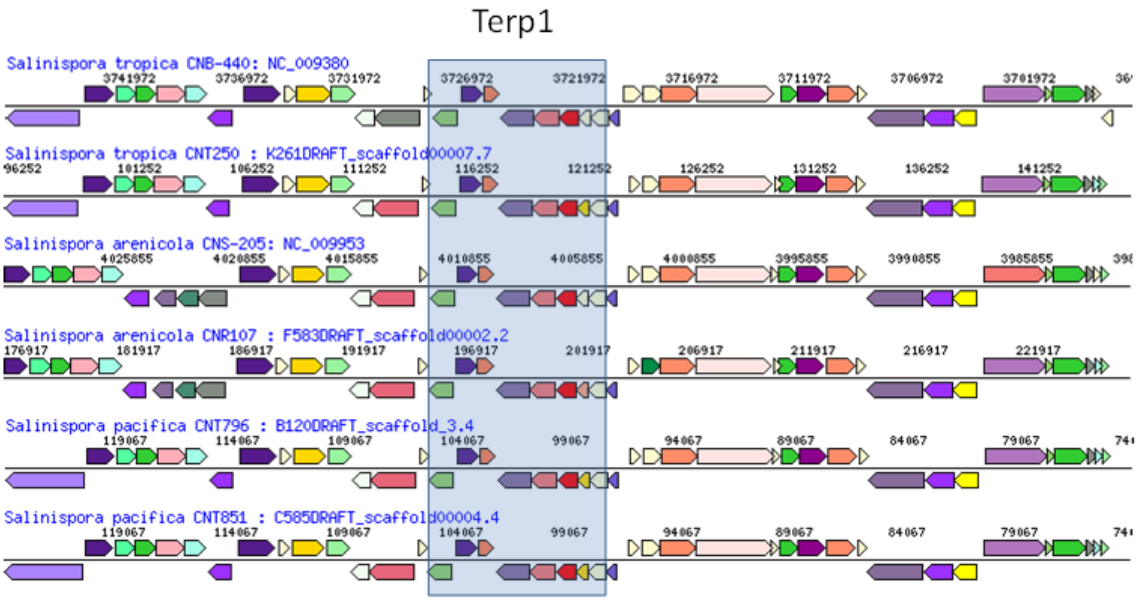


Figure 2.5: Alignments of *terp1* genes and gene neighborhoods from genomes of selected strains of *S. tropica*, *S. arenicola*, and *S. pacifica* showing conservation of genome organization across species.

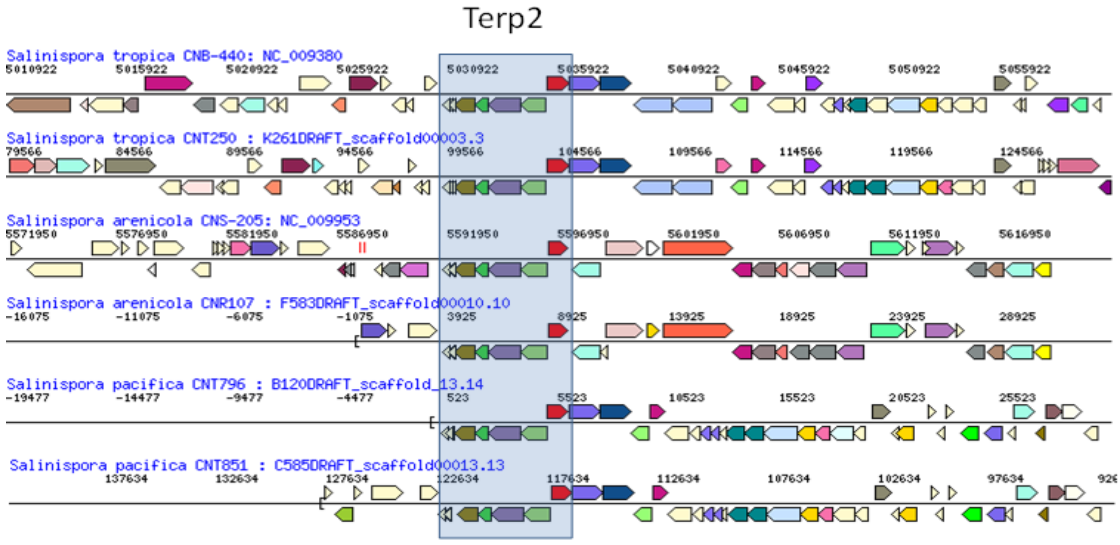


Figure 2.6: Alignments of *terp2* genes and gene neighborhoods from genomes of selected strains of *S. tropica*, *S. arenicola*, and *S. pacifica* showing conservation of genome organization across species.

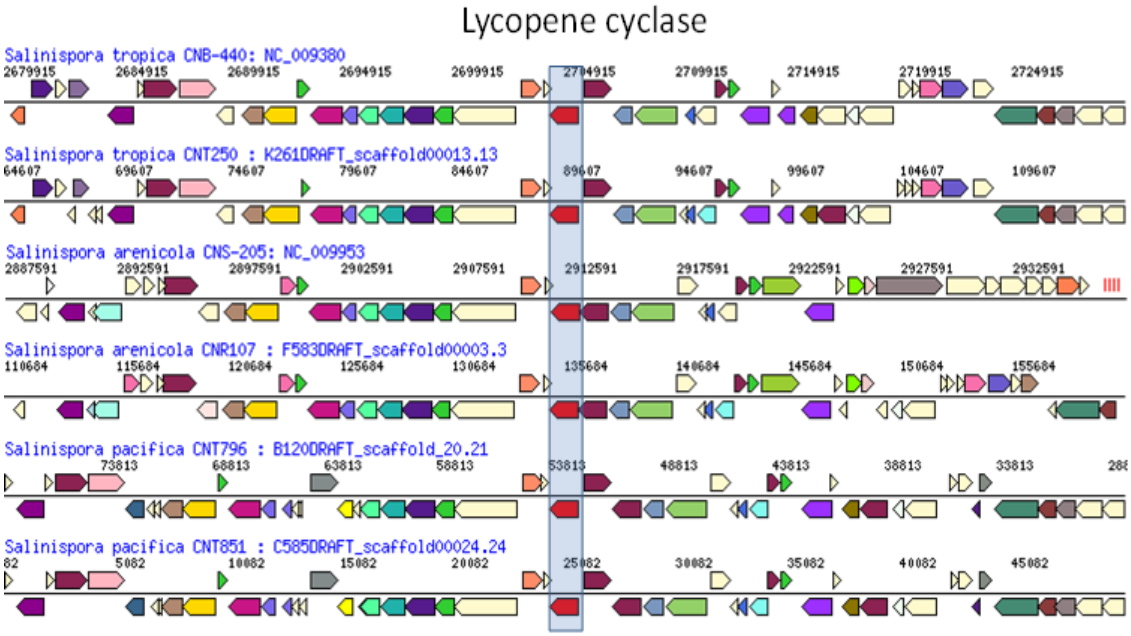


Figure 2.7: Alignments of lycopene cyclase genes and gene neighborhoods from genomes of selected strains of *S. tropica*, *S. arenicola*, and *S. pacifica* showing conservation of genome organization across species.

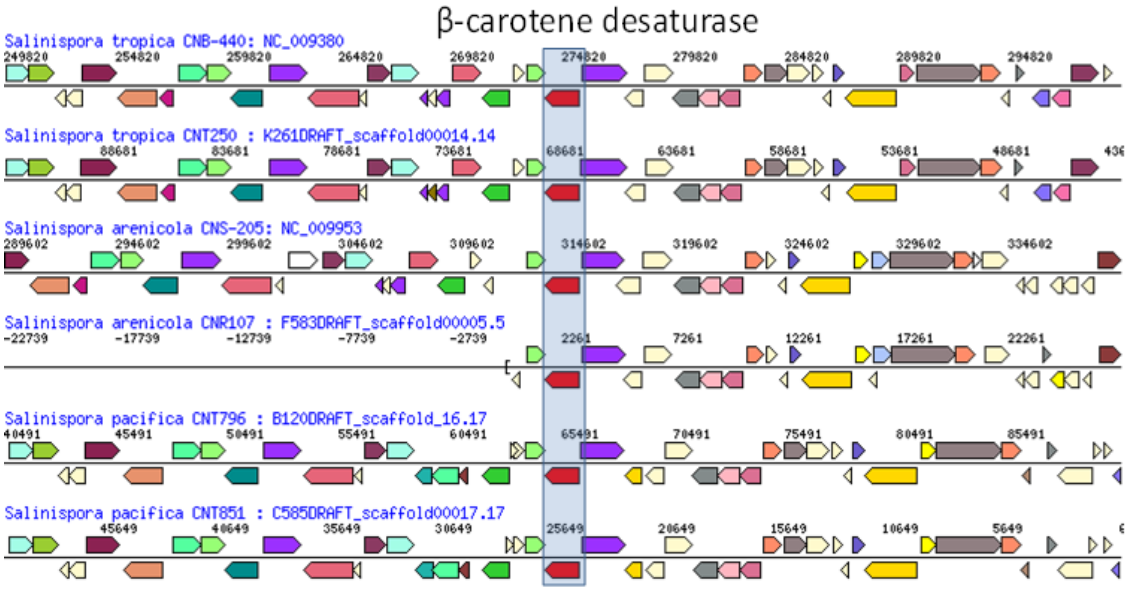


Figure 2.8: Alignments of β-carotene desaturase genes and gene neighborhoods from genomes of selected strains of *S. tropica*, *S. arenicola*, and *S. pacifica* showing conservation of genome organization across species.

The four regions of the genome will be discussed individually, beginning with the *terp1* and *terp2* gene clusters that were first identified during a survey of secondary metabolic gene clusters in *S. tropica*[16].

The *terp1* gene cluster, the region occupied by *strop3244 –strop3251*, includes a mixture of genes with high, low, and no homology to known carotenogenesis pathways. The borders of the cluster are loosely defined by predicted gene function and previous annotation[12,16]. Two genes, *strop3248* and *strop3251*, have high homology to genes in the lycopene biosynthesis pathway, and are predicted to be necessary to carotenoid biosynthesis (Table 2.1). Three other genes in this cluster have low homology to genes involved in structural modification of lycopene, specifically in carotenoid glycosylation (*strop3246* and *strop3247*) and ketolation/hydroxylation (*strop3244*). Query sequences for these glycosylation and ketolation/hydroxylation came from phylogenetically distant bacteria, which could decrease the strength of homology. The other genes in the *terp1* cluster had no predicted function in carotenoid biosynthesis, but were included in further analyses as they were situated between carotenogenesis genes and, therefore, may represent novel carotenoid biosynthetic functions.

Table 2.1: Results of BLAST analysis of the predicted *S. tropica* carotenogenesis genes.

Cluster	Locus	Gene annotation	Predicted carotenoid function	Genome	% identity
Terp1	3244	Spheroidene monooxygenase	*crtA*	*Rhodobacter sphaeroides*	33%
	3245	N-acetyltransferase	none		
	3246	Protein of unknown function	*cruF*	*Meiothermus ruber*	36%
	3247	Glycosyl transferase	*cruC*	*Meiothermus ruber*	42%
	3248	Phytoene dehydrogenase	*crtI* *crtD*	*Streptomyces griseus* *Meiothermus ruber*	31% 33%
	3249	Alcohol phosphatidyl transferase	none		
	3250	5,10- methylene tetrahydrofolate reductase	none		
	3251	Polyprenyl synthetase	*crtE*	*Streptomyces griseus*	42%
	3252	Hypothetical protein	none		
	3253	Protein kinase	none		
Terp2	4437	Regulatory protein	*merR*	*Frankia sp. Eul1c*	58%
	4438	Isopentyl diphosphate isomerase	*ipi*	*Nocardia brasiliensis*	66%
	4439	Phytoene dehydrogenase	*crtI*	*Streptomyces griseus*	54%
	4440	Polyprenyl synthetase	*crtE*	*Streptomyces griseus*	43%
	4441	Phytoene synthase	*crtB*	*Frankia sp. Eul1c*	50%
none	0241	Amine oxidase	*crtU*	*Pseudonocardia dioxanivorans*	51%
none	2408	Lycopene cyclase	*crtY*	*Streptomyces avermitilis*	47%

The *terp2* gene cluster comprises five genes, *strop4437--strop4441*, all of which have a predicted role in carotenoid biosynthesis. The cluster borders are well defined[12] and are bounded by tRNA genes on one side, while the other side was determined by the loss of synteny when compared to other *Salinispora* genomes. The *terp2* cluster is

comprised entirely of predicted carotenoid biosynthesis genes (Table 2.1). These genes are homologous to those involved in the early dedicated steps of carotenoid biosynthesis (Scheme 1.1), including the merR regulatory gene, common in these pathways[10]. Both *terp* clusters contain polyprenyl synthetase and phytoene dehydrogenase genes. Interestingly, among the predicted carotenogenesis genes, only the *terp2* cluster is located within a *Salinispora* genomic island[12].

Two additional predicted carotenogenesis genes were found elsewhere in the genome, independent of other related genes. These genes, *strop2408* and *strop0241*, are homologous to genes involved in lycopene cyclization and β-carotene desaturation/methylation, respectively. These genes are immediately surrounded by hypothetical genes, as well as those involved in protein biosynthesis/modification.

2.2.3 Inactivation of predicted genes and analysis of mutant phenotypes

As bacterial carotenoid biosynthetic genes are typically clustered based on the literature, the identification of non-clustered genes in multiple regions of the genome was unexpected. Their presence, however, was not necessarily indicative of their involvement in the pathway. In order to confirm the biosynthetic pathway of the *Salinispora* pigment, the predicted carotenogenesis genes were inactivated by PCR-directed mutagenesis in which the gene of interest was replaced with a gene encoding antibiotic resistance[17,18] and the resultant mutants analyzed for altered phenotypes.

Both the *terp1* and *terp2* clusters contain genes with predicted roles in lycopene biosynthesis. This isoprene, which is derived from the head-to-head condensation of two geranylgeranyl pyrophophtate (GGPP) molecules serves as the central precursor to the majority of carotenoids. Thus, phytoene synthase (*strop4441*), in *terp2*, and

polyprenyl synthetases (*strop4440* and *strop3251*), in both loci, were genetically inactivated, and extracts of the mutants were analyzed by HPLC. Inactivation of *strop4441* revealed a complete loss of orange pigmentation as demonstrated by both the visible phenotype, as well as the absence of compounds from the HPLC trace at 450 nm (Figure 2.9). The genetic interrogation of the two polyprenyl synthetases that share 67% sequence identity revealed that just the *strop4440* homolog in *terp2* is associated with visible carotenoid pigmentation as the *strop3251* knowckout mutant showed no alteration from the wild-type. These observations suggest that the *terp2* locus is largely responsible for the biosynthesis of carotenoid precursors in *Salinispora*. Further, these experiments demonstrate for the first time that carotenoids alone are responsible for vegetative pigmentation in these bacteria.

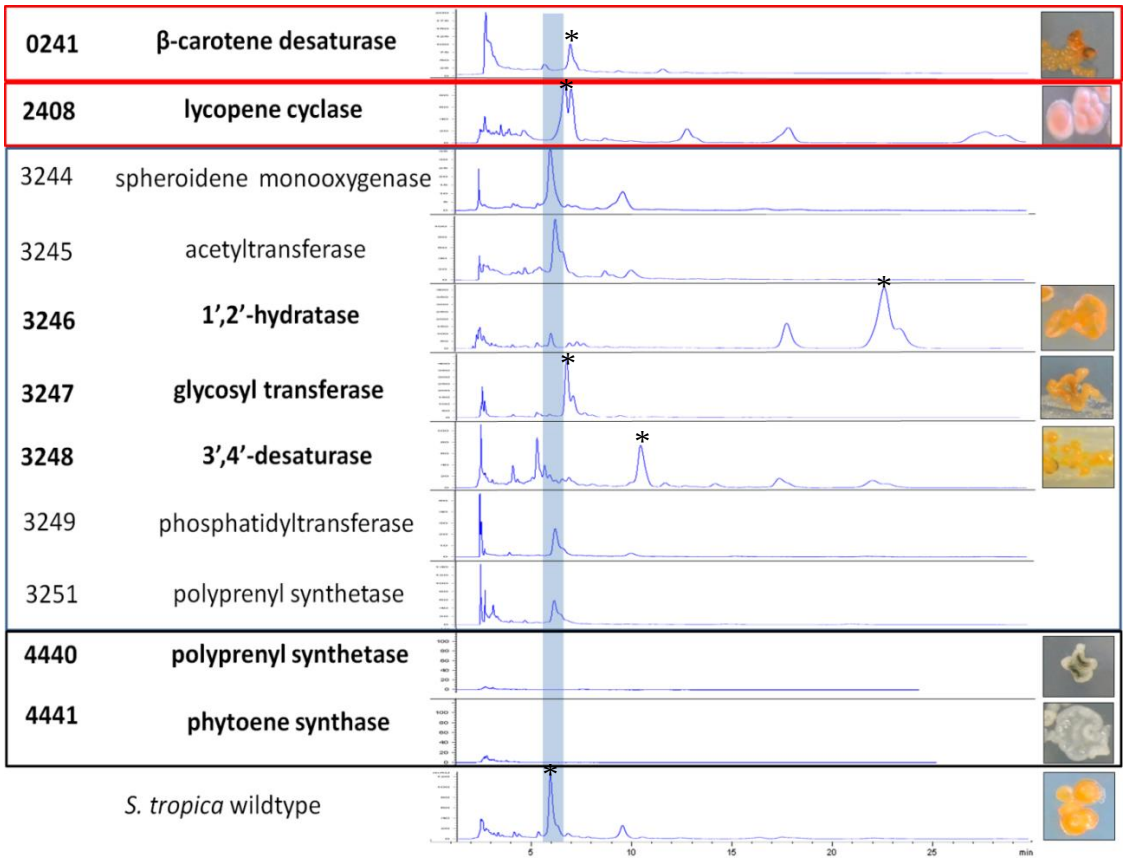


Figure 2.9: Results of gene inactivation experiments showing gene locus, predicted biosynthetic role, HPLC trace and photo of carotenoid mutant colonies. Vertical blue bar shows retention time of wild-type peak. Mutations that altered carotenoid production are highlighted in bold and are shown in the photographs. ***** shows major mutant carotenoid compound targeted for further investigation in Chapter 3.

The remainder of the genes in the *terp1* cluster were similarly interrogated to determine their roles in the biosynthetic pathway. Of the eight genes investigated in the *strop3244—strop3251* locus, carotenoid extracts of three gene knockout mutants showed compounds at different HPLC retention times and, in some cases, with different UV spectra than the wild-type (Figure 2.9). These observations indicate that although the mutants continue to synthesize carotenoids, the molecules produced are different from the two wild-type carotenoids, thereby confirming a role for these genes in the biosynthesis. These genes include a hydratase (*strop3246*) which is predicted to serve

as an attachment site for a glycosyl residue via the glycosyl transferase (*strop3247*). The *strop3248* is predicted to be a carotenoid desaturase. Since bacteria mutated at this gene were still pigmented, *strop3248* cannot be the phytoene dehydrogenase found in lycopene biosynthesis. This role is predicted to be encoded by *strop4439* (in *terp2*), which is was not interrogated genetically, but is the only other *S. tropica* CNS-440 gene predicted to perform this function. Thus, the *strop3248* likely results in an additional desaturation. The remaining genes in the cluster did not alter *Salinispora* pigmentation when inactivated and were deemed unnecessary to the primary pathway under native growth conditions.

One of these genes, *strop3244*, has homology (57% identity) to the known carotenoid gene *crtA* which is responsible for the 2' addition of either a hydroxyl or keto group. Three additional genes in the cluster (*strop3245, strop3249,* and *strop3250)* had no predicted role in carotenoid biosynthesis and inactivation of two (*strop3245* and *strop3249*) confirmed that they are not utilized. At this point, the functions of *strop3245* and *strop3249* in *Salinispora* metabolism are undetermined. No successful deletion could be created in the *strop3250* gene and thus its function in carotenoid biosynthesis is not confirmed. This gene is annotated as a 5,10-methylenetetrahydrofolate reductase, and homologs in *Streptomyces* are involved in the *met* pathway responsible for methionine biosynthesis[19]. Inactivation of genes in these pathways are often lethal or result in spontaneous reversion/mutation to the wild-type phenotype[19]. This observation may explain why it was not possible to obtain genetic mutants in this region despite several attempts.

Extracts of the non-clustered gene (*strop0241* and *strop2408*) mutants also showed an altered HPLC profile (Figure 2.9), suggesting their potential for participating in modification of the *Salinispora* carotenoid. In addition to the HPLC evidence, the *strop2408* deletion mutant had a visibly altered phenotype with the production of pink, rather than orange, colonies. The production of pink colonies has been shown previously when there is a loss of function of the lycopene cyclase gene in otherwise orange *Streptomyces*, thought to be due to the accumulation of lycopene instead of the wild-type isoreneiratene[20].

Gene inactivation and HPLC analysis of the resulting mutant strain extracts confirmed the involvement of four genome regions in the biosynthesis of a single major carotenoid. The *terp2* cluster predicted to be responsible for the first dedicated steps in carotenoid biosynthesis, building a linear, conjugated molecule from isoprene units (Scheme 1.1). The non-clustered and *terp1* clustered genes then perform further modifications. *Terp1*, however, appears to be a mixture of carotenoid biosynthetic genes and others not involved in this pathway. The genes *strop3244* and *strop3251*, despite their homology to carotenoid biosynthetic genes, appear to be inactive, and could either be remnants of carotenoid evolution in this bacterium or are participating in other pathways. Loss of function in these genes, as with the other successful mutations, resulted in no obvious changes in growth rate or phenotype other than pigmentation. These compounds, while likely to be adaptive in nature, are not necessary for growth or development under normal laboratory growth conditions.

To date only two examples of a non-clustered gene arrangement have been described in bacterial carotenoid biosynthesis outside of the cyanobacteria. They were

identified in the extremophilic Deinococcus-Thermus group (pathway has been explored genetically in *Deinococcus radiodurans* and proposed via sequence homology in *Meiothermus ruber*) as well as in *Gemmatimonas aurantiaca* a Gram-negative bacterium isolated from activated sludge (Figure 2.10)[21,22]. In both of these cases, the biosynthetic pathway was incomplete and many of the proposed enzymes lacked characterization.

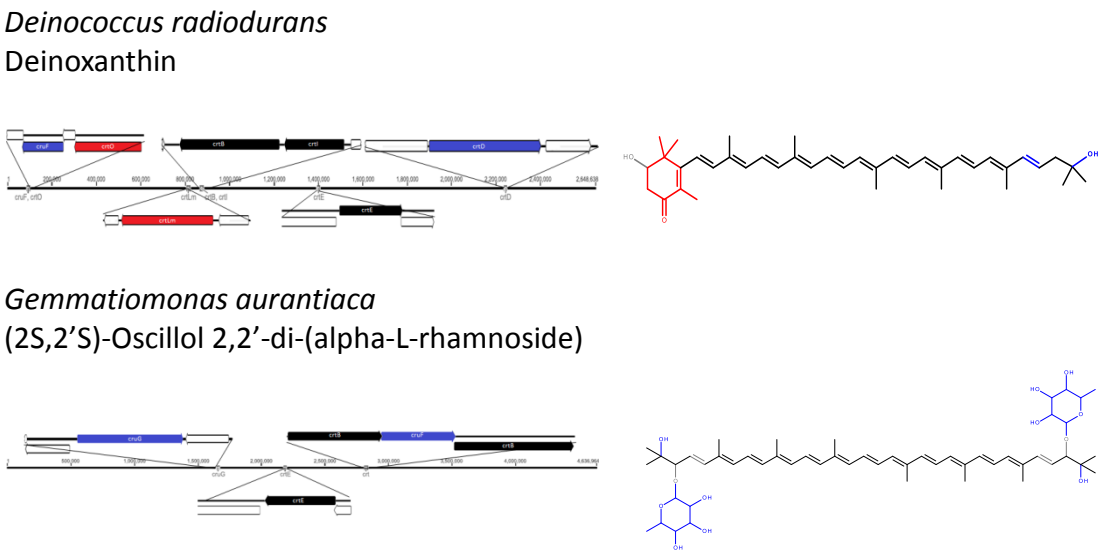


Figure 2.10: Biosynthetic pathway and resulting carotenoid in the two published examples of non-clustered bacterial carotenoid biosynthesis. Gene color corresponds to the region of the structure it is thought to encode[22,23].

2.2.4 Bioinformatic predictions of carotenoid structure

In general, carotenoid biosynthesis has been well studied and, although new enzyme functions continue to be uncovered, much is known about the relationship between gene sequence and function that enable us to make predictions about the structure of the *Salinispora* carotenoid[24-28].

As shown in Scheme 1.1, carotenoid biosynthesis is a conserved process that begins with the condensation of isoprene units to form the longer chain molecules farnesyl pyrophosphate (FPP, C_{15} product) or geranylgeranyl pyrophosphate (GGPP, C_{20} product). Two molecules of either FPP or GGPP undergo a head-to-head condensation to produce a C_{30} dehydrosqualene or a C_{40} phytoene using a dehydrosqualene synthase or phytoene synthase, respectively. Since these synthase enzymes are conserved, the carbon chain length of the resulting molecule can be predicted by sequence similarity. As Figure 2.11 describes, the *strop4441* encoded synthase protein clades more closely with the phytoene synthase CrtB than it does with the dehydrosqualene synthase CrtM, suggesting that *Salinispora* produces a 40-carbon carotenoid molecule (Scheme 2.1).

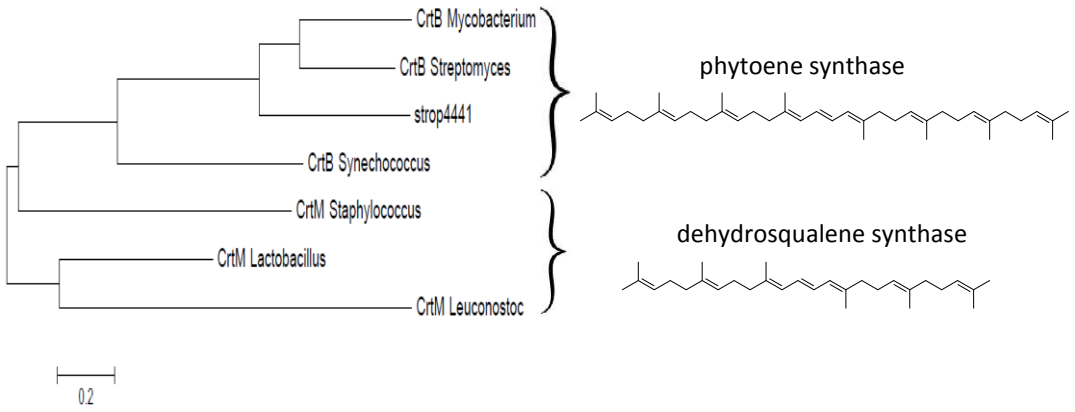
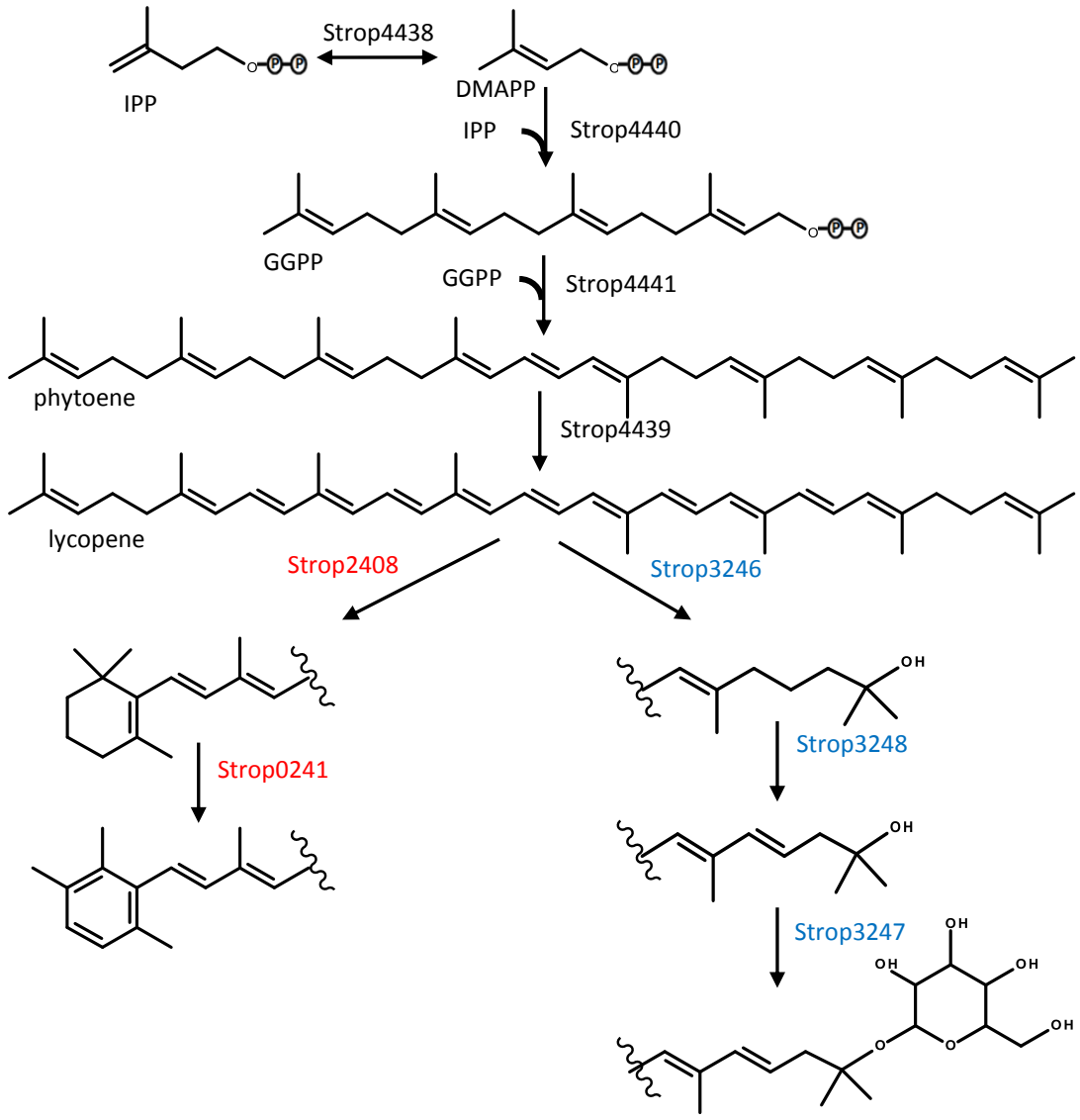


Figure 2.11: Clade of *strop4441* and known carotenoid synthase sequences. Enzyme products are shown below the titles.



Scheme 2.1: Proposed biosynthesis of *S. tropica* carotenoid based on predicted gene functions showing *terp2* genes in black, *terp1* genes in blue, and non-clustered genes in red.

The next stage in carotenoid biosynthesis is dehydrogenation of the linear molecule, which lengthens the chromophore, enabling the carotenoid's pigmentation and antioxidative properties. Several desaturases are known in carotenoid biosynthesis to be responsible for the formation of lycopene or neurosporene, as well as for further

downstream modification steps. Three of the genes shown to be involved in *Salinispora* carotenoid biosynthesis were predicted to function as desaturases (*strop4439*, *strop3248*, and *strop0241*) with two (*strop4439* and *strop3248*) annotated to perform the same phytoene dehydrogenase step. A clade of these protein sequences with known carotenoid desaturases, however, suggests that they have independent roles (Figure 2.12). *Strop4439* clades most closely with the 4-step phytoene dehydrogenases, as is found in lycopene biosynthesis, while *strop3248* likely acts after lycopene formation in the additional 3',4'-desaturation (Scheme 2.1). Finally, *strop0241* most closely resembles a β-carotene desaturase/methyl transferase which is known to convert a β-ring to an aromatic ring (Scheme 2.1), as is found in isorenieratene.

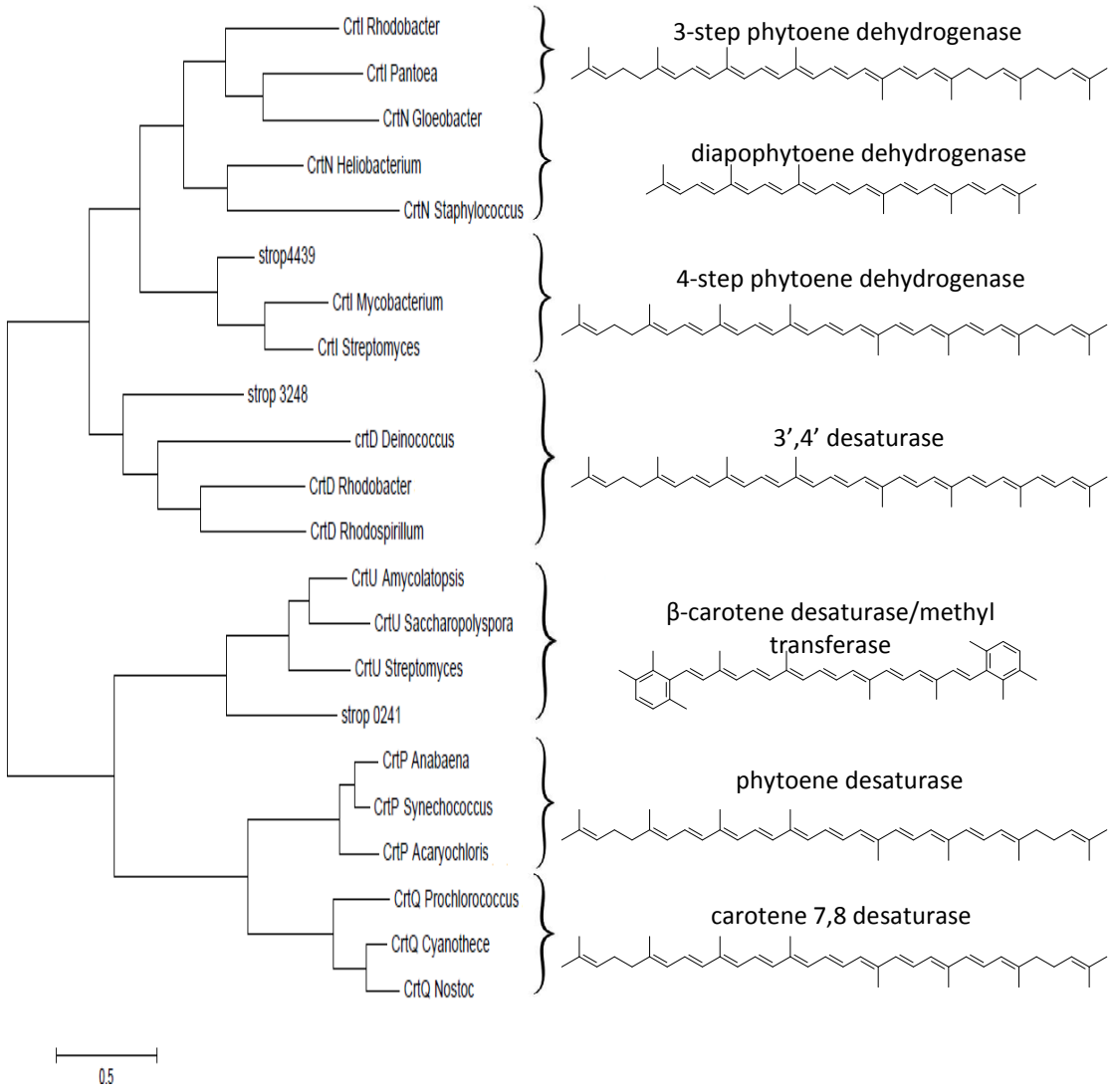


Figure 2.12: Clade of *S. tropica* desaturases and known carotenoid desaturase sequences. Enzyme products are shown below the titles.

Thus based on the assortment of carotenoid biosynthesis genes confirmed through mutagenesis to be involved in the construction of the *Salinispora* carotenoid, we can predict its chemical structure as shown in Figure 2.13. The *terp2* gene cluster makes lycopene, a 40-carbon linear molecule with four desaturation steps. *Strop2408* forms a β-ring on one side of the molecule, which is converted in to an aromatic ring by

the desaturase/methyl transferase action of *strop0241* (Scheme 2.1). On the other end of the molecule, *strop3246*, a CruF-like protein adds a 1' hydroxyl, which is glycosylated by the *strop3247* glycosyl transferase. Addtionally, the 3',4-- linkage is desaturated by a CrtD homolog, encoded by *strop3248*. This structure, if accurate, would represent a novel carotenoid.

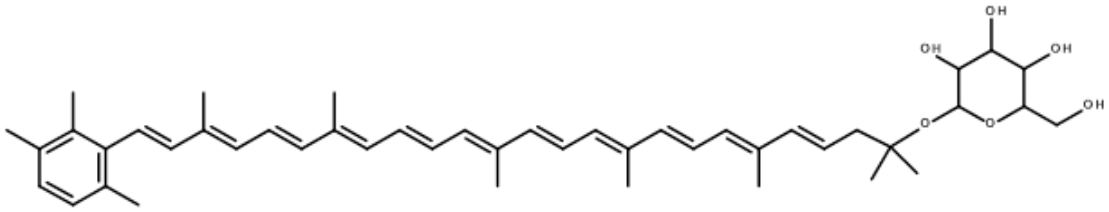


Figure 2.13: Proposed structure of the *Salinispora* carotenoid structure based on the predicted functions of the genes in the pathway.

Finding evidence of carotenoid glycosylation in *Salinispora* was unexpected, as glycosylated carotenoids are rare in actinomycetes. Only eight actinomycetes spread over 6 genera are known to produce glycosylated carotenoids[29], suggesting a role for horizontal gene transfer in the evolution of these compounds. As many of these species lack a genome sequence, little is known about their biosynthesis or the biological function of glycosylated carotenoids.

2.3 Conclusions

Orange pigmentation in marine bacteria of the genus *Salinispora* is the result of the accumulation of a carotenoid molecule, likely with a novel glycosylated structure. The biosynthesis of this compound involves genes in four distinct regions of the genome, involving two gene clusters and two individual genes. This scattered genomic arrangement is unusual in bacterial secondary metabolism and particularly for carotenoid biosynthesis. The *Salinispora* carotenoid gene pathway utilizes genes that

were initially overlooked due to low homology to known carotenoid genes. This emphasizes the need for molecular characterization of biosynthetic genes in order to make accurate structural and biochemical predictions. The unexpected composition and arrangement of this biosynthetic pathway pushes the boundaries of what is understood about bacterial genome organization, specifically in regards to secondary metabolism. This has impacts on the ability to accurately predict secondary metabolite structure as well as pathway function. In addition, it raises questions about the roles of genomic rearrangement and horizontal gene transfer in pathway evolution.

2.4 Materials and Methods

2.4.1 Identification of candidate carotenoid biosynthetic genes in *Salinispora*

Two terpene clusters, named *terp1* and *terp2*, were previously identified in a global analysis of secondary metabolic clusters in *S. tropica*, and had been predicted to be involved in carotenoid biosynthesis-based gene annotation[16]. Genes in *terp2*, *strop3251*, *strop3248*, and *strop2408* were previously annotated as carotenoid biosynthesis genes based on domain function, BLAST query searches of these genes against the nr database were used to confirm homology. Additional genes were identified via their sequence homology to known genes. Protein sequences for known carotenoid biosynthetic genes were gathered from NCBI and used as BLASTx query sequences against the *Salinispora tropica* CNB-440 and *Salinispora arenicola* CNS-205 closed genomes. Query sequences came from actinomycete genomes where possible. In the initial search, positive gene hits were those which had greater than 35% sequence identity with more than 50% query coverage. A secondary search, concerning genes involved in carotenoid glycosylation, had less stringent sequence identity

requirements to accommodate the large phylogenetic distance between *Salinispora* and the query-containing genome strains. Genes had to be present in both *Salinispora* genomes to be considered a part of the carotenoid pathway.

2.4.2 Analysis of carotenoid biosynthetic gene synteny among *Salinispora* genomes

Eighty-eight *Salinispora* genomes are available on the Integrated Microbial Genomes database hosted by the Joint Genome Institute[30]. Synteny of the carotenoid gene clusters and neighborhoods was investigated by performing a sequence homology search of the gene of interest against the *Salinispora* genomes and visually examining the top hit and the surrounding region in the Gene Ortholog Neighborhood tool.

2.4.3 Phylogenetic trees for function prediction of the putative biosynthetic genes

Protein sequences were gathered from Genbank on NCBI from representative phytoene/squalene synthases and carotenoid desaturases to be compared with the *S. tropica* CNS-440 putative homologs (*strop4441* and *strop4439* vs synthases; *strop3248* and *strop0241* vs desaturases). Analyses were performed in the bioinformatics software package MEGA (Molecular Evolution Genetics Analysis)[31]. Sequences were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation)[32] and then used in a maximum-likelihood tree following the default settings.

2.4.4 Bacterial strains and growth conditions

Wild-type *Salinispora* were grown in A1 liquid (per liter: 10 g starch, 4 g yeast extract, 2 g peptone, and 28 g Instant Ocean Marine Salts) and on solid (A1 with addition of 18 g agar/liter) media at 30°C[33]. Liquid cultures were grown as 50 ml cultures in 250 ml Erlenmeyer flasks containing a spring with shaking at 200 rpm. All bacterial cultures were grown without light. Mutant *Salinispora* strains were grown as

above, but with the addition of apramycin (50 µg/ml) as a selection marker and naladixic acid (100 µg/ml).

Escherichia coli strains were grown in LB (Luria-Bertani) broth and on solid LB-agar media with the following antibiotics added as appropriate: carbenicillin (100 µg/ml), chloramphenicol (12.5 µg/ml), and apramycin (50 µg/ml). *E. coli* strains were grown at 37°C, except for strain BW25113, which was grown at 30°C when necessary to maintain the temperature-sensitive pKD20 plasmid.

2.4.5 Inactivation of putative carotenoid genes in *Salinispora*

Gene inactivation experiments were carried out in *Salinispora tropica* CNB-440 using PCR-directed gene mutagenesis, as previously described[17,18]. Briefly, the apramycin resistance cassette (*aac(3)IV)* from pIJ773 was PCR amplified and extended using tailed PCR primers (Table 2.2). The extended sequence region was homologous to the sequence surrounding the gene of interest and provided a binding site for recombination. The appropriate pCCFOS based fosmid and extended resistance cassette were introduced in to *E. coli* BW25113/pkD20 via electroporation. The mutant fosmid was inserted in to *S. tropica* CNB-440 through conjugation using *E. coli* S17-1. Mutations were confirmed by colony PCR or PCR of the extracted *S. tropica* DNA and sequencing of the PCR product.

Table 2.2: Sequences of the tailed PCR primers used in the PCR-directed mutagenesis of *S. tropica* genes. Apramycin resistance cassette binding site is shown in bold

Primer name	Primer sequence
0241F	TTCTCAACACCGGCATATCTCCGGCAAACTGGTGCGATG**ATTCCGGGGATCCGTCGACC**
0241R	GGGATCGCCGGCTTCCGGTCAGTGTCCCCGGGCGAGCAG**TGTAGGCTGGAGCTGCTTC**
2408F	CGGGGCGGCCGGTCCGGCCGGTGACCCCGACCATGGGTG**ATTCCGGGGATCCGTCGACC**
2408R	GCTGGACGATGCAGCACCCGGCCCGGCGACGGGCGGTCA**TGTAGGCTGGAGCTGCTTC**
3244F	GACTTGTCGAACGAGCCACGCCCGGTCAACCGCCGCGCA**ATTCCGGGGATCCGTCGACC**
3244R	CGTCCCCACCTTCGTTCCGCCGTCCGTGGTCCAGCCCAG**TGTAGGCTGGAGCTGCTTC**
3245F	CTGGTGCGGTGGACACCGGACGACCTCGTTCGGCGGCTG**ATTCCGGGGATCCGTCGACC**
3245R	GGGGGCTCTTCCGGAGGCGGCGGGTGGAGGGAGCGGCAG**TGTAGGCTGGAGCTGCTTC**
3246F	CGGCCCACCGCCCTGCCGAGGCGTACCCGCCAGACCGTC**ATTCCGGGGATCCGTCGACC**
3246R	CGGTGCCCCACTTCGTCGGGGGCCGGCGGTGAGACGTCG**TGTAGGCTGGAGCTGCTTC**
3247F	CTGTTGCTGGCGGTGCTGGCCGGCGTGGCCGCGCTGACC**ATTCCGGGGATCCGTCGACC**
3247R	TGCCTCCTAGACGACCGGACGGCCCCGCCAACTCAGGCG**TGTAGGCTGGAGCTGCTTC**
3248F	GGCACACCATGGCGCGGATCGTGATCGTCGGCGCCGGGG**ATTCCGGGGATCCGTCGACC**
3248R	ATCGCGACACCGCGGGCGGGCGTCGGCGGACCGGGCCTA**TGTAGGCTGGAGCTGCTTC**
3249F	GCCACGACACCGCGCGATCGGTGCGACGGTTGATCTGTG**ATTCCGGGGATCCGTCGACC**
3249R	TCGCGATTCCCGTGCGCCTCCGGTGGGCCGGGAGCGTCA**TGTAGGCTGGAGCTGCTTC**
3250F	CTAGGGATAGCGGCCGTGCGCCGCTAGAGTCATGGCGTG**ATTCCGGGGATCCGTCGACC**
3250R	CGTCGCACCGATCGCGCGGTGTCGTGGCGCGTCTGGTCA**TGTAGGCTGGAGCTGCTTC**
3251F	GCACTGGCCGACTTCCTGGTGACCCGACGCGCCTGGATG**ATTCCGGGGATCCGTCGACC**
3251R	CCCCGTCCGGTCGAGGTCTGGCCGGATGAGGCGGGGTCA**TGTAGGCTGGAGCTGCTTC**
4440F	TCGACGCAACTTGTAACTCTAGGGTATGCTCCCCACATG**ATTCCGGGGATCCGTCGACC**
4440R	CTGGCTACCTCCTCGAGAAAACGCGCCCATCATGCTCGC**TGTAGGCTGGAGCTGCTTC**
4441F	GTTGCGTCGATTCATGCGTCGAATTGAGGAGGATCGGTG**ATTCCGGGGATCCGTCGACC**
4441R	CGGTCAGCCTAGGGAAGGGCCGGGTGTGGGTGTCAACA**GTGTAGGCTGGAGCTGCTTC**

2.5.6 Extraction and analysis of mutant and wild-type carotenoids

Liquid cultures (50 ml) of *Salinispora* were harvested after 7-10 days growth by centrifugation at 8000 rpm for 20 minutes. The media was discarded and the cell pellet was soaked in approximately 20 ml acetone for carotenoid extraction. Acetone was passed through a filter to remove cell debris and dried *in vacuo*. Extraction with acetone was repeated until the solvent no longer gained color from the cell pellet, approximately three times.

Crude extracts were resuspended in acetonitrile and analyzed on an Agilent 1200 series analytical HPLC using reversed-phase conditions, detecting at 450 nm. Extracts were analyzed on a Luna 5µm C18 column 100 x 4.6 mm, with 98% acetonitrile and 2% water under isocratic conditions. The HPLC comparing *S. tropica* and *S. arenicola* crude extracts used a solvent system of 80% acetonitrile, 15% methanol and 5% isopropanol isocratically on a Luna 5 µm C18 150 x 4.6mm column.

2.6: References

1 Maldonado, L. A., Fenical, W., Jensen, P. R., Kauffman, C. A., Mincer, T. J., Ward, A. C., Bull, A. T. & Goodfellow, M. *Salinispora arenicola* gen. nov., sp nov and *Salinispora tropica* sp nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *Int. J. Syst. Evol. Micr.***55**, 1759-1766, (2005).

2 Ahmed, L., Jensen, P. R., Freel, K. C., Brown, R., Jones, A. L., Kim, B. Y. & Goodfellow, M. *Salinispora pacifica* sp nov., an actinomycete from marine sediments. *Anton. Leeuw. Int. J. G.* **103**, 1069-1078, (2013).

3 Davis, N. K. & Chater, K. F. Spore color in *Stretptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Mol. Microbiol.* **4**, 1679-1691, (1990).

4 Yu, T. W. & Hopwood, D. A. Ectopic expression of the *Streptomyces coelicolor whiE* genes for polyketide spore pigment synthesis and their interaction with the *act* genes for actinorhodin biosynthesis. *Microbiol.-Uk* **141**, 2779-2791 (1995).

5 Kelemen, G. H., Brian, P., Flardh, K., Chamberlin, L., Chater, K. F. & Buttner, M. J. Developmental regulation of transcription of whiE, a locus specifying the polyketide spore pigment in *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **180**, 2515-2521 (1998).

6 Krubasik, P. & Sandmann, G. A carotenogenic gene cluster from *Brevibacterium linens* with novel lycopene cyclase genes involved in the synthesis of aromatic carotenoids. *Mol. Gen. Genet.* **263**, 423-432 (2000).

7 Krugel, H., Krubasik, P., Weber, K., Saluz, H. P. & Sandmann, G. Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *BBA-Mol. Cell Biol. L.* **1439**, 57-64, (1999).

8 Provvedi, R., Kocincova, D., Dona, V., Euphrasie, D., Daffe, M., Etienne, G., Manganelli, R. & Reyrat, J. M. SigF controls carotenoid pigment production and affects transformation efficiency and hydrogen peroxide sensitivity in *Mycobacterium smegmatis*. *J. Bacteriol.* **190**, 7859-7863, (2008).

9 Takano, H., Obitsu, S., Beppu, T. & Ueda, K. Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol.* **187**, 1825-1832, (2005).

10 Takano, H., Asker, D., Beppu, T. & Ueda, K. Genetic control for light-induced carotenoid production in non-phototrophic bacteria. *J. Ind. Microbiol. Biotechnol.* **33**, 88-93, (2006).

11 Pezzoni, M., Costa, C. S., Pizarro, R. A. & Oppezzo, O. J. The relationship between carotenoids and sunlight response in members of the family Micrococcaceae. *J. Basic Microbiol.* **51**, 325-329, (2011).

12 Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S. & Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME Journal* **3**, 1193-1203, (2009).

13 Kuzuyama, T., Takahashi, S., Dairi, T. & Seto, H. Detection of the mevalonate pathway in *Streptomyces* species using the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene. *J. Antibiot. (Tokyo)* **55**, 919-923 (2002).

14 Armstrong, G. A. Eubacteria show their true colors - genetics of carotenoid pigment biosynthesis from microbes to plants. *J. Bacteriol.* **176**, 4795-4802 (1994).

15 Ballouz, S., Francis, A. R., Lan, R. & Tanaka, M. M. Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* **6**, (2010).

16 Udwary, D. W., Zeigler, L., Asolkar, R. N., Singan, V., Lapidus, A., Fenical, W., Jensen, P. R. & Moore, B. S. Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 10376-10381, (2007).

17 Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 1541-1546, (2003).

18 Eustaquio, A. S., Pojer, F., Noe, J. P. & Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **4**, 69-74, (2008).

19 Blanco, J., Coque, J. J. R. & Martin, J. F. The folate branch of the methionine biosynthesis pathway in *Streptomyces lividans*: Disruption of the 5,10-methylenetetrahydrofolate reductase gene leads to methionine auxotrophy. *J. Bacteriol.* **180**, 1586-1591 (1998).

20 Myronovskyi, M., Tokovenko, B., Broetz, E., Rueckert, C., Kalinowski, J. & Luzhetskyy, A. Genome rearrangements of *Streptomyces albus* J1074 lead to the carotenoid gene cluster activation. *Appl. Microbiol. Biotechnol.* **98**, 795-806, (2014).

21 Tian, B. & Hua, Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol.* **18**, 512-520, (2010).

22 Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2 '-dirhamnoside. *Microbiol-Sgm* **156**, 757-763, (2010).

23 Tian, B. & Hua, Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol.* **18**, 512-520, (2010).

24 Klassen, J. L. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS One* **5**, e11257, (2010).

25 Armstrong, G. A. & Hearst, J. E. Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. *FASEB J.* **10**, 228-237 (1996).

26 Phadwal, K. Carotenoid biosynthetic pathway: molecular phylogenies and evolutionary behavior of crt genes in eubacteria. *Gene* **345**, 35-43, (2005).

27 Maresca, J. A., Graham, J. E., Wu, M., Eisen, J. A. & Bryant, D. A. Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 11784-11789, (2007).

28 Sandmann, G. Evolution of carotene desaturation: The complication of a simple pathway. *Arch. Biochem. Biophys.* **483**, 169-174, (2009).

29 Britton, G., Liaaen-Jensen, S. & Pfander, H. *Carotenoids handbook.* (Birkhäuser Verlag, 2004).

30 Markowitz, V. M., Chen, I. M., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., Ratner, A., Huang, J., Woyke, T., Huntemann, M., Anderson, I., Billis, K., Varghese, N., Mavromatis, K., Pati, A., Ivanova, N. N. & Kyrpides, N. C.. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res.* **42**, D560-567, (2014).

31 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731-2739 (2011).

32 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797, (2004).

33 Beer, L. L. & Moore, B. S. Biosynthetic Convergence of Salinosporamides A and B in the Marine Actinomycete Salinispora tropica. *Org. Lett.* **9**, 845-848, (2007).

3 Characterization and structure elucidation of the novel carotenoid sioxanthin

3.1 Introduction

Carotenoid chemical structures provide valuable insight into their biosynthesis and function. These prolific pigmented compounds have a common evolutionary background and consist of certain common chemical features: a polyisoprenoid construction, a long, conjugated chain at the central core, and near symmetry on either side of the central double bond[1]. Structural modifications on either side of the central core contribute to the diversity of carotenoid molecules found in nature. The diversity of structural modifications is reflective of diversity in the evolution of carotenoid biosynthesis as well as in the functions that these molecules can play in cells[1]. Analysis of carotenoid structures, however, can be difficult as these compounds are unstable in the presence of oxygen, high temperature, acid, and light. Thus, multiple modes of analysis are often required for identification[2,3]. However, difficulties in separation and identification are further compounded by the structural similarities in carotenoids and the presence of interfering compounds[4].

The key unifying feature of carotenoid molecular structures is the chromophore. The chromophore is a conjugated linear chain and refers to the series of alternating double and single carbon-carbon bonds that make up the core of the carotenoid molecule. This portion of the molecule determines both the geometry and the chemical reactivity of carotenoids[1]. Due to steric hindrance, carotenoids favor the all-*trans* conformation, which influences their interactions with proteins and membrane structures[1]. Furthermore, the conjugated system results in delocalized π-electrons that give carotenoids their characteristic pigmentation as well as their antioxidative

properties[1]. The wavelength of light absorbed is determined by the length of the chromophore and can be measured by UV/Vis spectroscopy[5]. At least seven C—C double bonds must be in conjugation in order for the molecule to be pigmented. Increasing chromophore length makes it easier to excite the π-electrons and results in a longer wavelength emission (shifted toward the red light end of the visible spectrum)[1]. The alkene region of the chromophore can be identified by nuclear magnetic resonance (NMR), though individual protons can sometimes to be difficult to distinguish in this rather homogenous region.

The diversity of the carotenoid structures comes from the modifications of the linear core including the addition of functional groups[6]. Carotenoids come in two basic types: carotenes, comprised of only carbon and hydrogen, and xanthophylls, which have oxygen-containing functional groups[7]. Carotenoids are generally extremely hydrophobic molecules, though the addition of functional groups can alter the regional polarity of the molecule[1]. Polarity influences how a carotenoid interacts with the membrane and has implications for chemical analyses of these molecules such as retention times on high-performance liquid chromatography (HPLC)[8]. Functional groups also impact the UV/Vis spectra of the compound by lengthening the chromophore or influencing the spectral fine structure[5] and can also determine the amenability of the compound to ionization for mass spectrometric (MS) analyses. Gas chromatographic (GC)-MS techniques, a favorite for analysis of nonpolar analytes, are unsuitable for carotenoids, which degrade under high temperatures required for this technique[4]. Therefore, atmospheric pressure chemical ionization (APCI) and electrospray ionization (ESI) are the favored MS methods for carotenoid studies for

determination of molecular mass and formula[2 6]. These methods have been shown to produce a parent ion as well as characteristic fragmentations[2,6]. However, carotenoid ionization is often poor and can be easily overshadowed by other compounds in an impure sample[4].

Knowledge of the structural modifications is necessary to confirm the predicted biosynthesis and to assign function. In this chapter, several analytical chemistry techniques are combined to determine the structure of the *Salinispora* carotenoid. This structure is predicted in Chapter 2 to be a novel carotenoid with both aryl and glycosylated end groups. Structures are also determined for pathway mutants to establish the function of each gene in the biosynthetic pathway.

3.2 Results and Discussion

3.2.1 Separation and purification of pigments from wild-type *S. tropica*

Carotenoids were isolated from six liters of wild-type *S. tropica* cultures and purified via high performance liquid chromatography (HPLC). Considering the poor stability of carotenoids, exposure to light, heat, and oxygen was limited. Reversed-phase HPLC solvents were buffered with triethylamine and ammonium acetate, which has been shown to improve carotenoid yields during HPLC purification[9]. Furthermore, as the structure in Chapter 2 shows, the *S. tropica* structure contains a sugar residue. Structures are also determined for pathway mutants to confirm the function of each gene in the biosynthesis.

The reversed-phase HPLC (C18; 94.05% methanol, 5% dichloromethane, 0.05% triethylamine, and 0.05M ammonium acetate; 3 ml/min; λ=450nm) results showed a major peak at around 21 minutes, with four other minor peaks (Figure 3.1). The minor

peaks are likely conformational isomers, as upon isolation they re-equilibrated back to multiple peaks. Mass spectral analysis by ESI and APCI were attempted, but no parent ion could be identified for the carotenoid. Analysis using ^{1}H NMR showed that the carotenoid sample was overwhelmed by impurities that were not UV-active. To remove these impurities, the isolated carotenoid underwent a second round of HPLC purification, this time under normal-phased (silica; 93% hexanes, 5% DCM, 2% isopropanol; 3 ml/min; λ=450nm) conditions. The major peak was isolated again (Figure 3.2) and reanalyzed by ^{1}H NMR, which showed a single purified compound (Figure 3.3).

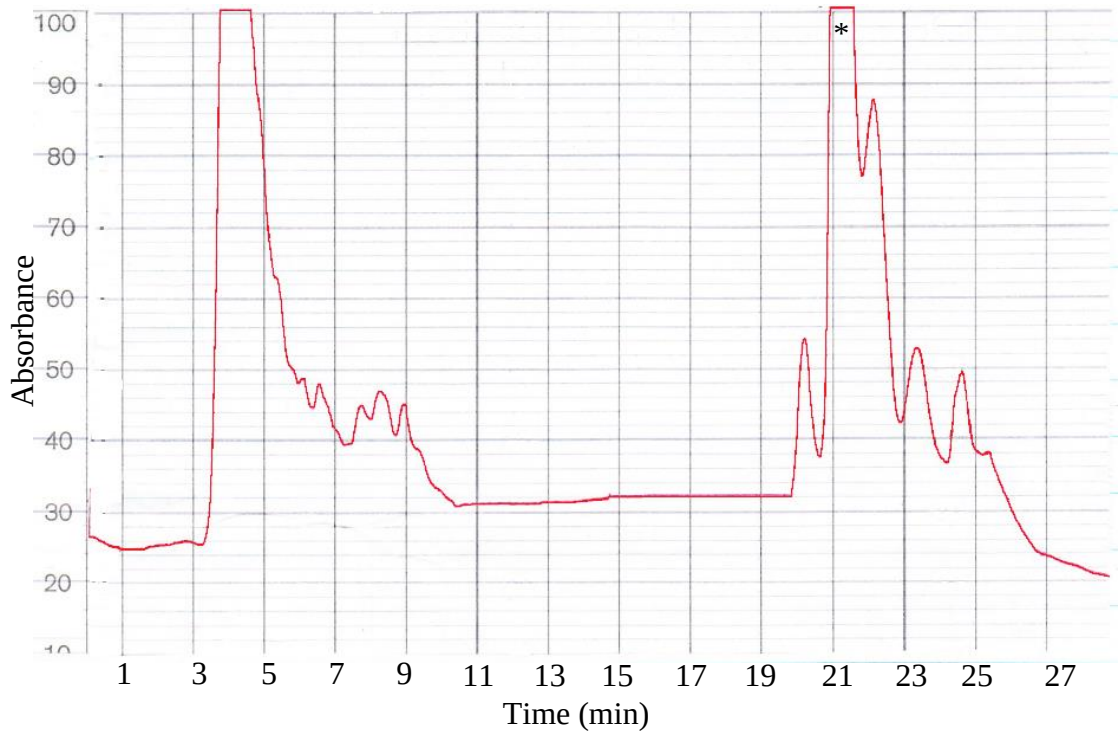


Figure 3.1: Reversed phased HPLC of the acetylated crude extract of the wild-type *S. tropica*. * shows the collected peak.

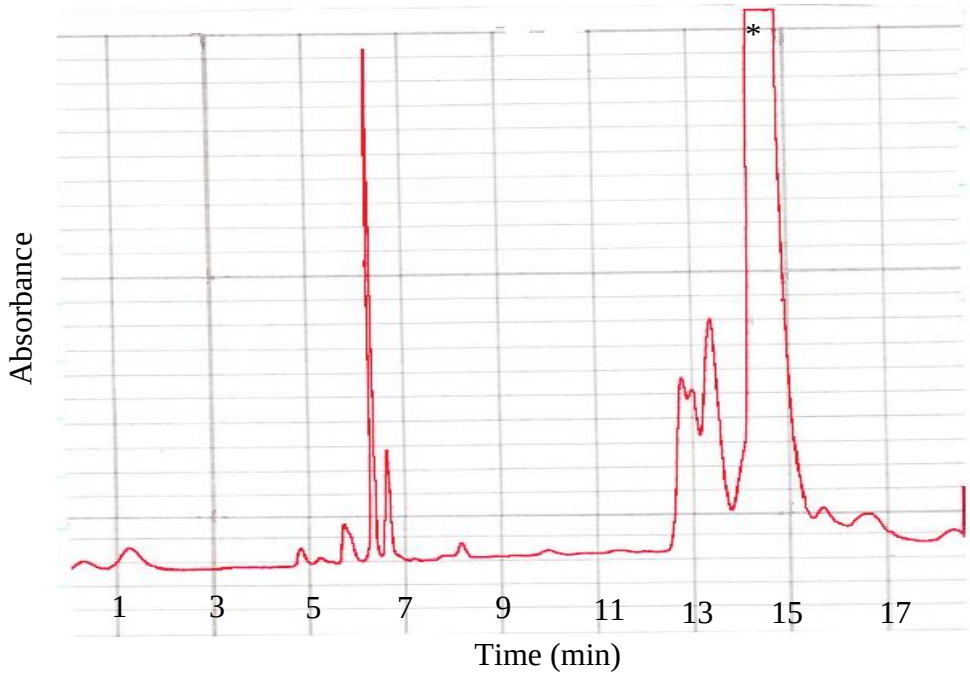


Figure 3.2: Normal phased HPLC of the peak isolated in Figure 3.1. * indicates peak that was captured.

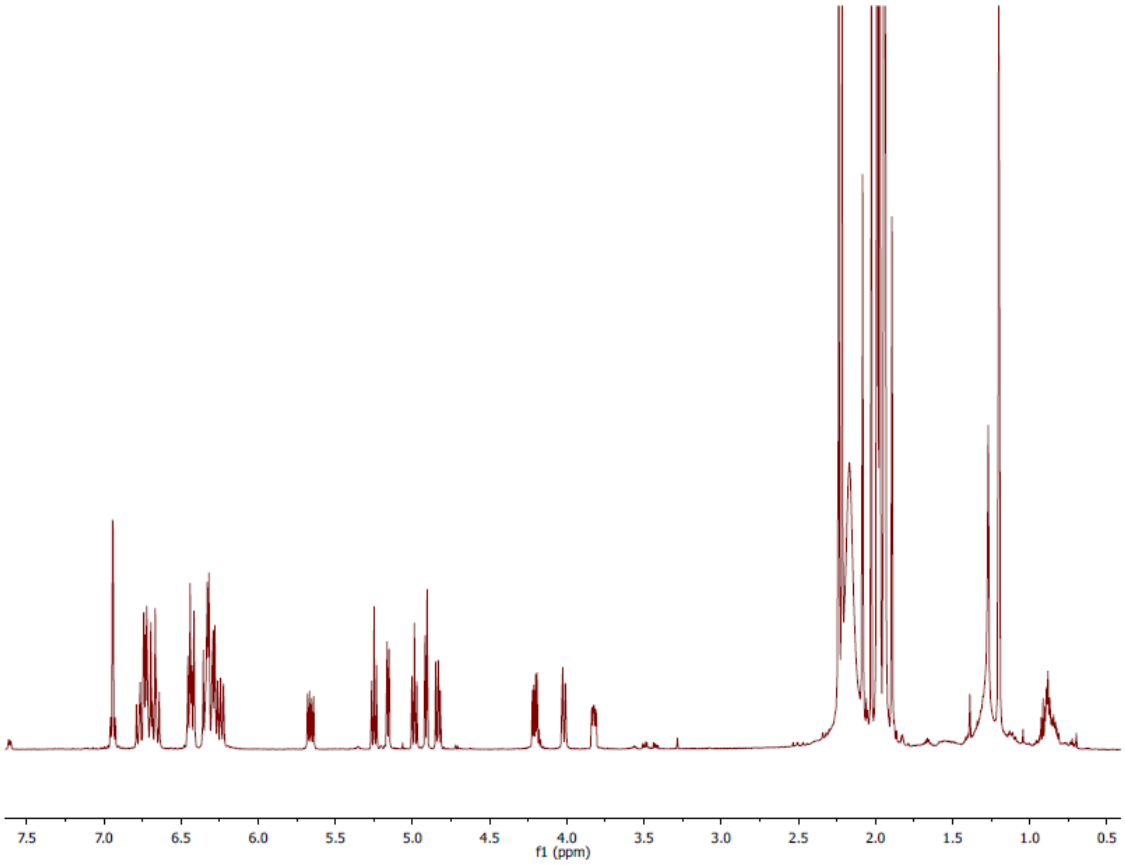
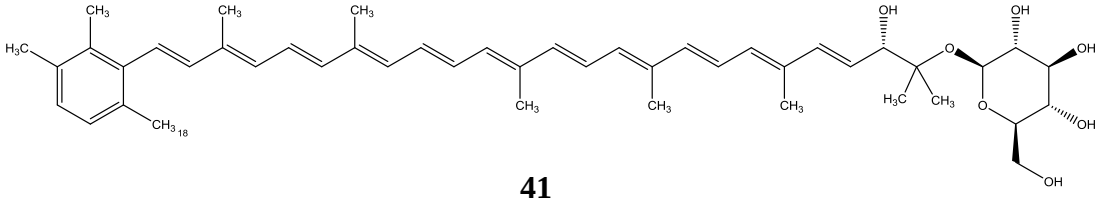


Figure 3.3: ^{1}H NMR (CD_3CN; 600MHz) of the purified carotenoid from *Salinispora tropica*.

3.2.2. Structure elucidation of the *Salinispora* carotenoid

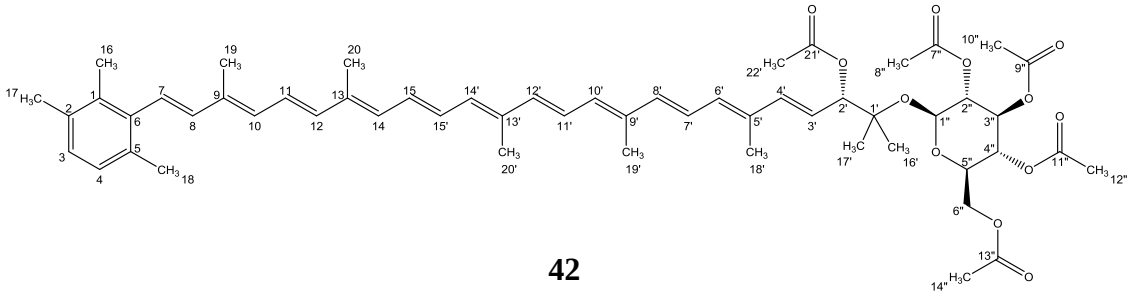
The purified compound was analyzed by electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) mass spectrometry (MS). Poor ionization by ESI again resulted in no useable mass data. APCI returned an exact mass of 937.5008 m/z $(M+H)^+$ corresponding to a molecular formula of $C_{56}\ H_{73}\ O_{12}$ (δ = 1.0 ppm). For further structure elucidation, the purified compound underwent comprehensive 2-dimensional NMR analyses, including COSY (Suppl Figure 3.1), TOCSY (Suppl Figure 3.2), ROESY (Suppl Figure 3.3), NOESY (Suppl Figure 3.4), HSQC (Suppl Figure 3.5), and HMBC (Suppl Figure 3.6). The combined NMR data confirmed the structure of a glycosylated, aromatic carotenoid structure as shown in

Figure 3.4. The structure represents a novel carotenoid (2'*S*)-1'-(β-D-glucopyranosyloxy)-3',4'-didehydro-1',2'-dihydro-φ,ψ-caroten-2'-ol **(41)**, which we named "sioxanthin". The confirmed structure is similar to the structured predicted in Chapter 2 with the addition of a hydroxyl group attached at the 2' position. Structure elucidation took place on an acetylated derivative, and resulted in structure **(42)** as shown in Figure 3.5.



41

Figure 3.4: Structure of the *S. tropica* carotenoid, sioxanthin (**41**).



42

Figure 3.5: Penta-acetylated sioxanthin structure used in analysis and structure elucidation (**42**).

The ^{1}H NMR spectrum has signals in the predicted regions for a glycosylated carotenoid. It shows methyl groups, including those from the added acetyl groups, in the δ1.0-2.5 region. The structural features on the right side of the molecule, including the glycosyl and hydroxyl groups are located in the δ3.5-6.0 region. The alkene chain is represented in the δ6-6.75 region. Far downfield, at δ6.95, the aromatic ring protons overlap. Carbon assignments were made using HMBC and HSQC data (Table 3.1).

Table 3.1: Carbon and proton NMR shifts and HMBC correlations (CD_3CN, 600MHz). "x" indicates atoms that could not be distinguished. Other missing values are due to overlapping signals in the spectra and are indistinguishable.

C#	δ_C	δ_H, mult.	COSY	HMBC	NOESY
1	134.7				
3	129[x]	6.95, d (7.6)[x]		18	
4	128.1[x]	6.93, d (7.6)[x]		18	
6	138.1				
7	126.5	6.76, d (3.5)			
16	17.1	2.21, s			
17x	20.65	2.24, s		1, 3, 4, 6,	
18x	20.65	2.24, s		1, 3, 4, 6	
19	12.8	2.08, s			
20	12.8				
1'	79.9				
2'	80.8	5.16, d (8.0)	5.66	1', 3', 4', 16', 21'	
3'	123.8	5.66, dd (15.6, 8.0)	6.33, 5.16	2', 5'	
4'	139.7	6.33	5.66	2', 3', 5', 6', 18'	
5'	135.2				
6'	133.9	6.28		7'	
7'	125.86	6.67		5'	
16'	22.3x	1.20, s		1', 2', 17'	
17'	24.0x	1.19, s		1', 2', 16'	
18'	12.7	1.89, s		4', 5', 6'	
19'	12.8	1.99		7'	
20'	12.8				
21'	170.6				
1"	95.9	4.91, d (8.0)	4.83	1'	3.82
2"	72.0	4.83, dd (9.8, 8.0)	4.91, 5.25	1", 3", 7"	4.99
3"	73.3	5.25, dd (9.8)	4.83, 4.99	2", 4", 9"	3.82
4"	69.5	4.99, dd (9.8)	5.25, 3.82	3", 5", 11"	4.83, 4.03
5"	72.1	3.82, ddd (8.0, 5.6, 2.5)	4.99, 4.20, 4.03	1", 4"	4.91, 5.25, 4.20
6"	62.8	4.20x, dd (6.5, 5.2) 4.03x, dd (12.2, 2.6)	3.82 3.82	4", 5", 13" 5", 13"	3.82
7"	16.99				
8"	20.8	1.97, s		7"	
9"	170.7				
10"	21.0	2.03, s		9"	
11"	170.3				
12"	21.17	2.02			
13"	170.9				
14"	12.7	1.99, s		13"	

The structural features on the polar end of the molecule were clearly identified using the 2D NMR techniques. COSY confirmed the structure of a hexose as well as the 2'-hydroxyl and the 3',4'-desaturation (Figure 3.6). This desaturation showed correlations to the rest of the alkene chain, identifying the linkage between the sugar residue and the rest of the molecule. This linkage was confirmed by TOCSY data which identified the glycosyl group and the alkene chain as separate spin systems (Figure 3.6). The 2'-acetylated hydroxyl and the 3,4'-desaturation were confirmed by the COSY correlations as well as the splitting patterns in the ^{1}H NMR.

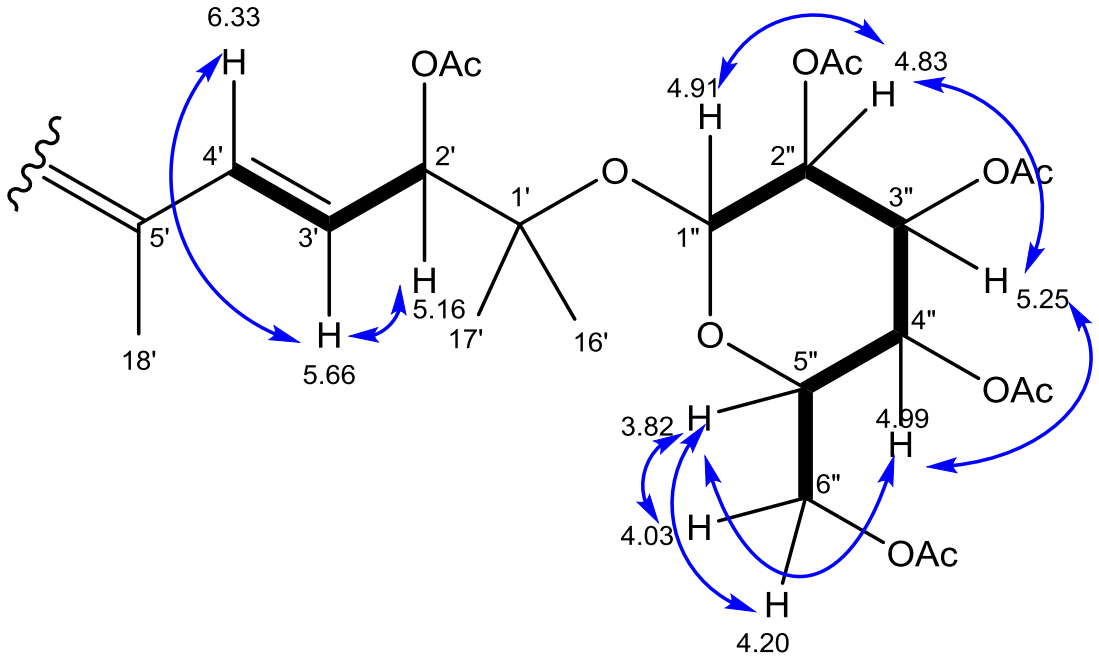


Figure 3.6: COSY and TOCSY correlations confirm the structural modifications on the right side of the molecule. Arrows depict COSY correlations. TOCSY spin systems are shown in bold bonds.

The linkage between the glycosyl group and the alkene chain was further confirmed through HMBC correlations (Figure 3.7). HMBC data also confirmed the presence of a total of five acetyl groups, corresponding to the five hydroxyl functional groups present in the molecule (Figure 3.8).

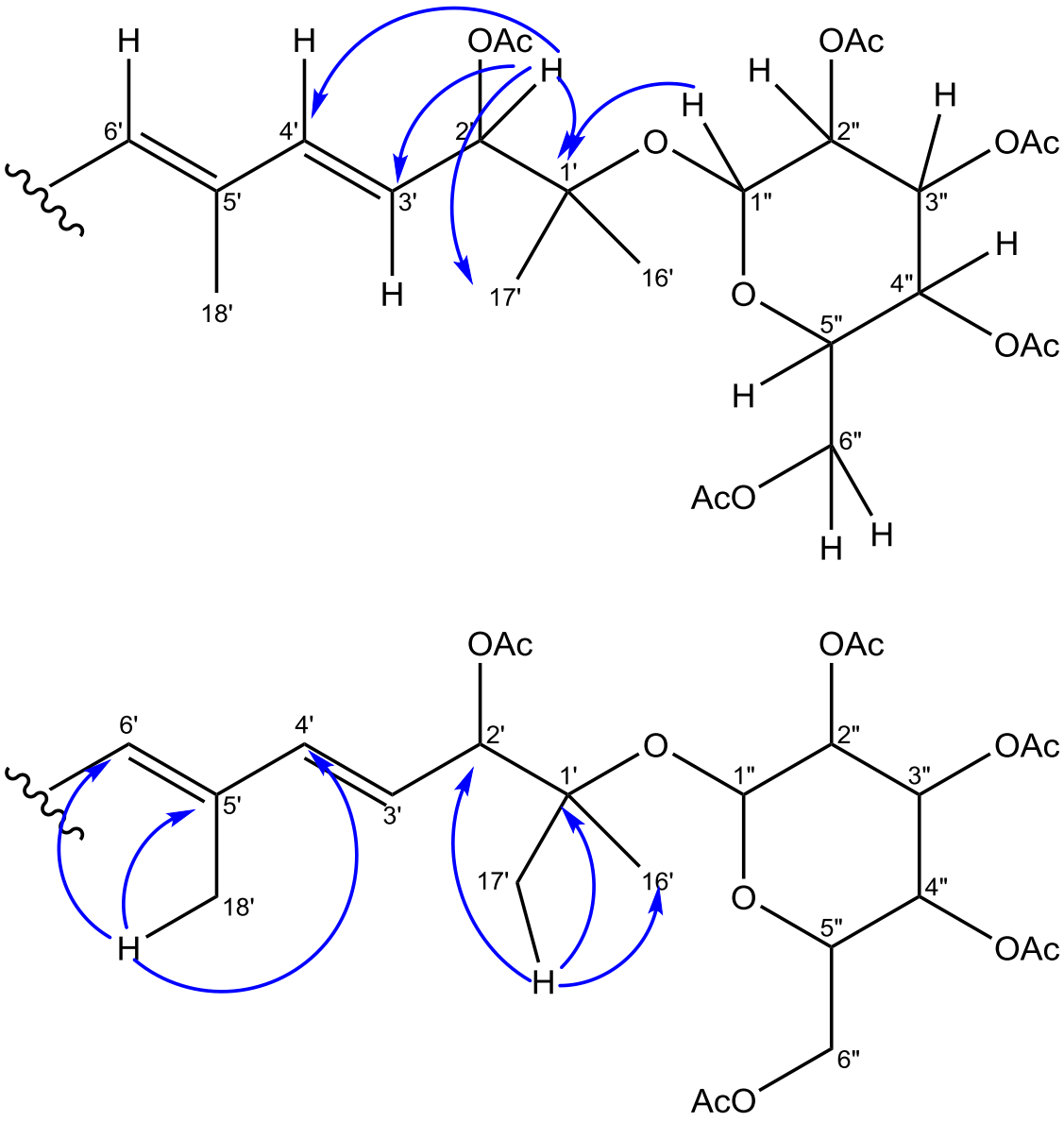


Figure 3.7: HMBC correlations confirming the region between TOCSY spin systems, linking the functional groups to the remainder of the molecule.

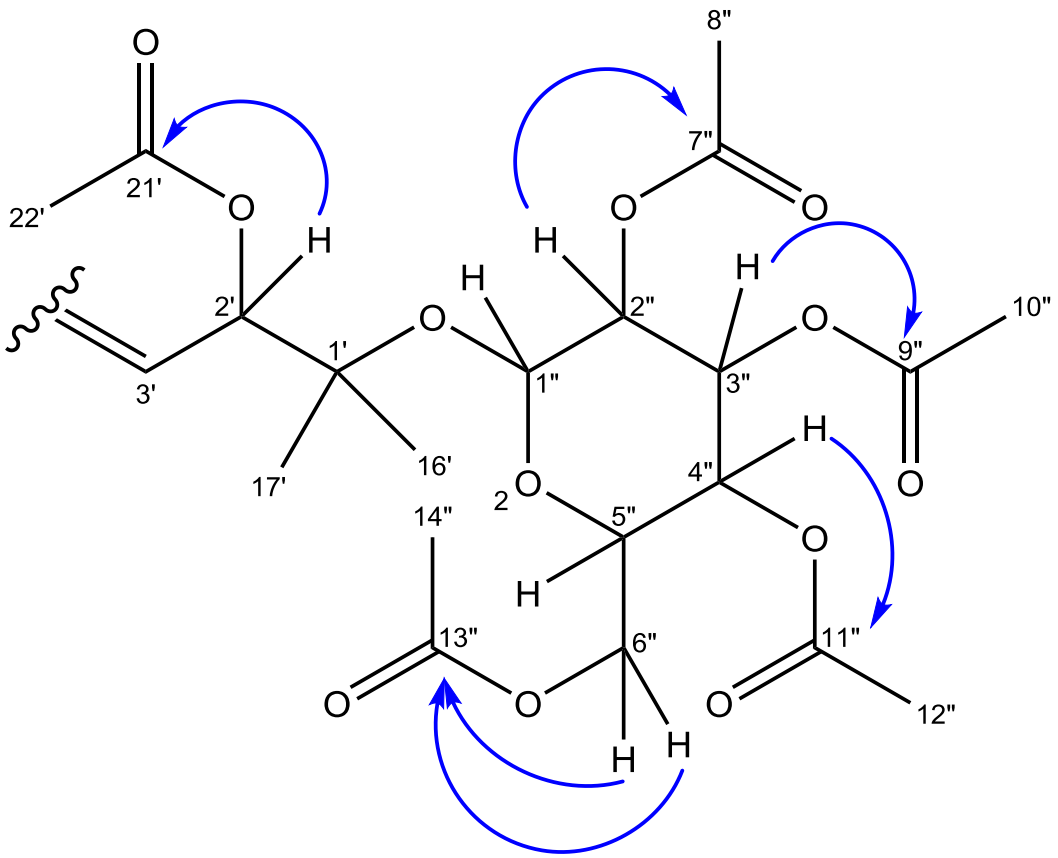


Figure 3.8: HMBC correlations showing correlations to the carbonyl carbons in each acetyl group.

The repetitive nature of the molecule makes it difficult to confidently assign every carbon and hydrogen atom. Though alkene and methyl groups can easily be identified, the individual isoprene units have overlapping carbon and hydrogen shifts, making them difficult to distinguish from one another. Comparisons to published NMR shifts aided in the assignment[10]. All methyl carbons in this region have a shift of about δ12.8. The confidence in the structure of the alkene chain portion is enhanced by the UV/Vis and MS data, which adhere to the literature describing carotenoid structures[10].

The aromatic end of the compound was confirmed through the presence of aromatic hydrogens at carbons 3(δ6.93) and 4 (δ6.95). Although these two positions are indistinguishable, they are consistent with published NMR shifts of a 1,2,5,6-

tetrasubstituted benzene that is a common structure element in carotenes (Figure 3.9)[10]. The HMBC data show correlations to a methyl group carbon, referring to either carbon 17 or 18.

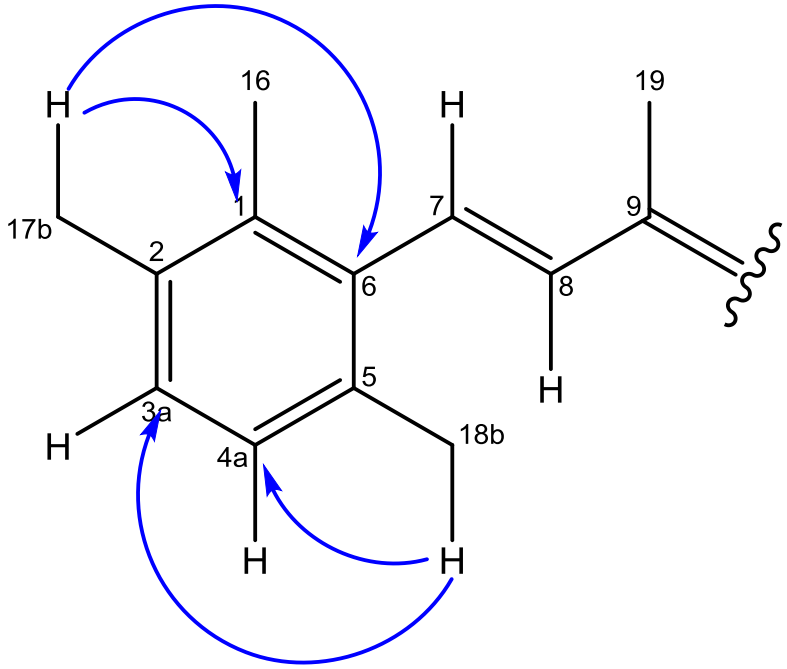


Figure 3.9: HMBC correlations of the aromatic ring. Letter labels refer to carbons that cannot be distinguished.

Stereochemistry of the hexose was determined through NOESY and ROESY correlations, which identified the sugar as glucose (Figure 3.10). This identification of glucose was unsurprising as D-glucose is the most common sugar known in bacterial glycosylated carotenoids[11].

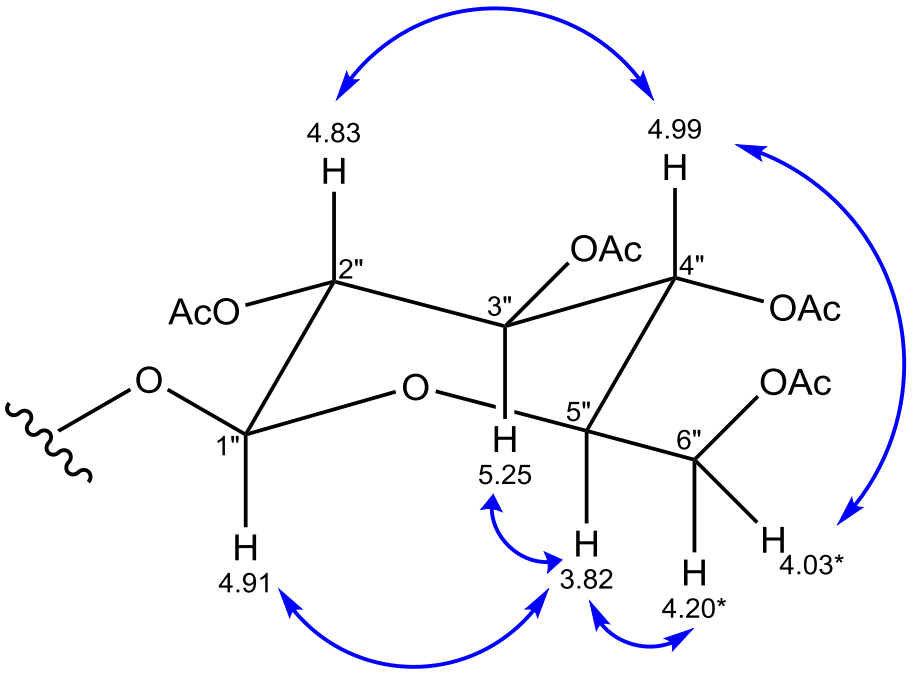


Figure 3.10: NOESY correlations of the sugar residue demonstrating stereochemistry.

Stereochemistry of the 2'-hydroxyl was determined through circular dichroism (CD) of the wild-type and the *strop3247* (glycosyl transferase) mutant compound. Both compounds showed the same CD spectral patterns and determined that the hydroxyl is in the *S* configuration through comparisons with previously established standards (Figure 3.11)[12]. The molecule showed little optical rotation, $[\alpha]^{25}_{p}$ = 0.0015 deg (c = 0.0008 deg; methanol). As with CD measurements, this value was compared with the aglycone compound from the *strop3247* mutant. The mutant compound had a small and opposite rotation $[\alpha]^{25}_{p}$ = - 0.0035 deg (c = 0.0008 deg, methanol). The presence of an attached glucose may have an impact on the overall optical rotation of the molecule.

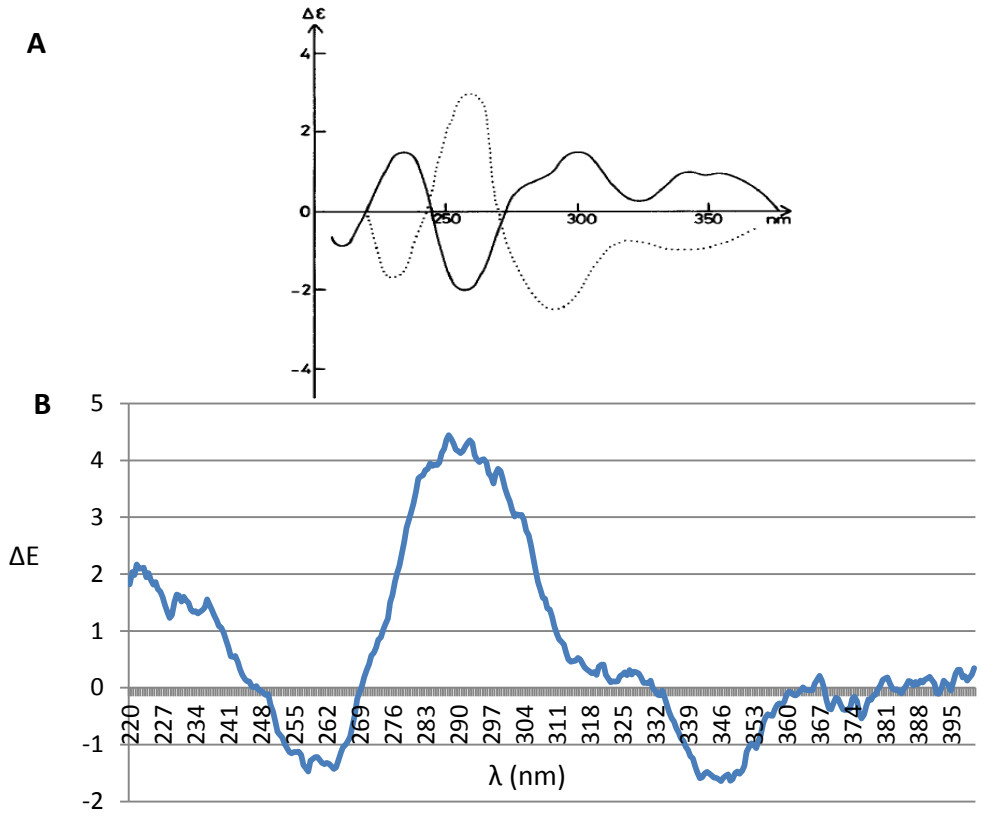


Figure 3.11: CD spectra of the carotenoid showing the chirality at the C-2'. A) The established CD spectra for the C-2' position in similar molecules, solid line shows *S* configuration, dotted line shows *R* configuration (adapted from Ronnenberg et al 1985 [12]). B) CD spectrum of the aglycone sioxanthin intermediate.

The UV/Vis spectrum for sioxanthin has peaks at 446, 472, and 504 nm (with a fine structure of %III/II 84), showing a chromophore similar to that of lycopene (Figure 3.12). Though sioxanthin has never before been described, the modifications on both ends of the molecule are known in other carotenoids. The aromatic ring portion is known, for example, in isorenieratene, the primary carotenoid in the *Streptomyces*[13]. The other end of the molecule contains a 1'glycosylation, a 2' hydroxyl, and a 3',4' desaturation, all of which are known in molecules such as pheixanthophyll, found in a single strain of *Mycobacterium phlei*[14]. Sioxanthin represents the first example of all of these carotenoid modifications taking place in a single molecule.

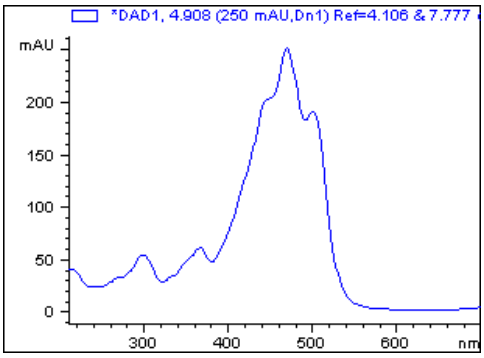


Figure 3.12: UV/Vis spectrum of sioxanthin showing the characteristic carotenoid peak structure.

3.2.3 Structures of the mutant compounds

As with the wild-type extracts, extracts from the mutant bacteria (developed in Chapter 2) were stabilized by acetylation and purified by two rounds of HPLC. In both rounds, the major peak was isolated. Extractions and purifications of the carotenoid mutant bacteria resulted in much lower yields. Despite using ten liters of bacterial cultures, only the *strop3247* mutant yielded enough purified material for NMR analysis. The reason for the lower yields is unknown, but may be due to altered efficiency in growth, production or extraction. Mutant bacteria tend to have lowered growth efficiency due to the alteration in genotype and phenotype and, presumably the presence of antibiotics as a selective marker. Compound production may further be impeded by missing steps in the biosynthetic pathway. While little is known about the distribution of enzymes in the carotenoid biosynthetic pathway beyond their localization to the membrane, the formation of a complex in these proteins would help explain the reduced production efficiency as one portion of the complex must be bypassed. Furthermore, each mutant produces a modified structure with slightly altered polarity, which could impact the efficiency of the established extraction method.

The purified mutant extracts, with the exception of the *strop0241* mutant, provided sufficient material for high resolution mass spectrometry (HR-MS). Mutant extracts were characterized using HR-MS, which provided a molecular formula (Table 3.2). This information, along with UV/Vis (describing the length of the chromophore) (Table 3.2) and HPLC retention time (indicating the presence or absence of polar functional groups) data allowed for the predicted structures (Figure 3.13). Although the purified compounds from the *strop0241* mutant were visibly orange, they had very low yields. Masses obtained from this mutant were not consistent with carotenoid molecules and were likely due to impurities.

Table 3.2: UV/Vis and MS data gave rise to the predicted structures of carotenoids from mutant bacterial extracts.

Compound	Source	UV/Vis	%III/II	Measured mass (M+H)	Formula	Error (ppm)
41	wild-type	446, 472, 504	84			
42	pentacetate sioxanthin			937.5088	$C_{56}H_{72}O_{12}$	1.0
43	*strop2408* mt	446, 470, 502	89	555.4563	$C_{40}H_{58}O$	0.5
44	*strop3246* mt	440, 462, 492	83	533.4145	$C_{40}H_{53}$	0.3
45	*strop3247* mt	446, 472, 504	233	607.4140	$C_{42}H_{54}O_3$	1.8
46	*strop3248* mt	436, 456, 488	88	551.4252	$C_{40}H_{54}O$	0.9

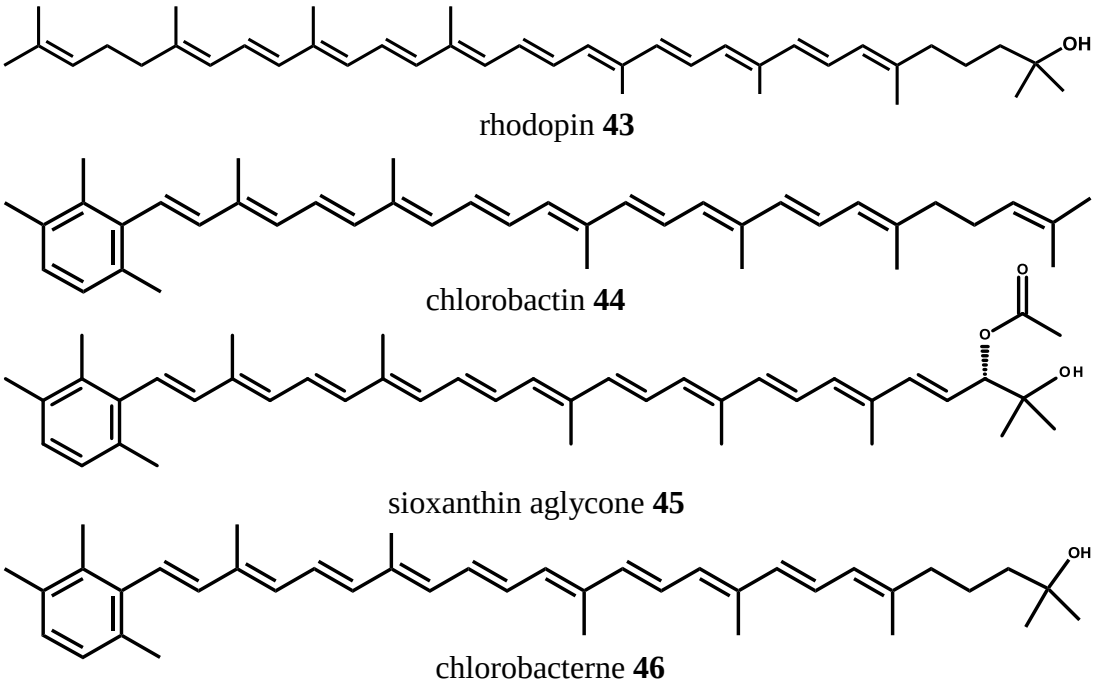


Figure 3.13: Proposed structures of major carotenoids from the biosynthetic mutants.

Further structure elucidation was done on the *strop3247* mutant compound, which underwent ^{1}H NMR analysis. The NMR confirms that the structure **(45)** lacks the glucose functional group, but is in all other ways identical to the wild-type structure (Figure 3.14).

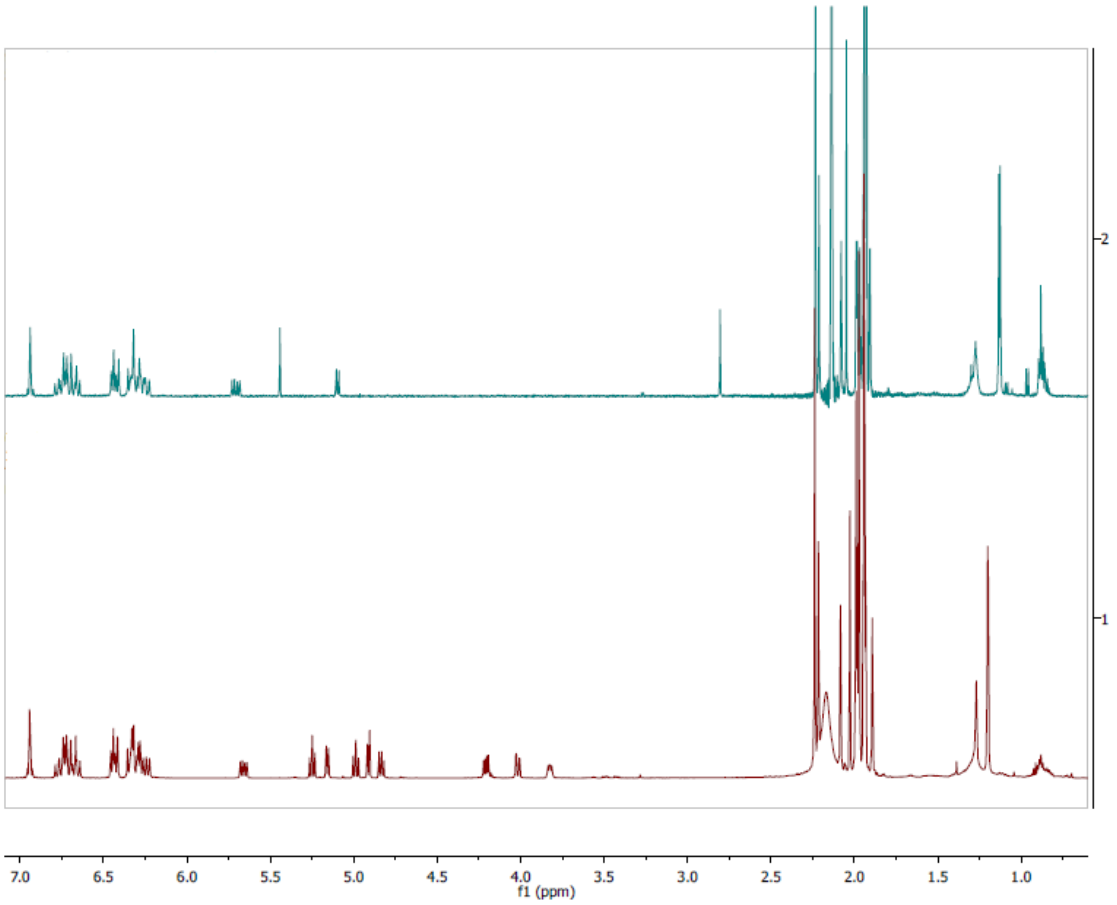


Figure 3.14: ^{1}H NMR (CD_3CN; 600MHz) spectrum of the *strop3247* mutant compound **45** (blue) compared with the spectrum of the wild-type sioxanthin acetate derivative **42** (red). Glucose signals are found in the δ3.5-5.5 ppm region.

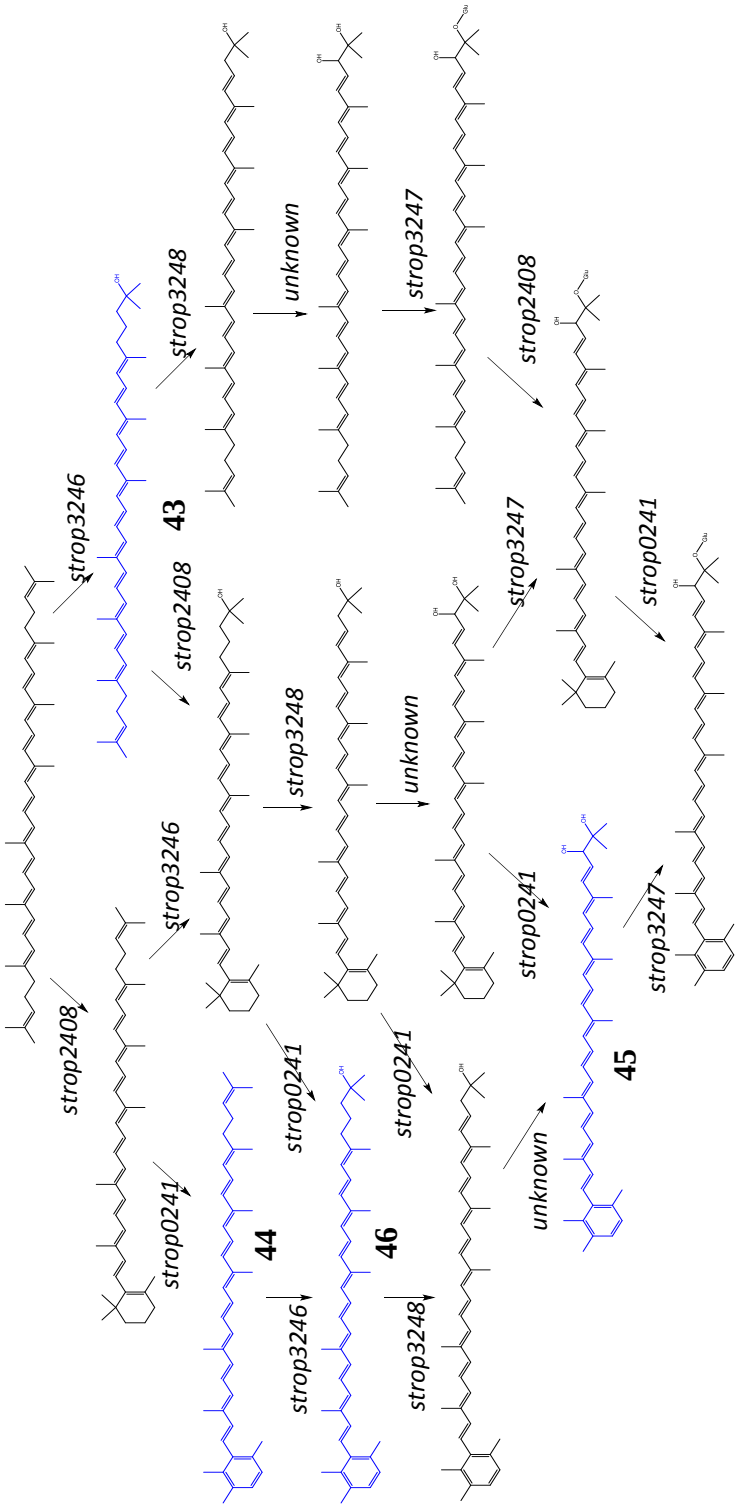
3.2.4 Structural information of mutant carotenoids improves understanding of the biosynthetic pathway

Analyses of the mutant carotenoid structures confirmed the putative roles of the genes first assigned by informatics and *in vivo* mutagenesis in Chapter 2. Structures confirmed that the inactivated genes performed their predicted functions by resulting in compounds with the corresponding structural feature missing (Table 3.2, Figure 3.13). Therefore, *strop2408* encodes a lycopene cyclase (*crtY*) that forms a cyclized end, *strop3246* encodes a 1',2'-hydratase (*cruF*) which adds a 1'-hydroxyl, *strop3247*

encodes a glycosyl transferase (*cruC*) which adds a glucose to the 1'-hydroxy, and *strop3248* encodes a 3',4'-desaturase which adds a double bond. Though it could not be confirmed by MS data, *strop0241* is predicted to be the desaturase/isomerase responsible for the conversion of the β-ring in to an aryl functional group. The only biosynthetic step that was not accounted for in this study is that for the addition of the 2'-hydroxy group. The biosynthetic gene responsible for the addition of this functional group is unknown, as there is currently no known enzyme that performs this function in carotenoid biosynthesis. A 2,2'-hydroxylase is also missing from the biosynthetic pathways of (2S,2'S)-oscillol 2,2'-di-(alpha-L-rhamnoside), from *Gemmatiomonas aurantiaca,* and deinoxanthin, from *Deinococcus radiodurans*[15,16]. Both of these pathways have a non-clustered arrangement, as is found in *Salinispora*[15,16]. Cyanobacteria are also known to produce carotenoids with 2,2'-hydroxy group additions, but no gene has been found to be responsible for this enzymatic reaction[6,17]. The closest to a 2'-hydroxylase is the *crtA* gene. Though primarily known as a ketolase, there are known homologs of this gene that instead result in a hydroxyl in the 2,2' position[18]. However, the *crtA* homolog in *Salinispora* (*strop3244*) was shown not to participate in the biosynthesis of sioxanthin (Fig 2.8).

Identification of these structural features provides evidence of the order in which enzymes act during the biosynthesis of sioxanthin. The intermediates isolated from the biosynthetic mutants show that both ends of the compound are biosynthesized independently (Scheme 3.1), though in each of these ends, biosynthesis follows a certain order. Unsurprisingly, on the left side of the molecule, cyclization takes place prior to desaturation/isomerization. The right side of the molecule is less obvious, but

can be determined by intermediates. The 1',2'-hydration occurs first, resulting in a 1'-hydroxy group as well as a saturation of the 1',2' bond. Next is the desaturation of the 3',4' bond, followed by the addition of the 2'-hydroxy, and finally, the glycosylation of the 1'-hydroxyl. Searches of the chemical structure databases confirm that all of the identified intermediate compounds have been previously identified, with the exception of the *strop3247* mutant compound (**45**).



Scheme 3.1: Proposed biosynthetic pathway of sioxanthin with the genes responsible for each biosynthetic step in italics. Blue structures are intermediates that have been confirmed by MS and/or NMR

3.3 Conclusions

The *S. tropica* carotenoid structure was determined using a combination of UV/Vis, MS, NMR, and CD data. This compound, assigned the trivial name sioxanthin, is a novel carotenoid with both aryl and glucose functional groups. Each end of the molecule has been previously identified, which aided in the structural elucidation. Carotenoids from biosynthetic mutants were determined using NMR and MS techniques, which enabled structural predictions. These intermediate structures confirm the function of the inactivated genes as well as provide insight into the sequence of biosynthesis.

3.4 Materials and methods

3.4.1 Growth conditions of *Salinispora* cultures.

S. tropica wild-type and mutants were grown in 50 mL starter cultures as described in Section 2.4.4. Apramycin and naladixic acid were used in the growth of mutant strains. After five days of growth, 5 mL of the starter culture were transferred under sterile conditions to 2.5 L flasks containing 1 L A1 liquid medium. Cultures were allowed to continue growing for 10-14 days at 30° C and shaking at 220 rpm. All cultures were checked for contamination by monitoring growth of 10 µL of the culture on an A1 agar media plate grown in the dark at 30° C. Contaminated cultures were excluded from further analysis.

3.4.2 Harvesting cells and pigment extractions

Cultures were centrifuged in a Beckman-Coulter Avanti J-E centrifuge at 11,000 x g for 30 minutes at 4° C. Media was discarded and cell pellets from 6-10 liters of cell culture were combined for extraction. Pigment extracts were obtained by soaking cell

pellets in approximately 100 mL of acetone for several hours with occasional stirring. The colored solvent was passed through cheesecloth to remove cell debris. Acetone extractions were repeated until the solvent was no longer colored after exposure to the cells. Between extractions the cell pellets were stirred with deionized water, as extraction yield was higher from wet cells. Extractions were combined and dried by first partitioning with brine, followed by use of magnesium sulfate or sodium sulfate, removal of salts by passing through a filter, and finally dried *in vacuo*.

3.4.3 Derivitization of the crude extract

The crude pigment extracts were acetylated to improve stability during the purification steps. The dry extracts were resuspended in 1-3 mL dichloromethane (DCM). For each mL of DCM, 250 µl triethylamine (Et_3N), 250 µl acetic anhydride, and 0.1 mg of 4-dimethylaminopyridine were added. The reaction was allowed to sit at room temperature for at least four hours. The pigments were extracted from the DCM solution using ethyl acetate and partitioning with brine. The ethyl acetate was transferred and dried with sodium sulfate. Salts were removed by passing the solution through a glass wool filter and the resulting solution was dried under nitrogen. The conversion was checked by resuspending the extract in acetonitrile and analyzing the solution on HPLC. The new major peak had a longer retention time, indicating a less polar compound, and the wild-type compound peak had been eliminated, suggesting complete conversion.

3.4.4 Purification of the *S. tropica* carotenoid

Purification of the primary pigment required two rounds of HPLC peak isolation using a Hewlett Packard Series II 1090 Liquid Chromatograph. HPLC traces were

monitored at 450 nm and major carotenoid peaks were confirmed by examining the UV/Vis spectrum. Care was taken to keep the samples in low light conditions, in cool temperatures and under nitrogen, where possible. Crude extracts were first purified on a reversed-phase system using a Phenomenex Luna 5 µm C18 100A 250 x 10mm column. The solvent used for the wild-type and most mutant extracts was 94.05% methanol, 5% DCM, 0.05% Et_3N, and 0.05M ammonium acetate run under isocratic conditions at 3 mL/min. The conditions were slightly altered for better separation of the *strop3246* mutant extracts and consisted of 85.45% methanol, 9.5% acetonitrile, 5% DCM, 0.05% Et_3N, and 0.05M ammonium acetate, also under isocratic conditions. The major HPLC peak of each crude extract was collected and dried under nitrogen.

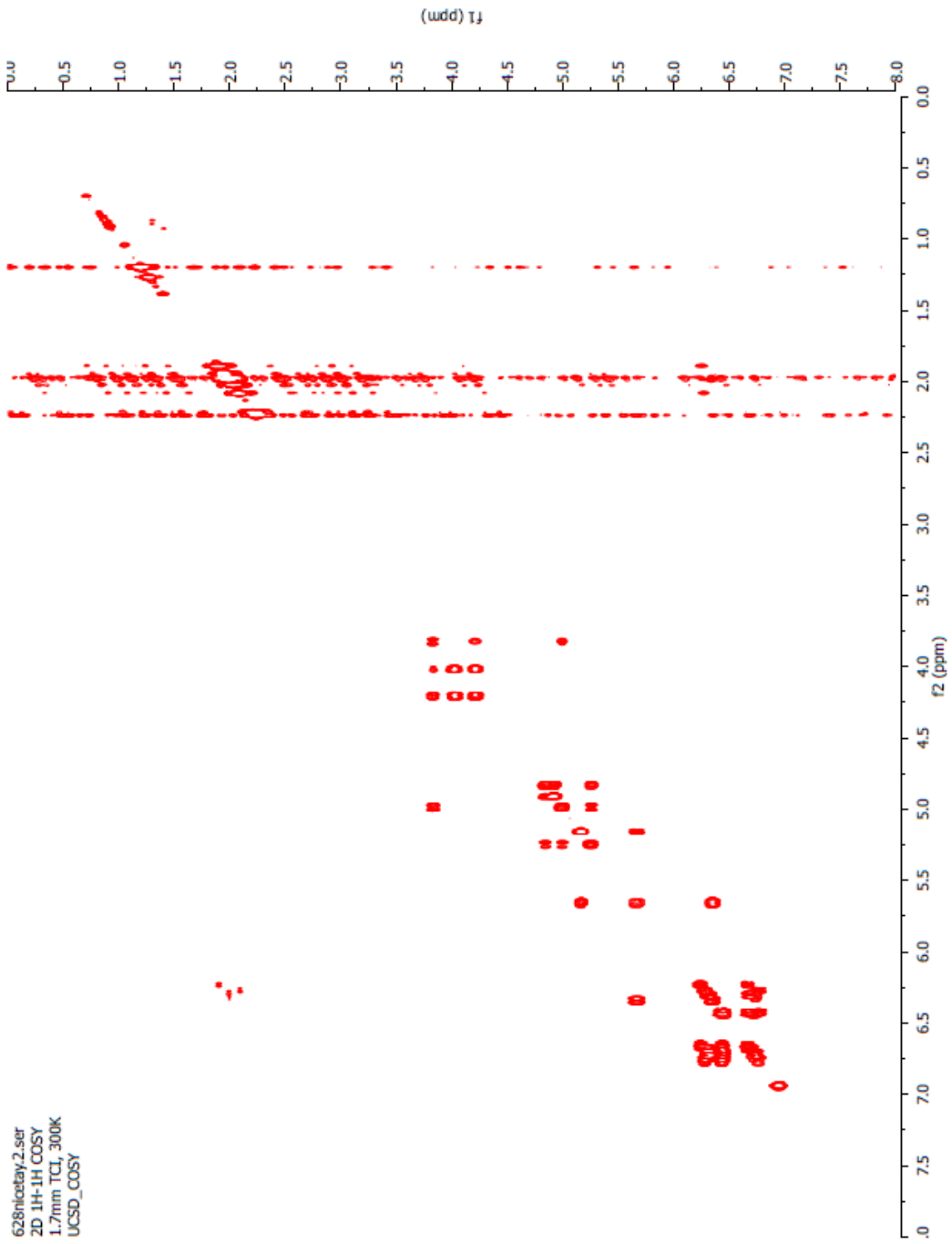
Semi-pure extracts were then re-injected on the same instrument under normal phase conditions. Samples were separated on a Phenomenex Luna 5 µm Silica (2) 100A 250 x 10 mm column. The solvent system was run isocratically with a flow rate of 3 mL/min and for most samples consisted of the following: 93% hexanes, 5% DCM, 2% isopropanol. The one exception was the *strop2408* mutant extract which used a solvent system of 91% hexanes, 5% DCM, and 4% isopropanol. Purified samples were dried under nitrogen and stored at -20° C. Sample purity was confirmed via ^{1}H NMR. All solvents were of HPLC grade purity.

3.4.5 Structure elucidation of the carotenoid compounds

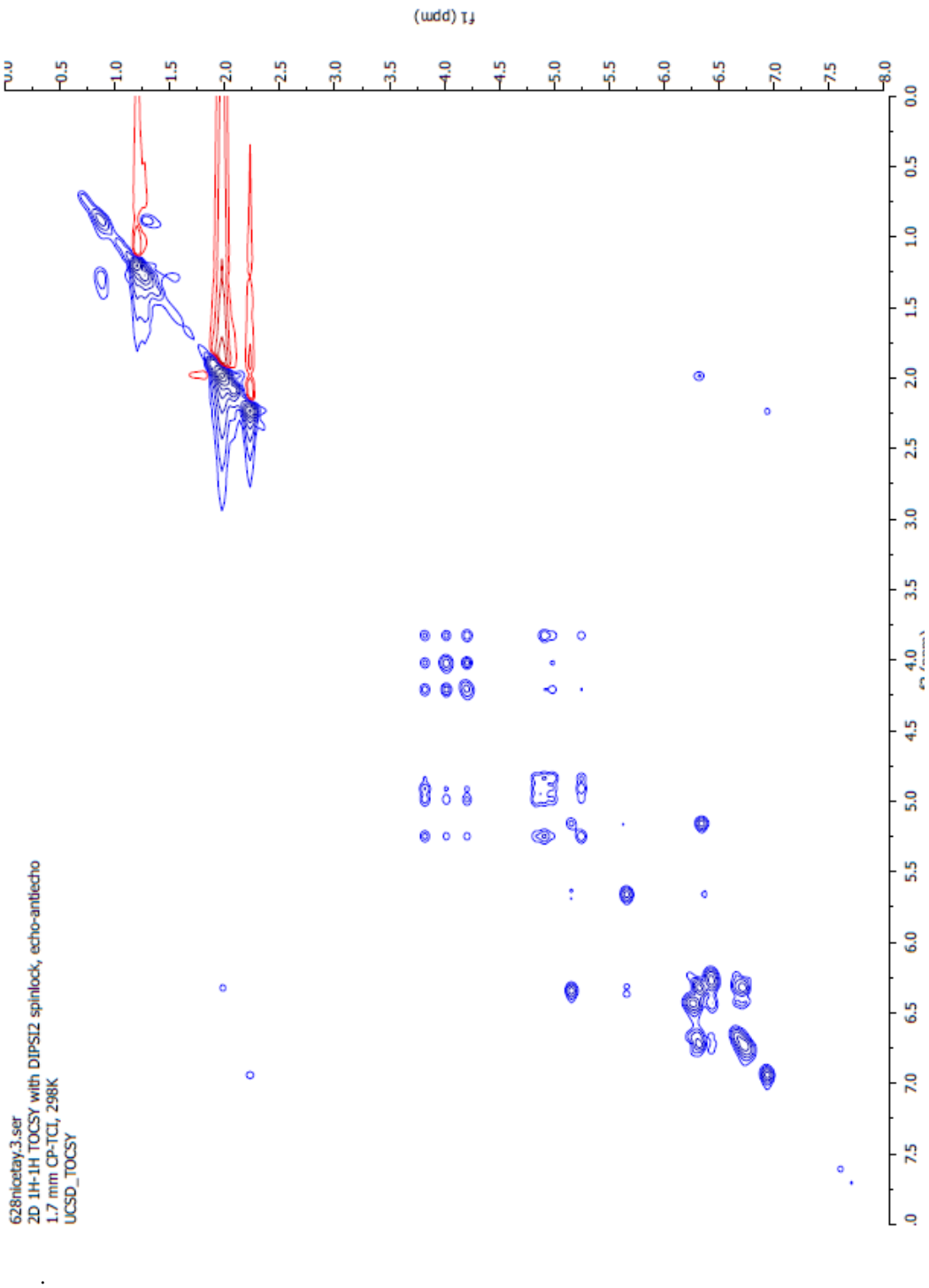
Purified compound was analyzed by NMR at the UC San Diego Skaggs School of Pharmacy and Pharmaceutical Sciences NMR Facility on a Varian NPA600 MHz NMR fitted with a 1.7 mm inverse detection triple resonance (H-C/N/D) cryoprobe. The wild-type sample in CD_3CN was analyzed by ^{1}H NMR as well as COSY, TOCSY,

NOSY, HMBC, and HMQC spectroscopy. Mutant samples were analyzed by ^{1}H NMR. High resolution mass spectrometry was performed at the UCSD Chemistry and Biochemistry Molecular MS Facility using an Agilent 6230 APCI-TOFMS in positive mode. UV/Vis was determined on the HPLC and confirmed on an Agilent Cary60 UV-Vis in methanol. Circular dichroism was performed in distilled methanol on a Jasco 810 spectropolarimeter in quartz cells with a 1 mm path length and scanning from 220-400 nm. Optical rotation was measured on a Jasco P210 polarimeter in distilled methanol.

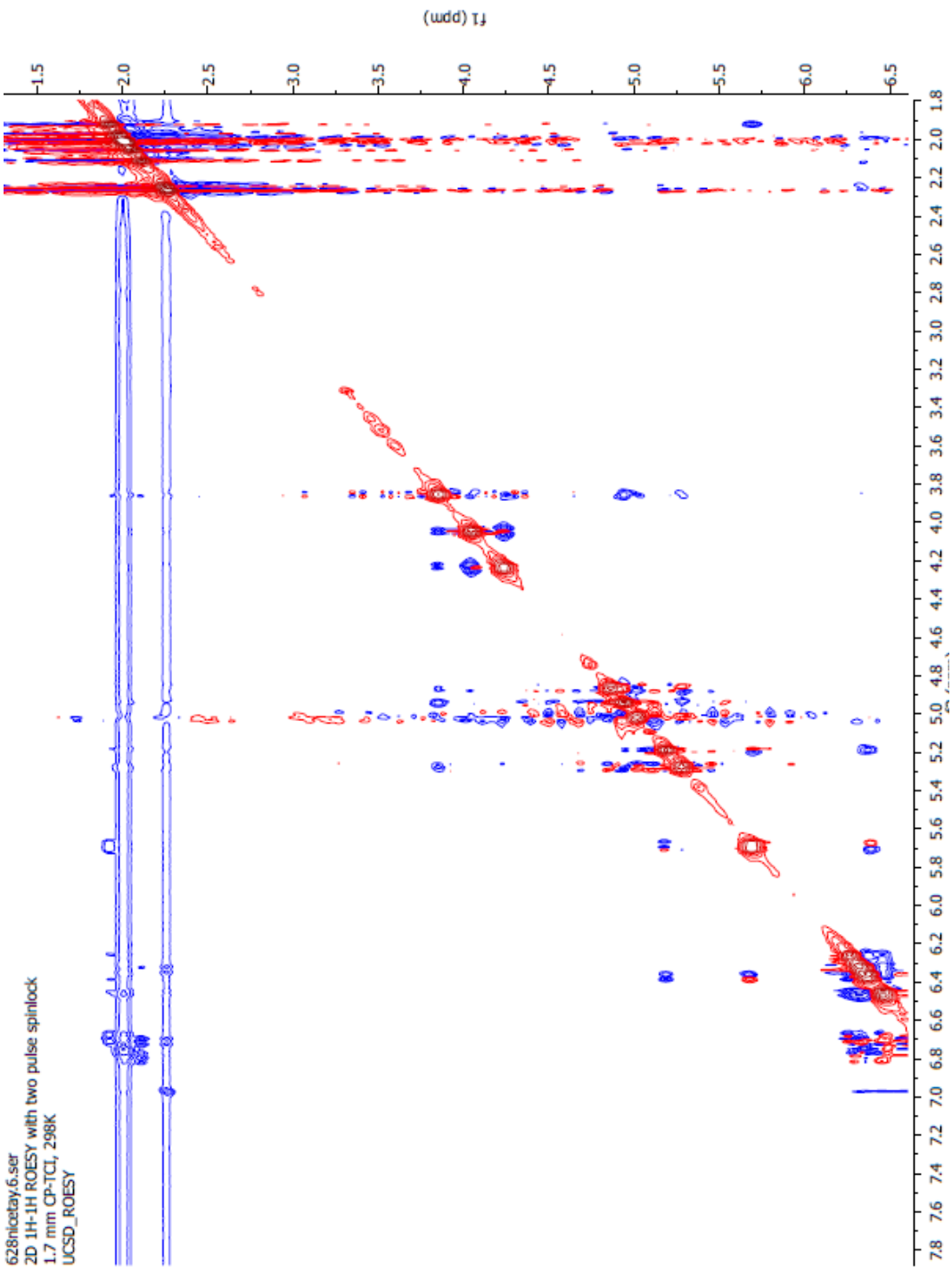
3.5 Supplemental Information



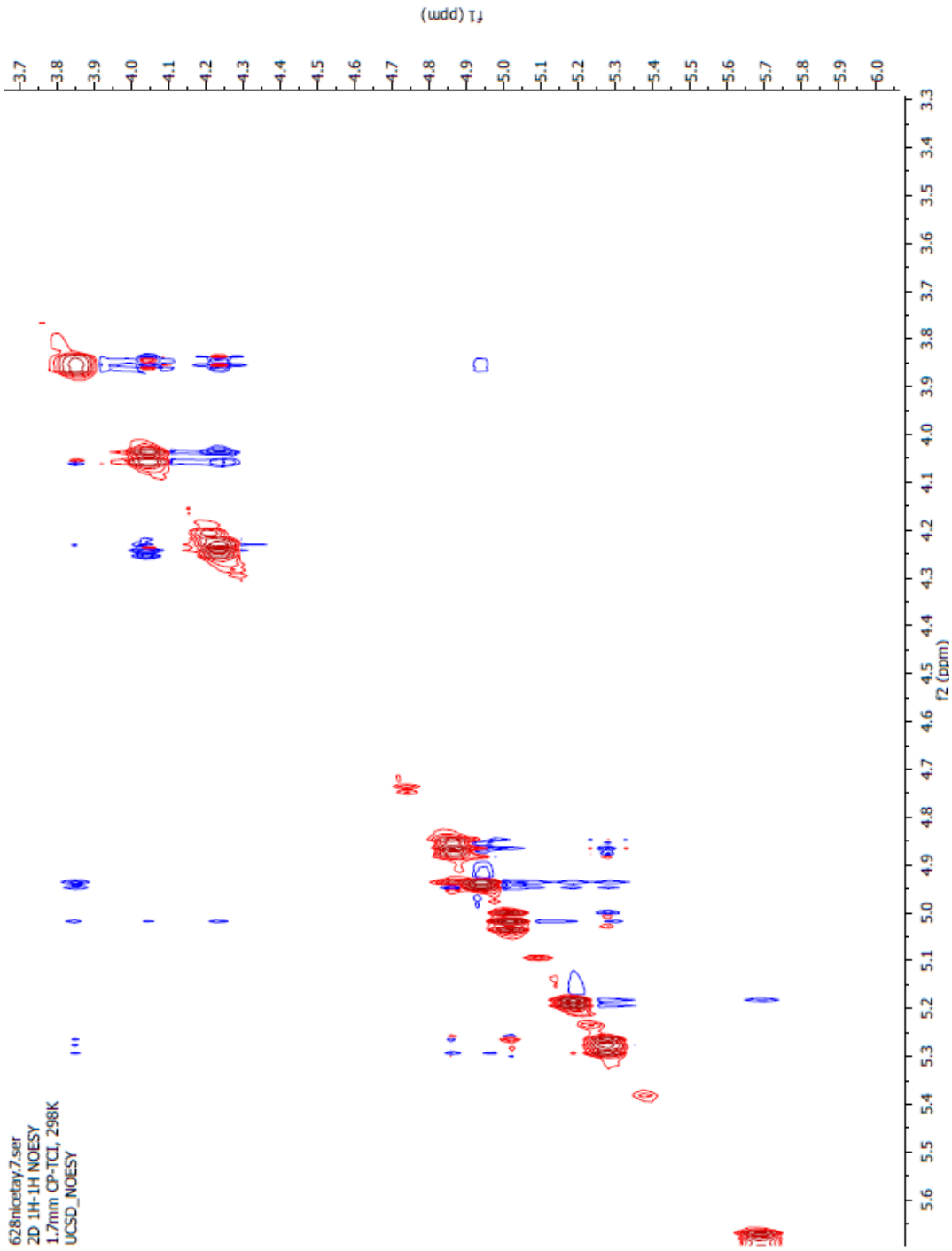
Suppl Figure 3.1: COSY (CD_3CN; 600MHz) spectrum of sioxanthin.



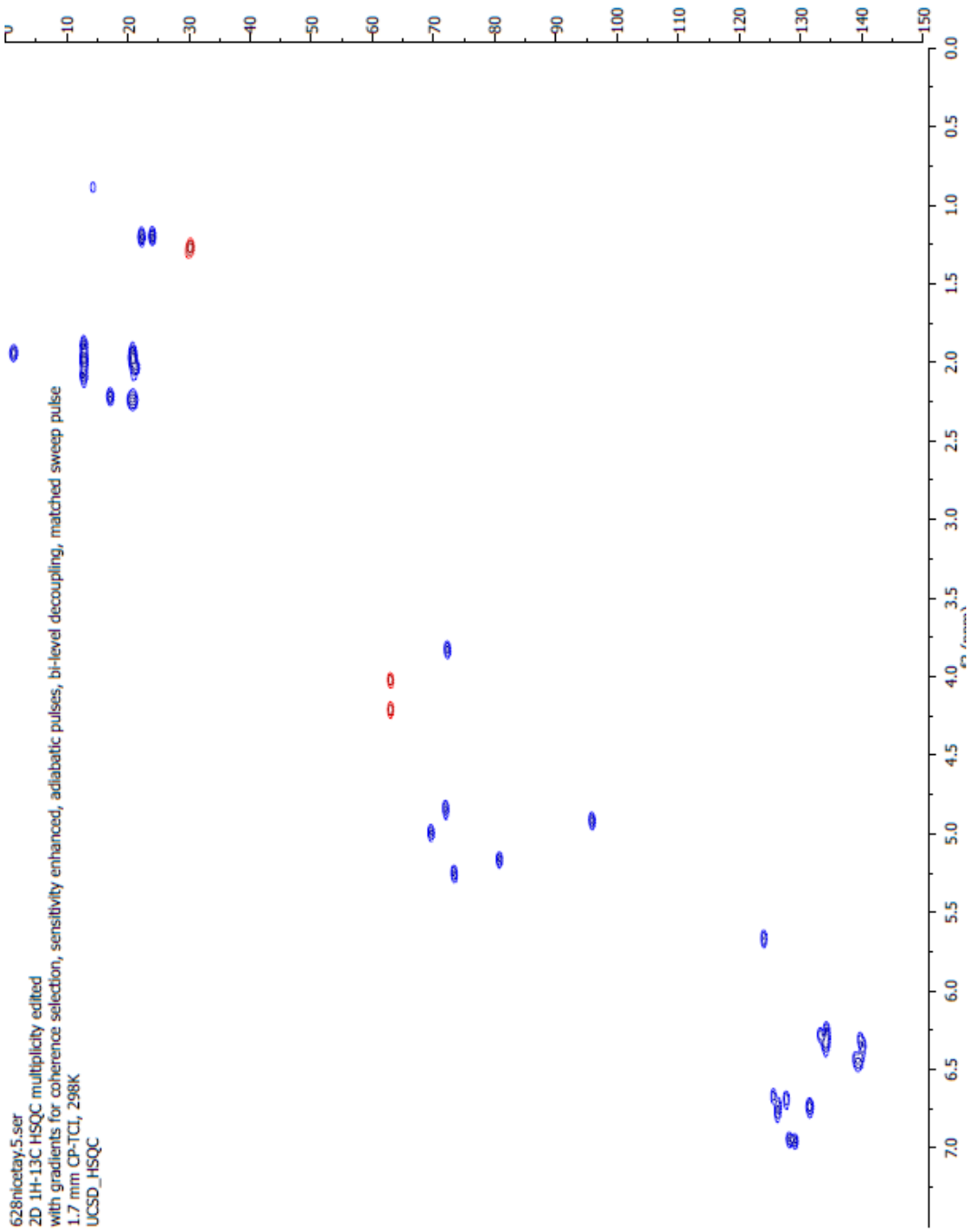
Suppl Figure 3.2: TOCSY (CD_3CN; 600MHz) spectrum of sioxanthin



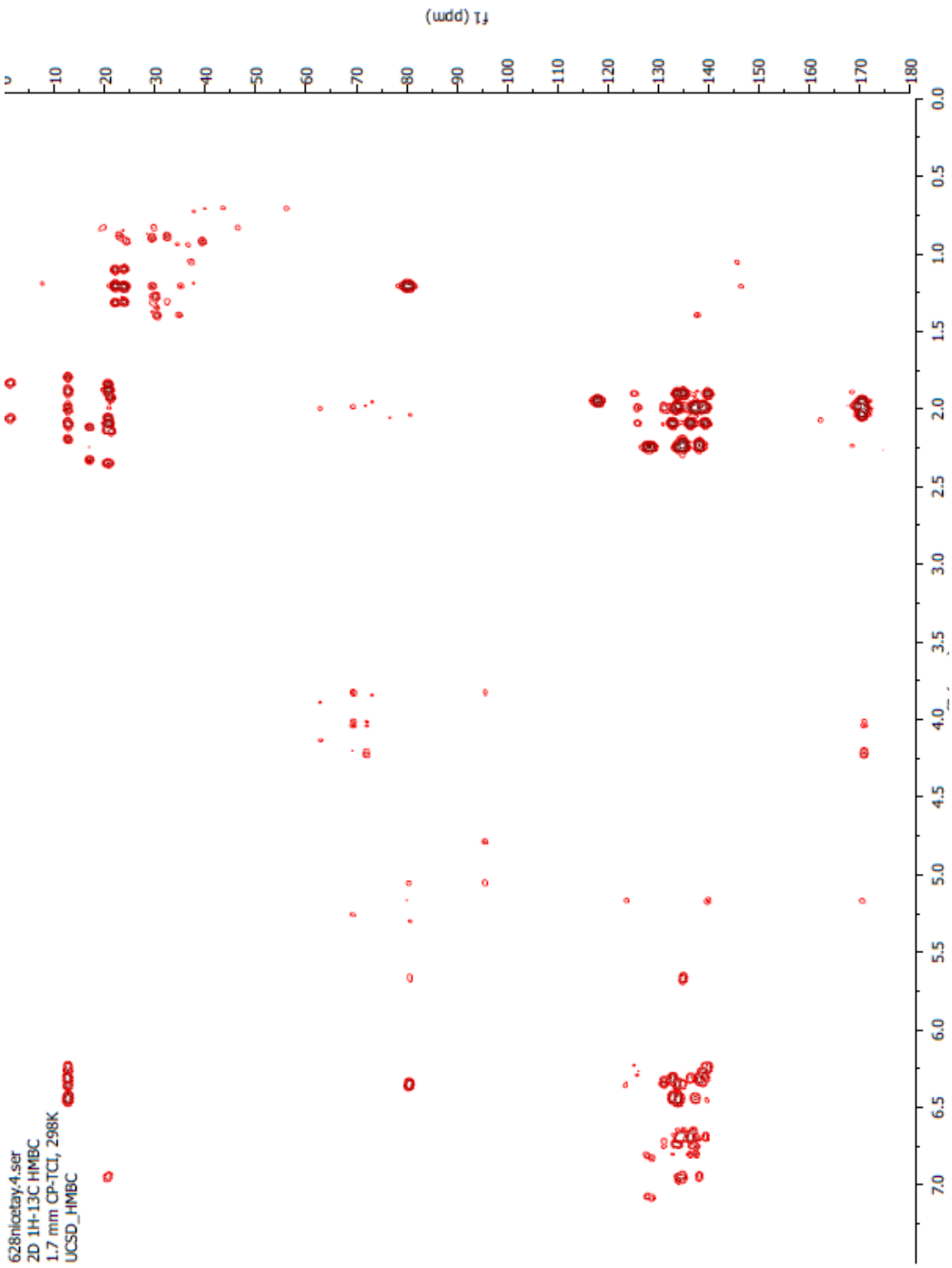
Suppl Figure 3.3: ROESY (CD_3CN; 600MHz) spectrum of sioxanthin.



Suppl Figure 3.4: NOESY (CD_3CN; 600MHz) spectrum of sioxanthin.



Suppl Figure 3.5: HSQC (CD_3CN; 600MHz) spectrum of sioxanthin.



Suppl Figure 3.6: HMBC (CD_3CN; 600MHz) spectrum of sioxanthin.

3.6 References

1 Britton, G. Structure and properties of carotenoids in relation to function. *FASEB J.* **9**, 1551-1558, (1995).

2 vanBreemen, R. B. Liquid chromatography mass spectrometry of carotenoids. *Pure Appl. Chem.* **69**, 2061-2066, (1997).

3 Schiedt, K. L.-J., Synnove. in *Carotenoids: Isolation and Analysis* Vol. 1A *Carotenoids* (ed G.; Liaan-Jensen Britton, S.; Pfander, H.) Ch. 5, 81-109 (Birkhauser Verlag, 1995).

4 vanBreemen, R. B. Innovations in carotenoid analysis using LC/MS. *Anal. Chem.* **68**, A299-A304, (1996).

5 Britton, G. in *Carotenoids: Spectroscopy* Vol. 1B (eds G. Britton, S. Liaaenjensen, & H. Pfander) Ch. 2, 13-62 (Birkhauser Verlag, 1995).

6 Britton, Liaaen-Jensen, S., Pfander, H. *Carotenoids handbook*. (Birkhäuser Verlag, 2004).

7 Armstrong, G. A. in *Annu. Rev. Microbiol.* Vol. 51 *Annual Review of Microbiology* (ed L. N. Ornston) 629-659 (Annual Reviews Inc., 1997).

8 Sander, L. C., Sharpless, K. E., Craft, N. E. & Wise, S. A. Development of engineered stationary phases for the separation of carotenoid isomers. *Anal. Chem.* **66**, 1667-1674, (1994).

9 Hart, D. J. & Scott, K. J. Development and evaluation of an HPLC method for the analysis of carotenoids in foods, and the measurement of the carotenoid content of vegetables and fruits commonly consumed in the UK. *Food Chem.* **54**, 101-111, (1995).

10 Englert, G. in *Carotenoids: Spectroscopy* Vol. 1B (ed G.; Liaan-Jensen Britton, S.; Pfander, H.) Ch. 6, 147-259 (Birkhauser, 1995).

11 Choi, S. K., Osawa, A., Maoka, T., Hattan, J., Ito, K., Uchiyama, A., Suzuki, M., Shindo, K. & Misawa, N. 3-beta-Glucosyl-3'-beta-quinovosyl zeaxanthin, a novel carotenoid glycoside synthesized by Escherichia coli cells expressing the Pantoea ananatis carotenoid biosynthesis gene cluster. *Appl. Microbiol. Biotechnol.* **97**, 8479-8486, (2013).

12 Ronneberg, H., Andrewes, A. G., Borch, G., Berger, R. & Liaaen-Jensen, S. CD correlation of C-2' substituted monocyclic carotenoids. *Phytochemistry* **24**, 309-319, (1985).

13 Krugel, H., Krubasik, P., Weber, K., Saluz, H. P. & Sandmann, G. Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of

isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *BBA-Mol. Cell Biol. L.* **1439**, 57-64, (1999).

14 Hertzber.S & Liaaen-Jensen, S. Bacterial carotenoids .20. Carotenoids of Mycobacterium phlei strain Vera .2. Structures of phlei-xanthophylls - 2 novel tertiary glucosides. *Acta Chem. Scand.* **21**, 15-&, (1967).

15 Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2 '-dirhamnoside. *Microbiol.-Sgm* **156**, 757-763, (2010).

16 Tian, B. & Hua, Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol.* **18**, 512-520, (2010).

17 Graham, J. E. & Bryant, D. A. The biosynthetic pathway for myxol-2 ' fucoside (myxoxanthophyll) in the cyanobacterium *Synechococcus* sp strain PCC 7002. *J. Bacteriol.* **191**, 3292-3300, (2009).

18 Rahlert, N., Fraser, P. D. & Sandmann, G. A *crtA*-related gene from Flavobacterium P99-3 encodes a novel carotenoid 2-hydroxylase involved in myxol biosynthesis. *FEBS Lett.* **583**, 1605-1610, (2009).

4 Carotenoid biosynthesis gene clustering and evolutionary history of sioxanthin

4.1 Introduction

The increased availability of sequenced bacterial genomes has furthered the investigation of secondary metabolite biosynthetic genes and their organization in bacterial genomes. Secondary metabolic genes in bacterial genomes are typically organized into clusters, in which genes in a common pathway occupy neighboring space in the genome[1,2]. Gene clustering is thought to confer several advantages for the regulation, maintenance, and exchange of these accessory pathways[1,2]. Similarly, with the exception of cyanobacteria, gene clustering is the common organization for carotenoid biosynthetic genes in bacterial genomes (Figure 4.1)[3]. As shown in Chapter 2, the arrangement of carotenoid biosynthetic genes in *S. tropica* does not fit this standard.

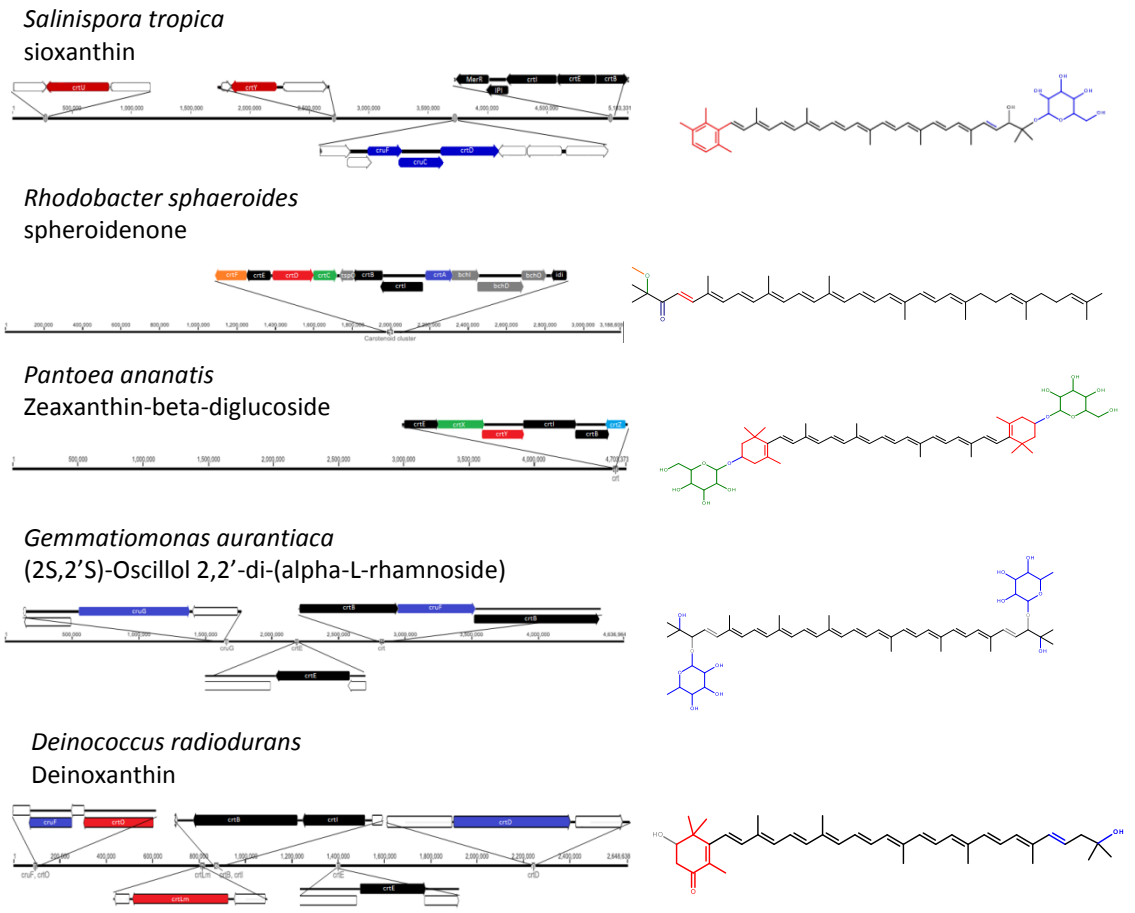


Figure 4.1: Examples of genome arrangements of various carotenoid biosynthetic pathways and their products. It shows the *Salinispora tropica* architecture in comparison to two clustered pathways and two non-clustered pathways. Gene color corresponds to region of the region of the structure that the gene is responsible for. Gray genes in *Rhodobacter sphaeroides* are responsible for the biosynthesis if bacteriochlorophyll[4].

Only two examples have been shown in which genes required for carotenoid biosynthesis are dispersed throughout the genome[5,6] (Figure 4.1). This makes the arrangement in *Salinispora* unusual, though not unique. In both of these published cases, gene assignment is based on bioinformatics predictions, without biochemical characterization, and both pathways are missing a gene responsible for 2,2'-hydroxylation as is the case in *S. tropica*.

Mechanisms that contribute to genome architecture have long been of interest in bacteria, and have been a focus of study in the genus *Salinispora*. The *Salinispora* genome, and in particular its secondary metabolic arsenal, is the result of frequent horizontal gene transfer[7,8]. The majority of secondary metabolic pathways are localized in genomic islands, regions of divergence between species genomes[7]. These genomic islands are also enriched in mobile genetic elements, suggesting that these regions are hot spots for genetic exchange and rearrangement[7]. Even outside of the genomic islands, *Salinispora* genomes contain CRISPR regions, remnants of phage integration, that implicate transduction as a mechanism of genetic acquisition[7].

Carotenoid biosynthetic genes in *Salinispora* are dispersed throughout the genome, occupying at least four independent regions. These regions include two gene clusters, two non-clustered genes, and one gene whose identity remains unknown. While there are many hypotheses pertaining to the benefits of development and maintenance of gene clusters in secondary metabolism[1], there is a lack of understanding of this alternative, dispersed genome arrangement. As sioxanthin biosynthesis appears to be constitutive, there is no requirement of coordinated regulation driving the maintenance or formation of a single cluster. There are two main hypotheses regarding the development of this arrangement: that the sioxanthin gene cluster was unified in the past and spread apart over time, or that the different regions of the pathway were acquired separately through horizontal gene transfer. Horizontal exchange of biosynthetic genes is known to play a role in the diversification of carotenoid chemical structure, particularly among the Actinobacteria[9-11]. Even outside of bacteria, horizontal exchange of genes is an important mechanism in the expansion of carotenoid

biosynthesis. Recent studies have identified a few species of arthropods with carotenoid biosynthetic genes in their genomes inherited from fungi, providing an example of HGT of these genes and making them the first known animals capable of carotenoid biosynthesis[12,13]. Carotenoid biosynthesis genes are known to be horizontally transferred as whole clusters, either as standalone carotenoid pathways or as a part of gene clusters encoding for photosynthesis machinery[14,15]. Few studies have looked at the history of individual genes within a carotenoid biosynthetic gene cluster and have indicated that at times, only particular components of pathways are transferred[9-11].

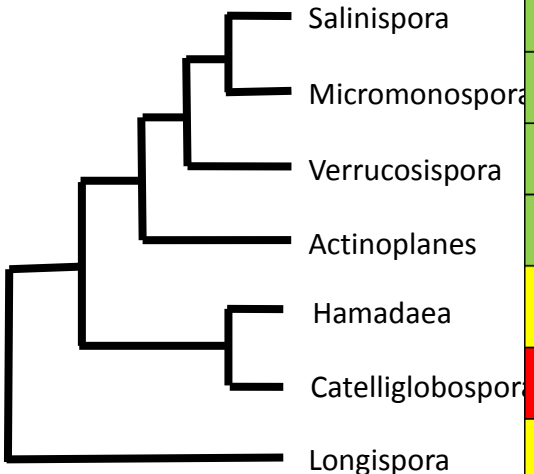
In this chapter, the role of horizontal gene transfer in the sioxanthin biosynthetic pathway genome arrangement is explored. The prevalence of this genome arrangement in carotenoid biosynthesis beyond *S. tropica* is investigated to determine the uniqueness of this system. Furthermore, the phylogenetic history of each region is examined to explore the evolution of the sioxanthin molecule which enables a discussion of the role of horizontal gene transfer in the expansion of chemical diversity.

4.2 Results and Discussion

4.2.1 Sioxanthin biosynthesis outside of *Salinispora*

The sioxanthin biosynthetic pathway and genome arrangement are not localized to the genus *Salinispora*. Other bacteria in the bacterial family Micromonosporaceae produce an orange pigment, likely associated with carotenoid biosynthesis. However, no work has been done to identify the particular compound. The sioxanthin pathway genes, as they are found in *Salinispora*, are also present in the genera *Micromonospora*, *Verrucosispora* as well as some species in the genus *Actinoplanes* (Figure 4.2, Figure

4.3, Figure 4.4). It is difficult to predict the biosynthesis of sioxanthin in these bacteria with certainty, as there is still an unidentified gene in the pathway and the presence of a gene does not necessitate its transcription. However, the high degree of pathway synteny between these genera (Figure 4.3, Figure 4.4) provides some confidence that sioxanthin contributes to their orange pigmentation. The *Hamadea* and *Longispora* contain genes for both clusters, though usually in a different conformation than what is found in *Salinispora* (Figure 4.3, Figure 4.4). As neither of these genera harbors lycopene cyclase or β-carotene desaturase genes, they are not predicted to produce sioxanthin (Figure 4.2). Sioxanthin biosynthesis appears to be a feature of *Salinispora* and closely related genera, though not of Micromonosporaceae as a whole. This suggests that the sioxanthin pathway developed in a common ancestor of a few genera, but after the split of the Micromonosporaceae family.



Terp1	Terp2	Lycopene cyclase	Desaturase
Yes Syntenic	Yes Syntenic	Yes	Yes
Yes Syntenic	Yes Syntenic	Yes	Yes
Yes Syntenic	Yes Syntenic	Yes	Yes
Yes Syntenic	Yes Syntenic	Some	Yes
Yes Not syntenic	Yes Syntenic	No	No
No	No	No	No
Yes Not syntenic	Yes Not Syntenic	No	No

Figure 4.2: Summary of the presence of sioxanthin pathway genes in other Micromonosporaceae genera. The dendrogram shows relative relationships of the Micromonosporaceae genera while the table describes the presence of genes in the region and whether they are syntenic. Green shows complete and syntenic gene regions, red shows absence of genes, and yellow shows either genes present but not syntenic or genes not present in all species.

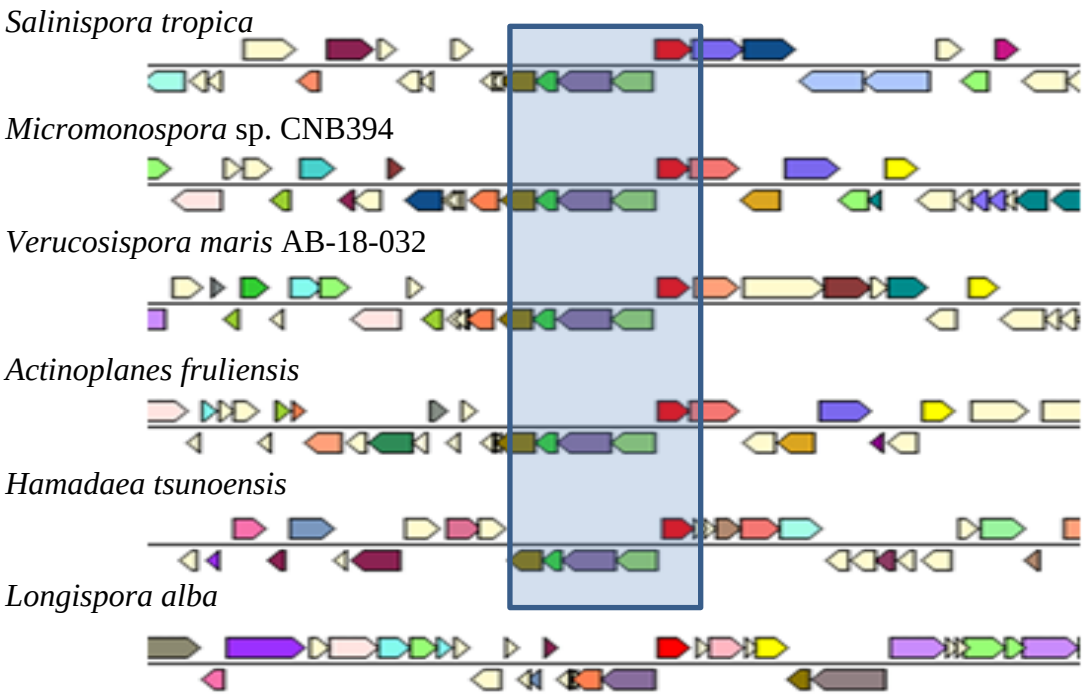


Figure 4.3: Comparison of the gene cluster responsible for lycopene biosynthesis (*terp2* homologs) in members of the Micromonosporaceae. Genes of a with a similar COG function are colored alike. The blue box highlights the *terp2* cluster (*strop4437-strop4441*) in *Salinisipora* and genera for which this cluster is syntenic.

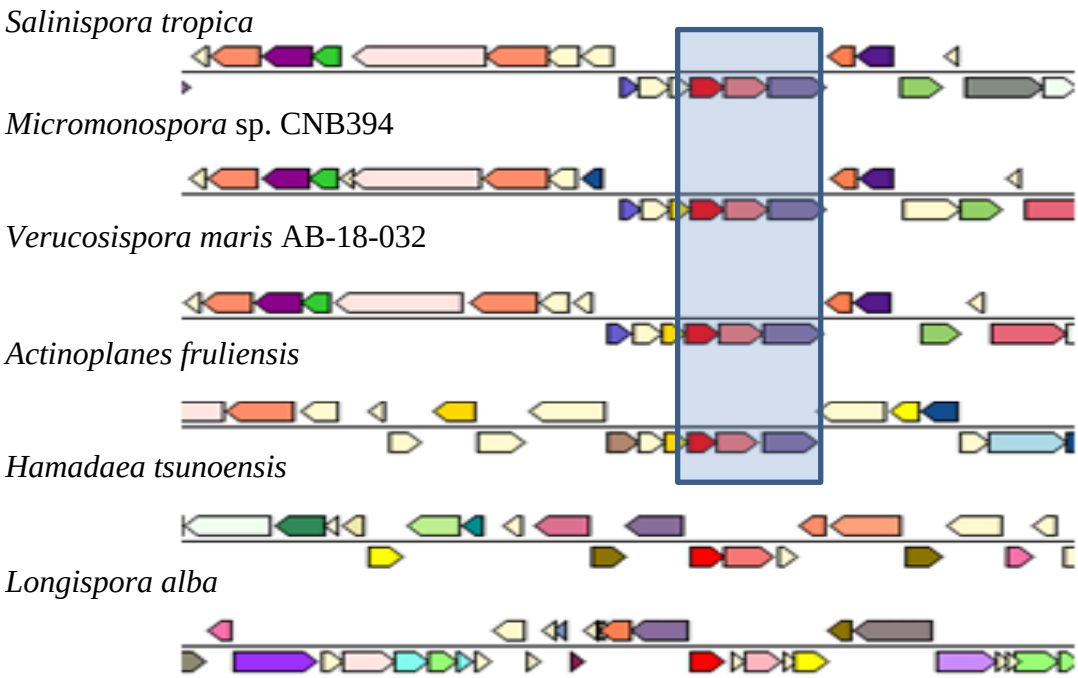


Figure 4.4: Comparison of the gene cluster responsible for sioxanthin glycosylation and 3', 4'-desaturation in members of the Micromonosporaceae. Genes of a with a similar COG function are colored alike. The blue box highlights the portion of the *terp1* cluster (*strop3246-strop3248*) necessary to sioxanthin biosynthesis in *Salinisipora* and genera for which this cluster is syntenic.

The potential for sioxanthin production was explored in other actinomycetes. No actinomycetes outside of the Micromonosporaceae contained the full suite of carotenogenesis genes as found in *Salinispora.* Three actinomycete genera (*Gordonia, Amycolatopsis*, and *Nocardia*) possess a complimentary set of genes that could code for sioxanthin biosynthesis (Figure 4.5). They contain an alternative lycopene cyclase and multiple desaturases, which may be involved in desaturation of the 3',4' and the β-cycle desaturation steps. Two of these genera, *Gordonia* and *Nocardia*, also have a gene for carotenoid ketolation. Ketone-containing carotenoids have been confirmed in *Gordonia*[16], making it unlikely that sioxanthin is produced in these other genera.

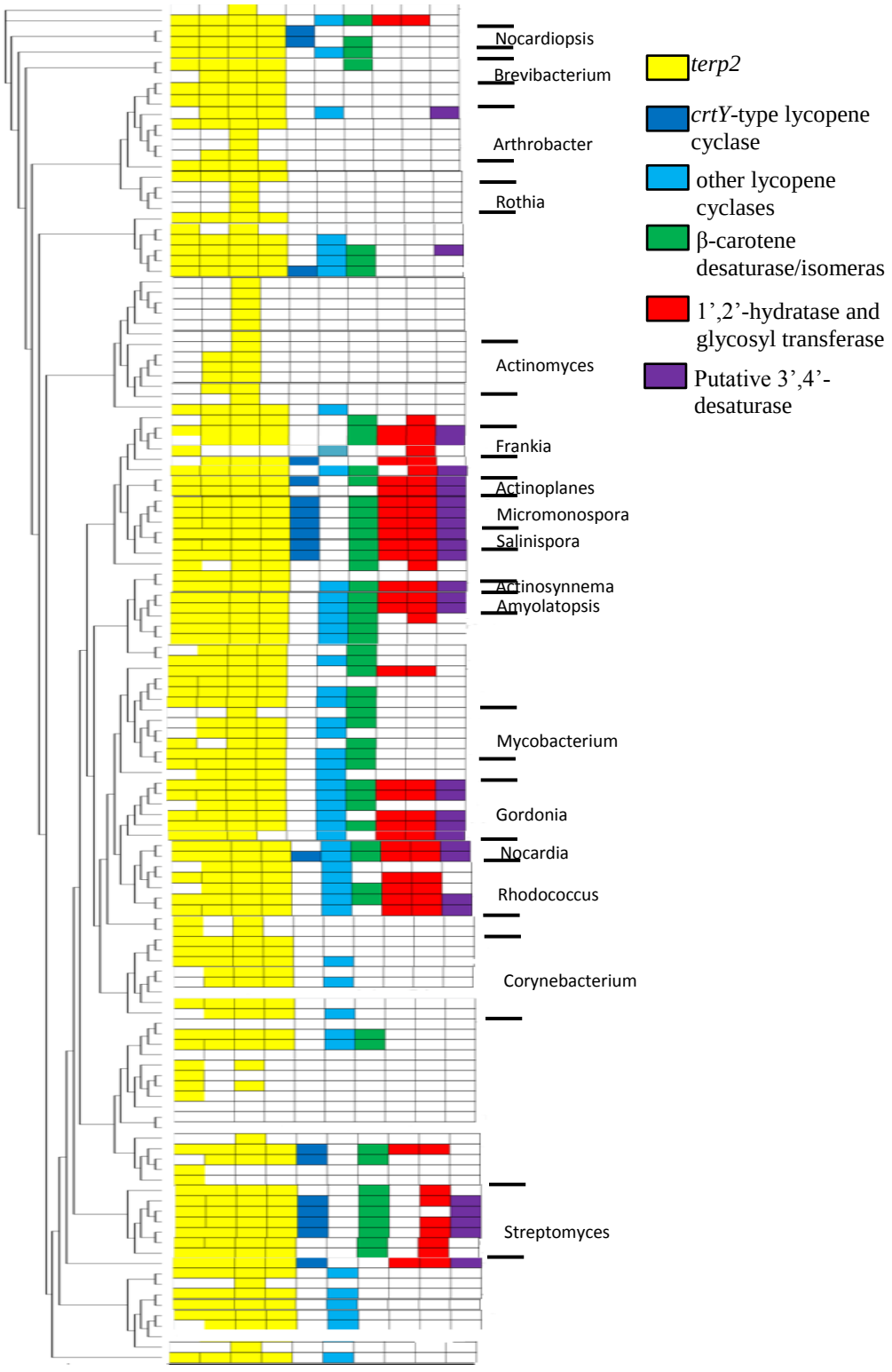


Figure 4.5: Neighbor-joining tree of 16S rRNA sequences showing the presence or absence of sioxanthin biosynthesis genes elsewhere in the actinobacteria. elect genera are labelled on the tree.

This investigation of sioxanthin biosynthetic genes reveals interesting trends regarding carotenoid biosynthesis among the actinobacteria. For example, it confirms that structural features such as glycosyl groups are more common among the actinomycetes than originally anticipated and the distribution may be the result of horizontal gene transfer. In actinomycetes, nine glycosylated carotenoids have been identified from eight genera[16,17], including sioxanthin biosynthesis in *Salinsipora*. Most of these carotenoids are found in single species or single strains that do not currently have sequenced genomes, so biosynthetic genes are not known. The glycosylated end of sioxanthin most closely resembles that of phleixanthophyll, isolated from *Mycobacterium phlei* strain Vera. No glycosylated carotenoids have been found or are predicted to be made in any other strain of this genus. In addition to the occasional glycosylated carotenoid isolated from actinomycetes, a search for sioxanthin biosynthetic gene homologues identified other bacteria with a predicted capacity to biosynthesize glycosylated carotenoids. In many cases, the genes responsible for glycosylation were overlooked by earlier bioinformatics predictions[9]. An analysis of patterns in the ecological niche of these diverse bacteria could aid in the prediction of a biological role for the addition of a glucose functional group, as well as other structural features, of carotenoid molecules.

It should be noted that it is difficult to predict carotenoid production without fine-scale analyses of individual genomes due to errors in assigning genes with related functions and overlooking genes necessary to the pathway due to low sequence homologies. This is evident in the early predictions made regarding carotenoid biosynthesis in *Salinispora*, which overlooked the contribution of many of the *terp1*

genes[9] and misannotated the function of the 3',4'-desaturase. Furthermore, as this work has reiterated, there remain carotenoid biosynthetic enzymes whose gene identity is unknown reaffirming the need for more work in understanding carotenoid biosynthesis.

4.2.2 Non-clustered carotenoid biosynthetic genes in bacteria is more common than anticipated

Gene clustering is considered to be the standard configuration of bacterial secondary metabolic genes in the same pathway[1]. Carotenoid biosynthesis falls in to this category to the point that textbooks describing carotenoid biosynthesis state that the pathways are clustered in all bacteria outside of cyanobacteria[3,18]. Three examples are now known in bacteria in which genes for the biosynthesis of a single carotenoid are arranged in subclusters and scattered throughout the genome. These bacteria are from three distantly related phyla, *Gemmatimonas aurantiaca* (Gemmatimonadetes)[6], *Deinococcus radiodurans* (Deinococcus-Thermus)[5], and now *Salinispora tropica* (Actinobacteria). This observation shows that the non-clustered phenomenon is phylogenetically widespread, though the prevalence is unknown.

As published studies of carotenoid biosynthesis have overwhelming uncovered clustered genomes, attention was focused on bacterial genomes in which carotenoid biosynthesis has not been studied. This was accomplished by searching for carotenoid biosynthesis genes in publically available bacterial genome sequences. In most of the cases, a carotenoid compound has not been identified, meaning that all of the genes in the biosynthetic pathway were unknown. Therefore, two carotenoid biosynthesis genes were used as a proxy for the pathway. Genomic locations of phytoene synthase genes, which encode the first dedicated step in carotenoid biosynthesis and are vital to the

production of these compounds, were compared with the locations of lycopene cyclase genes, responsible for the most common modification in carotenoid biosynthesis and a gene that is not clustered in *Salinispora*. Only bacteria with both of these genes were investigated giving a total of 86 pathways from 75 genomes in 5 bacterial classes. The sequences were a mixture of closed and draft genomes, but the genome completion status had no impact on the clustering of genes. The results showed that these two genes were clustered in only about half of these genomes (Figure 4.6). This observation could not be explained by phylogeny, as the pattern held true for individual classes with the exception of the gammaproteobacteria, where only five genomes were investigated (Figure 4.6). Some genome sequences had multiple gene copies. In these cases, there were no patterns to clustering -- two genomes had no clustering in these two genes, one genome (*Streptomyces griseus* subsp *griseus*) had three sets of clustered genes, while the others contained a mix of clustered and non-clustered genes.

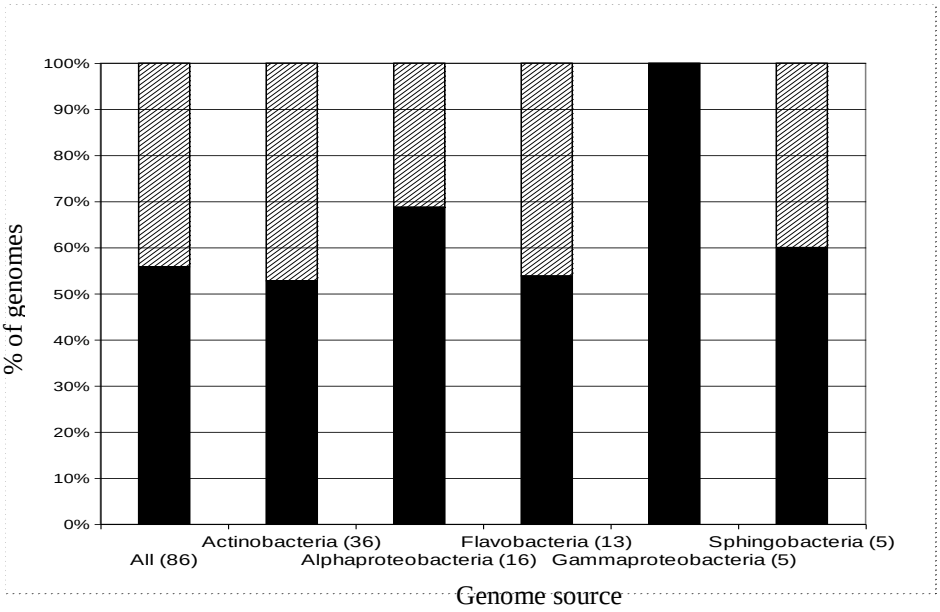


Figure 4.6: Chart of percentage of clustered (solid) and non-clustered (hashed) carotenoid biosynthetic pathways in selected bacterial genomes. Total genomes and individual phylogenetic grouping is shown, total number of genomes in each category is in parenthesis.

The gene clustering pattern could also not be explained by environment, where both clustered and non-clustered arrangements were found in marine, freshwater, soil, and host-association environments as well as in both high and low temperature extremes (Figure 4.7). Though it cannot yet be shown that all of these genes are active in their predicted biosynthetic pathways, these results do suggest that clustering of carotenoid biosynthetic genes is greatly overestimated. The mechanisms for development and maintenance of this genome architecture are not yet known, but may provide insights towards bacterial genome organization and evolution.

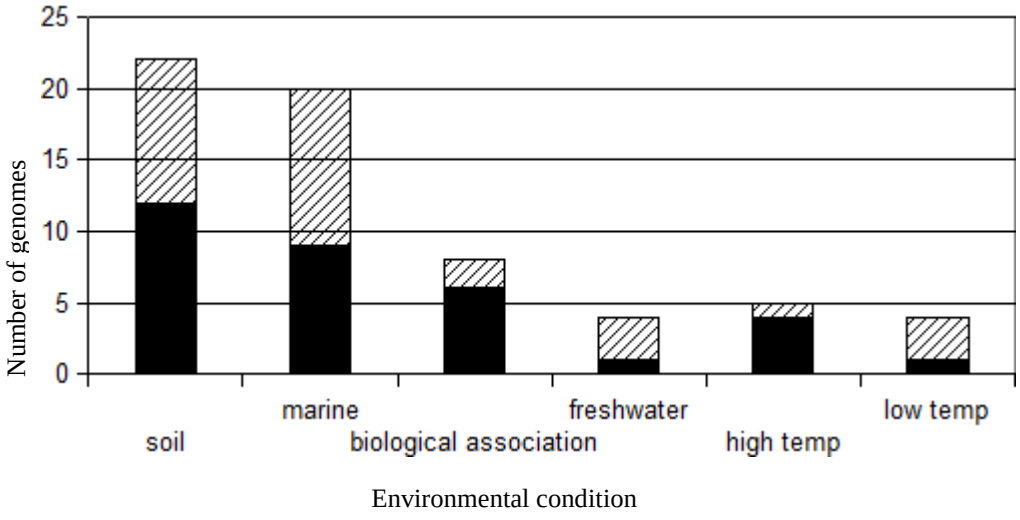


Figure 4.7: Chart of number of genomes with clustered (solid) and non-clustered (hashed) carotenoid biosynthetic genes arranged by environment in which the host organism was isolated.

4.2.3 Patterns of organization within the non-clustered arrangement

The observation that single gene clusters are not required in carotenoid biosynthesis led to questions regarding the patterns of subclustering in these pathways. In sioxanthin biosynthesis, the pathway has a pseudomodular arrangement in which individual genomic regions are responsible for different carotenoid structural features.

Modules refer to a set of standardized, individual units that can be combined to create a more complex structure. In secondary metabolism, modularity is common in polyketide synthases (PKS), in which a relatively small number of enzymes can be combined and rearranged to produce a large chemical diversity[19]. Modularity has also been implicated in carotenoid biosynthesis, as a means of explaining the large chemical diversity and the relatively low number of known biosynthetic genes[20]. It has been shown that exchange and extension of genes in bacterial carotenoid pathways result in the biosynthesis of new compounds[20,21]. This observation suggests that carotenoid biosynthetic pathways have the flexibility to alter their final product in response to changes in their gene composition, and that, like PKS systems, carotenoid structural diversity may arise from the exchange or addition of modules.

In *S. tropica*, the functional role of the carotenoid biosynthetic genes are somewhat correlated with their genomic arrangement. As shown in Scheme 2.1, the genes in *terp2* work together to produce lycopene, which provides the backbone for further modification. The *terp1* genes encode enzymes that modify only the right side of the molecule, contributing the glycosylated functional group as well as the additional desaturation. The left side of the molecule, the result of cyclization and aromatization, are encoded in the non-clustered genes. This pattern is suggestive of horizontally transferred genes in which each gene or cluster of genes were added to the genome separately and were then incorporated in to the biosynthesis.

However, this pseudomodular pattern is not found in the other two known non-clustered pathways (Figure 4.1) from *Deinococcus* and *Gemmatomonas*. In these cases, the genes associated with particular parts of the molecule are not co-clustered. Rather

they are mixed together in small clusters that are dispersed throughout the genome. More work would need to be done to understand the mechanisms driving the genomic architecture in these cases and the impacts on regulation. It is worth pointing out that these analyses of genome architecture fail to account for the three-dimensional confirmation that a bacterial chromosome occupies in nature. Work in *E. coli* transcriptomes has demonstrated a periodicity in the bacterium's genome showing that co-transcribed genes occupy neighboring regions in a folded genome, even though they do not physically reside adjacent to one another[22]. Finding periodicity in these bacteria with non-clustered carotenoid pathways would require extensive transcriptome analyses and understanding of primary metabolism.

4.2.4 Insights in to the development of the *terp1* gene cluster

Patterns in the *Salinispora* gene clustering were further investigated through MultiGeneBlast analysis of the *terp1* cluster. This cluster is particularly interesting because it includes genes that are not involved in sioxanthin biosynthesis, despite the fact that some have a predicted role in carotenogenesis. It seemed possible, then, that this cluster has an alternative function in another bacterium, in which all of the genes are required. To investigate this idea, the *terp1* sequence was used as a query in a MultiGeneBlast against the NCBI bacterial genome database. The results show that the *terp1* cluster is not found outside of the Micromonosporaceae. However, the constituent genes are found dispersed throughout another gene cluster with unknown function in several *Frankia* genomes (Figure 4.8).

The MultiGeneBlast results do suggest that there are common groupings of genes found within the *terp1* cluster. Homologs of *strop3246 (cruF), strop3247 (cruC)*

and *strop3248 (crtD)* are commonly found together as part of other carotenoid biosynthesis pathways (Figure 4.8). In other genomes, homologs of *strop3245* (acetyltransferase), *strop3250* (tetrahydrofolate reductase) and *strop3251* (putative *crtE*) are found together without associations with other carotenoid biosynthetic genes (Figure 4.8). In fact, in some genomes a variety of additional genes are found between the acetyltransferase and the tetrahydrofolate reductase, suggesting that this region is a common location for the transfer of genes (Figure 4.8). It is likely, then, that the *terp1* cluster in *Salinispora* is, in fact, a composite of multiple gene transfers in this genomic region. This is most apparent in the *Actinoplanes*, which contains the complete *terp1* cluster, in addition of a succinate metabolism gene cluster which has been inserted upstream of the tetrahydrofolate reductase (Figure 4.8). The hypothesis that this gene cluster is a composite of multiple sub-clusters is further supported by the differing gene trees of *strop3245* and *strop3247* (Figure 4.17 and Figure 4.18).

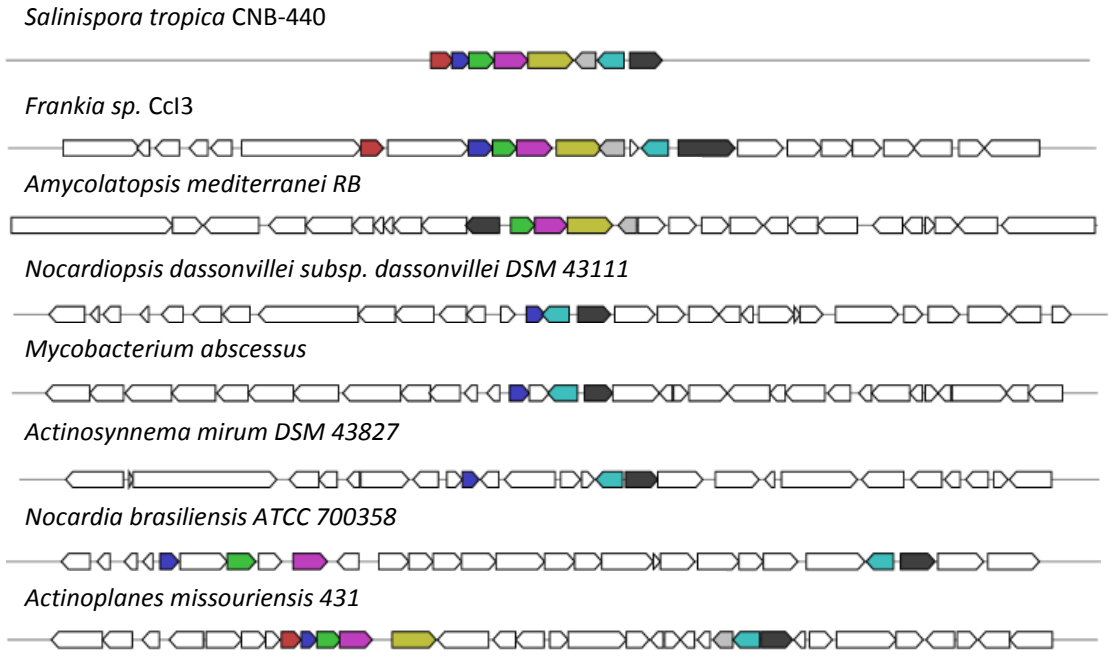


Figure 4.8: Results of the MultiGene BLAST of the *S. tropica terp1* cluster.

4.2.5 Evolutionary history of each region of the sioxanthin biosynthetic pathway

One explanation for the distributed genomic organization of the sioxanthin biosynthetic pathway is that each portion transferred to the genome from different places and/or at different times in *Salinispora* evolution. Character trees showing the presence and absence of sioxanthin gene homologs were analyzed to show the gene traits of ancestral species. Two genes from each cluster (*strop4440* and *strop4441* from *terp2* and *strop3245* and *strop3247* from *terp1*) as well as each of the non-clustered genes were used as character traits. The trees show that these genes do have different distributions among the actinomycetes and that there is a differential appearance of these genes in *Salinispora* ancestors. This relationship suggests complexity among carotenoid biosynthesis in actinomycetes and, particularly, the development of the sioxanthin biosynthetic pathway that cannot be explained purely by vertical

transmission of genes. The non-clustered genes *crtU* (*strop0241* homologs) and *crtY* (*strop2408* homologs) have different distributions throughout actinomycetes, with *crtU* being much more prevalent (Figure 4.9 and Figure 4.10). Lycopene cyclases appear to be rare in actinomycetes, but several types of lycopene cyclase are known and this analysis covers only the *crtY*-type, the homolog found in the sioxanthin pathway. In both cases, the parsimony suggests that the distribution of these genes is due to the acquisition of genes at different points in actinomycete evolution. This includes a single instance of gene gain within the Micromonosporaceae. Neither *crtU* nor *crtY* is universal within the Micromonosporaceae, as the genera *Longispora* and *Catelliglobospora* are both lacking these genes. Furthermore, there appear to be instances of gene loss of the *crtY* homolog within the *Actinoplanes*, as there is a mixture of species that do and do not have this homolog.

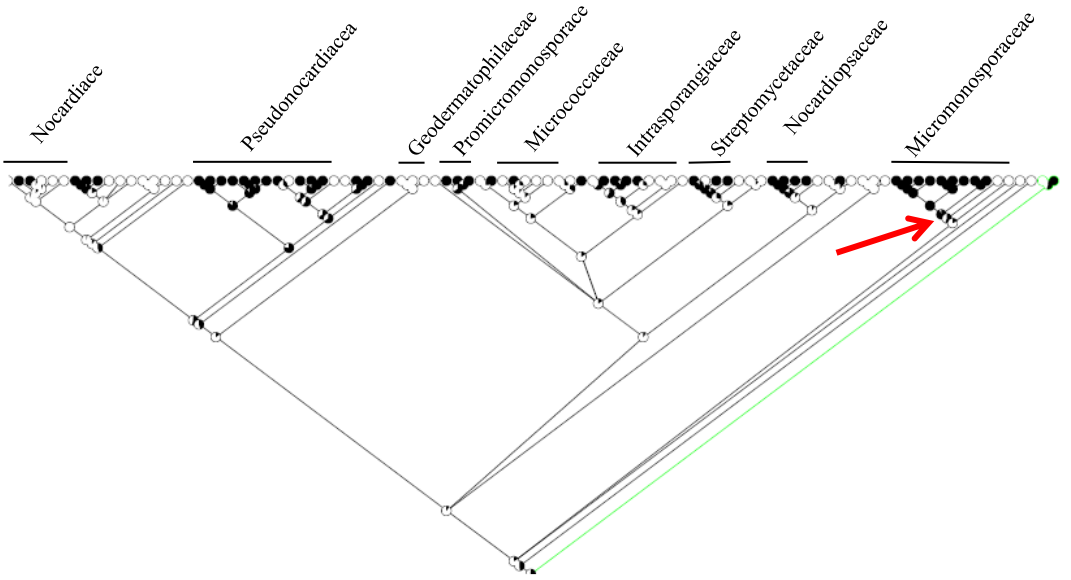


Figure 4.9: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of *crtU* homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

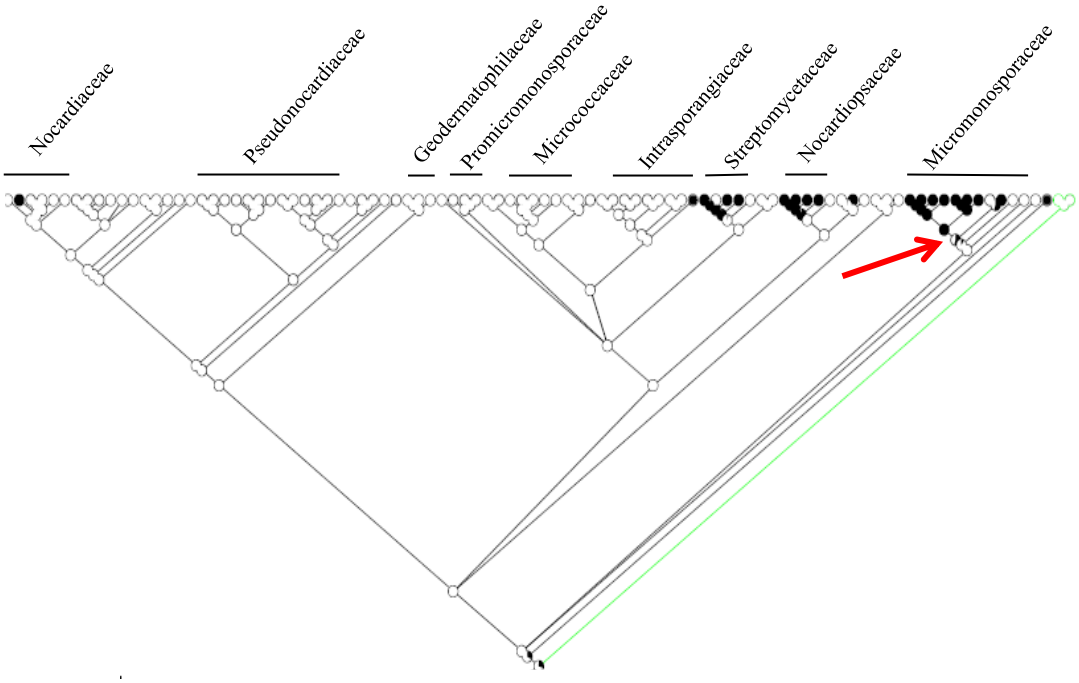


Figure 4.10: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of *crtY* homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

Genes from the *terp1* cluster also showed different distributions among the actinomycetes. Though both *strop3245* (an acetyltransferase) and *strop3247* (*cruC*) were predicted to be part of the same cluster, only *strop3247* is active in the sioxanthin biosynthetic pathway, serving as the glycosyl transferase. The acetyltransferase gene was likely present in an early ancestor of the actinomycete order and experienced loss from a few families over time (Figure 4.11). The glycosyl transferase, on the other hand, was acquired late in the development of a few families, including within the Micromonosporaceae (Figure 4.12). The differing distribution of these two genes is

evident of independent evolutionary histories. This is particularly interesting considering that these genes are co- clustered.

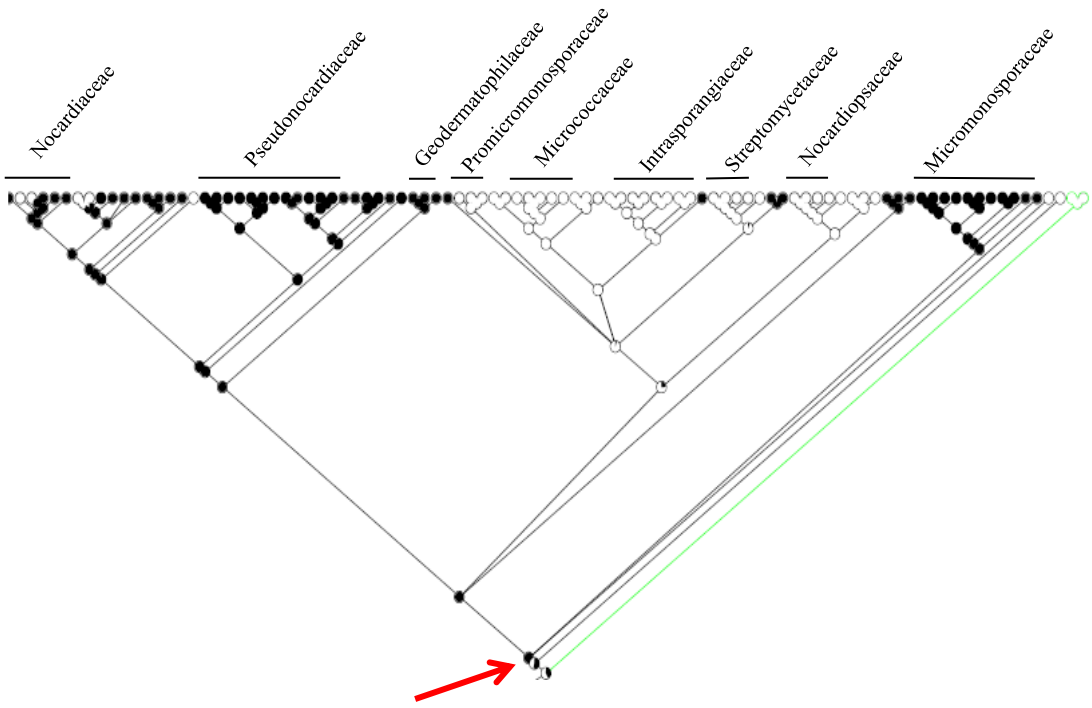


Figure 4.11: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of acetyltransferase homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

Figure 4.12: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of *cruC* homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

The two genes from the *terp1* cluster, unsurprisingly, have a similar phylogentic distribution and history. The *crtE* (*strop4440*) gene encodes a polyprenyl synthetase. This gene is vital to all carotenoid biosynthetic pathways and is closely related to enzymes that are involved in other isoprenoid biosynthetic pathways, making it widely distributed. Polyprenyl synthetase genes are clearly common and deeply rooted within the actinomycetes (Figure 4.13). Similarly, *crtB* (*strop4441*) encodes a phytoene synthase, another necessary step in carotenoid biosynthesis. Homologs of this gene are expected in every carotenoid-containing bacterium, thus explaining its wide distribution and its long history within the actinomycetes (Figure 4.14). Some families, however,

appear to be missing this gene. The Micrococcaceae, for example, are known to produce carotenoids but are missing this gene homolog, suggesting that they have a more divergent enzyme providing a similar function[23].

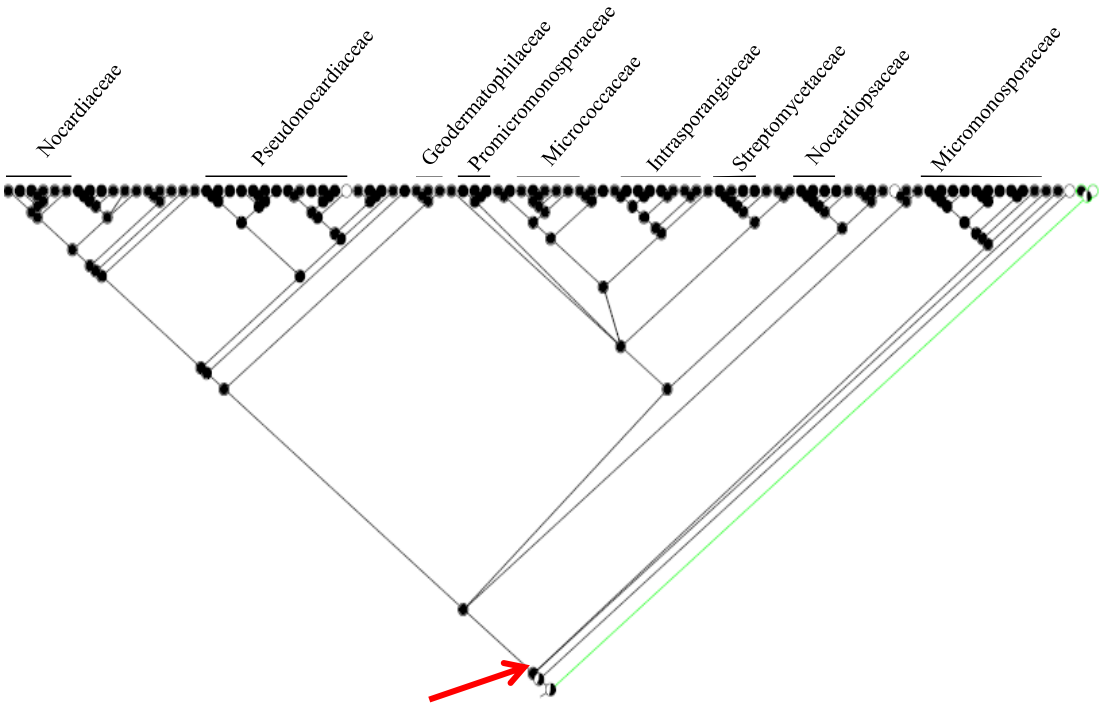


Figure 4.13: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of *crtE* homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

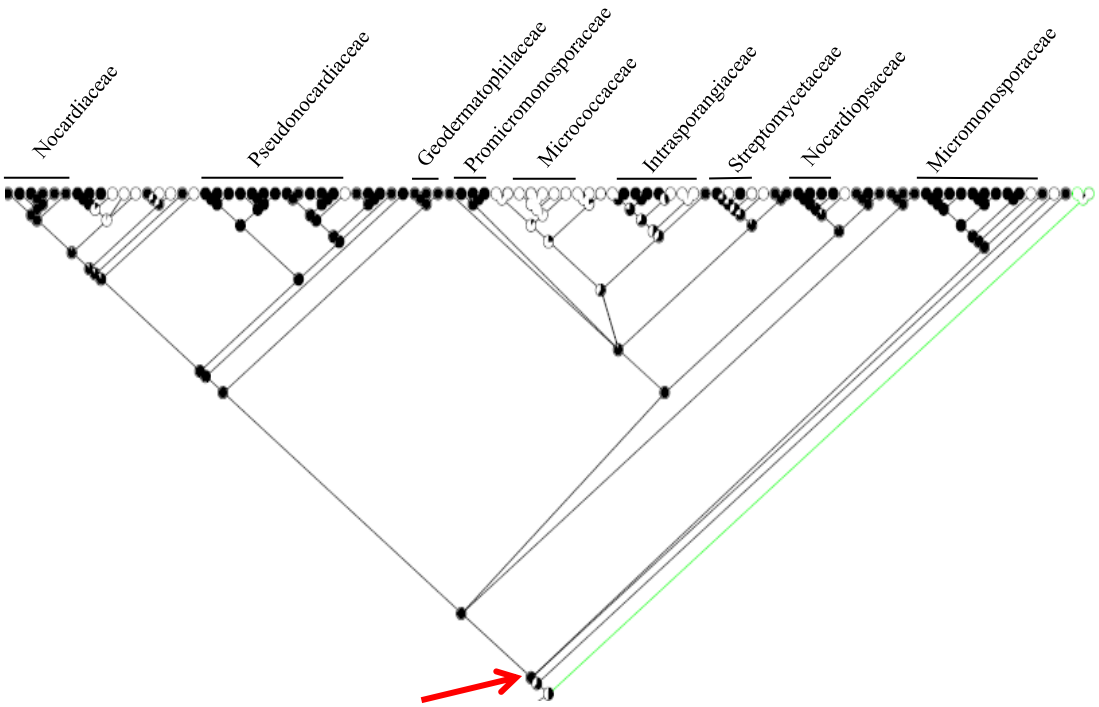


Figure 4.14: Character tree (maximum-likelihood using *rpoB* sequences) showing the distribution of *crtB* homologs among the actinomycetes as well as the predicted traits of ancestral species. A cyanobacterial outgroup is shown in green. Presence of a gene is indicated by a black circle and likelihood of presence in an ancestral species is shown in the proportion of black in an ancestral node. Red arrow indicates ancestor that likely acquired the gene.

To further investigate the origins of these genes, the phylogenies of the individual genes were compared to the species trees in which the genes are found, built from the *rpoB* sequences. Gene trees built from *rpoB* sequences alone are imperfect measures of phylogeny, making these trees most useful for analyses at the family level. *rpoB* was chosen over 16S because of its greater phylogenetic reliability in closely related species[24], such as actinomycetes. All of the gene trees diverge from the species tree, though they do so to varying degrees. The *crtU* homolog shared by the Micromonosporaceae and other members of the actinomycetes (*Pseudonocardia* and *Actinomycetospora*) is more closely related to homologs found in the cyanobacteria and proteobacteria, a large divergence from the phylogenetic relationship of these groups in

the species tree (Figure 4.15). This divergence further supports the idea that this gene was horizontally transferred, likely by a member of the cyanobacteria or proteobacteria. Similarly, the *crtY* tree shows a great deal of divergence from the species tree, with a great deal of exchange between and within bacterial families (Figure 4.16). In this case, the *crtY* gene is shared among the actinobacteria, though the gene tree suggests that this gene originated in the Bacteriodetes (such as *Pedobacter*) and was then shared throughout actinomycetes. The acetyltransferase gene (*strop3245*) shows a great deal of exchange among bacteria in general and there have been several points of acquisition within actinobacteria. Within Micromonosporaceae and related actinomycetes, there is little variation in this gene, though the tree does suggest some early exchange with a member of the proteobacteria, and later with cyanobacteria (Figure 4.17). Glycosyl transferases in *Salinispora* are most closely related to other actinobacterial homologs, however, these are most closely related to the Chloroflexi (green non-sulfur bacteria) than vertical transmission would predict (Figure 4.18). Perhaps unsurprisingly, the gene trees of those from the *terp2* gene cluster diverge the least from their species trees (Figure 4.19, Figure 4.20). They are not, however, identical and it appears that most of the divergence happens within the actinomycetes themselves. It is interesting to note that genes that show the lowest incidence of horizontal gene exchange (*strop4440* and *strop4441* in *terp2*) are the only ones that are located within a genomic island[7].

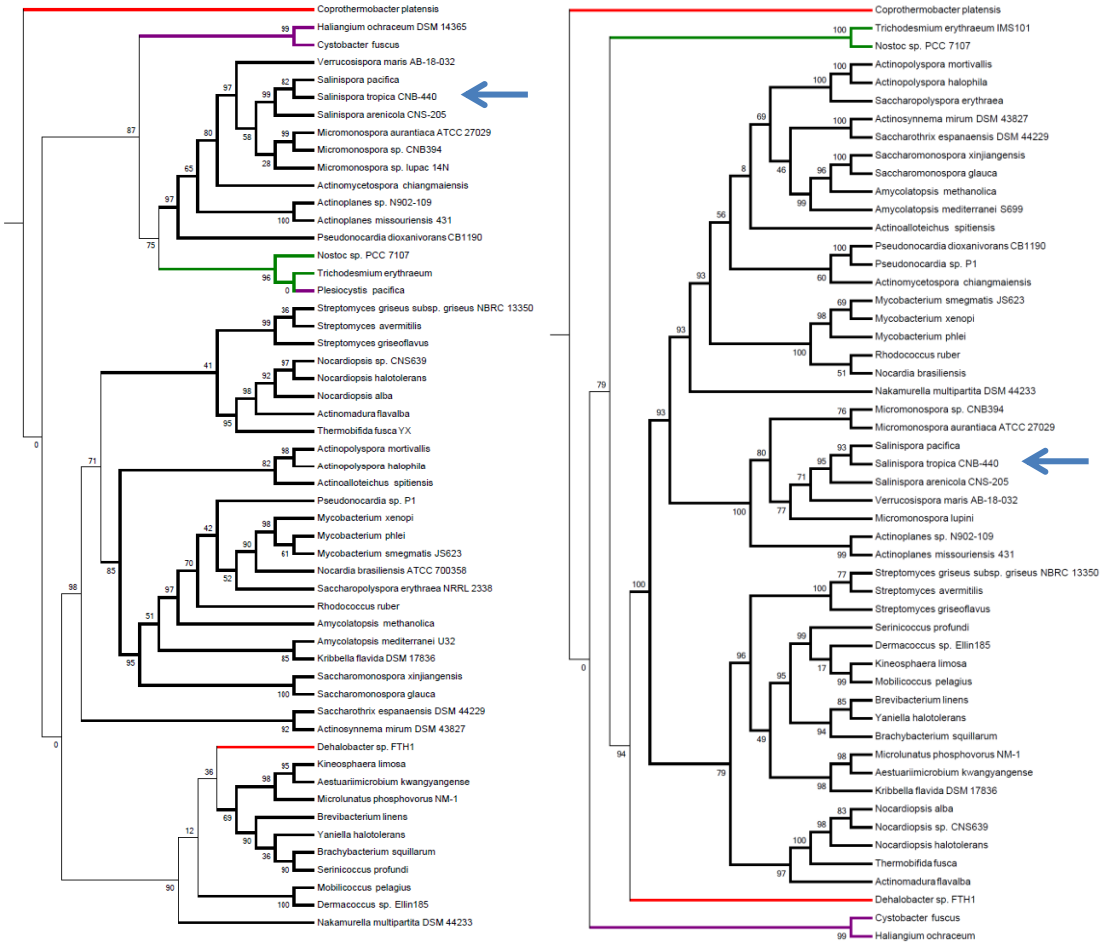


Figure 4.15: Comparison of the *crtU* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica.* Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

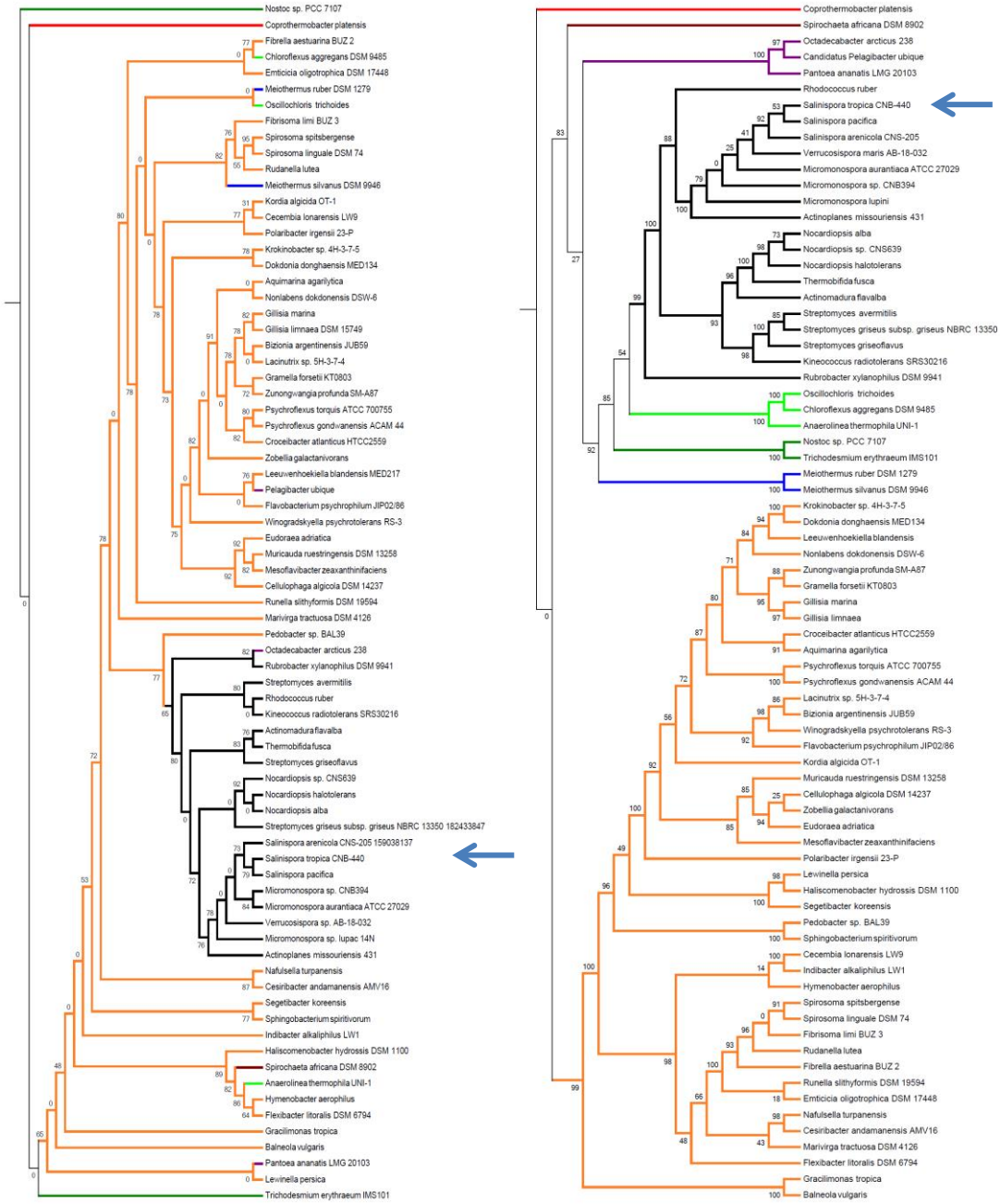


Figure 4.16: Comparison of the *crtY* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica*. Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

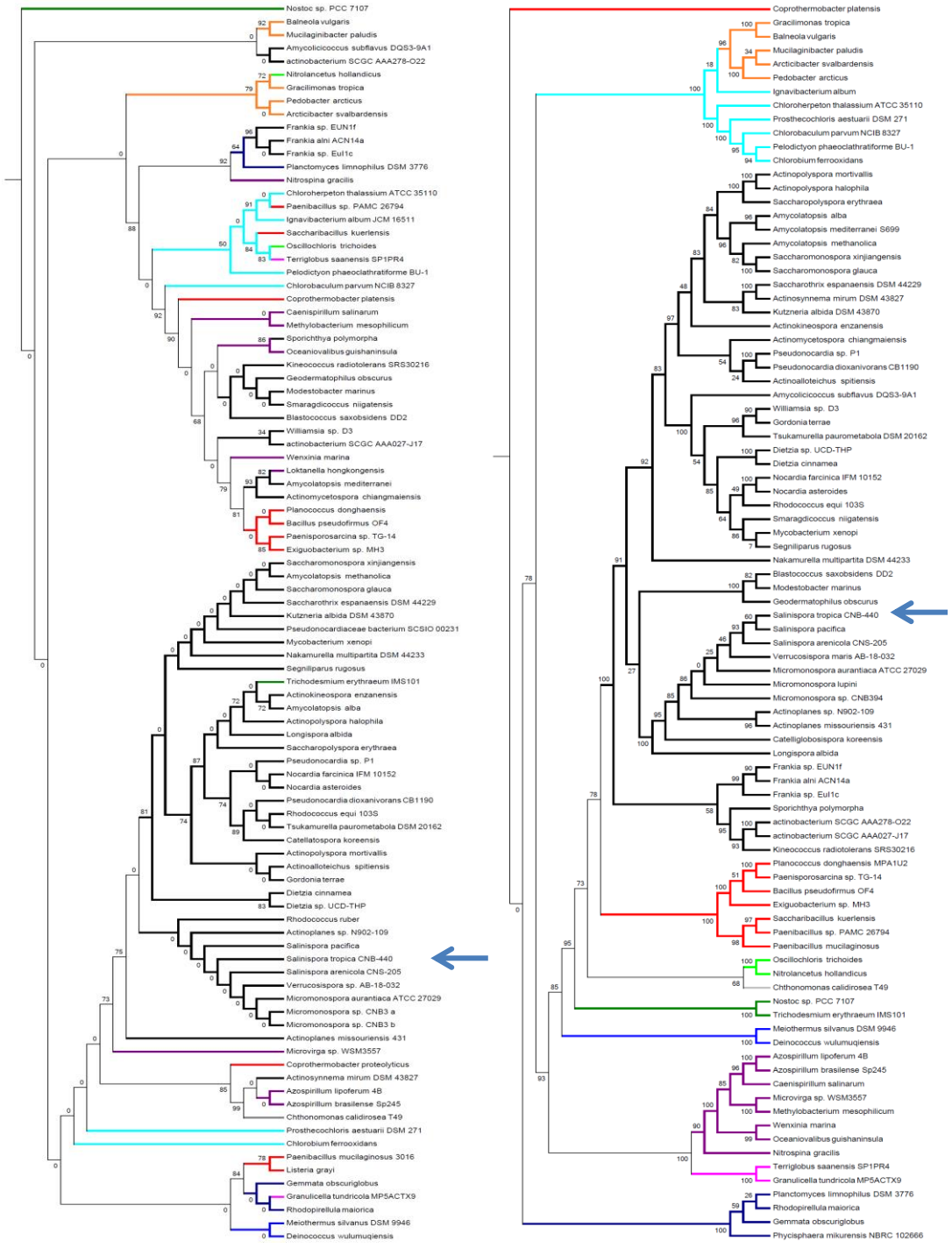


Figure 4.17: Comparison of the *strop3245* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica.* Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

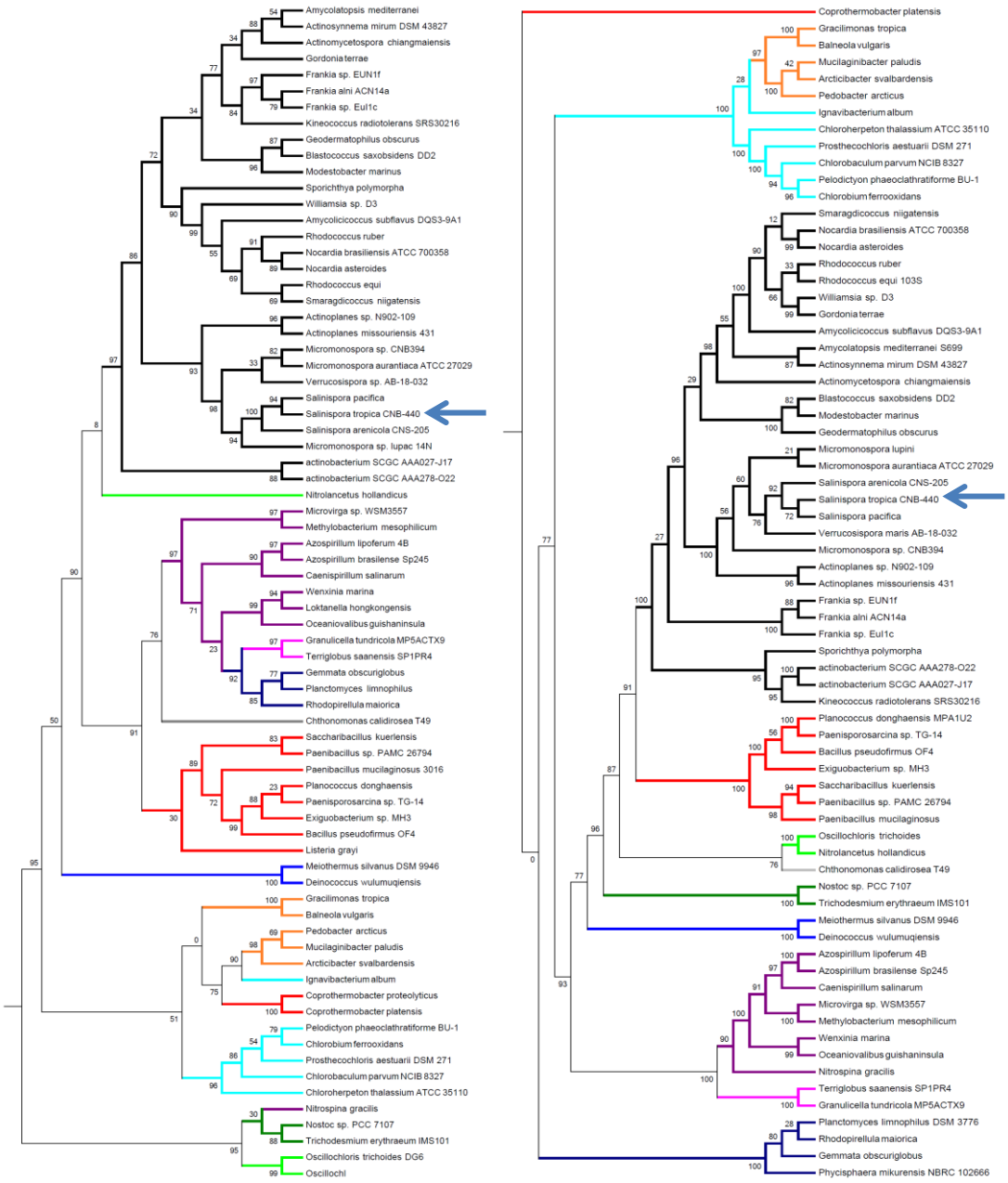


Figure 4.18: Comparison of the *cruC* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica.* Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

Figure 4.19: Comparison of the *crtE* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica*. Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

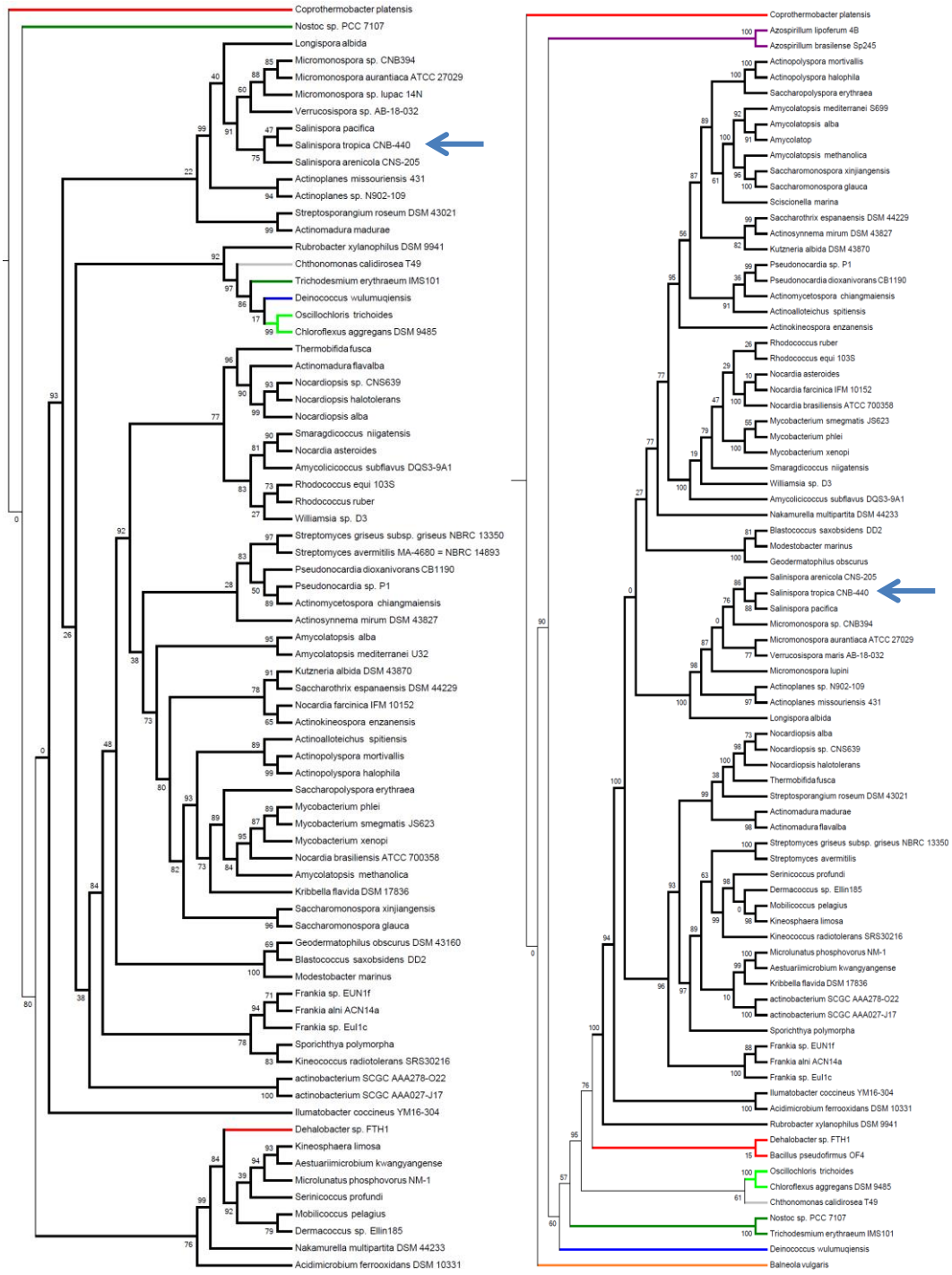


Figure 4.20: Comparison of the *crtB* maximum-likelihood gene tree (left) to the corresponding *rpoB* tree (right). Blue arrows show the location of *Salinispora tropica.* Colors on the tree correspond to bacterial phyla and are as follows: Firmicutes in red, Deinococcus-Thermus in blue, Actinobacteria in black, Chloroflexi in green, Armatimonadetes in gray, Cyanobacteria in dark green, Spirochaetes in brown, Proteobacteria in purple, Acidobacteria in pink, Bacteriodetes in orange, Chlorobi in light blue, and Planctomyces in dark blue.

Comparisons of the gene trees to species trees confirm what is seen in the ancestral traits regarding the evolutionary history of the sioxanthin biosynthetic pathway. The genes responsible for carotenoid biosynthesis are common in actinomycetes and were introduced early in their evolutionary history. The genes required for modifications, responsible for the structural diversity among carotenoids, arrived in the genome much later, following the divergence of the Micromonosporaceae. Furthermore, the *crtU*, *crtY*, and *cruC* genes all have different phylogenetic histories that suggest horizontal gene transfer from different hosts. This utilization of genes from different sources is likely what contributed to the unusual biosynthetic pathway and unique chemical structure demonstrated by sioxanthin.

The sioxanthin genes were not listed in the Darkhorse database as being the product of horizontal gene transfer[25]. However, this could be a result of the fact that Darkhorse does not predict horizontally transferred genes more ancient than the genus level. Several genes near the lycopene cyclase (*strop2408*) do appear in the Darkhorse database showing horizontal gene transfer from several bacterial phyla (Table 4.1), implying that it is a region of frequent gene exchange.

Table 4.1: Darkhorse database matches to predicted horizontally transferred genes near sioxanthin biosynthetic genes.

Salinispora gene locus	Gene function	Best match genus	Best match phylum
Strop_2401	Hypothetical	*Opitutus*	Verrucomicrobia
Strop_2402	ABC transferase	*Oceanolica*	Proteobacteria
Strop_2403	Hypothetical	*Nostoc*	Cyanobacteria
Strop_2404	Membrane transporter	*Roseoflexus*	Chloroflexi
Strop_2415	Hypothetical	*Anabaena*	Cyanobacteria

The evolution of the sioxanthin biosynthetic pathway can be summarized as shown in Figure 4.21. Three of the genes investigated (*crtE* and *crtB* of *terp2* and the acetyltransferase of *terp1*) have a long history within this group of bacteria and were vertically inherited from a common ancestor of the Order Actinomycetales. The other three genes (*cruC* of *terp1* and the unclustered *crtU* and *crtY*) are much more recent additions, having been acquired after the split of the Micromonosporaceae from other bacterial families, but before the divergence of the *Salinispora*, *Micromonospora*, *Verrucosispora*, and *Actinoplanes* genera. This acquisition was then likely followed by gene loss of *crtY* homologs in several of the *Actinoplanes* species.

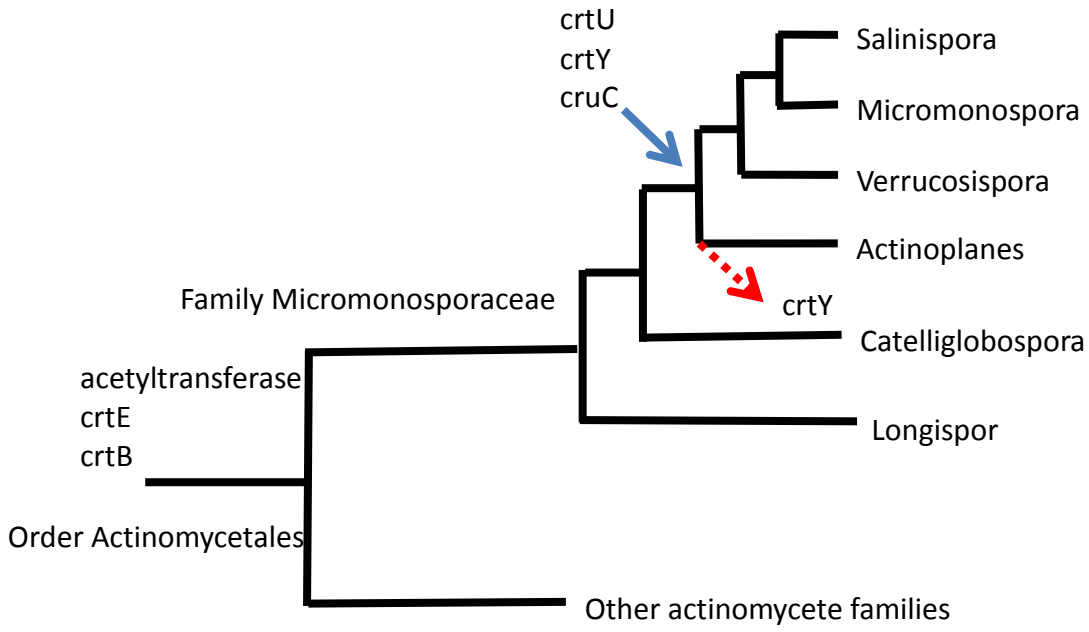


Figure 4.21: Tree showing relative relationship of the Micromonosporaceae genera and other actinomycetes, highlighting the time of entry of sioxanthin genes in to the genomes. Three genes were present prior to the subdivision of the order actinomycetales. Three others entered later (blue arrow) followed by some gene loss (red arrow).

4.3 Conclusions

Gene clustering of carotenoid biosynthetic genes is believed to be the standard arrangement in bacterial genomes. However, despite the abundance of studies that reveal this architecture, a search of bacterial genome databases show that sub-clustering or non-clustering of genes may be just as common. The sub-clustering of genes in a secondary metabolic pathway may have implications on the ability to identify chemical compounds associated with orphan pathways, as it may lead to overlooking important parts of a pathway. This non-clustered arrangement may be the result of differential horizontal gene transfer in which partial pathways are exchanged, potentially leading to novel compounds. Sioxanthin biosynthesis is an example of this phenomenon. Not unique to *Salinispora*, the sioxanthin pathway was found in other closely-related genera of the bacterial family Micromonosporaceae. Actinomycetes have long had the capacity to produce carotenoids, but members of the Micromonosporaceae likely obtained additional genes that enabled a specific suite of carotenoid modifications. Sioxanthin biosynthesis, then, may be the result of the acquisition of genes from three separate sources that altered the carotenoid pathway already present.

4.4 Materials and Methods

4.4.1 Analysis of sioxanthin biosynthesis genes in the Micromonosporaceae and other actinobacteria

Gene neighborhood and syntenty analyses in the Micromonosporaceae were performed using Integrated Microbial Genomes Expert Review (img.jgi.doe.gov) where the newly sequenced actionmycete genomes were deposited. Excluding the *Salinispora* there were 15 genomes from the Micromonosporaceae family. Genes were found by running a BLAST analysis of the sioxanthin genes against the Micromonosporaceae

genomes. The BLAST hits were then used to explore the gene neighborhoods with the same top COG (Cluster of Orthologous Genes) hit.

The potential for sioxanthin production in other actinomycetes was investigated by searching for the required genes within actinomycete genomes. The 16S rRNA gene sequences were gathered from 200 actinomycete species and used to build a neighbor-joining tree in the program Geneious version 5.1.7 (Biomatters, www.geneious.com). The tree gives relative relationships of actinomycete families. BLASTx searches were performed to identify the sioxanthin biosynthetic gene homologs in each genome. Protein sequences with a greater than 35% identity over 75% of the length of the query sequence were considered homologous.

4.4.2 Carotenoid biosynthesis gene cluster analysis in sequenced bacteria

The *Salinispora tropica* lycopene cyclase protein sequence (strop_2408) was used as a BLASTp query sequence against the nr database to find bacterial genomes with crtY-type lycopene cyclase genes. The genome and gene locus were noted for each hit. The genomes were then reanalyzed with BLAST to using strop_4441 as a query sequence to find the locations of the phytoene synthase genes. Genes were determined to be clustered if they were in close proximity to each other and were separated by no more than ten genes unrelated to carotenoid biosynthesis genes. Graphs were made in Excel to show the proportion of pathways in each category with a particular genome architecture. The NCBI database provided phylogenetic information for the bacterial genomes, and literature searches were done to identify the natural environment.

4.4.3 Gene phylogenies and ancestral state

Gene trees were developed for six genes (*strop0241, strop2408, strop3245, strop3247, strop4440* and *strop4441*) in the sioxanthin pathway. Homologues were gathered through BLASTp searches first on the Phylogeny.fr website (www.phylogeny.fr)[26]. A second round of BLASTp searches were performed through NCBI which excluded actinomycete genomes from the search, allowing for a greater phylogenetic coverage. Gene sequences were determined to be homologs when they had a sequence identity of greater than 35% over 50% of the sequence length and an e-value less than 1^-20. Gene trees were pruned to reduce the number of species represented in each genus that showed phylogenetic similarity. Species trees were built using *rpoB* sequences of the genomes represented in the gene trees. All sequence alignment and tree building were done on the phylogeny.fr website (www.phylogeny.fr)[26]. Sequences were aligned using MUSCLE [27]. Alignment curation was done by Gblocks[28] allowing for smaller final blocks, gap positions within the final blocks, and less strict flanking positions. Maximum-likelihood trees were built using PhyML 3.0 with an SH-like Approximate Likelihood-Ratio Test[29]. Trees were visualized and edited in MEGA 5.2[30].

The species tree in the ancestral state was built by combining *rpoB* sequences of all of the actinomycetes in the individual species trees. Cyanobacterial sequences were used as an outgroup. The ancestral node was inferred using the trace character history function implemented in Mesquite v2.75[31]. A character matrix was created for each gene homolog and likelihood calculations were performed using an Mk1 model. Likelihood scores of greater than 50% on the ancestral nodes were used to infer the points of gene acquisition.

4.4.4 MultiGene BLAST analysis of *terp1* cluster

The MultiGene BLAST program was obtained from the developer's website (http://multigeneblast.sourceforge.net/). Query sequences were Genbank files created on the NCBI website (http://www.ncbi.nlm.nih.gov/) using *S. tropica* CNB-440 genome region 3720230-3727993 for *terp1* and 5031196-5036389 for *terp2* query sequences. The query sequences were searched against the bacterial genome database, downloaded from the NCBI website.

4.5 References

1 Ballouz, S., Francis, A. R., Lan, R. & Tanaka, M. M. Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* **6**, (2010).

2 O'Brien, J. & Wright, G. D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **22**, 552-558, (2011).

3 Armstrong, G. A. & Hearst, J. E. Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. *FASEB J.* **10**, 228-237, (1996).

4 Lang, H. P., Cogdell, R. J., Takaichi, S. & Hunter, C. N. Complete DNA sequence, specific Tn5 insertion map, and gene assignment of the carotenoid biosynthesis pathway of Rhodobacter sphaeroides. *J. Bacteriol.* **177**, 2064-2073, (1995).

5 Tian, B. & Hua, Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol.* **18**, 512-520, (2010).

6 Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2 '-dirhamnoside. *Microbiol.-Sgm* **156**, 757-763, (2010).

7 Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S., Moore, B. S. & Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J.* **3**, 1193-1203, (2009).

8 Ziemert, N., Lechner, A., Wietz, M., Millan-Aguinaga, N., Chavarria, K. L. & Jensen, P. R. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. U. S. A.*, (2014).

9 Klassen, J. L. Phylogenetic and evolutionary patterns in microbial carotenoid biosynthesis are revealed by comparative genomics. *PLoS One* **5**, e11257, (2010).

10 Klassen, J. L. Pathway evolution by horizontal transfer and positive selection is accommodated by relaxed negative selection upon upstream pathway genes in purple bacterial carotenoid biosynthesis. *J. Bacteriol.* **191**, 7500-7508, (2009).

11 Phadwal, K. Carotenoid biosynthetic pathway: molecular phylogenies and evolutionary behavior of crt genes in eubacteria. *Gene* **345**, 35-43, (2005).

12 Cobbs, C., Heath, J., Stireman, J. O., 3rd & Abbot, P. Carotenoids in unexpected places: gall midges, lateral gene transfer, and carotenoid biosynthesis in animals. *Mol. Phylogenet. Evol.* **68**, 221-228, (2013).

13 Moran, N. A. & Jarvik, T. Lateral Transfer of Genes from Fungi Underlies Carotenoid Production in Aphids. *Science* **328**, 624-627, (2010).

14 Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K. & Nagashima, K. Horizontal Transfer of the Photosynthesis Gene Cluster and Operon Rearrangement in Purple Bacteria. *J. Mol. Evol.* **52**, 333-341, (2001).

15 Giraud, E., Hannibal, L., Fardoux, J., Jaubert, M., Jourand, P., Dreyfus, B., Sturgis, J. N. & Vermeglio, A. Two distinct crt gene clusters for two different functional classes of carotenoid in Bradyrhizobium. *J. Biol. Chem.* **279**, 15076-15083, (2004).

16 Takaichi, S., Maoka, T., Akimoto, N., Carmona, M. L. & Yamaoka, Y. Carotenoids in a Corynebacterineae, *Gordonia terrae* AIST-1: Carotenoid glucosyl mycoloyl esters. *Biosci. Biotech. Bioch.* **72**, 2615-2622, (2008).

17 Britton, G., Liaaen-Jensen, S. & Pfander, H. *Carotenoids handbook*. (Birkhäuser Verlag, 2004).

18 Armstrong, G. A. Genetics of eubacterial carotenoid biosynthesis: a colorful tale. *Annu. Rev. Microbiol.* **51,** 629-659 (1997).

19 Fischbach, M. A. & Walsh, C. T. Assembly-Line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanisms. *Chem. Rev.* **106**, 3468-3496, (2006).

20 Garcia-Asua, G., Lang, H. P., Cogdell, R. J. & Hunter, C. N. Carotenoid diversity: a modular role for the phytoene desaturase step. *Trends Plant Sci.* **3**, 445-449, (1998).

21 Kim, S. H., Park, Y. H., Schmidt-Dannert, C. & Lee, P. C. Redesign, Reconstruction, and Directed Extension of the Brevibacterium linens C-40 Carotenoid Pathway in Escherichia coli. *Appl. Environ. Microbiol.* **76**, 5199-5206, (2010).

22 Wright, M. A., Kharchenko, P., Church, G. M. & Segrè, D. Chromosomal periodicity of evolutionarily conserved gene pairs. *P Natl Acad Sci USA* **104**, 10559-10564, (2007).

23 Pezzoni, M., Costa, C. S., Pizarro, R. A. & Oppezzo, O. J. The relationship between carotenoids and sunlight response in members of the family Micrococcaceae. *J. Basic Microbiol.* **51**, 325-329, (2011).

24 Case, R. J., Boucher, Y., Dahllof, I., Holmstrom, C., Doolittle, W. F. & Kjelleberg, S. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl. Environ. Microbiol.* **73**, 278-288, (2007).

25 Podell, S., Gaasterland, T. & Allen, E. E. A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics* **9**, (2008).

26 Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J. F., Guindon, S., Lefort, V., Lescot, M., Claverie, J. M. & Gascuel, O. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465-469, (2008).

27 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797, (2004).

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

29 Guindon, S., Dugayard, J. F., Lefort, V., Anisimova, M., Hordijik, W. & Gascuel, O. New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321, (2010).

30 Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. & Kumar, S.. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731-2739, (2011).

31 Mesquite: a modular system for evolutionary biology analysis v. 2.75 (http://mesquiteproject.org, 2011).

5 Conclusions and future directions

5.1 Conclusions

Natural products provide a rich source of chemical diversity with potential applications for human health and biotechnology. Perhaps equally important is the knowledge that secondary metabolites provide regarding evolution and ecology of the producer organism. In addition to the potential medical applications, analyses of natural chemical structures are vital to understanding how an organism adapts to and interacts with its surrounding environment. The genus *Salinispora* has been a prolific source of secondary metabolites, thus providing new tools to combat disease as well as novel insights in the biosynthesis, evolution, biological function, and biogeography of these compounds[1-7].

This dissertation describes the identification and biosynthesis of a carotenoid natural product responsible for the characteristic orange pigmentation of the genus *Salinispora*. Using a variety of purification methods and comprehensive NMR analyses, Chapter 3 discusses the complete structure elucidation revealed a novel carotenoid molecule. The structure of this novel compound, now referred to as "sioxanthin", unmasked some surprising structural features, such as an aromatic ring and an attached glucose residue.

The biosynthesis of this compound is similarly interesting, employing several distant regions of the genome to produce a single compound. This is divergent from what is commonly believed to be a standard arrangement for microbial carotenoid biosynthetic genes, and indeed most secondary metabolites in general, in that biosynthetic genes cluster together in a single genome neighborhood[8-11]. In Chapter 2,

carotenogenesis homologous genes were identified in four distinct genomic regions by sequence homology to characterized genes. It was confirmed that all four regions of the genome are required for sioxanthin biosynthesis through a series of gene inactivation experiments and chemical analyses of the mutant bacterial chemotypes. Non-clustered carotenoid clusters are poorly described in literature. Thus, *Salinipsora* provided a special opportunity to apply genetic techniques to the study of this type of biochemical pathway. The exploration of this pathway uncovered previously overlooked genes in carotenoid biosynthesis in actinomycetes and may allow for improved structural prediction in other species.

This project also provided an opportunity to explore the evolution of *Salinispora* carotenoid biosynthesis through phylogenetic comparisons of the disjointed regions of the pathway. Chapter 4, discusses the biosynthetic pathway for sioxanthin and its likely evolution via the transfer of carotenoid biosynthetic genes from multiple sources at different times in *Salinispora* evolution. This type of partial exchange in modular pathways may be a driving force in the expansion of chemical diversity in secondary metabolism.

5.2 Future directions

5.2.1 Biological function of sioxanthin

It is well known that the biological function of a chemical compound is dependent on its structure[11,12]. Much work has been done on the structure-function relationships in carotenoids[12-16]. Many of these studies have focused on the role of the carotenoid, such as providing protection against oxidative and UV damage to non-photosynthetic organisms[17-22]. Similar functions were explored in *Salinispora tropica*,

in which wild-type and pigmentless mutants were exposed to UV irradiation and hydrogen peroxide. UV exposure, in both liquid and solid media, showed no apparent impact on the growth of either the wild-type or the mutant strains. Thus, it seems likely that protection against UV damage is not the primary function of carotenoids in *Salinispora*. Similarly, as *Salinispora* is a primarily sediment-dwelling bacterial genus, it is unknown how much exposure to UV it receives in the natural environment[23,24].

Exposure of the wild-type and mutant strains to hydrogen peroxide did show a differential response. Sterile filters with 0.1% and 0.2% hydrogen peroxide in water (sterile water was used as a control) were placed on agar plates of either the wild-type or pigmentless mutant bacteria. Pigmented *S. tropica* producing sioxanthin had a much smaller zone of inhibition (Figure 5.1) suggesting that sioxanthin might serve a protective role against oxidative stress. There are several known sources of oxidative stress in nature. In addition to environmental sources, natural products produced by neighboring microbes or by *Salinispora* itself may result in oxidative stress. It has been shown that antibiotics induce a generalized oxidative stress in the target bacterium, in addition to the more targeted response[25,26]. Moreover, carotenoid production is itself sometimes correlated with biosynthesis of other bioactive compounds[27]. This led to the hypothesis that via sioxanthin, *Salinispora* may be protecting itself from its own secondary metabolites. However, there is no apparent change in growth between the wild-type and mutant species. Analyses of changes in secondary metabolic profiles with the loss of sioxanthin production did not correlate to the decreased production of any other metabolites. A more in depth analysis may reveal a link between sioxanthin production and the biosynthesis of any other secondary metabolites, but it is likely that

protection from other *Salinispora*-produced secondary metabolites is not the primary function of sioxanthin.

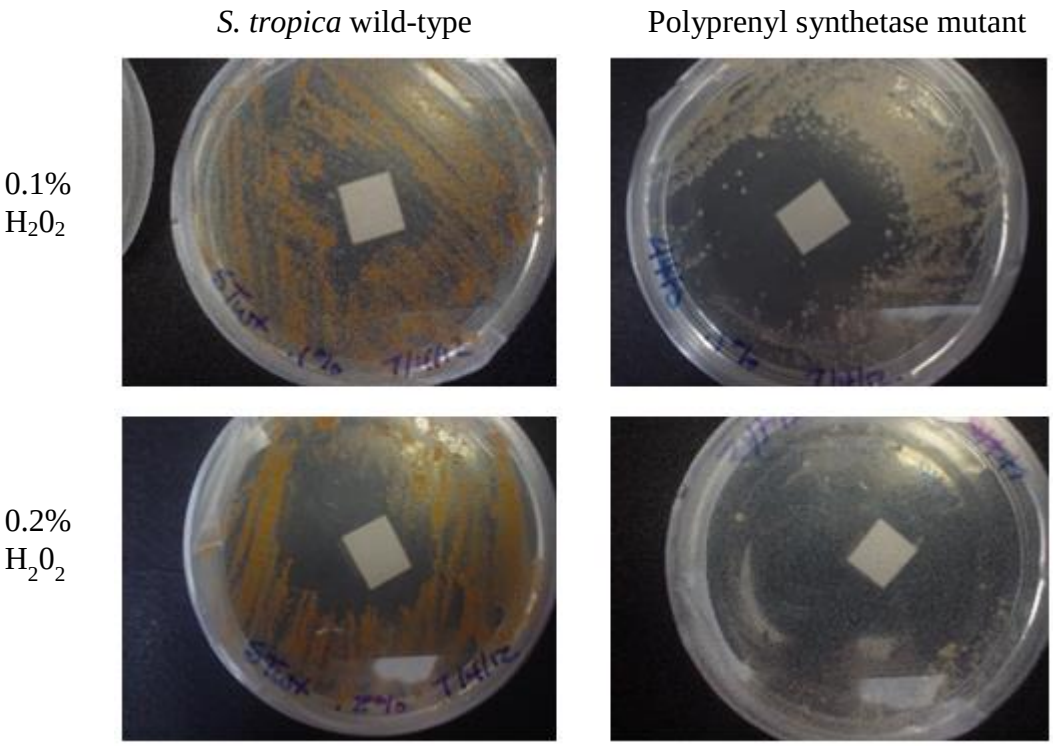


Figure 5.1: Bacterial lawns exposed to filter paper soaked in different concentrations of hydrogen peroxide showing different zones of inhibition between pigmented and non-pigmented *S. tropica.*

The observation that sioxanthin protects *Salinispora* against oxidative stress is not surprising, as it is a common function of highly conjugated molecules such as sioxanthin[12]. For a carotenoid to serve as an antioxidant, it only requires a chromophore[12]. Functionalization has been shown to alter the antioxidative efficiency of carotenoid molecules, particularly those that alter the length of the chromophore[12]. This has not been tested in *Salinispora*, but could be done using the sioxanthin intermediates made by pathway mutants generated during this study.

Another possible role for the unique structural features of sioxanthin may relate to their interaction with the cell membrane[12,16]. Carotenoids can serve as structural molecules, attaching themselves to proteins or, more commonly, integrating within lipid membranes[12]. Their interactions with lipid membranes are mediated by the presence of functional groups that modify the compound's polarity[12]. Carotenes, made up entirely of carbon and hydrogen atoms, are nonpolar and thus localize completely within the hydrophobic region of the membrane, somewhat perpendicular to the lipid tails (Figure 5.2)[16]. Xanthophylls that have oxygen-containing functional groups on both ends of the molecule are able to span the lipid membrane, remaining parallel to the lipid tails and exposing their hydrophilic functional groups to the lipid phosphate moieties (Figure 5.2). These carotenoid interactions with the lipid membrane have opposite effects on membrane fluidity: carotenes increase membrane fluidity while xanthophylls with two polar end groups increase membrane rigidity[16]. The sioxanthin structure presents an unusual dilemma having one nonpolar end group comprising of the aromatic ring, and the other end group being polar as it comprises a hydroxyl group and a glucose residue. No studies were found which investigated the membrane interactions of carotenoids that have one polar and one non-polar end. Sioxanthin allows such an opportunity to study this unusual structural phenomenon. If carotenoids in *Salinispora* impact membrane fluidity, this may be studied by the impact of temperature on the growth of the wild-type and pigment mutant bacteria. Furthermore, membrane fluidity could be tested by spectrometric analyses of the purified pigment in artificial membranes[12,16].

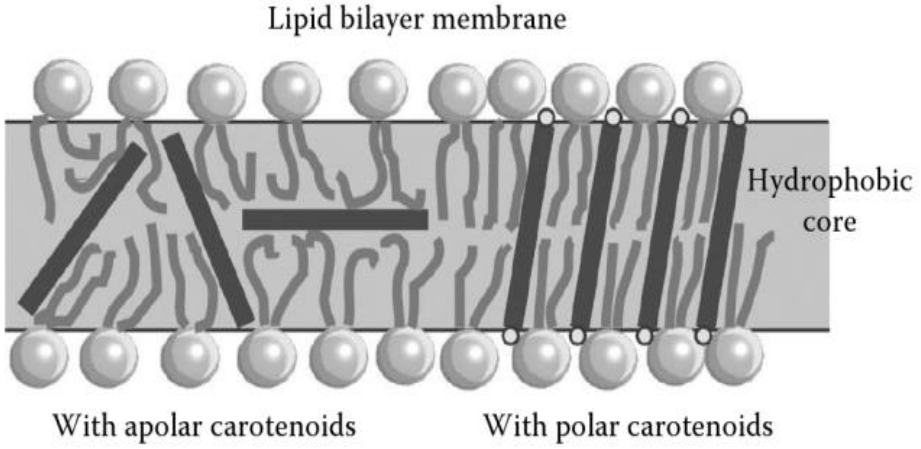


Figure 5.2: The orientation of polar and nonpolar carotenoids in lipid membranes [16].

Glycosylation is an interesting structural feature that may provide insight to the compound's biological function. Glucosyl groups can assist in the structural role of carotenoids. This has been demonstrated for cyanobacteria where glycosylated carotenoids are common[28]. Studies have shown that glycosylated carotenoids are localized to the thylakoid membrane and mutants that lack the glycosyl functional groups show a loss of thylakoid structure[29]. It seems that the glycosyl moiety serves as a binding motif that enables the proper folding and stacking of the thylakoid membrane[29]. *Salinispora* are heterotrophic and, therefore, lack thylakoid[23]. It can be presumed that sioxanthin is localized to the cell membrane; a hypothesis which could be verified through partition of cell structures in density gradient centrifugation. However, the role of the glycosyl residue in the cell membrane may be similar to that of carotenoids in thylakoid membranes in that it extends beyond the membrane and serves as some sort of anchor or attachment site. The 2'-hydroxyl may provide enough polarity for that portion of the molecule to be imbedded in the phosphate groups leaving the nonpolar portion of the molecule to rest in the hydrophobic region of the membrane.

It is unknown whether the glycosyl group faces in the interior or the exterior of the cell, or for what it would serve as an attachment site. Further studies exploring the role of this functional group could be performed using the glycosyl transferase mutant generated in this study that produces an aglycone intermediate of sioxanthin. Glycosylated carotenoids are of interest to food and nutrition industries for their increased uptake in mammalian tissues[30]

Though the *Salinispora* genus is undergoing extensive investigations regarding its chemical potential, biogeography, and phylogeny, little is known about its growth in nature or its role in their environmental niche. The biological function of sioxanthin would give insight in how the Micromonosporaceae, and bacteria in general, experience and interact with their environment.

5.2.3 Identification of the missing biosynthetic gene

The sioxanthin pathway, like that of carotenoids in the other known non-clustered pathways, contains a 2'-hydroxylase whose genetic identity remains unknown[17,31]. Identifying candidate genes for this role may be difficult, as a hydroxyl at this position is rare in carotenoids outside of the cyanobacteria[28]. The difficulty is further compounded by the fact that related bacteria known to have similar carotenoid structures (like *Mycobacterium phlei* strain Vera) are lacking a sequenced genome[32]. Searches for this gene, however, may be possible through genome comparisons. Cyanobacteria tend to have very highly divergent carotenoid biosynthetic genes[8], but domain searches may uncover candidate genes that could then be verified by gene inactivation experiments in *Salinispora.*

5.2.2 Formation of gene cluster architecture

It is plausible that multiple instances of horizontal gene transfer drive the evolution of structural diversity in carotenoids and the development of compounds such as sioxanthin. The possession of multiple gene clusters in the sioxanthin biosynthetic pathway may be the result of such gene exchanges. However, it is unclear what drives the maintenance of this genomic architecture over a single gene cluster, as is common in secondary metabolite biosynthesis. There must be a lack of selective pressure on the genomes of these non-clustered systems in areas such as regulation. Coordinated gene regulation is suggested as one of the driving forces in the development and maintenance of gene clusters[10]. In *Salinispora*, as well as the other known non-clustered pathway containing bacteria, carotenoids appear to be continuously produced, implying that its biosynthetic genes are constitutively expressed. *S. tropica* strains were grown in light and dark conditions on a variety of media (including high and low nutrient) without change in pigments. This eliminates the need for complex coordinated expression and removes regulation as a driving force for the formation of a gene cluster. To explore this hypothesis, *Salinispora* transcriptional data, ideally under multiple conditions, would be necessary to confirm that the genes are indeed constitutively expressed. Furthermore, the idea that gene clustering is related to regulation would need to be explored in other systems in which the pathway architecture is compared to the conditions under which carotenoids are expressed. In *Streptomyces*, for example, carotenoid biosynthesis is the result of a single gene cluster that is light induced[21]. While this dissertation briefly explored the relationship between gene clustering and environmental niche, it would be interesting to expand such a search to include carotenoid expression conditions related to gene clustering.

5.3 Final thoughts

This work describes the discovery and elucidation of a new carotenoid compound, sioxanthin, from *Salinispora*, an important marine bacterial genus for the production natural products. I uncovered the sioxanthin biosynthetic pathway, identifying genes in multiple distant regions of the genome and provide insight in to the evolution of the biosynthesis and chemical structure. In addition, my work lays the groundwork for further studies on the evolution of secondary metabolites, biological function of carotenoids, and the ecology of *Salinispora*.

5.4 References

1 Feling, R. H., Buchanan, G. O., Mincer, T. J., Kauffman, C. A., Jensen, P. R. & Fenical, W. Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew Chem Int Edit* **42**, 355-+, (2003).

2 Jensen, P. R., Mincer, T. J., Williams, P. G. & Fenical, W. Marine actinomycete diversity and natural product discovery. *Anton Leeuw Int J G* **87**, 43-48, (2005).

3 Jensen, P. R. & Mafnas, C. Biogeography of the marine actinomycete *Salinispora*. *Environ Microbiol* **8**, 1881-1888, (2006).

4 Jensen, P. R., Williams, P. G., Oh, D.-C., Zeigler, L. & Fenical, W. Species-specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl. Environ. Microbiol.* **73**, 1146-1152, (2007).

5 Freel, K. C., Nam, S.-J., Fenical, W. & Jensen, P. R. Evolution of secondary metabolite genes in three closely related marine actinomycete species. *Appl. Environ. Microbiol.* **77**, 7261-7270, (2011).

6 Penn, K., Jenkins, C., Nett, M., Udwary, D. W., Gontang, E. A., McGlinchey, R. P., Foster, B., Lapidus, A., Podell, S., Allen, E. E., Moore, B. S. & Jensen, P. R. Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *ISME J* **3**, 1193-1203, (2009).

7 Ziemert, N., Lechner, A., Wietz, M., Millan-Aguinaga, N., Chavarria, K. L. & Jensen, P. R. Diversity and evolution of secondary metabolism in the marine actinomycete genus *Salinispora*. *Proc. Natl. Acad. Sci. U. S. A.*, (2014).

8 Armstrong, G. A. in *Annu. Rev. Microbiol.* Vol. 51 *Annual Review of Microbiology* (ed L. N. Ornston) 629-659

9 Fischbach, M. A., Walsh, C. T. & Clardy, J. The evolution of gene collectives: How natural selection drives chemical innovation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 4601-4608, (2008).

10 Ballouz, S., Francis, A. R., Lan, R. & Tanaka, M. M. Conditions for the evolution of gene clusters in bacterial genomes. *PLoS Comput. Biol.* **6**, (2010).

11 Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829-837, (2004).

12 Britton, G. Structure and properties of carotenoids in relation to function. *FASEB J.* **9**, 1551-1558, (1995).

13 Lutnaes, B. F., Strand, A., Petursdottir, S. K. & Liaaen-Jensen, S. Carotenoids of thermophilic bacteria - *Rhodothermus marinus* from submarine Icelandic hot springs. *Biochem. Syst. Ecol.* **32**, 455-468, (2004).

14 Dieser, M., Greenwood, M. & Foreman, C. M. Carotenoid pigmentation in Antarctic heterotrophic bacteria as a strategy to withstand environmental stresses. *Arct Antarct Alp Res.* **42**, 396-405, (2010).

15 Kocher, S., Breitenbach, J., Muller, V. & Sandmann, G. Structure, function and biosynthesis of carotenoids in the moderately halophilic bacterium *Halobacillus halophilus*. *Arch. Microbiol.* **191**, 95-104, (2009).

16 Gruszecki, W. in *Carotenoids Physical, Chemical, and Biological Functions and Properties* (ed John T. Landrum) Ch. 2, (CRC Press, 2009).

17 Tian, B. & Hua, Y. Carotenoid biosynthesis in extremophilic Deinococcus-Thermus bacteria. *Trends Microbiol.* **18**, 512-520, (2010).

18 Galbis-Martinez, M., Padmanabhan, S., Murillo, F. J. & Elias-Arnanz, M. CarF mediates signaling by singlet oxygen, generated via photoexcited protoporphyrin IX, in *Myxococcus xanthus* light-induced carotenogenesis. *J. Bacteriol.* **194**, 1427-1436, (2012).

19 Liu, G. Y. & Nizet, V. Color me bad: microbial pigments as virulence factors. *Trends Microbiol.* **17**, 406-413, (2009).

20 Pezzoni, M., Costa, C. S., Pizarro, R. A. & Oppezzo, O. J. The relationship between carotenoids and sunlight response in members of the family Micrococcaceae. *J. Basic Microbiol.* **51**, 325-329, (2011).

21 Takano, H., Obitsu, S., Beppu, T. & Ueda, K. Light-induced carotenogenesis in *Streptomyces coelicolor* A3(2): Identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster. *J. Bacteriol.* **187**, 1825-1832, (2005).

22 Takano, H., Asker, D., Beppu, T. & Ueda, K. Genetic control for light-induced carotenoid production in non-phototrophic bacteria. *J. Ind. Microbiol. Biotechnol.* **33**, 88-93, (2006).

23 Maldonado, L. A., Fenical, W., Jensen, P. R., Kauffman, C. A., Mincer, T. J., Ward, A. C., Bull, A. T. & Goodfellow, M. *Salinispora arenicola* gen. nov., sp nov and *Salinispora tropica* sp nov., obligate marine actinomycetes belonging to the family Micromonosporaceae. *Int. J. Syst. Evol. Micr.* **55**, 1759-1766, (2005).

24 Mincer, T. J., Jensen, P. R., Kauffman, C. A. & Fenical, W. Widespread and persistent populations of a major new marine actinomycete taxon in ocean sediments. *Appl. Environ. Microbiol.* **68**, 5005-5011, (2002).

25 Belenky, P. & Collins, J. J. Antioxidant strategies to tolerate antibiotics (vol 334, pg 915, 2011). *Science* **334**, (2011).

26 Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A common mechanism of cellular death induced by bactericidal antibiotics. *Cell* **130**, 797-810, (2007).

27 Vynne, N. G., Mansson, M., Nielsen, K. F. & Gram, L. Bioactivity, chemical profiling, and 16S rRNA-based phylogeny of *Pseudoalteromonas* strains collected on a global research cruise. *Mar Biotechnol* **13**, 1062-1073, (2011).

28 Britton, G. Liaaen-Jensen, S. & Pfander, H. *Carotenoids handbook*. (Birkhäuser Verlag, 2004).

29 Mohamed, H. E., van de Meene, A. M. L., Roberson, R. W. & Vermaas, W. F. J. Myxoxanthophyll is required for normal cell wall structure and thylakoid organization in the cyanobacterium, *Synechocystis* sp strain PCC 6803. *J. Bacteriol.* **187**, 6883-6892, (2005).

30 Sy, C., Gleize, B., Chamot, S., Dangles, O., Carlin, F., Caris Veyrat, C. & Borel, P. Glycosyl carotenoids from marine spore-forming *Bacillus* sp. strains are readily bioaccessible and bioavailable. *Food Res Int* **51**, 914-923, (2013).

31 Takaichi, S., Maoka, T., Takasaki, K. & Hanada, S. Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscillol 2-rhamnoside, and proposed biosynthetic pathway of oscillol 2,2 '-dirhamnoside. *Microbiol.-Sgm* **156**, 757-763, (2010).

32 Hertzber.S & Jensen, S. L. Bacterial carotenoids .20. Carotenoids of *Mycobacterium phlei* strain Vera .2. Structures of phlei-xanthophylls - 2 novel tertiary glucosides. *Acta Chem. Scand.* **21**, 15-&, (1967).