

University of Alberta

Carotenoid diversity in novel *Hymenobacter* strains isolated from Victoria Upper
Glacier, Antarctica, and implications for the evolution of microbial carotenoid
biosynthesis

by

Jonathan Lee Klassen

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Microbiology and Cell Biotechnology

Department of Biological Sciences

©Jonathan Lee Klassen

Fall 2009

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

Examining Committee

Dr. Julia Foght, Department of Biological Science

Dr. Phillip Fedorak, Department of Biological Sciences

Dr. Brenda Leskiw, Department of Biological Sciences

Dr. David Bressler, Department of Agriculture, Food and Nutritional Science

Dr. Jeffrey Lawrence, Department of Biological Sciences, University of Pittsburgh

Abstract

Many diverse microbes have been detected in or isolated from glaciers, including novel taxa exhibiting previously unrecognized physiological properties with significant biotechnological potential. Of 29 unique phlotypes isolated from Victoria Upper Glacier, Antarctica (VUG), 12 were related to the poorly studied bacterial genus *Hymenobacter* including several only distantly related to previously described taxa. Further study of these microorganisms revealed genotypic, phenotypic, morphological and chemotaxonomic divergence from named species and suggested that they likely represent novel *Hymenobacter* species. These studies also indicated, however, that the systematics of *Hymenobacter* and related microorganisms is more complex than previously realized, and may exhibit poorly defined species boundaries due to cosmopolitan dispersal, significant rates of horizontal gene transfer and reintroduction of archived genotypes, e.g., from glacial ice. These processes are reflected in the carotenoid composition of *Hymenobacter* and related organisms, which includes several novel methyl- and xylosyl-derivatives of 2'-hydroxyflexixanthin with distributions indicative of horizontal gene transfer or differential gain and/or loss of terminal biosynthetic pathway steps. These processes have been previously underappreciated in assessments of microbial carotenoid diversity and suggest the need for fine-scale phylogenetic study of carotenoid distribution in other microbial taxa. Further comparative genomics-based evaluation of microbial carotenoid biosynthesis indicated its wide phylogenetic distribution and diversification, controlled by several lineage-specific modes of evolution

including horizontal transfer, *de novo* enzyme evolution followed by differential gene loss, co-evolution with biochemically associated structures and elevated mutation rates. The latter especially interacts with horizontal transfer depending on metabolic pathway topology, exemplified by the evolution of purple bacterial carotenoid biosynthesis. Exploration of VUG microbial diversity, therefore, not only revealed novel taxa and biotechnologically interesting compounds but also spurred broader evaluation of the mechanisms of metabolic pathway evolution applicable to many other taxa and biochemical pathways.

Acknowledgements

First and foremost, I thank my wife Verity and children Joel and Alice for their love, patience and support throughout my graduate studies. Without your backing, nothing that follows would have occurred. I also thank my parents for their support throughout the long process that was my education.

Much thanks and gratitude to my supervisor Dr. Julia Foght for her support throughout this project, especially in encouraging me to pursue my own research interests outside of the original project scope and for the many ways you have taught me so much during my PhD program. Thanks also to all other members of the Microbiology RIG, especially, Dr. P. Fedorak, for your material and intellectual contributions to my research. I especially thank my lab mates past and present for their friendship, listening ears, and excellent technical assistance (after the requisite “you want to do/broke what?!?”): A. Adebusuyi, M. Baker, C. Bentz, S. Cheng, D. Coy, S. Ebert, Dr. B. Hearn, J. Hulecki, Dr. R. Kumaraswamy, Dr. K. Landry, C. Li, X. Lim, Dr. C. Nesbø, T. Penner, A. Scott, K. Semple, Dr. T. Siddique, A. Wong and R. Young.

Thanks to Dr. J. Barker and Dr. M. Sharp for their provision of VUG samples. Also appreciated has been the technological assistance of many people at the University of Alberta: Dr. R. McKay (National High Field NMR Centre); W. Moffat (Department of Chemistry Analytical and Instrumental Laboratory); R. Whittal (Department of Chemistry Mass Spectrometry Laboratory); and all of the helpful people in the Department of Biological Sciences Molecular Biology Service Unit. I thank for their provision of bacterial strains: Dr. J. Aislabie (Landcare Research, New Zealand); Prof. H.-J. Busse (University of Vienna, Austria); and C. Fang (China Center for Type Culture Collection, China).

This research has been funded by an NSERC post-graduate research scholarship (Doctoral) and NSERC operating grants held by J. Foght, a Province of Alberta Scholarship and a Province of Alberta Graduate Scholarship (Masters), a Walter H. Johns Scholarship and a Mary Louise Imrie Graduate Student Award from the University of Alberta Faculty of Graduate Studies and Research and a Travel Subsidy and teaching support from the Department of Biological Sciences.

Table of Contents

1. Introduction.....	1
1.1. The genus <i>Hymenobacter</i>	1
1.2. Carotenoids	4
1.2.1. Structure and nomenclature	5
1.2.2. Physiological function	6
1.2.3. Biotechnological application	10
1.3. Horizontal gene transfer.....	12
1.4. Thesis overview and research objectives.....	14
1.5. Additional research conducted outside of the scope of my PhD program	15
1.6. Literature cited.....	15
2. Isolation of Bacteria from Victoria Upper Glacier, Antarctica Glacial Ice .	25
2.1. Introduction.....	25
2.2. Materials and methods	27
2.2.1. Study site and sampling	27
2.2.2. Bacterial culture and isolation	28
2.2.3. Bacterial identification using 16S rRNA gene sequencing.....	29
2.3. Results.....	30
2.3.1. Enumeration of cultured heterotrophic bacteria	30
2.3.2. Identification of isolated organisms.....	31
2.4. Discussion.....	33
2.5. Literature cited.....	35
3. Genotypic and Phenotypic Characterization of Novel <i>Hymenobacter</i> Species	41
3.1. Introduction.....	41
3.2. Materials and methods	42
3.2.1. Strain selection.....	42
3.2.2. Genotypic analysis	42
3.2.3. Phenotypic characterization	44
3.2.4. Fatty acid determination	46

3.2.5.	Carbon source utilization	47
3.3.	Results.....	47
3.3.1.	Phylogenetic analysis of <i>Hymenobacter</i> species and related strains using the 16S rRNA gene and <i>gyrB</i>	47
3.3.2.	Phenotypic characterization of <i>Hymenobacter</i> species and related VUG strains	51
3.3.3.	Characterization of the fatty acid composition of <i>Hymenobacter</i> species and related strains.....	57
3.3.4.	Biolog characterization of carbon source utilization profiles of <i>Hymenobacter</i> species and related strains	63
3.4.	Discussion.....	68
3.5.	Literature cited.....	73
4.	Differences in Carotenoid Composition Among <i>Hymenobacter</i> and Related Strains Support a Tree-Like Model of Carotenoid Evolution.....	78
4.1.	Introduction.....	78
4.2.	Materials and methods.....	79
4.2.1.	Bacterial strains and growth conditions.....	79
4.2.2.	16S rRNA gene sequencing and phylogenetic analysis.....	79
4.2.3.	Carotenoid extraction and analysis	80
4.2.4.	Carotenoid purification and mass spectrometry	81
4.2.5.	Nucleotide sequence accession numbers	82
4.3.	Results.....	82
4.3.1.	Phylogenetic analysis of the genus <i>Hymenobacter</i> and related VUG strains	82
4.3.2.	Carotenoid content in <i>Hymenobacter</i> and related strains	83
4.3.3.	Chemical characterization of carotenoids.....	86
4.3.4.	The effect of culture age on carotenoid composition in <i>Hymenobacter</i> and related strains.....	89
4.4.	Discussion.....	91
4.5.	Literature cited.....	93
5.	1'-Xylosyl and 2'-Methyl derivatives of 2'-Hydroxyflexixanthin are Major Carotenoid Pigments of <i>Hymenobacter</i>	97
5.1.	Introduction.....	97
5.2.	Materials and methods	98

5.2.1.	General experimental procedures	98
5.2.2.	Bacterial material and isolation procedures.....	99
5.3.	Results and discussion	100
5.4.	Chemical characteristics of <i>Hymenobacter</i> carotenoids.....	106
5.5.	Literature cited.....	107
6.	Microbial Carotenoid Diversity and Lineage-Specific Evolutionary Mechanisms are Revealed by Comparative Genomics: Implications for the Rational Study of Pathway Diversity.....	111
6.1.	Introduction.....	111
6.1.1.	General introduction and study rationale.....	111
6.1.2.	Known microbial carotenoid biosynthetic pathways.....	113
6.2.	Materials and methods	117
6.2.1.	Dataset construction.....	117
6.2.2.	Phylogenetic methods	120
6.2.3.	Statistical methods	121
6.3.	Results.....	122
6.3.1.	Distribution of microbial carotenoid biosynthetic pathways.....	122
6.3.2.	Broad evolutionary patterns are revealed by core carotenoid biosynthetic protein phylogenies	124
6.3.3.	Carotenoid biosynthesis in Proteobacteria that produce β -bicyclic xanthophylls.....	133
6.3.4.	Carotenoid production in Proteobacteria that produce linear xanthophylls.....	139
6.3.5.	Carotenoids as precursors for retinal biosynthesis.....	140
6.3.6.	Carotenoid biosynthesis in the Bacteroidetes	141
6.3.7.	Carotenoid biosynthesis in the Actinobacteria	146
6.3.8.	C30 carotenoid biosynthesis	147
6.3.9.	Carotenoid biosynthesis in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage	148
6.3.10.	Evolutionary selection in carotenoid biosynthesis pathways.....	153
6.3.11.	Recombination in carotenoid biosynthetic pathway evolution..	157
6.4.	Discussion.....	158
6.5.	Conclusions.....	163

6.6.	Literature cited.....	164
7.	Pathway Evolution by Horizontal Transfer and Positive Selection is Accommodated by Relaxed Negative Selection Upon Upstream Pathway Genes in Purple Bacterial Carotenoid Biosynthesis	171
7.1.	Introduction.....	171
7.2.	Materials and methods	174
7.3.	Results.....	175
7.3.1.	Identification of carotenoid biosynthetic pathway distribution in the purple bacteria and horizontally transferred carotenoid biosynthesis genes	175
7.3.2.	Horizontally transferred <i>crtA</i> genes in <i>Rv. gelatinosus</i> and <i>H. phototrophica</i> are under positive selection.....	179
7.3.3.	Other carotenoid biosynthetic genes in <i>Rv. gelatinosus</i> and <i>H. phototrophica</i> are under relaxed negative selection	183
7.3.4.	Two selection regimes exist for purple bacteria carotenoid biosynthetic genes.....	184
7.4.	Discussion.....	185
7.5.	Literature cited.....	190
8.	Synthesis and Conclusions.....	197
8.1.	Literature Cited	200
	Appendix A – Supplemental Data for Chapter 1	201
	Appendix B – Supplemental Data for Chapter 3	210
	Appendix C – Supplemental Data for Chapter 5	213
	Appendix D – Supplemental Data for Chapter 6.....	220
	Appendix E – Supplemental Data for Chapter 7	273

List of Tables

Table 1.1. Described <i>Hymenobacter</i> species and the sources of their isolation.	3
Table 2.1. R2A plate counts of melted VUG ice incubated at one of three temperatures for 6 weeks.	30
Table 2.2. Identification of representative VUG isolates by nearly full-length 16S rRNA gene sequencing and comparison to related isolates using BLAST searches of the GenBank nr database.	32
Table 3.1. Sequences for primers used in this study.	43
Table 3.2. Phenotypic characterization of <i>Hymenobacter</i> and reference VUG strains.	52
Table 3.3. Colony and cell morphology of <i>Hymenobacter</i> and related VUG strains.	54
Table 3.4. Whole cell fatty acid composition of <i>Hymenobacter</i> and related isolates.	58
Table 3.5. Presumptive identifications of detected fatty acids based on comparison to the BAME standard, literature values and/or derivation from their molecular masses.	62
Table 3.6. Biolog GN-2 carbon source utilization by selected <i>Hymenobacter</i> and related strains.	63
Table 4.1. Percent abundance of carotenoids present in VUG and reference <i>Hymenobacter</i> strains.	85
Table 4.2. UV-Vis absorption maxima, high-resolution MALDI molecular weights, deduced chemical formulae, and presumptive identifications of carotenoids purified from strains VUG-A141a (carotenoids 1-5) and VUG-A42aa (carotenoids 6-9).	88
Table 5.1. Experimental Parameters for 2D- ¹ H, ¹ H-TOCSY NMR Experiments.	99
Table 5.2. ¹ H-NMR spectral assignments for carotenoids 3 and 4, isolated from strain VUG-A141a, and 5 and 6, isolated from strain VUG-A42aa, determined using 1D- ¹ H-NMR and 2D- ¹ H-TOCSY spectra.	102
Table 6.1. Known microbial carotenoid biosynthetic proteins used for <i>in silico</i> carotenoid biosynthetic pathway reconstruction, their synonyms and biochemical functions.	115
Table 6.2. Inferred positive selection on carotenoid biosynthetic genes as determined from elevated d _n /d _s for specific sequences compared to their phylogenetic neighbors.	156
Table 6.3. Summary of lineage-specific evolutionary mechanisms identified in this study.	160

Table 7.1. Mean d_n/d_s substitution ratios with associated standard deviations and comparison to spirilloxanthin-producing bacteria (excluding <i>Rv. gelatinosus</i> and <i>H. phototrophica</i>), or to spheroidenone-producing bacteria for <i>crtA</i> , using the Mann-Whitney U test.	181
Table 7.2. Comparisons of differences between mean d_n/d_s values of carotenoid biosynthetic pathway genes from spheroidenone-producing bacteria (excluding <i>Rv. gelatinosus</i> and <i>H. phototrophica</i>).	184

List of Figures

Figure 1.1. Environments from which <i>Hymenobacter</i> and related isolates (white bars) and clones (grey bars) have been recovered, scored by the number of studies in which they have been reported.	2
Figure 1.2. Chemical structures of lycopene, isoprene and bacterial carotenoid end group configurations. Carbon numbers are indicated for β and ψ -end groups.	5
Figure 2.1. Distribution of phylotypes isolated at different temperatures.	31
Figure 3.1. Neighbor-joining phylogenetic tree of near full-length (\approx 1500 nt) 16S rRNA gene sequences for VUG (bolded) and other described <i>Hymenobacter</i> strains.	49
Figure 3.2. Maximum-likelihood phylogeny of partial (\approx 1200 nt) <i>gyrB</i> nucleotide sequences.	50
Figure 4.1. Near full-length 16S rRNA gene phylogeny of cultured <i>Hymenobacter</i> and <i>Hymenobacter</i> -like VUG strains (in boldface) constructed by neighbor-joining. The major carotenoid content of strains is indicated.	83
Figure 4.2. Representative HPLC chromatograms for (a) VUG-A42aa and (b) VUG-A141a and (c) the proposed structure of the <i>trans</i> -isomer of carotenoid 5, 2'-hydroxyflexixanthin.	84
Figure 4.3. Possible chemical structures of (a) carotenoid 7 and (b) carotenoids 8 and 9.	87
Figure 4.4. Relative carotenoid composition of (a) <i>H. gelipurpurascens</i> and (b) VUG-A141a during incubation for 26 d on parallel R2A plates incubated in the dark at 18°C.	90
Figure 5.1. The major carotenoids in <i>Hymenobacter</i> and their proposed biosynthetic relationships.	101
Figure 5.2. CD spectrum (200-400 nm) in ethanol at room temperature of carotenoid 5 isolated from strain VUG-A141a.	103
Figure 6.1. (next page) Known carotenoid biosynthetic pathways.	113
Figure 6.2. Distribution of carotenoid biosynthetic pathways in genome sequences of the IMG database, version 2.4.	123
Figure 6.3. Phylogenetic tree of CrtB and CrtM protein sequences constructed using RAxML.	125
Figure 6.4. Phylogenetic tree of CrtI protein sequences constructed using RAxML.	133
Figure 6.5. Phylogenetic trees constructed from nearly full-length 16S rRNA genes from carotenoid-producing members of (A) the <i>Bacteroidetes</i> ; (B) <i>Actinobacteria</i> and (C) <i>Cyanobacteria</i> constructed using RAxML.	143

Figure 6.6. Phylogenetic tree of the paralogs CrtP and CrtQ (PDS and ZDS in eukaryotes) protein sequences constructed using RAxML.....	148
Figure 6.7. Distributions of pair-wise d_n/d_s values, rounded to one decimal place, for <i>Synechococcus</i> , bicyclic xanthophyll-producing γ - <i>Proteobacteria</i> , C ₄₀ carotenoid-producing <i>Actinobacteria</i> and myxobacteria.....	154
Figure 7.1. The carotenoid biosynthetic pathway in purple bacteria.....	173
Figure 7.2. (See also previous page) RAxML phylogenetic tree of (a) CrtD and (b) CrtA amino acid sequences in purple bacteria and (for CrtA) the <i>Bacteroidetes</i>	178
Figure 7.3. Plots of pair-wise estimated d_n and d_s substitution rates for primarily spheroidenone-producing bacteria (circles) and primarily spirilloxanthin-producing bacteria (squares). Values from comparisons involving <i>H. phototrophica</i> and <i>Rv. gelatinosus</i> are indicated in orange and blue, respectively.	180

List of Abbreviations

ARDRA – Amplified ribosomal DNA restriction analysis

BAME – Bacterial acid methyl ester

CD – Circular dichroism

CFU – Colony forming units

DOC – Dissolved organic carbon

DNA – Deoxyribonucleic acid

ERIC-PCR – Enterobacterial repetitive intergenic consensus-PCR

ESI – Electrospray ionization

FAME – Fatty acid methyl ester

GC-MS – Gas chromatography-mass spectrometry

HPLC – High performance liquid chromatography

MALDI – Matrix-assisted laser desorption/ionization

ML – Maximum Likelihood

NANUC – National High Field NMR Center

NMR – Nuclear magnetic resonance

PCR – Polymerase chain reaction

PDA – Photodiode array

REP-PCR – Repetitive extragenic palindromic-PCR

TOCSY – Total correlation spectroscopy

UV-Vis – Ultraviolet-visible

VUG – Victoria Upper Glacier

1. Introduction

As originally envisioned, my thesis project was to study the microbial diversity in ice collected from Victoria Upper Glacier (VUG), Antarctica. This research direction was a continuation of the Foght lab's interest in glacier microbial ecology (Skidmore et al. 2000, Foght et al. 2004, Skidmore et al. 2005, Bhatia et al. 2006, Cheng and Foght 2007). Studying the microbial ecology of an Antarctic glacier permanently frozen at its base allows comparison with the more temperate Arctic and alpine systems studied previously. Initial culture-based evaluation of microbial diversity at VUG suggested that low numbers of bacteria were present (Chapter 2), precluding the use of non-culture-based ecological methods due to the restricted sample volumes available. Remarkable among VUG isolates, however, were many strains loosely related to the genus *Hymenobacter*. The poorly studied nature of this genus (only three species were formally described as of 2005: Hirsch et al. 1998, Collins et al. 2000, Buczolits et al. 2002) and the abundance and diversity of my VUG isolates prompted my further study of this genus to better understand its basic physiological properties facilitating survival in and adaptation to cold and otherwise oxidatively-stressed environments.

1.1. The genus *Hymenobacter*

The genus *Hymenobacter* is a deep-branching lineage of the *Flexibacteraceae* (*Bacteroidetes*). *Hymenobacter*-like organisms have a cosmopolitan distribution, reflected in their frequent detection in a wide range of environments (Figure 1.1). They are most often found in environments typified by oxidative and desiccation stress including aerosols, dry soils, irradiated meat and permanently cold environments. These latter systems experience increased water solubility of oxygen due to lower temperatures, liquid water scarcity due to freezing and concomitant lower water activity due to solute extrusion from ice crystals (Krembs and Demming 2008, Kuhn 2008). Because *Hymenobacter* strains living in dry soils and cold environments were particularly adapted to

survive stresses similar to those encountered in aerosols, they may be amenable to this mode of transport. This hypothesis has yet to be specifically tested.

The biogeography of *Hymenobacter*, i.e., cosmopolitan but enriched in oxidatively-stressed and desiccated environments, especially corresponds to the Baas-Becking hypothesis: “Everything is everywhere, but the environment selects” (Baas Becking 1934). Allopatric diversification is therefore expected to be minimal for this genus, but again, this has not been explicitly tested. One possible complicating factor concerning *Hymenobacter* biogeography might be the periodic reintroduction of ice-bound genotypes into the global *Hymenobacter* population during melt events (Rogers et al. 2004). This mechanism might archive in the ice matrix genotypes which are adaptive under modern selection regimes but which were selected against earlier in history, perhaps to extinction. Further study of *Hymenobacter* biogeography, therefore, may enlighten the understanding of microbial evolution, particularly the influence of

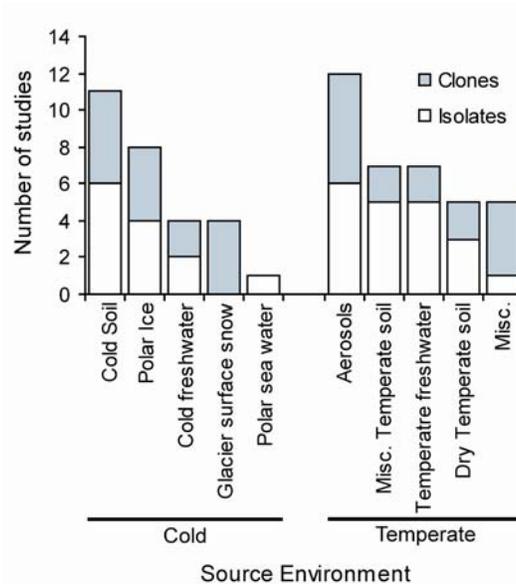


Figure 1.1. Environments from which *Hymenobacter* and related isolates (white bars) and clones (grey bars) have been recovered, scored by the number of studies in which they have been reported. *Hymenobacter*-like sequences in the GenBank nr database (accessed March 2009) were identified using BLAST (Altschul et al. 1990), and the isolation source was extracted from their annotations. See Supplemental Tables A1 and A2 for source data.

island biogeography (temporal or otherwise) on organisms with otherwise extremely large and widely-distributed effective population sizes (Berg and Kurland 2002).

As currently defined in March 2009, the genus *Hymenobacter* consists of 12 species (Table 1.1) and includes organisms formerly classified as “*Taxeobacter*” (Buczolits et al. 2006); the latter strains were previously described only informally and incompletely (Reichenbach 1992). All described *Hymenobacter* species are Gram negative, short, rounded bacilli that form bright red colonies on solid media. Many additional *Hymenobacter*-like strains have been isolated (according to annotations of their 16S rRNA gene sequences in GenBank; Supplemental Table A1), but further study of their physiology or taxonomy is rare. Many *Hymenobacter*-like 16S rRNA gene sequences cloned directly from environmental DNA are also present in GenBank (Supplemental Table A2), further highlighting diversity within this phylogenetic lineage.

Table 1.1. Described *Hymenobacter* species and the sources of their isolation.

Species	Source	Reference	GenBank Accession Number
<i>H. actinosclerus</i> CCUG 39621	Irradiated pork	(Collins et al. 2000)	Y17356
<i>H. aerophilus</i> I/26-Cor1 (DSM 13606)	Italian museum aerosol	(Buczolits et al. 2002)	EU155008
<i>H. chitinivorans</i> Txc1 ^a (DSM 11115)	Cretan soil	(Reichenbach 1992, Buczolits et al. 2006)	Y18837
<i>H. deserti</i> ZLB-3 (CCTCC AB 207171)	Chinese desert soil	(Zhang et al. 2009)	EU325941
<i>H. gelipurpurascens</i> Txg1 ^a (DSM 11116)	Soil, unknown geographical locale	(Reichenbach 1992, Buczolits et al. 2006)	Y18836
<i>H. norwichensis</i> NS/50 (DSM 15439)	UK museum aerosol	(Buczolits et al. 2006)	AJ549285
<i>H. ocellatus</i> Myx 2105 ^a (DSM 11117)	South African soil	(Reichenbach 1992, Buczolits et al. 2006)	Y18835
<i>H. rigui</i> WPCB131 (KCTC 12533)	Korean freshwater wetland	(Baik et al. 2006)	DQ089669
<i>H. roseosalivarius</i> AA718 (DSM 11622)	Antarctic McMurdo Dry Valley soil	(Hirsch et al. 1998)	Y18833
<i>H. psychrotolerans</i> Tibet-IIU11 (DSM 18569)	Chinese permafrost sediment	(Zhang et al. 2008)	DQ177475
<i>H. soli</i> PB17 (KCTC 12607)	Korean grassland soil	(Kim et al. 2008)	AB251884
<i>H. xinjiangensis</i> X2-1g (CCTCC AB 206080)	Irradiated Chinese desert sand	(Zhang et al. 2007)	DQ888329

^aOriginally described as “*Taxeobacter*” species (Reichenbach 1992)

Phenotypically, all known *Hymenobacter* species are aerobic heterotrophs which in some cases can also reduce nitrate (Hirsch et al. 1998, Buczolits et al. 2006, Zhang et al. 2008, Zhang et al. 2009). Most strains were isolated on media containing relatively low concentrations of nutrients (e.g., R2A) and are only mildly salt tolerant, if at all. Growth typically occurs within the range 4-28°C; temperature optima have not been rigorously determined for any *Hymenobacter* species due to a dearth of reported growth rates. Many carbon sources have been reported to support growth of *Hymenobacter* species, with differences occurring between strains, and also probably due to the different methods used to determine growth or activity. Similarly, polymer hydrolysis has been reported for several strains but not tested systematically.

Hymenobacter species exhibit chemotaxonomic characteristics typical of the phylum *Bacteroidetes* (see Table 1.1 for references). Branched fatty acids are predominant in *Hymenobacter* membranes, but their phospholipid head groups (of which there are several) remain unidentified. All *Hymenobacter* species possess a G+C content ranging from 55-65%, *sym*-homospermine as a major polyamine, red-pink pigmentation and menaquinone-7 as the major respiratory quinone, with menaquinone-6 sometimes present in smaller amounts. All described species are catalase and oxidase positive and are sensitive to most classes of antibiotics. These common characteristics support the inclusion of all currently described *Hymenobacter* species into a single genus, in accordance with phylogenies of their 16S rRNA genes (e.g., Zhang et al. 2009).

1.2. Carotenoids

The most striking morphological feature of *Hymenobacter* is undoubtedly their bright red-pink pigmentation resulting from the presence of carotenoids (Bircher and Pfander 1997). Because carotenoids are biotechnologically interesting compounds (Section 1.2.3) and are known to play a role in bacterial oxidative stress resistance and cold adaptation (Section 1.2.2), further investigation of carotenoid diversity, particularly in my VUG *Hymenobacter*

isolates and related reference strains, became a major focus of my thesis research.

1.2.1. Structure and nomenclature

Carotenoids are isoprenoid compounds produced by many bacteria, fungi and all photosynthetic eukaryotes. Over 600 structurally distinct carotenoids are known (Britton et al. 2004), a number which is certainly an underestimate. Carotenoids are formed by sequential condensations of the five-carbon compound isoprene (Figure 1.2), a monomer produced by all organisms for the biosynthesis variously of steroids and hopanoids, archaeal lipids, isoprenoid quinones, many secondary metabolites and other less well-known compounds (Boucher and Doolittle 2000). In most bacteria, isoprene is produced via the 1-deoxy-D-xylulose-5-phosphate pathway proceeding from pyruvate and D-glyceraldehyde-3-phosphate, whereas all archaea, eukaryotes and some bacteria utilize the mevalonate pathway beginning with acetoacetyl-CoA and acetyl-CoA; some actinomycetes possess both pathways (Boucher and Doolittle 2000). Biotechnological exploitation of these pathways for isoprenoid production has been the focus of many studies (see Section 1.2.3).

Carotenoid biosynthesis begins with the head-to-head condensation of two molecules of either the C₁₅ farnesyl pyrophosphate (three isoprene units) or the C₂₀ geranylgeranyl pyrophosphate (four isoprene units) (Sandmann and Misawa 1992, Wieland et al. 1994). All carotenoids, therefore, begin as symmetrical C₃₀

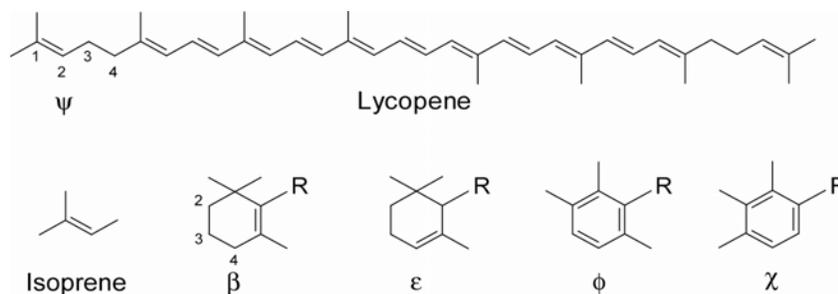


Figure 1.2. Chemical structures of lycopene, isoprene and bacterial carotenoid end group configurations. Carbon numbers are indicated for β and ψ-end groups. “R” refers to the remainder of the carotenoid structure.

or C₄₀ compounds which are sequentially modified at each end, often by different biochemical processes. Carotenoids with other carbon numbers also occur, such as C₄₅ and C₅₀ carotenoids which are produced by (di)prenylation of a C₄₀ carotenoid (Krubasik et al. 2001). Carotenoid cleavage may also produce two unique apocarotenoids, one derived from each half of the parent molecule (Auldridge et al. 2006). The exception is cleavage at the central 15,15' double bond, which may produce two molecules of the same apocarotenoid.

Undoubtedly the most important structural modification of carotenoids is desaturation (Sandmann 2009). The net result of desaturation is a chain of conjugated double bonds across the carotenoid backbone ranging from 3 to 13 double bonds, such as in lycopene (Figure 1.2). The lack of rotational freedom for carbon atoms along this chain gives carotenoids a largely planar shape (Britton 1995). Carotenoid absorption spectra are mostly defined by conjugated double bond chain length (Schmidt et al. 1994); in phototrophs, this directly affects the spectral region available for light-harvesting (Stomp et al. 2007). Similar effects of the degree of conjugation on antioxidative activity have been noted (Stahl and Sies 2003), although complicating factors such as the degree of hydroxylation also occur (Albrecht et al. 2000). Aside from conjugation of the carotenoid backbone, most other types of carotenoid modifications (e.g., cyclization, ketolation, hydroxylation, etc.) occur at carotenoid end groups. Specific nomenclature, therefore, exists to differentiate between carotenoid end group types (Britton 1995, Britton et al. 2004); examples relevant to this thesis are shown in Figure 1.2). See Section 6.1.2 for a more comprehensive review of carotenoid structural and biosynthetic pathway diversity.

1.2.2. Physiological function

Carotenoids play diverse physiological roles depending on the carotenoid type, organism and its associated cellular structures. Perhaps best studied is their antioxidative activity, a result of extensive charge delocalization facilitated by the conjugated double bond chain, which, given its membrane localization, especially protects membrane lipids and proteins (Stahl and Sies 2003, Krinsky and Johnson

2005). Carotenoid radicals can be formed by oxidation ($\text{Car} + \text{R}^{+\bullet} \rightarrow \text{Car}^{+\bullet} + \text{R}$), reduction ($\text{Car} + \text{e}^- \rightarrow \text{Car}^{\bullet-}$), abstraction of hydrogen from the carotenoid ($\text{Car-H} + \text{R}^\bullet \rightarrow \text{Car}^\bullet + \text{RH}$) or addition of a radical-containing molecule to the carotenoid, forming an adduct ($\text{Car} + \text{ROO}^\bullet \rightarrow \bullet\text{Car-OOR}$; Britton 1995, Krinsky and Johnson 2005). Radical quenching occurs by carotenoid cleavage or epoxidation. Transfer of radicals from other molecules, including other antioxidants such as vitamin C and tocopherol, is also possible (Stahl and Sies 2003, Krinsky and Johnson 2005), as is deleterious, pro-oxidant radical transfer to other non-antioxidative cellular components at high oxidant concentrations (Krinsky and Johnson 2005).

In addition to radicals, carotenoids also protect cells from energetically excited singlet oxygen ($^1\text{O}_2$), a highly reactive and destructive molecule (Stahl and Sies 2003, Krinsky and Johnson 2005). In any photosynthetic reaction center, light energy may excite porphyrins to their excited triplet state, from which energy can readily be passed to ground-state oxygen, forming $^1\text{O}_2$. This mechanism occurs both during aerobic photosynthesis and in oxygen-stressed anaerobic phototrophs, and is lethal in the absence of carotenoids. Carotenoids have an excited triplet state much lower than that of both excited triplet-state porphyrins and $^1\text{O}_2$ and can therefore readily accept energy from either (Frank and Cogdell 1996, Fraser et al. 2001, Frank and Brudvig 2004). This energy is dissipated as heat ($\text{Car} + ^1\text{O}_2 \rightarrow ^3\text{Car}^* + \text{O}_2 \rightarrow \text{Car} + \text{heat}$) because the low-energy excited triplet state of carotenoids prevents passing of energy to other molecules from this state; carotenoids acting in this manner are therefore antioxidants. In most photosynthetic eukaryotes, quenching of excited chlorophyll molecules during high light stress may also occur by de-epoxidation of the carotenoid violaxanthin to form zeaxanthin; violaxanthin is reformed under dark conditions, thereby forming the “xanthophyll cycle” (Young et al. 1997). In the opposite direction, the energetic transition from the carotenoid ground state to either its first or second excited singlet state (e.g., by light) is highly favorable. This energy can be readily passed to (bacterio)chlorophyll, as its ground state is lower than both the first and second carotenoid singlet states (Frank and Cogdell 1996,

Fraser et al. 2001). In this way carotenoids act in all photosynthetic organisms as accessory pigments, harvesting light in the 400-500 nm region. Carotenoids may also function in all photosynthetic organisms as “molecular wires” to transfer electrons between reaction center components (Frank and Brudvig 2004) and are necessary to physically organize photosynthetic reaction center assembly (Fraser et al. 2001, Cuttriss et al. 2007).

Whereas many carotenoids interact with proteins, particularly the membrane-bound photosynthetic reaction center and soluble carotenoid-binding proteins (Kerfeld 2004), others exist freely in the membrane, with which they associate due to their extensive hydrophobic character. Carotenoids and other terpenoid lipids can themselves form membranes and may have evolved prior to fatty acids in this role (Ourisson and Nakatani 1994). Carotenoid orientation in a membrane depends on its structure: carotenes (which are exclusively hydrocarbon) partition to the fatty acyl zone whereas xanthophylls (which are amphipathic) span the membrane so that the hydrocarbon portion interacts with fatty acyl chains and the oxygenated portion interacts with phospholipid head groups (Gruszecki 1999, Gruszecki and Strzalka 2005). The only known exception is lutein (β,ϵ -carotene) which, due to rotational freedom about the 6,7 position generated by the ϵ -end group, can either span the membrane or lie flat along one leaflet (Gruszecki 1999, Gruszecki and Strzalka 2005). The length of most C_{40} xanthophylls and membrane bilayers is approximately equal (~ 3 nm), with slightly longer carotenoids spanning the membrane at an angle (Gruszecki 1999, McNulty et al. 2008). The degree of carotenoid “fit” in a membrane correlates with its antioxidative activity (McNulty et al. 2008).

The presence of carotenoids in a lipid membrane has several ramifications. Xanthophylls, due to their membrane-spanning orientation, act as “molecular rivets”, holding membrane monolayers in close contact; this is particularly true of their glycosyl derivatives (Ourisson and Nakatani 1994, Gruszecki 1999). Xanthophylls also restrict permeability through the middle of the membrane by their interactions with fatty acyl chains (Gruszecki 1999). Xanthophyll head

groups reduce lateral molecular motion of surrounding fatty acids (and therefore membrane fluidity) due to interactions with phospholipid head groups, but also slightly increase permeability of the adjacent membrane region due to polar repulsive forces (Gruszecki 1999). Net decreases in membrane permeability and fluidity caused by xanthophylls are remarkably analogous to those caused by sterols in eukaryotes (Ourisson and Nakatani 1994, Gruszecki 1999, Gruszecki and Strzalka 2005). Carotenes, in contrast, increase membrane fluidity by disrupting fatty acyl chain packing due to their horizontal orientation (Gruszecki 1999, Gruszecki and Strzalka 2005). Membrane carotenoids can be altered to modulate membrane fluidity in response to low temperatures or high osmolarity, either by increasing the overall carotenoid concentration (Chattopadhyay et al. 1997, Fong et al. 2001) and/or decreasing carotenoid polarity (Jagannadham et al. 2000).

Membrane carotenoid solubility depends on membrane composition, carotenoid type and concentration (Gruszecki 1999). Carotenoid precipitates can accumulate beneficially in photosynthetic eukaryote chromoplasts, which are photoprotective, non-photosynthetic chloroplast derivatives exhibiting high pigment stability and antioxidative function (Cuttriss et al. 2007). Carotenoid precipitates may also form in localized membrane domains of appropriate biophysical character (Gruszecki 1999). Particularly significant is their effect on membrane phase transitions, whereby discrete changes between membrane states (e.g., fluid to crystalline) are abolished and replaced with a continuum of membrane fluidity due to the differential solubility of carotenoids in different membrane domains (Gruszecki 1999). This phenomenon may have particular relevance to freeze-thaw and desiccation tolerance in carotenoid-containing organisms.

In addition to the phenomena described above, it recently has been suggested that some small charged molecules, such as hydronium ions, might specifically be allowed through the membrane by interaction with a membrane-spanning xanthophyll polyene chain (Kupisz et al. 2008). While not yet

rigorously demonstrated, this mechanism could allow proton flux down a concentration gradient similar in magnitude to proton pumping by bacteriorhodopsin (Kupisz et al. 2008); the physiological ramifications or adaptations towards this phenomenon (if it truly exists *in vivo*) are unknown. Also postulated (but not demonstrated) is the ability of carotenoids to transfer radicals between cellular compartments by oxidation and reduction at opposite ends of a membrane-spanning carotenoid (Johnson 2009). Additionally, and better studied *in vivo*, a unique mechanism of carotenoid function has been described recently in *Staphylococcus aureus* whereby carotenoids protect bacteria from oxidative killing in phagocytes, thereby acting as a virulence factor (Liu et al. 2004, Liu et al. 2005, Clauditz et al. 2006). Specific inhibition of carotenoid production has been achieved fortuitously using known squalene synthase inhibitors (Liu et al. 2008) developed for lowering human cholesterol levels and has been prominently advocated as a treatment against *S. aureus* infections (Daum 2008, Walsh and Fischbach 2008).

1.2.3. Biotechnological application

Carotenoids and apocarotenoids have significant biotechnological potential, including as human nutraceuticals (see below), natural pigments (Mortensen 2006), aroma compounds (e.g., in tea and wine; Rodríguez-Bustamante and Sánchez 2007), plant hormones (Auldridge et al. 2006) and antibiotics (Ikawa et al. 2001). Production of some volatile apocarotenoids have been linked to deteriorating odor and taste of freshwater (Jüttner 1984), the control of which is a continuing problem. Applications generally exploit multiple carotenoid functions; supplementation of the diet of farmed fish with astaxanthin-rich algae, for example, improves both their flesh color and nutritional properties (Lorenz and Cysewski 2000).

Carotenoid biotechnology has been spurred recently by the linkage of carotenoid consumption to the prevention in humans of many types of cancer, osteoporosis and cardiovascular, ocular and neurodegenerative diseases, among others (reviewed by Stahl and Sies 2003, Fraser and Bramley 2004, Krinsky and

Johnson 2005, Perera and Yen 2007). Carotenoids also function in humans as precursors for Vitamin A and additionally modulate cellular growth, expression of P450-type detoxifying enzymes and cell-cell communication (Stahl et al. 2002). Specific carotenoid function and tissue distribution is dependent upon many factors, including differential release from the food matrix and differential absorption and transport (e.g., in different lipoprotein types) throughout the body; this is dependent on carotenoid structure, particularly hydrophobicity (Perera and Yen 2007). Note, however, that the unequivocal benefit to humans of carotenoid supplementation has yet to be demonstrated. Most prominently, a double-blind clinical trial failed to show any protective effect of prophylactic carotenoid supplementation (The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group 1994) and one prominent study even showed a pro-oxidant effect of carotenoids in tobacco smokers, facilitating the development of lung cancer (Hennekens et al. 1996). Undoubtedly, the structural diversity and complex chemistry (particularly regarding oxidation) of carotenoids makes optimizing their use in human nutrition an ongoing challenge.

Primarily because of their role as human micronutrients, the biotechnological market for carotenoids is extensive, with an estimated value by 2010 exceeding one billion US dollars (cited by Del Campo et al. 2007). Whereas the majority of this demand is met through chemical synthesis, these processes disadvantageously lack control of stereochemistry, have difficulty producing complex structures, require metallic catalysts and organic solvents during production and have relatively low bioavailability (Ausich 1997). Because of consumer demand for natural products and the disadvantages of chemical synthesis, recent research has focused on natural production of carotenoids. Algal sources are particularly well developed, including *Dunaliella* for β -carotene (Del Campo et al. 2007, Raja et al. 2007, Lamers et al. 2008) and *Haematococcus pluvialis* for astaxanthin (Lorenz and Cysewski 2000, Del Campo et al. 2007). However, the use of contamination-prone open ponds or expensive photobioreactors and the cost effectiveness of chemical synthesis have allowed

these biological processes to capture only a fraction of the carotenoid market (Ausich 1997, Lorenz and Cysewski 2000).

Due to the relatively low production costs of well-developed fermentation technologies, recent research has concerned the recombinant production of carotenoids in well-understood and commercially-acceptable (e.g., food-grade) host organisms. These approaches have been fostered by the adoption of carotenoid biosynthesis as a model system for metabolic pathway engineering because of their easily observable phenotype (e.g., by UV-vis spectroscopy of whole cell extracts). Two different approaches are commonly used. First, recombinant biosynthetic pathway construction combines genes from multiple organisms in a suitable host (typically *Escherichia coli*) to produce new or rare biosynthetic pathways, especially coupled to *in vitro* enzyme evolution (reviewed by Sandmann et al. 1999, Schmidt-Dannert 2000, Sandmann 2002, Umeno et al. 2005, Wang et al. 2007). Some of the resulting carotenoids exhibit enhanced biotechnological properties such as antioxidant activity (Albrecht et al. 2000, Nishida et al. 2005). Secondly, carotenoids may be overproduced (again, primarily in *E. coli*) by metabolic engineering of precursor pathways or alteration of host metabolism to increase metabolic flux towards carotenoid production (reviewed by Das et al. 2007, Klein-Marcuschamer et al. 2007, Ajikumar et al. 2008). This approach encompasses a wide range of metabolic engineering strategies, many of which have found subsequent application in the production of other biotechnologically interesting compounds.

1.3. Horizontal gene transfer

Horizontal gene transfer is a, if not the, major mechanism of microbial evolution (Koonin et al. 2001, Gogarten and Townsend 2005). Essentially, horizontal gene transfer is the acquisition of a gene by one organism which has evolved in another. A horizontally transferred gene can: (i) be novel in the recipient lineage, thereby providing a novel function; (ii) be orthologous to a preexisting gene in the recipient lineage (i.e., xenologous in the new host; (Koonin et al. 2001)), thereby allowing subfunctionalization of the two proteins;

or (iii) replace an ancestrally present homolog (Abby and Daubin 2007). At a population level, horizontal transfer of closely related genes between related organisms is essentially equivalent to recombination and serves to limit genetic diversity at that locus (Lawrence and Retchless 2009). In contrast, successful horizontal transfer of genes more distantly related to the recipient host causes diversifying evolution of biochemical networks over time and has led to the current extent of biochemical diversity.

Current opinion suggests horizontal gene transfer to be extremely widespread, likely affecting all genes at some point in their history (Dagan and Martin 2007, Dagan et al. 2008, Dagan and Martin 2009). Horizontal gene transfer is thought to be especially important at the earliest stages of metabolic pathway and cellular evolution and in determining the present gene content of the various phylogenetic lineages (Woese 1998, 2002, Boucher et al. 2003). This is congruent with an ancient evolution of the major protein fold families with their subsequent evolution by subfunctionalization (Caetano-Anollés et al. 2007) and the “complexity hypothesis”, whereby the propensity of a gene to be horizontally transferred is inversely related to its (or its product’s) biochemical network connectivity (Jain et al. 1999). Some researchers have even argued (controversially) that horizontal gene transfer has rendered microbial taxonomy (as reflecting evolution) incapable of being described according to Darwin’s classic concept of a bifurcating tree (Doolittle 1999, Baptiste and Boucher 2008, Doolittle 2009). Clearly, it is difficult to overstate the impact of horizontal gene transfer and its importance in understanding microbial phylogenetics and evolution.

Genes can be horizontally transferred by one of three mechanisms (Zaneveld et al. 2008): (i) conjugation, whereby conjugative plasmid-encoded genes are shared between (typically related) organisms via a transfer mechanism encoded by that plasmid; (ii) transformation, whereby DNA is taken up directly from the environment; and (iii) transduction, whereby DNA from a bacteriophage-infected host is packaged into the phage head during phage

assembly and is introduced into a new host during infection by that misassembled phage. Which of these mechanisms is predominant in a particular organism remains debatable and may be taxon-specific (Zaneveld et al. 2008). Whether or not a horizontally-transferred gene is integrated into a particular genome principally depends the ability of the encoded gene to be transcribed and to encode a selectively favorable phenotype conferrable to the recipient host (i.e., integratable into the host metabolic network), while minimizing the possible toxic effects of horizontally transferred gene integration. These factors may be directly related to the relatedness of the donor and recipient organisms (Gogarten and Townsend 2005). Selection governing fixation of a horizontally transferred gene is discussed more extensively in Chapter 7.

1.4. Thesis overview and research objectives

My research project originally aimed to characterize the microbial diversity of VUG glacial ice. Whereas many interesting strains were isolated (Chapter 2), overall bacterial abundance was low and required greater sample volumes than those available for further ecological study using non-culture-based techniques. Remarkable among the VUG isolates were many diverse *Hymenobacter*-like strains, likely representing novel species. The physiological and taxonomic characterization of these isolates subsequently became a major focus of my thesis research (Chapter 3). Analysis of the carotenoid pigmentation of VUG and reference *Hymenobacter* strains revealed that their diversity was substantially broader than realized previously (Chapter 4, Klassen and Foght 2008) and included novel compounds (Chapter 5, Klassen et al. 2009). Particularly intriguing was the implication of carotenoid evolution by horizontal transfer and differential gene gain and/or loss, a hitherto underappreciated phenomenon (Chapter 4, Klassen and Foght 2008). These evolutionary analyses were expanded beyond *Hymenobacter* using extensive bioinformatics analyses of all carotenoid biosynthetic proteins. From this analysis I suggest lineage-specific carotenoid distributions and evolutionary mechanisms and highlight the most profitable directions for future study (Chapter 6). A portion of this latter dataset

was subjected to further evolutionary analyses, revealing selective mechanisms by which horizontal transfer was facilitated to evolve carotenoid biosynthesis in the purple bacteria (Chapter 7, Klassen, submitted). This thesis, therefore, not only describes novel microbes and chemical compounds, but extrapolates the implications of these results to more generally understand the generation of microbial diversity.

1.5. Additional research conducted outside of the scope of my PhD program

In addition to the thesis-related research presented here, I have also collaborated with Dr. Jackie Aislabie (Landcare Research, NZ) to study the bacterial diversity of Antarctic ornithogenic soils. This research has recently been published (Aislabie et al. 2009).

1.6. Literature cited

- Abby, S., and Daubin, V. 2007. Comparative genomics and the evolution of prokaryotes. *Trends Microbiol.*, 15: 135-141.
- Aislabie, J., Jordan, S., Ayton, J., Klassen, J.L., Barker, G.M., and Turner, S. 2009. Bacterial diversity associated with ornithogenic soil of the Ross Sea region, Antarctica. *Can. J. Microbiol.*, 55: 21-36.
- Ajikumar, P.K., Tyo, K., Carlsen, S., Mucha, O., Phon, T.H., and Stephanopoulos, G. 2008. Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms. *Mol. Pharmaceut.*, 5: 167-190.
- Albrecht, M., Takaichi, S., Steiger, S., Wang, Z.-Y., and Sandmann, G. 2000. Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*. *Nat. Biotechnol.*, 18: 843-846.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.
- Auldridge, M.E., McCarty, D.R., and Klee, H.J. 2006. Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr. Opin. Plant Biol.*, 9: 315-321.

- Ausich, R.L. 1997. Commercial opportunities for carotenoid production by biotechnology. *Pure Appl. Chem.*, 69: 2169-2173.
- Baas Becking, L.G.M. 1934. *Geobiologie of inleiding tot de milieukunde* W.P. Van Stockum & Zoon, The Hague, the Netherlands.
- Baik, K.S., Seong, C.N., Moon, E.Y., Park, Y.-D., Yi, H., and Chun, J. 2006. *Hymenobacter rigui* sp. nov., isolated from wetland freshwater. *Int. J. Syst. Evol. Microbiol.*, 56: 2189-2192.
- Bapteste, E., and Boucher, Y. 2008. Lateral gene transfer challenges principles of microbial systematics. *Trends Microbiol.*, 16: 200-207.
- Berg, O.G., and Kurland, C.G. 2002. Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.*, 19: 2265-2276.
- Bhatia, M., Sharp, M., and Foght, J. 2006. Distinct bacterial communities exist beneath a high arctic polythermal glacier. *Appl. Environ. Microbiol.*, 72: 5838-5845.
- Bircher, C., and Pfander, H. 1997. Synthesis of carotenoids in the gliding bacteria *Taxeobacter*: (all-*E*,2'*R*)-3-deoxy-2'-hydroxyflexixanthin. *Helv. Chim. Acta*, 80: 832-837.
- Boucher, Y., and Doolittle, W.F. 2000. The role of lateral gene transfer in the evolution of isoprenoid biosynthesis pathways. *Mol. Microbiol.*, 37: 703-716.
- Boucher, Y., Douady, C.J., Papke, R.T., Walsh, D.A., Boudreau, M.E.R., Nesbø, C.L., Case, R.J., and Doolittle, W.F. 2003. Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.*, 37: 283-328.
- Britton, G. 1995. Structure and properties of carotenoids in relation to function. *FASEB J.*, 9: 1551-1558.
- Britton, G., Liaaen-Jensen, S., and Pfander, H. 2004. *Carotenoids handbook* Birkhäuser Verlag, Basel, Switzerland.
- Buczolits, S., Denner, E.B.M., Vybiral, D., Wieser, M., Kämpfer, P., and Busse, H.J. 2002. Classification of three airborne bacteria and proposal of *Hymenobacter aerophilus* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 52: 445-456.
- Buczolits, S., Denner, E.B.M., Kämpfer, P., and Busse, H.-J. 2006. Proposal of *Hymenobacter norwichensis* sp. nov., classification of '*Taxeobacter ocellatus*', '*Taxeobacter gelipurpurascens*' and '*Taxeobacter chitinivorans*'

as *Hymenobacter ocellatus* sp. nov., *Hymenobacter gelipurpurascens* sp. nov. and *Hymenobacter chitinivorans* sp. nov., respectively, and emended description of the genus *Hymenobacter* Hirsch *et al.* 1999. *Int. J. Syst. Evol. Microbiol.*, 56: 2071-2078.

- Caetano-Anollés, G., Shin Kim, H., and Mittenthal, J.E. 2007. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Nat. Acad. Sci. USA*, 104: 9358-9363.
- Chattopadhyay, M.K., Jagannadham, M.V., Vairamani, M., and Shivaji, S. 1997. Carotenoid pigments of an Antarctic psychrotrophic bacterium *Micrococcus roseus*: temperature dependent biosynthesis, structure, and interaction with synthetic membranes. *Biochem. Biophys. Res. Commun.*, 239: 85-90.
- Cheng, S.M., and Foght, J.M. 2007. Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.*, 59: 318-330.
- Clauditz, A., Resch, A., Wieland, K.-P., Peschel, A., and Götz, F. 2006. Staphyloxanthin plays a role in the fitness of *Staphylococcus aureus* and its ability to cope with oxidative stress. *Infect. Immun.*, 74: 4950-4953.
- Collins, M.D., Hutson, R.A., Grant, I.R., and Patterson, M.F. 2000. Phylogenetic characterization of a novel radiation-resistant bacterium from irradiated pork: description of *Hymenobacter actinosclerus* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 50: 731-734.
- Cuttriss, A.J., Mimica, J.L., Howitt, C.A., and Pogson, B.J. 2007. Carotenoids. *In* The structure and function of plastids. R.R. Wise and J.K. Hooper. (ed.) Springer, Dordrecht, The Netherlands. pp. 315-334.
- Dagan, T., and Martin, W. 2007. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc. Nat. Acad. Sci. USA*, 104: 870-875.
- Dagan, T., Artzy-Randrup, Y., and Martin, W. 2008. Modular networks and cumulative impact of lateral transfer in prokaryote genome evolution. *Proc. Nat. Acad. Sci. USA*, 105: 10039-10044.
- Dagan, T., and Martin, W. 2009. Getting a better picture of microbial evolution en route to a network of genomes. *Phil. Trans. Royal Soc. B*, 364: 2187-2196.
- Das, A., Yoon, S.-H., Lee, S.-H., Kim, J.-Y., Oh, D.-K., and Kim, S.-W. 2007. An update on microbial carotenoid production: application of recent metabolic engineering tools. *Appl. Microbiol. Biotechnol.*, 77: 505-512.

- Daum, R.S. 2008. Removing the golden coat of *Staphylococcus aureus*. N. Engl. J. Med., 359: 85-87.
- Del Campo, J.A., García-González, M., and Guerrero, M.G. 2007. Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. Appl. Microbiol. Biotechnol., 74: 1163-1174.
- Doolittle, W.F. 1999. Phylogenetic classification and the universal tree. Science, 284: 2124-2128.
- Doolittle, W.F. 2009. The practice of classification and the theory of evolution, and what the demise of Charles Darwin's tree of life hypothesis means for both of them. Phil. Trans. Royal Soc. B, 364: 2221-2228.
- Foght, J., Aislabie, J., Turner, S., Brown, C.E., Ryburn, J., Saul, D.J., and Lawson, W. 2004. Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. Microb. Ecol., 47: 329-340.
- Fong, N.J.C., Burgess, M.L., Barrow, K.D., and Glenn, D.R. 2001. Carotenoid accumulation in the psychrotrophic bacterium *Arthrobacter agilis* in response to thermal and salt stress. Appl. Microbiol. Biotechnol., 56: 750-756.
- Frank, H.A., and Cogdell, R.J. 1996. Carotenoids in photosynthesis. Photochem. Photobiol., 63: 257-264.
- Frank, H.A., and Brudvig, G.W. 2004. Redox functions of carotenoids in photosynthesis. Biochemistry, 43: 8607-8615.
- Fraser, N.J., Hashimoto, H., and Cogdell, R.J. 2001. Carotenoids and bacterial photosynthesis: the story so far... Photosynth. Res., 70: 249-256.
- Fraser, P.D., and Bramley, P.M. 2004. The biosynthesis and nutritional uses of carotenoids. Prog. Lipid Res., 43: 228-265.
- Gogarten, J.P., and Townsend, J.P. 2005. Horizontal gene transfer, genome innovation and evolution. Nat. Rev. Microbiol., 3: 679-687.
- Gruszecki, W.I. 1999. Carotenoids in Membranes. In The photochemistry of carotenoids. H.A. Frank, A.J. Young, G. Britton, and R.J. Cogdell. (ed.) Kluwer Academic Publishers, New York, NY. pp. 363-379.
- Gruszecki, W.I., and Strzalka, K. 2005. Carotenoids as modulators of lipid membrane physical properties. Biochim. Biophys. Acta, 1740: 108-115.

- Hennekens, C.H., Buring, J.E., Manson, J.E., Stampfer, M., Rosner, B., Cook, N.R., Belanger, C., LaMotte, F., Gaziano, J.M., Ridker, P.M., Willett, W., and Peto, R. 1996. Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. *N. Engl. J. Med.*, 334: 1145-1149.
- Hirsch, P., Ludwig, W., Hethke, C., Sittig, M., Hoffmann, B., and Gallikowski, C.A. 1998. *Hymenobacter roseosalivarius* gen. nov., sp. nov. from continental Antarctic soils and sandstone: bacteria of the *Cytophaga/Flavobacterium/Bacteroides* line of phylogenetic descent. *Syst. Appl. Microbiol.*, 21: 374-383.
- Ikawa, M., Sasner, J.J., and Haney, J.F. 2001. Activity of cyanobacterial and algal odor compounds found in lake waters on green alga *Chlorella pyrenoidosa* growth. *Hydrobiologia*, 443: 19-22.
- Jagannadham, M.V., Chattopadhyay, M.K., Subbalakshmi, C., Vairamani, M., Narayanan, K., Rao, C.M., and Shivaji, S. 2000. Carotenoids of an Antarctic psychrotolerant bacterium, *Sphingobacterium antarcticus*, and a mesophilic bacterium, *Sphingobacterium multivorum*. *Arch. Microbiol.*, 173: 418-424.
- Jain, R., Rivera, M.C., and Lake, J.A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Nat. Acad. Sci. USA*, 96: 3801-3806.
- Johnson, J.D. 2009. Do carotenoids serve as transmembrane radical channels? *Free Radical Bio. Med.*, 47: 321-323.
- Jüttner, F. 1984. Dynamics of the volatile organic substances associated with cyanobacteria and algae in a eutrophic shallow lake. *Appl. Environ. Microbiol.*, 47: 814-820.
- Kerfeld, C.A. 2004. Water-soluble carotenoid proteins of cyanobacteria. *Arch. Biochem. Biophys.*, 430: 2-9.
- Kim, K.-H., Im, W.-T., and Lee, S.-T. 2008. *Hymenobacter soli* sp. nov., isolated from grass soil. *Int. J. Syst. Evol. Microbiol.*, 58: 941-945.
- Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.
- Klassen, J.L., McKay, R., and Foght, J.M. 2009. 2'-Methyl and 1'-xylosyl derivatives of 2'-hydroxyflexixanthin are major carotenoids of *Hymenobacter* species. *Tetrahedron Lett.*, 50: 2656-2660.

- Klein-Marcuschamer, D., Ajikumar, P.K., and Stephanopoulos, G. 2007. Engineering microbial cell factories for biosynthesis of isoprenoid molecules: beyond lycopene. *Trends Biotechnol.*, 25: 417-424.
- Koonin, E.V., Makarova, K.S., and Aravind, L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.*, 55: 709-742.
- Krembs, C., and Demming, J.W. 2008. The role of exopolymers in microbial adaptation to sea ice. *In Psychrophiles: from biodiversity to biotechnology.* R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 247-264.
- Krinsky, N.I., and Johnson, E.J. 2005. Carotenoid actions and their relation to health and disease. *Mol. Aspects Med.*, 26: 459-516.
- Krubasik, P., Kobayashi, M., and Sandmann, G. 2001. Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C₅₀ carotenoid formation. *Eur. J. Biochem.*, 268: 3702-3708.
- Kuhn, M. 2008. The climate of snow and ice as boundary condition for microbial life. *In Psychrophiles: from biodiversity to biotechnology.* R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 3-15.
- Kupisz, K., Sujak, A., Patyra, M., Trebacz, K., and Gruszecki, W.I. 2008. Can membrane-bound carotenoid pigment zeaxanthin carry out a transmembrane proton transfer? *Biochim. Biophys. Acta*, 1778: 2334-2340.
- Lamers, P.P., Janssen, M., De Vos, R.C.H., Bino, R.J., and Wijffels, R.H. 2008. Exploring and exploiting carotenoid accumulation in *Dualiella salina* for cell-factory applications. *Trends Biotechnol.*, 26: 631-638.
- Lawrence, J.G., and Retchless, A.C. 2009. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. *In Horizontal Gene Transfer.* M.B. Gogarten, J.P. Gogarten, and L. Olendzenski. (ed.) Humana Press, New York, NY. pp. 29-53.
- Liu, C.-I., Liu, G.Y., Song, Y., Yin, F., Hensler, M.E., Jeng, W.-Y., Nizet, V., Wang, A.H.-J., and Oldfield, E. 2008. A cholesterol biosynthesis inhibitor blocks *Staphylococcus aureus* virulence. *Science*, 319: 1391-1394.

- Liu, G.Y., Doran, K.S., Lawrence, T., Turkson, N., Puliti, M., Tissi, L., and Nizet, V. 2004. Sword and shield: linked group B streptococcal β -hemolysin/cytolysin and carotenoid pigment function to subvert host phagocyte defense. *Proc. Natl. Acad. Sci. U. S. A.*, 101: 14491-14496.
- Liu, G.Y., Essex, A., Buchanan, J.T., Datta, V., Hoffman, H.M., Bastian, J.F., Fierer, J., and Nizet, V. 2005. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.*, 202: 209-215.
- Lorenz, R.T., and Cysewski, G.R. 2000. Commercial potential for *Haematococcus* microalgae as a natural source of astaxanthin. *Trends Biotechnol.*, 18: 160-167.
- McNulty, H., Jacob, R.F., and Mason, R.P. 2008. Biologic activity of carotenoids related to distinct membrane physiochemical interactions. *Am. J. Cardiol.*, 101[suppl]: 20D-29D.
- Mortensen, A. 2006. Carotenoids and other pigments as natural colorants. *Pure Appl. Chem.*, 78: 1477-1491.
- Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., Komemushi, S., Miki, W., and Misawa, N. 2005. Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'- β -hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanthophylls. *Appl. Environ. Microbiol.*, 71: 4286-4296.
- Ourisson, G., and Nakatani, Y. 1994. The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chem. Biol.*, 1: 11-23.
- Perera, C.O., and Yen, G.M. 2007. Functional properties of carotenoids in human health. *Int. J. Food Prop.*, 10: 201-230.
- Raja, R., Hemaiswarya, S., and Rengasamy, R. 2007. Exploitation of *Dunaliella* for β -carotene production. *Appl. Microbiol. Biotechnol.*, 74: 517-523.
- Reichenbach, H. 1992. *Taxeobacter*, a new genus of the *Cytophagales* with three new species. In *Advances in the taxonomy and significance of Flavobacterium, Cytophaga and related bacteria: proceedings of the 2nd international symposium on Flavobacterium, Cytophaga and related bacteria*. P.J. Jooste. (ed.) University of the Orange Free State, Bloemfontein, South Africa. pp. 182-185.
- Rodríguez-Bustamante, E., and Sánchez, S. 2007. Microbial production of C₁₃-

- norisoprenoids and other aroma compounds via carotenoid cleavage. *Crit. Rev. Microbiol.*, 33: 211-230.
- Rogers, S.O., Starmer, W.T., and Castello, J.D. 2004. Recycling of pathogenic microbes through survival in ice. *Med. Hypotheses*, 63: 773-777.
- Sandmann, G., and Misawa, N. 1992. New functional assignment of the carotenogenic genes *crtB* and *crtE* with constructs of these genes from *Erwinia* species. *FEMS Microbiol. Lett.*, 90: 253-258.
- Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, O., and Böger, P. 1999. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. *Trends Biotechnol.*, 17: 233-237.
- Sandmann, G. 2002. Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures. *ChemBioChem*, 3: 629-635.
- Sandmann, G. 2009. Evolution of carotene desaturation: the complication of a simple pathway. *Arch. Biochem. Biophys.*, 483: 169-174.
- Schmidt, K., Connor, A., and Britton, G. 1994. Analysis of pigments: carotenoids and related polyenes. *In* Chemical methods in prokaryotic systematics. M. Goodfellow and A.G. O'Donnell. (ed.) John Wiley & Sons, Chichester. pp. 403-461.
- Schmidt-Dannert, C. 2000. Engineering novel carotenoids in microorganisms. *Curr. Opin. Biotechnol.*, 11: 255-261.
- Skidmore, M., Anderson, S.P., Sharp, M., Foght, J., and Lanoil, B.D. 2005. Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Appl. Environ. Microbiol.*, 71: 6986-6997.
- Skidmore, M.L., Foght, J.M., and Sharp, M.J. 2000. Microbial life beneath a high arctic glacier. *Appl. Environ. Microbiol.*, 66: 3214-3220.
- Stahl, W., Ale-Agha, N., and Polidori, M.C. 2002. Non-antioxidant properties of carotenoids. *Biol. Chem.*, 383: 553-558.
- Stahl, W., and Sies, H. 2003. Antioxidant activity of carotenoids. *Mol. Aspects Med.*, 24: 345-351.
- Stomp, M., Huisman, J., Stal, L.J., and Matthijs, H.C.P. 2007. Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. *ISME J.*, 1: 271-282.

- The Alpha-Tocopherol Beta Carotene Cancer Prevention Study Group 1994. The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N. Engl. J. Med.*, 330: 1029-1035.
- Umeno, D., Tobias, A.V., and Arnold, F.H. 2005. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol. Mol. Biol. Rev.*, 69: 51-78.
- Walsh, C.T., and Fischbach, M.A. 2008. Inhibitors of sterol biosynthesis as *Staphylococcus aureus* antibiotics. *Angew. Chem. Int. Ed.*, 47: 5700-5702.
- Wang, F., Jiang, J.G., and Chen, Q. 2007. Progress on molecular breeding and metabolic engineering of biosynthesis pathways of C₃₀, C₃₅, C₄₀, C₄₅, C₅₀ carotenoids. *Biotechnol. Adv.*, 25: 211-222.
- Wieland, B., Feil, C., Gloria-Maercker, E., Thumm, G., Lechner, M., Bravo, J.-M., Poralla, K., and Götz, F. 1994. Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of *Staphylococcus aureus*. *J. Bacteriol.*, 176: 7719-7726.
- Woese, C.R. 1998. The universal ancestor. *Proc. Nat. Acad. Sci. USA*, 95: 6854-6859.
- Woese, C.R. 2002. On the evolution of cells. *Proc. Nat. Acad. Sci. USA*, 99: 8742-8747.
- Young, A.J., Phillip, D., Ruban, A.V., Horton, P., and Frank, H.A. 1997. The xanthophyll cycle and carotenoid-mediated dissipation of excess excitation energy in photosynthesis. *Pure Appl. Chem.*, 69: 2125-2130.
- Zaneveld, J.R., Nemergut, D.R., and Knight, R. 2008. Are all horizontal gene transfers created equal? Prospects for mechanism-based studies of HGT patterns. *Microbiology*, 154: 1-15.
- Zhang, G., Niu, F., Busse, H.-J., Ma, X., Liu, W., Dong, M., Feng, H., An, L., and Cheng, G. 2008. *Hymenobacter psychrotolerans* sp. nov., isolated from the Qinghai-Tibet Plateau permafrost region. *Int. J. Syst. Evol. Microbiol.*, 58: 1215-1220.
- Zhang, L., Dai, J., Tang, Y., Luo, X., Wang, Y., An, H., Fang, C., and Zhang, C. 2009. *Hymenobacter deserti* sp. nov., isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 59: 77-82.
- Zhang, Q., Liu, C., Tang, Y., Zhou, G., Shen, P., Fang, C., and Yokota, A. 2007. *Hymenobacter xinjiangensis* sp. nov., a radiation-resistant bacterium

isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*,
57: 1752-1756.

2. Isolation of Bacteria from Victoria Upper Glacier, Antarctica Glacial Ice

2.1. Introduction

Despite first impressions of their inhospitability, permanently frozen environments such as glaciers harbor diverse and often abundant microbes (Hodson et al. 2008). Glacial environments amenable to life include supraglacial snow (Miteva 2008), surface meltwater streams and lakes (Bhatia et al. 2006), cryoconite holes (Christner et al. 2003, Porazinska et al. 2004), ice veins (Price 2000, Mader et al. 2006, Krembs and Demming 2008) and crystals (Rohde and Price 2007), liquid water microfilms e.g., along clay grains (Tung et al. 2006) and liquid water and unfrozen sediments at their pressure-melting points along the glacial base (Sharp et al. 1999, Skidmore et al. 2000, Foght et al. 2004, Skidmore et al. 2005, Bhatia et al. 2006, Cheng and Foght 2007, Christner et al. 2008, Hodson et al. 2008). Of special interest in the latter category because of their inaccessibility and postulated novelty have been Antarctic subglacial lakes and rivers, subjects of numerous researchers (Karl et al. 1999, Priscu et al. 1999, Abyzov et al. 2001, Christner et al. 2001, Christner et al. 2006, Lavire et al. 2006, D'Elia et al. 2008, Lanoil et al. 2009).

Several motivations exist for studying glacial microbiology. First, subglacial microbes may be active along the glacier bed, especially during winter when other inputs into the subglacial environment are limited and oxygen becomes depleted. They are therefore responsible for subglacial geochemical transformations (e.g., rock dissolution), especially those involving nitrogen, sulfur, iron and methane (Skidmore et al. 2000, Wadham et al. 2004, Hodson et al. 2005, Skidmore et al. 2005, Wynn et al. 2006, Wadham et al. 2007). These biogeochemical reactions, previously considered unlikely or insignificant, substantially impact models of geochemical processes on a large scale during periods of significant glacial advance such as ice ages (Tranter et al. 2005, Wadham et al. 2008). Second, glaciers are excellent exobiology analogue sites,

especially for testing sampling methods (Abyzov et al. 1999, Juck et al. 2005). The recent discovery of ice and methane emissions from Mars further emphasizes the specific astrobiological relevance of icy ecosystems (Abyzov et al. 1998, Tung et al. 2005, Price 2007). Third, glacial microbiology impacts the interpretation of climatic records archived in glacial cores by changing the composition of entrained gases through *in situ* microbial activity (Miteva et al. 2007, Rohde et al. 2008) and/or by providing direct historical climatic information from the phylogenetic composition of microbes present at different depths (Yao et al. 2006, Miteva et al. 2009). Fourth, glaciers, like other cold environments, provide excellent systems for studying microbial biogeography because of the distance between sampling locations (e.g., Arctic versus Antarctic) and strong selection for specific, stress-tolerant organisms (e.g., Staley and Gosink 1999, Rodrigues et al. 2009). Fifth, due to their unique ecological characteristics, glaciers contain many novel taxa not commonly found in other environments (see Chapter 3, among many other examples). Glaciers are particularly known for harboring ancient microbes and have been used to test the limits of microbial survival (Price and Sowers 2004, Bidle et al. 2007, Stewart Johnson et al. 2007, Price 2009) and evolution under cold climatic conditions (Price 2007). Archived microbes may harbor adaptive genes driven to extinction elsewhere by selection; whether this includes entombed pathogens remains controversial (Dancer et al. 1997, Castello et al. 1999, Rogers et al. 2004). Finally, the presence of unknown strains and archived genotypes makes glaciers excellent sources for biotechnologically interesting proteins (e.g., cold-active lipases, proteases and hydrolases are used in detergents, food production, biofuel production and molecular biology; Huston 2008) and lipids (e.g., omega-3 and polyunsaturated fatty acids; Nichols et al. 1999), discovered using culture-based techniques and recently metagenomics (Simon et al. 2009).

Recent studies have demonstrated that microbes in the basal ice of an Arctic glacier are distinct from those found supraglacially and in adjacent environments (i.e., they are native to the glacial base; Bhatia et al. 2006, Cheng and Foght 2007) and that the microbial composition of these environments correlates with the

subglacial geochemistry (Skidmore et al. 2005). Regarding Antarctic systems, Blood Falls, a saline, Fe^{2+} -rich outfall of the Taylor Glacier (McMurdo Dry Valley complex, Antarctica) derived from a marine brine trapped beneath this glacier, has been particularly well-studied. Blood Falls hosts a chemolithotrophic, sulfate-reducing microbial community where sulfate is likely regenerated by Fe^{3+} reduction (Mikucki et al. 2004, Mikucki and Priscu 2007, Mikucki et al. 2009); which microbes are ice- or brine-derived, however, remains unclear. Other Antarctic glacial ice samples have been studied (Christner et al. 2000, Bidle et al. 2007), especially from the Lake Vostok ice core (Abyzov 1993, Abyzov et al. 2004) and the accreted lake ice at its bottom (Karl et al. 1999, Priscu et al. 1999, Abyzov et al. 2001, Christner et al. 2001, Christner et al. 2006, Lavire et al. 2006, D'Elia et al. 2008). However, most studies typically report samples cored from the surface and (except for Blood Falls and the Lake Vostok core, which reflect ancient marine and lacustrine sources, respectively) do not consider geochemical processes occurring at the glacial base, the objective of analysis reported here.

2.2. Materials and methods

2.2.1. Study site and sampling

Victoria Upper Glacier ($77^{\circ}16'S$, $161^{\circ}29'E$; hereafter VUG) is situated at the northwesternmost extreme of Victoria Valley, part of the well-studied McMurdo Dry Valley complex, Antarctica (Kelly et al. 2002). VUG does not form part of the Antarctic Ice Sheet but instead drains local precipitation into the moisture-deficient valley (Kelly et al. 2002). It is bordered at its terminus by the perennially ice-covered Victoria Upper Lake and two smaller, unnamed water bodies. Expansion of these lakes, due either to VUG advance or melting, is responsible for moraine structure and deposition of lacustrine algae-containing sediment throughout the northwest portion of Victoria Valley (Kelly et al. 2002).

Like many Antarctic glaciers, VUG is cold-based, i.e., frozen to its bed. At its terminus, the glacier face is 50 m high, of which the bottom 15 m is debris-rich basal ice, 10-15 m of which is occluded by calved glacial ice (Barker et al.

2006). Sampling of an exposed vertical transect across the VUG was conducted by Dr. Joel Barker (Department of Earth and Atmospheric Sciences, University of Alberta) in January 2003. All samples were taken aseptically using an ethanol-sterilized spatula to transfer ice into pre-sterilized Whirl-Pack bags (Nasco). Sample names refer to the distance from the transect top, with the 220-230 cm sample corresponding to a location immediately above the glacial-basal ice interface. All samples were transferred frozen to the lab and stored at -20°C until analyzed. Although collected during the same sampling campaign, the samples analyzed here are distinct from those reported elsewhere (Barker et al. 2006).

2.2.2. Bacterial culture and isolation

Prior to use, beakers and forceps were rinsed with 4 M hydrochloric acid and autoclaved twice. In a UV- and Roccol-sterilized biosafety hood, ice samples were aseptically removed from Whirl-Pack bags into a sterile glass beaker which was covered with foil and incubated at 4°C to allow sample melting. Immediately following melting (approximately 20 h), samples were moved into a fume hood and sparged with O₂-free N₂ for approximately 15 min using a twice-autoclaved 25 mL pipette. Approximately 100 mL of the sample was transferred aseptically to a sterile N₂-sparged 165 mL serum bottle, capped and moved to a Roccol-sterilized laminar flow hood. To construct a dilution series, 10 mL of meltwater was added to sterile N₂-flushed 165 mL serum bottles containing 90 mL of 0.1% sodium pyrophosphate (Fisher) and shaken vigorously by hand for 5 min. The remainder of the dilution series up to 10⁻⁴ was prepared similarly with inversion, and remaining meltwater was refrozen and returned to Dr. Joel Barker for further geochemical analysis.

Microbes were cultured by plating 0.1 mL from each dilution, in triplicate, onto R2A plates chilled to 4°C (Difco). Each dilution series was incubated at 4°C, 10°C or room temperature (approximately 20°C) in the dark for 6 weeks, at which point new colonies (scored weekly) no longer appeared. Colonies were picked based upon their unique morphology, both between incubation temperatures and replicate plates incubated at the same temperature, and time of appearance.

Colonies were restreaked onto R2A plates and incubated at their isolation temperature for 1 or 2 weeks, depending on the isolate. Most cultures were restreaked until morphologically consistent; some cultures that consistently displayed mixed colony types were considered pure after sub-culturing several times plus determination of a degeneracy-free 16S rRNA gene sequence.

2.2.3. Bacterial identification using 16S rRNA gene sequencing

For all isolates, genomic DNA was isolated using the bead-beating and chemical lysis procedure of Foght et al. (2004). Partial 16S rRNA gene sequences were amplified using the primers PB36F and PB38R (Foght et al. 2004; see section 3.2.1 for sequences) using a PCR amplification protocol reported previously (Cheng and Foght 2007). Unique PCR products were identified by amplified ribosomal DNA restriction analysis (ARDRA) using CfoI and HaeIII (Roche). Digestion was performed as a 1:1 mixture of enzyme preparation to PCR product and incubated at 37°C for 3 h. Bands were visualized using ethidium bromide and UV transillumination following 2% agarose electrophoresis at 96 V for 3 h. Unique restriction digestion patterns were determined by analysis of fragment sizes using the Gel-Pro Analyzer version 4.5 image analysis software (MediaCybernetics).

Cloned and amplified 16S rRNA genes with unique ARDRA patterns were purified using the Gel-pure kit (Roche) and sequenced by the dideoxy terminator method (AB Big Dye Terminator kit, Applied Biosystems) at the University of Alberta Department of Biological Sciences Molecular Biology Service Unit using an ABI 3700 sequencer and primer PB36F according to standard protocols. The resulting partial 16S rRNA gene sequence (typically ~ 500 base pairs) was compared to the GenBank nr nucleotide database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Nucleotide>) using BLAST (Altschul et al. 1990) and to other VUG partial 16S rRNA gene sequences using non-bootstrapped neighbor-joining trees generated using CLUSTALX (Thompson et al. 1997) and the DNADIST (Kimera 2-parameter substitution matrix) and NEIGHBOR programs of the PHYLIP package (Felsenstein 1989). Single

representative strains from each resulting sequence cluster (containing identical or nearly identical partial 16S rRNA gene sequences) were selected for nearly full-length double-stranded rRNA gene sequencing (typically ~1480 base pairs) by the same procedure as above using the primers PB38R (Foght et al. 2004), 16S1F, 16S2R, 16S3F, 16S4F and 16S5R (Cheng and Foght 2007; see section 3.2.1. for sequences). Sequences corresponding to PB36F and PB38R were removed prior to final BLAST comparison to the GenBank nr database and deposition in GenBank (accession numbers EU155008-EU155017, GQ454797-GQ454806 and GQ454841-GQ454859).

2.3. Results

2.3.1. Enumeration of cultured heterotrophic bacteria

The number of microbes recovered from VUG varied from $3 \pm 6 \times 10^0$ to $2.7 \pm 0.5 \times 10^3$ CFU/mL depending on the sample and incubation conditions used (Table 2.1). The VUG 220-230 cm sample, taken from immediately above the basal-glacial ice interface, had much higher microbial counts than the VUG 60-70 cm and VUG 10-20 cm samples, which both represent “clean” glacial ice. The highest microbial counts were obtained using incubation on R2A at 4°C, suggesting either that the microbes are cold-adapted or that cold temperatures permit a higher resuscitation rate. Because preliminary most-probable-number analyses of the same samples using modified R2A indicated greater numbers of recoverable bacteria under oxygen-reduced (1.5% agar) compared with oxygen-replete conditions (no agar added; data not shown), the numbers reported here are likely underestimates. Similar to the plate count data (Table 2.1), more

Table 2.1. R2A plate counts of melted VUG ice incubated at one of three temperatures for 6 weeks. Ten-fold dilutions to 10^{-4} were plated in triplicate; one standard deviation is reported. Samples are listed as the distance down a vertical transect of the VUG face from which they were taken.

VUG Sample	CFU/mL incubated at:		
	4°C	10°C	~20°C
10-20 cm	7 ± 6	3 ± 6	10 ± 10
60-70 cm	30 ± 0	7 ± 6	7 ± 12
220-230 cm	$2.7 \times 10^3 \pm 5.3 \times 10^2$	$9.2 \times 10^2 \pm 99$	$6.0 \times 10^2 \pm 21$

phylotypes were isolated at 4°C and 10°C than at room temperature (Figure 2.1). Unlike the plate counts, differences between 4°C and 10°C were minor, with slightly more phylotypes isolated at 10°C than 4°C.

2.3.2. Identification of isolated organisms

In total, 202 bacterial strains were isolated from VUG. ARDRA analysis of these strains revealed 62 unique CfoI and 59 unique HaeIII patterns which together divided the 202 strains into 99 restriction fragment patterns. Partial 16S rRNA gene sequencing of 70 of these patterns and phylogenetic analysis revealed 29 unique sequence groups (Table 2.2), some of which were related but distinct (data not shown). This clustering pattern suggests that ARDRA analysis revealed diversity at a relatively high level of phylogenetic resolution, approximately equivalent to only a few nucleotide substitutions between 16S rRNA gene sequences. Of the recovered sequence groups, 17 were related to the genera *Acidovorax*, *Arthrobacter*, *Brevundimonas*, *Deinococcus*, *Flavobacterium*, *Frigoribacterium*, *Janthinobacterium*, *Knoellia*, *Nocardiodes*, *Polaromonas*, *Sphingomonas* and *Variovorax*. The remaining 12 sequence groups were loosely related to the strains from the genus *Hymenobacter* (Table 2.2), albeit with significant levels of divergence from previously described isolates, especially those known when these results were generated in 2004.

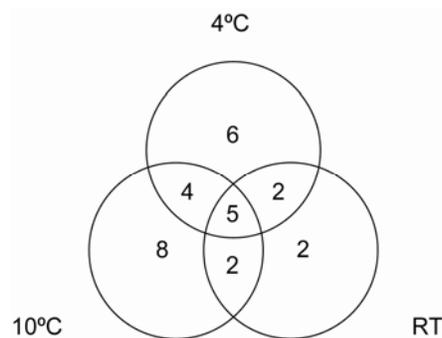


Figure 2.1. Distribution of phylotypes isolated at different temperatures.

Table 2.2. Identification of representative VUG isolates by nearly full-length 16S rRNA gene sequencing and comparison to related isolates using BLAST searches of the GenBank nr database (searched June 12, 2009)

Group (Isolate)	Percent Similarity (%)	Nearest neighbor	GenBank Accession Number
Alpha-1 (A51b)	99	Svalbard snow isolate <i>Sphingomonadaceae</i> bacterium N	DQ497241
	98	John Evans Glacier, Nunavut uncultured clone JEG.xsd4	DQ228414
	94	<i>Caulobacter leidyia</i> ATCC 15260	NR_025324
Alpha-2 (C10)*	99	Lake Vida, Antarctica ice uncultured clone ANTLV1_E12	DQ521492
	98	<i>Kaistobacter terrae</i> KCTC12630	AB258386
Alpha-3 (A41)	99	Lake <i>Sphingomonas</i> sp. HTCC503	AY444850
	95	<i>Sphingopyxis witflariensis</i> (activated sludge)	AJ416410
Arth-1 (A15)	99	Tibetan Permafrost isolate <i>Arthrobacter</i> sp. Tibet-IIVa3	DQ108397
	99	La Gorce Mountains, Antarctica soil isolate actinobacterium P19	DQ351734
Arth-2 (A21b)	99	<i>Arthrobacter agilis</i> LV7	AF134184
	100	<i>Arthrobacter</i> sp. Tibet-IIVa3 (Tibetan Permafrost)	DQ108397
	97	Actinobacterium P19 (La Gorce Mountains soil, Antarctica)	DQ351734
Beta-1 (A138)	97	<i>Arthrobacter agilis</i> strain LV7	AF134184
	99	Arctic sea ice floe melt pond uncultured clone ARKMP-77	AF468328
	98	John Evans Glacier, Nunavut uncultured clone JEG.g5	CP000529
Beta-2 (A90)	97	<i>Polaromonas naphthalenivorans</i> CJ2	DQ228409
	98	John Evans Glacier, Nunavut uncultured clone JEG.c8	DQ228397
	98	<i>Variovorax paradoxus</i> S110	CP001635
Beta-3 (A10)	99	Mount Everest surface snow uncultured clone EFS-71	EF190150
	99	Lake Vida, Antarctica ice uncultured clone ANTLV1_D07	DQ521485
	98	John Evans Glacier, Nunavut uncultured clone SOC A4(5)	DQ628932
Brev (A19)	98	<i>Polaromonas naphthalenivorans</i> CJ2	CP000529
	99	<i>Brevundimonas variabilis</i> ATCC 15255	AJ227783
	98	<i>Brevundimonas alba</i> DSM 4736	AJ227785
Deino (A62)	97	<i>Deinococcus radiopungans</i> ATCC 19172	NR_026403
	98	Svalbard soil isolate <i>Flavobacterium</i> sp. PR6-5	FJ889628
Flavo (A97)	97	John Evans Glacier, Nunavut isolate <i>Flavobacterium</i> sp. SOC A4(51)	DQ628945
	97	<i>Flavobacterium pectinovorum</i> DSM 6368	AM230490
Hym-1 (A106)	99	<i>Hymenobacter aerophilus</i> DSM 13606	EU155008
Hym-2 (A60a)	99	<i>Hymenobacter roseosalivarius</i> DSM 11622	Y18834
Hym-3 (A23a)	95	<i>Hymenobacter chitinivorans</i> DSM 11115	Y18837
Hym-4 (A124)	95	<i>Hymenobacter norwichensis</i> DSM 15439	AJ549285
	95	<i>Hymenobacter chitinivorans</i> DSM 11115	Y18837
Hym-5 (A112)	98	<i>Hymenobacter chitinivorans</i> DSM 11115	Y18837
Hym-6 (A48)	98	Arctic sea ice cryoconite hole clone ARKCRY2	AY198110
	95	<i>Hymenobacter soli</i> KTCT 12607	AB251884

Hym-7 (A58)	98 96	Lake Vida Ice, Antarctica clone ANTLV2_E05 <i>Hymenobacter soli</i> KTCT 12607	DQ521518 AB251884
Hym-8 (A49a)	99 96	Lake Vida Ice, Antarctica clone ANTLV2_E05 <i>Hymenobacter soli</i> KTCT 12607	DQ521518 AB251884
Hym-9 (A32a)	98 96	Lake Vida Ice, Antarctica clone ANTLV2_E05 <i>Hymenobacter soli</i> KTCT 12607	DQ521518 AB251884
Hym-10 (A141a)	98	<i>Hymenobacter roseosalivarius</i> DSM 13606	Y18834
Janth (A99a)	99 97	Canada Glacier, Antarctica cryoconite hole isolate bacterium CanDirty89 <i>Janthinobacterium agaricidamnosum</i> DSM 9628	AF479326 NR_026364
Knoel (A40a)	99 98 98	<i>Janibacter</i> -like ocean isolate sp. V4.BO.43 <i>Knoellia sinensis</i> DSM 12331 (Chinese cave) <i>Knoellia subterranean</i> DSM 12332 (Chinese cave)	AJ244674 AJ294412 AJ294413
Micro-1 (A93a)	99 96 96	Antarctic lake Fryxell mat bacterium R-8287 Guliya Glacier, China glacial ice isolate bacterium G200-C11 <i>Frigoribacterium faeni</i> DSM 10309	AJ440992 AF479342 AM410686
Micro-2 (A137)	99 96 96	Antarctic lake Fryxell mat bacterium R-8287 Guliya Glacier, China glacial ice isolate bacterium G200-C11 <i>Frigoribacterium faeni</i> DSM 10309	AJ440992 AF479342 AM410686
Nocard (A117)	97 95	<i>Nocardioides terrigena</i> Sajama Glacier, Bolivia, glacial ice bacterium SB12K-2-4	EF363712 AF479362
Sphing (A45b)*	97 95	Yukon River, Alaska uncultured clone YU201F09 John Evans Glacier, Nunavut uncultured clone JEG.b6	FJ694661 DQ228392

*Only partial sequence determined due to loss of culture viability

2.4. Discussion

Many different bacteria were isolated from VUG glacial ice (Table 2.2), particularly at low temperature, a condition which affected both the number and types of taxa recovered (Figure 2.1 and Table 2.1). Compared to glacial ice, particularly high numbers of bacteria were recovered from immediately above the basal-glacial ice interface (Table 2.1). This result corresponds well to the results obtained by Barker et al. (2006), whereby dissolved organic carbon (DOC) concentrations were relatively low throughout the VUG transect except for one spike approximately ten times greater in concentration immediately above the basal-glacial ice interface. Based upon fluorescence spectroscopy, the authors inferred from this DOC spike the presence of aromatic amino acids, reflecting recent biogenic activity (Barker et al. 2006). This is congruent with the isolation of relatively high numbers of viable microbes from the analogous 220-230 cm sample studied here (Table 2.1). Also present, especially in basal ice layers, were fluorescence spectra congruent with humic and fulvic acids, likely originating

from overridden sediment (Barker et al. 2006). The observed DOC and microbial abundance at the basal-glacial ice interface may be a result of upward pore water flow induced during the freezing-on of basal ice in this cold-based glacier (Souchez et al. 2004); why this resulted in a single DOC peak at VUG remains unclear.

The identities of the VUG isolates recovered in this study (Table 2.2) reveal both the potential of glacial ice to select for certain groups of microorganisms and the impact of local conditions on microbial provenance because of particular microbial abundance at that site or unique selective conditions (e.g., Skidmore et al. 2005, Mikucki and Priscu 2007). Several taxa isolated from VUG (*Sphingomonas*-like α -*Proteobacteria*, *Commomonas*-like β -*Proteobacteria*, *Arthrobacter*, *Flavobacterium*, *Frigoribacterium*, *Janthinobacterium*, *Kocuria*, *Microbacterium*) have also been isolated from other glaciers worldwide (Christner et al. 2000, Foght et al. 2004, Miteva et al. 2004, Miteva and Brenchley 2005, Skidmore et al. 2005, Xiang et al. 2005, Yao et al. 2006, Cheng and Foght 2007, Mikucki and Priscu 2007, Zhang et al. 2008b, Zhang et al. 2008c). In contrast, *Hymenobacter* and *Deinococcus* are much better known from Antarctic soils (e.g., Adams et al. 2006, Aislabie et al. 2006), and may represent microorganisms preferentially deposited specifically into or onto VUG. Similarly, abundant *Hymenobacter*-like organisms have been described from a Tibetan glacier (Zhang et al. 2008b), surrounded by deserts from which several *Hymenobacter* species have been described (Zhang et al. 2007, Zhang et al. 2008a, Zhang et al. 2009). In both cases, local *Hymenobacter* abundance is likely important for deposition into or onto glacial ice.

Unfortunately, insufficient sample availability (e.g., the lack of VUG basal ice samples) makes definitive determination of the effect of VUG basal ice geochemistry on its microbiology impossible. Furthermore, the low sample volumes available and low microbial abundance (assuming a high culturability of VUG microbes as reported previously using the same methods; Foght et al. 2004) suggest that application of non-culture-based microbiological techniques was not

feasible. The abundance and novelty of *Hymenobacter*-like isolates recovered from VUG, however, suggests the importance of their further study; this therefore became a major focus of my thesis research.

2.5. Literature cited

- Abyzov, S., Mitskevich, I.N., Poglazova, M.N., Barkov, N.I., Lipenkov, V.Y., Bobin, N.E., Kudryashov, B.B., and Pashkevich, V.M. 1999. Antarctic ice sheet as an object for methodological problems of exobiology. *Adv. Space Res.*, 23: 371-376.
- Abyzov, S.S. 1993. Microorganisms in the Antarctic ice. *In Antarctic microbiology*. E.I. Friedmann and A.B. Thistle. (ed.) Wiley-Liss, New York. pp. 265-295.
- Abyzov, S.S., Hoover, R.B., Imura, S., Mitskevich, I.N., Naganuma, T., Poglazova, M.N., and Ivanov, M.V. 2004. Use of different methods for discovery of ice-entrapped microorganisms in ancient layers of the Antarctic glacier. *Adv. Space Res.*, 33: 1222-1230.
- Abyzov, S.S., Mitskevich, I.N., Poglazova, M.N., Barkov, N.I., Lipenkov, V.Y., Bobin, N.E., Koudryashov, B.B., and Pashkevich, V.M. 1998. Antarctic Ice Sheet as a model in search of life on other planets. *Adv. Space Res.*, 22: 363-368.
- Abyzov, S.S., Mitskevich, I.N., Poglazova, M.N., Barkov, N.I., Lipenkov, V.Y., Bobin, N.E., Koudryashov, B.B., Pashkevich, V.M., and Ivanov, M.V. 2001. Microflora in the basal strata at Antarctic ice core above the Vostok Lake. *Adv. Space Res.*, 28: 701-706.
- Adams, B.J., Bardgett, R.D., Ayres, E., Wall, D.H., Aislabie, J., Bamforth, S., Bargagli, R., Cary, C., Cavacini, P., Connell, L., Convey, P., Fell, J.W., Frati, F., Hogg, I.D., Newsham, K.K., O'Donnell, A., Russell, N., Seppelt, R.D., and Stevens, M.I. 2006. Diversity and distribution of Victoria Land biota. *Soil Biol. Biochem.*, 38: 3003-3018.
- Aislabie, J.M., Chhour, K.-L., Saul, D.J., Miyauchi, S., Ayton, J., Paetzold, R.F., and Balks, M.R. 2006. Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. *Soil Biol. Biochem.*, 38: 3041-3056.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.
- Barker, J.D., Sharp, M.J., Fitzsimons, S.J., and Turner, R.J. 2006. Abundance and dynamics of dissolved organic carbon in glacier systems. *Arct. Antarct. Alp. Res.*, 38: 163-172.
- Bhatia, M., Sharp, M., and Foght, J. 2006. Distinct bacterial communities exist beneath a high arctic polythermal glacier. *Appl. Environ. Microbiol.*, 72:

5838-5845.

- Bidle, K.D., Lee, S., Marchant, D.R., and Falkowski, P.G. 2007. Fossil genes and microbes in the oldest ice on Earth. *Proc. Nat. Acad. Sci. USA*, 104: 13455-13460.
- Castello, J.D., Rogers, S.O., Starmer, W.T., Catranis, C.M., Ma, L.J., Bachand, G.D., Zhao, Y., and Smith, J.E. 1999. Detection of tomato mosaic tobamovirus RNA in ancient glacial ice. *Polar Biology*, 22: 207-212.
- Cheng, S.M., and Foght, J.M. 2007. Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.*, 59: 318-330.
- Christner, B.C., Kvitko II, B.H., and Reeve, J.N. 2003. Molecular identification of Bacteria and Eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles*, 7: 177-183.
- Christner, B.C., Mosley-Thompson, E., Thompson, L.G., and Reeve, J.N. 2001. Isolation of bacteria and 16S rDNAs from Lake Vostok accretion ice. *Environ. Microbiol.*, 3: 570-577.
- Christner, B.C., Skidmore, M.L., Priscu, J.C., Tranter, M., and Foreman, C.M. 2008. Bacteria in subglacial environments. *In Psychrophiles: from biodiversity to biotechnology*. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 51-71.
- Christner, B.C., Mosley-Thompson, E., Thompson, L.G., Zagorodnov, V., Sandman, K., and Reeve, J.N. 2000. Recovery and identification of viable bacteria immured in glacial ice. *Icarus*, 144: 479-485.
- Christner, B.C., Royston-Bishop, G., Foreman, C.M., Arnold, B.R., Tranter, M., Welch, K.A., Lyons, W.B., Tsapin, A.I., Studinger, M., and Priscu, J.C. 2006. Limnological conditions in Subglacial Lake Vostok, Antarctica. *Limnol. Oceanogr.*, 51: 2485-2501.
- Dancer, S.J., Shears, P., and Platt, D.J. 1997. Isolation and characterization of coliforms from glacial ice and water in Canada's High Arctic. *J. Appl. Microbiol.*, 82: 597-609.
- D'Elia, T., Veerapaneni, R., and Rogers, S.O. 2008. Isolation of microbes from Lake Vostok accretion ice. *Appl. Environ. Microbiol.*, 74: 4962-4965.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics*, 5: 164-166.
- Foght, J., Aislabie, J., Turner, S., Brown, C.E., Ryburn, J., Saul, D.J., and Lawson, W. 2004. Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb. Ecol.*, 47: 329-340.
- Hodson, A., Anesio, A.M., Tranter, M., Fountain, A., Osborn, M., Priscu, J., Laybourn-Parry, J., and Sattler, B. 2008. Glacial ecosystems. *Ecol. Monogr.*, 78: 41-67.

- Hodson, A.J., Mumford, P.N., Kohler, J., and Wynn, P.M. 2005. The High Arctic glacial ecosystem: new insights from nutrient budgets. *Biogeochemistry*, 72: 233-256.
- Huston, A.L. 2008. Biotechnological aspects of cold-adapted enzymes. *In Psychrophiles: from biodiversity to biotechnology*. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 347-363.
- Juck, D.F., Whissell, G., Steven, B., Pollard, W., McKay, C.P., Greer, C.W., and Whyte, L.G. 2005. Utilization of fluorescent microspheres and a green fluorescent protein-marked strain for assessment of microbiological contamination of permafrost and ground ice core samples from the Canadian High Arctic. *Appl. Environ. Microbiol.*, 71: 1035-1041.
- Karl, D.M., Bird, D.F., Björkman, K., Houlihan, T., Shackelford, R., and Tupas, L. 1999. Microorganisms in the accreted ice of Lake Vostok, Antarctica. *Science*, 286: 2144-2147.
- Kelly, M.A., Denton, G.H., and Hall, B.L. 2002. Late Cenozoic paleoenvironment in southern Victoria Land, Antarctica, based on a polar glaciolacustrine deposit in western Victoria Valley. *Geol. Soc. Am. Bull.*, 114: 605-618.
- Krembs, C., and Demming, J.W. 2008. The role of exopolymers in microbial adaptation to sea ice. *In Psychrophiles: from biodiversity to biotechnology*. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 247-264.
- Lanoil, B., Skidmore, M., Priscu, J.C., Han, S., Foo, W., Vogel, S.W., Tulaczyk, S., and Engerhardt, H. 2009. Bacteria beneath the West Antarctic Ice Sheet. *Environ. Microbiol.*, 11: 609-615.
- Lavire, C., Normand, P., Alekhina, I., Bulat, S., Prieur, D., Birrien, J.-L., Fournier, P., Hänni, C., and Petit, J.-R. 2006. Presence of *Hydrngenophilus thermoluteolus* DNA in accretion ice in the subglacial Lake Vostok, Antarctica, assessed using *rrs*, *cbb* and *hox*. *Environ. Microbiol.*, 8: 2106-2114.
- Mader, H.M., Pettitt, M.E., Wadham, J.L., Wolff, E.W., and Parkes, R.J. 2006. Glacial subsurface as a microbial habitat. *Geology*, 34: 169-172.
- Mikucki, J.A., and Priscu, J.C. 2007. Bacterial diversity associated with Blood Falls, a subglacial outflow from the Taylor Glacier, Antarctica. *Appl. Environ. Microbiol.*, 73: 4029-4039.
- Mikucki, J.A., Foreman, C.M., Sattler, B., Lyons, W.B., and Priscu, J.C. 2004. Geomicrobiology of Blood Falls: an iron-rich saline discharge at the terminus of the Taylor Glacier, Antarctica. *Aquat. Geochem.*, 10: 199-220.
- Mikucki, J.A., Pearson, A., Johnston, D.T., Turchyn, A.V., Farquhar, J., Schrag, D.P., Anbar, A.D., Priscu, J.C., and Lee, P.A. 2009. A contemporary microbially maintained subglacial ferrous "ocean". *Science*, 324: 397-400.

- Miteva, V. 2008. Bacteria in snow and glacier ice. *In* Psychrophiles: from biodiversity to biotechnology. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 31-50.
- Miteva, V., Sowers, T., and Brenchley, J. 2007. Production of N₂O by ammonia oxidizing bacteria at subfreezing temperatures as a model for assessing the N₂O anomalies in the Vostok ice core. *Geomicrobiol. J.*, 24: 451-459.
- Miteva, V., Teacher, C., Sowers, T., and Brenchley, J. 2009. Comparison of the microbial diversity at different depths of the GISP2 Greenland ice core in relationship to deposition climates. *Environ. Microbiol.*, 11: 640-656.
- Miteva, V.I., and Brenchley, J.E. 2005. Detection and isolation of ultrasmall microorganisms from a 120,000-year-old Greenland glacier ice core. *Appl. Environ. Microbiol.*, 71: 7806-7818.
- Miteva, V.I., Sheridan, P.P., and Brenchley, J.E. 2004. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl. Environ. Microbiol.*, 70: 202-213.
- Nichols, D., Bowman, J., Sanderson, K., Mancuso Nichols, C., Lewis, T., McMeekin, T., and Nichols, P.D. 1999. Developments with Antarctic microorganisms: culture collections, bioactivity screening, taxonomy, PUFA production and cold-adapted enzymes. *Curr. Opin. Biotechnol.*, 10: 240-246.
- Porazinska, D.L., Fountain, A.G., Nysten, T.H., Tranter, M., Virginia, R.A., and Wall, D.H. 2004. The biodiversity and biogeochemistry of cryoconite holes from McMurdo Dry Valley Glaciers, Antarctica. *Arct. Antarct. Alp. Res.*, 36: 84-91.
- Price, P.B. 2000. A habitat for psychrophiles in deep Antarctic ice. *Proc. Natl. Acad. Sci. U. S. A.*, 97: 1247-1251.
- Price, P.B. 2007. Microbial life in glacial ice and implications for a cold origin of life. *FEMS Microbiol. Ecol.*, 59: 217-231.
- Price, P.B. 2009. Microbial genesis, life and death in glacial ice. *Can. J. Microbiol.*, 55: 1-11.
- Price, P.B., and Sowers, T. 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc. Natl. Acad. Sci. U. S. A.*, 101: 4631-4636.
- Priscu, J.C., Adams, E.E., Lyons, W.B., Voytek, M.A., Mogk, D.W., Brown, R.L., McKay, C.P., Takacs, C.D., Welch, K.A., Wolf, C.F., Kirshtein, J.D., and Avci, R. 1999. Geomicrobiology of subglacial ice above Lake Vostok, Antarctica. *Science*, 286: 2141-2144.
- Rodrigues, D.F., Jesus, E.d.C., Ayala-del-Río, H.L., Pellizari, V.H., Gilichinsky, D., Sepulveda-Torres, L., and Tiedge, J.M. 2009. Biogeography of two cold-adapted genera: *Psychrobacter* and *Exiguobacterium*. *ISME J.*, 3: 658-665.

- Rogers, S.O., Starmer, W.T., and Castello, J.D. 2004. Recycling of pathogenic microbes through survival in ice. *Med. Hypotheses*, 63: 773-777.
- Rohde, R.A., and Price, P.B. 2007. Diffusion-controlled metabolism for long-term survival of single isolated microorganisms trapped within ice crystals. *Proc. Nat. Acad. Sci. USA*, 104: 16592-16597.
- Rohde, R.A., Price, P.B., Bay, R.C., and Bramall, N.E. 2008. *In situ* microbial metabolism as a cause of gas anomalies in ice. *Proc. Nat. Acad. Sci. USA*, 105: 8667-8672.
- Sharp, M., Parkes, J., Cragg, B., Fairchild, I.J., Lamb, H., and Tranter, M. 1999. Widespread bacterial populations at glacier beds and their relationship to rock weathering and carbon cycling. *Geology*, 27: 107-110.
- Simon, C., Herath, J., Rockstroh, S., and Daniel, R. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.*, 75: 2964-2968.
- Skidmore, M., Anderson, S.P., Sharp, M., Foght, J., and Lanoil, B.D. 2005. Comparison of microbial community compositions of two subglacial environments reveals a possible role for microbes in chemical weathering processes. *Appl. Environ. Microbiol.*, 71: 6986-6997.
- Skidmore, M.L., Foght, J.M., and Sharp, M.J. 2000. Microbial life beneath a high arctic glacier. *Appl. Environ. Microbiol.*, 66: 3214-3220.
- Souchez, R., Samyn, D., Lorrain, R., Pattyn, F., and Fitzsimons, S. 2004. An isotopic model for basal freeze-on associated with subglacial upward flow of pore water. *Geophys. Res. Lett.*, 31: L02401.
- Staley, J.T., and Gosink, J.J. 1999. Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu. Rev. Microbiol.*, 53: 189-215.
- Stewart Johnson, S., Hebsgaard, M.B., Christensen, T.R., Mastepanov, M., Nielsen, R., Munch, K., Brand, T., Glibert, M.T.P., Zuber, M.T., Bunce, M., Rønn, R., Gilichinsky, D., Froese, D., and Willerslev, E. 2007. Ancient bacteria show evidence of DNA repair. *Proc. Nat. Acad. Sci. USA*, 104: 14401-14405.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25: 4876-4882.
- Tranter, M., Skidmore, M., and Wadham, J. 2005. Hydrological controls on microbial communities in subglacial environments. *Hydrol. Process.*, 19: 995-998.
- Tung, H.C., Bramall, N.E., and Price, P.B. 2005. Microbial origin of excess methane in glacial ice and implications for life on Mars. *Proc. Natl. Acad. Sci. U. S. A.*, 102: 18292-18296.

- Tung, H.C., Price, P.B., Bramall, N.E., and Vrdoljak, G. 2006. Microorganisms metabolizing on clay grains in 3-km-deep Greenland basal ice. *Astrobiology*, 6: 69-86.
- Wadham, J.L., Bottrell, S., Tranter, M., and Raiswell, R. 2004. Stable isotope evidence for microbial sulphate reduction at the bed of a polythermal high Arctic glacier. *Earth Planet. Sci. Lett.*, 219: 341-355.
- Wadham, J.L., Cooper, R.J., Tranter, M., and Bottrell, S. 2007. Evidence for widespread anoxia in the proglacial zone of an Arctic glacier. *Chem. Geol.*, 243: 1-15.
- Wadham, J.L., Tranter, M., Tulaczyk, S., and Sharp, M. 2008. Subglacial methanogenesis: a potential climatic amplifier? *Global Biogeochem. Cy.*, 22: GB2021.
- Wynn, P.M., Hodson, A., and Heaton, T. 2006. Chemical and isotopic switching within the subglacial environment of a High Arctic glacier. *Biogeochemistry*, 78: 173-193.
- Xiang, S., Yao, T., An, L., Xu, B., and Wang, J. 2005. 16S rRNA sequences and differences in bacteria isolated from the Muztag Ata Glacier at increasing depths. *Appl. Environ. Microbiol.*, 71: 4619-4627.
- Yao, T., Xiang, S., Zhang, X., Wang, N., and Wang, Y. 2006. Microorganisms in the Malan ice core and their relation to climatic and environmental changes. *Global Biogeochem. Cy.*, 20: GB1004.
- Zhang, G., Niu, F., Busse, H.-J., Ma, X., Liu, W., Dong, M., Feng, H., An, L., and Cheng, G. 2008a. *Hymenobacter psychrotolerans* sp. nov., isolated from the Qinghai-Tibet Plateau permafrost region. *Int. J. Syst. Evol. Microbiol.*, 58: 1215-1220.
- Zhang, L., Dai, J., Tang, Y., Luo, X., Wang, Y., An, H., Fang, C., and Zhang, C. 2009. *Hymenobacter deserti* sp. nov., isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 59: 77-82.
- Zhang, Q., Liu, C., Tang, Y., Zhou, G., Shen, P., Fang, C., and Yokota, A. 2007. *Hymenobacter xinjiangensis* sp. nov., a radiation-resistant bacterium isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 57: 1752-1756.
- Zhang, X., Ma, X., Wang, N., and Yao, T. 2008b. New subgroup of *Bacteroidetes* and diverse microorganisms in Tibetan plateau glacial ice provide a biological record of environmental conditions. *FEMS Microbiol. Ecol.*, 67: 21-29.
- Zhang, X.F., Yao, T.D., Tian, L.D., Xu, S.J., and An, L.Z. 2008c. Phylogenetic and physiological diversity of bacterial isolated from Puruogangri ice core. *Microb. Ecol.*, 55: 476-488.

3. Genotypic and Phenotypic Characterization of Novel *Hymenobacter* Species

3.1. Introduction

As currently described, the genus *Hymenobacter* consists of thirteen identified species: *H. roseosalivarius* (Hirsch et al. 1998); *H. actinosclerus* (Collins et al. 2000); *H. aerophilus* (Buczolits et al. 2002); *H. norwichensis*, *H. chitinivorans*, *H. gelipurpurascens* and *H. ocellatus* (Reichenbach 1992, Buczolits et al. 2006); *H. rigui* (Baik et al. 2006); *H. xinjiangensis* (Zhang et al. 2007); *H. soli* (Kim et al. 2008); *H. psychrotolerans* (Zhang et al. 2008a); *H. deserti* (Zhang et al. 2009b); and *H. daecheongensis* (Xu et al. 2009). *H. chitinivorans*, *H. gelipurpurascens* and *H. ocellatus* were originally classified informally as “*Taxeobacter*” species (Reichenbach 1992), but subsequent and more rigorous analysis indicated their similarity with known *Hymenobacter* species and led to their reclassification within this genus (Buczolits et al. 2006). Several unnamed strains have also been isolated; those relevant to this study are NS/2 (Buczolits et al. 2006), P3 (Aislabie et al. 2006) and 35/26 (Saul et al. 2005). Most *Hymenobacter* species have been isolated from cold or otherwise oxidatively stressed and desiccated environments such as aerosols (Section 1.1).

According to 16S rRNA gene phylogenies, all *Hymenobacter* species form a monophyletic cluster within the *Sphingobacteria* sister to the *Flexibacteraceae* with high bootstrap support (Rickard et al. 2005, Suresh et al. 2006). *Adhaeribacter* branches at the base of this lineage (Rickard et al. 2005, Zhang et al. 2009a), with *Effluviibacter* (Suresh et al. 2006) and *Pontibacter* (Nedashkovskaya et al. 2005, Zhou et al. 2007, Zhang et al. 2008b) branching as sister taxa to *Hymenobacter*. Little is known about these microorganisms aside from their taxonomy, but their deep branching position relative to all other *Flexibacteraceae*, suggests their physiological and ecological divergence from better studied taxa, assuming that their 16S rRNA gene phylogeny accurately reflects their evolution

Most studies of *Hymenobacter* and related microorganisms have been conducted in isolation: i.e., comparisons are made to literature reports but not to parallel experiments. These deficiencies argue for further systematic study of these taxa. Previous research (Chapter 2) described the isolation from Victoria Upper Glacier, Antarctica (VUG) of 12 unique *Hymenobacter*-related sequence groups defined by ARDRA fingerprinting and preliminary 16S rRNA gene sequence analysis. Further systematic characterization of these VUG strains, 11 of the 13 described *Hymenobacter* species and strains NS/2, P3 and 35/26 was the objective of this study.

3.2. Materials and methods

3.2.1. Strain selection

In total, 58 *Hymenobacter*-like strains were isolated from VUG (Chapter 2). Unfortunately, the viability of several of these strains was lost during the course of my thesis research; because of this no isolate from sequence group Hym-11 was available for further study. To ensure that only unique strains were further characterized, enterobacterial repetitive intergenic consensus (ERIC)- and repetitive extragenic palindromic (REP)- (De Bruijn 1992) PCR analyses were used to discriminate between highly related strains. Genomic DNA was extracted according to Foght et al. (2004) and used as templates for PCR reactions using the primers ERIC1R and ERIC2 for ERIC-PCR and REP1R-I and REP2-I for REP-PCR (see Table 3.1 for sequences). Reaction mixtures and PCR conditions were those of De Bruijn (1992). The resulting PCR products were visualized by staining with SYBR-Green (Molecular Probes) premixed with PCR products prior to loading, according to the manufacturer's instructions, with fluorescence excitation at 473 nm and detection using a Fujifilm FLA-5000 scanner with a long pass filter following 2% agarose electrophoresis at 96 V for 2.5 h. Only strains with visually unique PCR product banding patterns in both methods were used for further study.

3.2.2. Genotypic analysis

As described previously (Chapter 2), 16S rRNA gene sequences were determined according to Foght et al. (2004) and Cheng and Foght (2007) using primers PB36F and PB38R for PCR amplification and the internal primers 16S1F, 16S2R, 16S3F, 16S4R and 16S5R for sequencing (Table 3.1). Part of the *gyrB* gene encoding the DNA gyrase β subunit was also amplified by PCR using the primers UP-1 and UP-2r (Yamamoto and Harayama 1995), the same reaction mixture used for the 16S rRNA gene (Foght et al. 2004) and the PCR reaction program of Yamamoto and Harayama (1995). Sequencing of *gyrB* amplicons was conducted at MBSU using the Big Dye method (Applied Biosystems) according to standard protocols using the primers UP-1S, UP-2Sr (Yamamoto and Harayama 1995), *gyrB*intF and *gyrB*intR (Table 3.1). GenBank accession numbers for the *gyrB* sequences determined here are GQ454807-GQ454840.

All sequences were assembled using the PREGAP v1.5 and GAP4 v4.10 programs of the Staden package (Dear and Staden 1991), ensuring that the *gyrB* sequence encoded an open reading frame. Assembled partial 16S rRNA gene sequences were aligned using CLUSTAL_X v2.0 (Thompson et al. 1997). Partial *gyrB* sequences were aligned as translated peptides using CLUSTALW

Table 3.1. Sequences for primers used in this study.

Primer name	Primer sequence (5' → 3')	Reference
ERIC1R	ATGTAAGCTCCTGGGGATTCAC	(De Bruijn 1992)
ERIC2	AAGTAAGTGA CTGGGGTGAGCG	(De Bruijn 1992)
REP1R-I	IIICGICGICATCIGGC	(De Bruijn 1992)
REP2-I	ICGICTTATCIGGCCTAC	(De Bruijn 1992)
PB36F	AGRGTTTGATCMTGGCTCAG	(Foght et al. 2004)
PB36R	GKTACCTTGTTACGACT	(Foght et al. 2004)
16S1F	ACTCCTACGGGAGGCAGCAG	(Cheng and Foght 2007)
16S2R	GTATTACCGCGGCTGCTGGCA	(Cheng and Foght 2007)
16S3F	GGATTAGATACCKGGTAGTCC	(Cheng and Foght 2007)
16S4R	GGTTAAGTCCCGCAACGAGC	(Cheng and Foght 2007)
16S5R	GCTCGTTGCGGGACTTAACC	(Cheng and Foght 2007)
UP-1	GAATCATCATGACCGTTCTGCAYGCNGG NAARTTY	(Yamamoto and Harayama 1995)
UP-2r	AGCAGGGTACGGATGTGCGAGCCRTCNG CRTCNGCRTCNGTC	(Yamamoto and Harayama 1995)
UP-1S	GAATCATCATGACCGTTCTGGA	(Yamamoto and Harayama 1995)
UP-2Sr	AGCAGGGTACGGATGTGCGAGCC	(Yamamoto and Harayama 1995)
<i>gyrB</i> intF	GCSGTSAGRCCYTCSGGAAGTC	This study
<i>gyrB</i> intR	AACTACAAYAAAY TGCATC	This study

(Thompson et al. 1994) as implemented in MEGA v4.0 (Tamura et al. 2007). All

sequences were trimmed to eliminate primer sequences and to ensure homogeneous sequence starting and ending positions. Insertions were observed using GENDOC (<http://www.nrbsc.org/gfx/genedoc/>) but removed prior to tree construction.

Bootstrapped maximum likelihood trees were generated using RAxML (Stamatakis et al. 2008) as implemented through the CIPRES web portal (<http://www.phylo.org/>). Trees were rooted using *Cytophaga hutchinsonii* ATCC 33406 (genome accession number: NC_008255) as an outgroup and the number of bootstrap replicates was determined automatically. Because the maximum likelihood 16S rRNA gene tree had unusually low overall bootstrap values (typically <60%), a neighbor joining tree for this gene was also constructed using PHYLIP v.3.65 (Felsenstein 1989); see Section 4.4.2 for a more detailed description of this method. Trees generated using either method were essentially congruent with each other.

3.2.3. Phenotypic characterization

All *Hymenobacter* strains were grown on R2 agar (R2A; Difco) at 18°C for 1 week in the dark; these conditions were approximately optimal for growth of all VUG *Hymenobacter*-like strains. R2A contains yeast extract, proteose peptone, casamino acids, glucose, soluble starch (all 0.5 g/L), sodium pyruvate, dipotassium phosphate (both 0.3 g/L), magnesium phosphate (0.05 g/L) and agar (15 g/L); the pH was approximately 7.2. Reference cultures typically grew more quickly than environmental isolates. Liquid culture techniques were not used because of the poor growth of many *Hymenobacter* and related strains, particularly those isolated from VUG, in all liquid media tested.

Growth-permissive temperatures were assessed using R2A incubated in the dark at 4, 10, 18, 28 or 37°C. Growth on more nutrient-rich media was examined using Luria-Bertani agar (contains 10 g tryptone/L (Difco), 10 g NaCl/L (EMD), 5 g yeast extract/L (Difco) and 15 g agar/L (Difco); pH 7.5) and trypticase soy agar (TSA; contains 30 g trypticase soy broth/L (BBL) and 15 g agar/L; pH not adjusted). Anaerobic and microaerophilic growth was assayed

using aerobically-prepared R2A plates incubated in sealed jars containing Anaerocult A and C packages (BBH), respectively. Halotolerance was tested using R2A supplemented with 0.5, 1, 2, 3 or 4% NaCl, and pH tolerance was tested using R2A agar adjusted to pH 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 \pm 0.2 using 0.22 μ m Millipore filter-sterilized HCl or NaOH. The pH of these media was determined following autoclaving; R2A having a pH of 10, 11 or 12 was discolored after autoclaving, but this color change was not observed during pH adjustment of neutral pH media following autoclaving. Sterility of these plates was ensured by incubation at room temperature for several days with no observed colony formation. Except for temperature tolerance experiments, all incubations were conducted at 18°C in the dark and scored after 1 and 2 weeks, using as inocula cultures grown on R2A at 10°C in the dark for 1 week (by which time robust growth was just beginning to appear) to ensure their active growth. This is hereafter referred to as “standard growth conditions”.

Colony morphologies were determined on R2A incubated and inoculated as above. These same cultures were used for microscopic examination using an AxioStar Plus microscope equipped with A-plan objective lenses and an AxioCam ICc 1 digital camera (Zeiss). Images were recorded using the AxioVisionLE software v.4.6.1.0 (Zeiss), which was also used for automatic adjustment of the white balance and image capture parameters. Cell size was estimated using the length measurement tool in the AxioVisionLE software. Gram stains were conducted according to standard methods using the BD Gram stain kit. The KOH test was conducted by suspending cells in 3% potassium hydroxide (EM Science) spotted on glass microscope, repeated stirring and testing for stringiness of the suspension when touched with a loop. The presence of catalase was tested by resuspending cells grown as above in 3% hydrogen peroxide (BDH) on a glass microscope slide with a positive test scored as the production of bubbles. The oxidase test was conducted using the same cultures and 1% *N,N,N',N'*-tetramethyl-*p*-phenylenediamine dihydrochloride (Eastman Organic Chemicals) spotted on Whatman #1 filter paper. A positive test was scored when the spotted

reagent turned purple within 5 sec when a loopful of colony was streaked on the paper.

3.2.4. Fatty acid determination

Whole-cell fatty acid composition was determined as fatty acid methyl esters (FAMES) essentially according to Embley and Wait (1994) in triplicate using standard growth conditions. For each extraction, a generous loopful of cells was inoculated into a 15 x 150 mm screw-capped tube with a Teflon liner containing 1 mL of 15% (w/v) NaOH in 50% (v/v) aqueous HPLC-grade methanol (Fisher) and incubated for 30 min at 100°C with vortexing after 5 min. Tubes were cooled to room temperature, 2 mL of 6M HCl in 50% aqueous methanol was added, incubated at 80°C for 10 min and cooled rapidly on ice. Once cooled, 1.25 mL of 1:1 hexanes/methyl-*tert*-butyl ether (both HPLC-grade; Fisher) was added, the tubes briefly shaken by hand and further shaken for 10 min at room temperature on an open top shaker. The bottom (aqueous) layer, including any white precipitate, was removed from these tubes, then 3 mL of 1.2% NaOH was added and the tubes mixed as above but for only 5 min; all remaining white precipitate dissolved during this step. The top (organic) layer was transferred into 2-dram vials, dried under N₂ gas and stored at -20°C until analyzed.

Prior to analysis, FAMES were dissolved in 200 µL of HPLC-grade dichloromethane (Fisher) and transferred to vials containing glass small-volume inserts. Extracts were analyzed by gas chromatography-mass spectrometry (GC-MS) using an Agilent 6890N gas chromatograph with an Agilent HP-5MS 0.25 mm x 30 m x 0.25 µm column and an Agilent 5973 mass spectrometer. The inlet temperature was 250°C and gas flow was 28.8 mL/min at 10.7 psi. The temperature program was: 100°C held for 1 min; a 15°C/min increase to 170°C; a 1°C/min increase to 190°C; and a 25°C/min increase to 290°C which was held for 1 min. Integration parameters were: initial peak reject = 0; initial peak reject = 0.033; shoulder detection = off; initial threshold = 15; and integrator OFF = 26.000 min. Fatty acid methyl esters were identified by comparison to the

bacterial acid methyl ester standard mix (BAME; Supelco) and values from the literature. Note that hydroxylated FAMES in the BAME standard were either degraded or not resolved by the GC-MS protocol used; their contribution to the fatty acid profiles determined in this research remain unknown.

3.2.5. Carbon source utilization

The Biolog GN2 MicroPlate (Biolog Inc., Hayward CA) was used to determine the ability of *Hymenobacter* and related VUG strains to metabolize a wide range of carbon substrates. Using a sterile cotton swab, cultures grown under standard growth conditions were resuspended in GN/GP-IF inoculating fluid (Biolog Inc.) by rubbing on the tube sides until a turbidity of approximately 50% was reached (recorded using a turbidimeter; Biolog Inc.). Care was taken both to ensure dispersion of all visible cell clumps and to avoid introducing bubbles into the tube, which would artificially increase the apparent turbidity. Cell suspensions were poured into a 15 x 100 mm Petri plate and 150 μ L was used to inoculate each well of the Biolog plate, again taking care to avoid obvious clumps. Plates were read following 1 and 2 weeks incubation in the dark at 18°C using a Molecular Devices Emax plate reader (Biolog Inc.) by recording absorbance at 590 and 750 nm. In some cases, plates were read after 10 and 14 days because the plate reader was not working on day 7. Data were processed using default parameters using the proprietary Biolog database software attached to the plate reader. Following the day 14 analysis, plates were checked visually to ensure that evaporation from the wells was minimal and that positive results were not exclusively the result of cell clumping and capsule metabolism (explained further below).

3.3. Results

3.3.1. Phylogenetic analysis of *Hymenobacter* species and related strains using the 16S rRNA gene and *gyrB*

Of the 58 *Hymenobacter*-like strains isolated from VUG, 20 unique ERIC- and/or *rep*-PCR banding pattern types were observed (Supplemental Figure B1).

Representatives from each of these pattern types were selected for further characterization. Partial 16S rRNA gene sequences from these VUG and reference *Hymenobacter* strains were determined and analyzed phylogenetically (Figure 3.1). As befits their taxonomic standing, all *Hymenobacter* species and related VUG strains form a single monophyletic clade internal to the related genera *Adhaeribacter*, *Effluviibacter* and *Pontibacter*. The monophyletic clustering of *Hymenobacter* relative to the other analyzed clades is confirmed by the presence of 16S rRNA gene insertion “a” (Figure 3.1) present in these strains and absent in the outgroups. (As an aside, *Effluviibacter roseus* clusters with other *Pontibacter* strains, as reported previously (Zhang et al. 2008b), suggesting the need for reevaluation of the taxonomic standing of this genus.)

Hymenobacter and related VUG strains formed three well-supported clades in the 16S rRNA gene phylogeny: (i) one containing *H. ocellatus* and *H. deserti*, further supported by the unique presence of 16S rRNA gene insertion “b” in both strains; (ii) one containing most VUG sequences, *H. soli* and P3; and (iii) one containing all remaining VUG and most described *Hymenobacter* species. This latter clade was further bifurcated into two sequence groups, one containing *H. chitinivorans*, *H. daecheongensis*, *H. norwichensis*, *H. roseosalivarius* and strains 35/26, VUG-A23a, VUG-A60a, VUG-A112, VUG-A124, VUG-A141a and VUG-A142, and the other containing *H. actinosclerus*, *H. aerophilus*, *H. gelipurpurascens*, *H. psychrotolerans*, *H. rigui*, *H. xinjiangensis* and strains NS/2 and VUG-A106. Notable well-supported relationships within these groups include: (i) strains VUG-A23a, VUG-A124 and VUG-A142; (ii) strain VUG-A112 and *H. chitinivorans*; (iii) strains 35/26, VUG-A60a, VUG-A141a and *H. roseosalivarius*; and (iv) strain VUG-A106, *H. actinosclerus* and *H. aerophilus*.

Surprisingly, the phylogeny generated using partial *gyrB* gene sequences differed substantially from those generated using the 16S rRNA gene (Figure 3.2). Several clades occurred in both trees, containing: (i) strains 35/26, VUG-A60a, VUG-A141a and *H. roseosalivarius*; (ii) strain VUG-A106, *H. actinosclerus* and *H. aerophilus*; and (iii) *H. soli*, strain P3 and most VUG strains, also supported by the common presence of *gyrB* insertion “a”. Strongly supported clades present

in the *gyrB* but not the 16S rRNA gene tree contained: (i) strain VUG-A112 with VUG -A124 and VUG-A142 but neither strain VUG-A23a or *H. chitinivorans*; (ii) *H. chitinivorans*, *H. norwichensis*, *H. psychrotolerans*, *H. rigui* and strains NS/2 and VUG-A23a, further supported by the common presence of *gyrB* insertion “b”; (iii) *H. actinosclerus*, *H. aerophilus*, *H. gelipurpurascens*, *H. xinjiangensis* and strain VUG-A106, further supported by the common presence of *gyrB* insertion “c”; and (iv) the relationship suggested by the common presence

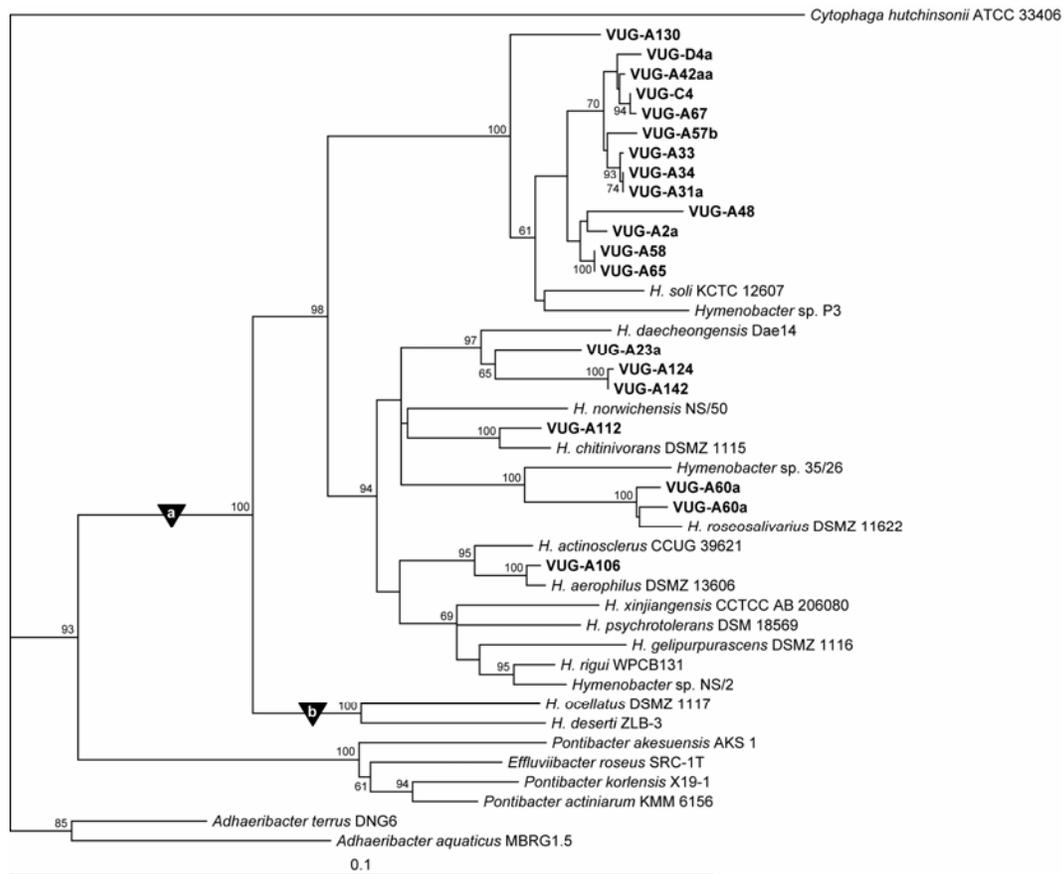


Figure 3.1. Neighbor-joining phylogenetic tree of near full-length (≈ 1500 nt) 16S rRNA gene sequences for VUG (bolded) and other described *Hymenobacter* strains. *C. hutchinsonii* and species from the genera *Adhaeribacter*, *Effluviibacter* and *Pontibacter* were included as outgroups. Only bootstrap values ≥ 60 are shown, and the scale bar represents 10% sequence divergence. Phylogenetically conserved insertions are shown by triangles on top of the insertion-containing branch; different letters indicate unique insertions. Those insertions present in only one sequence are not shown, and all were excluded from the alignment used to generate the phylogenetic tree.

of *gyrB* insertion “c” between clade (iii), *H. ocellatus* and the clade containing *H. soli*, strain P3 and most VUG strains. The deep branching of *H. ocellatus* suggested by 16S rRNA gene phylogenies was not present in the *gyrB* tree. These incongruencies suggest that either the true evolutionary history of *Hymenobacter* and related strains differs substantially from that suggested by 16S rRNA gene phylogenies or that *gyrB* has been frequently transferred horizontally within this genus.

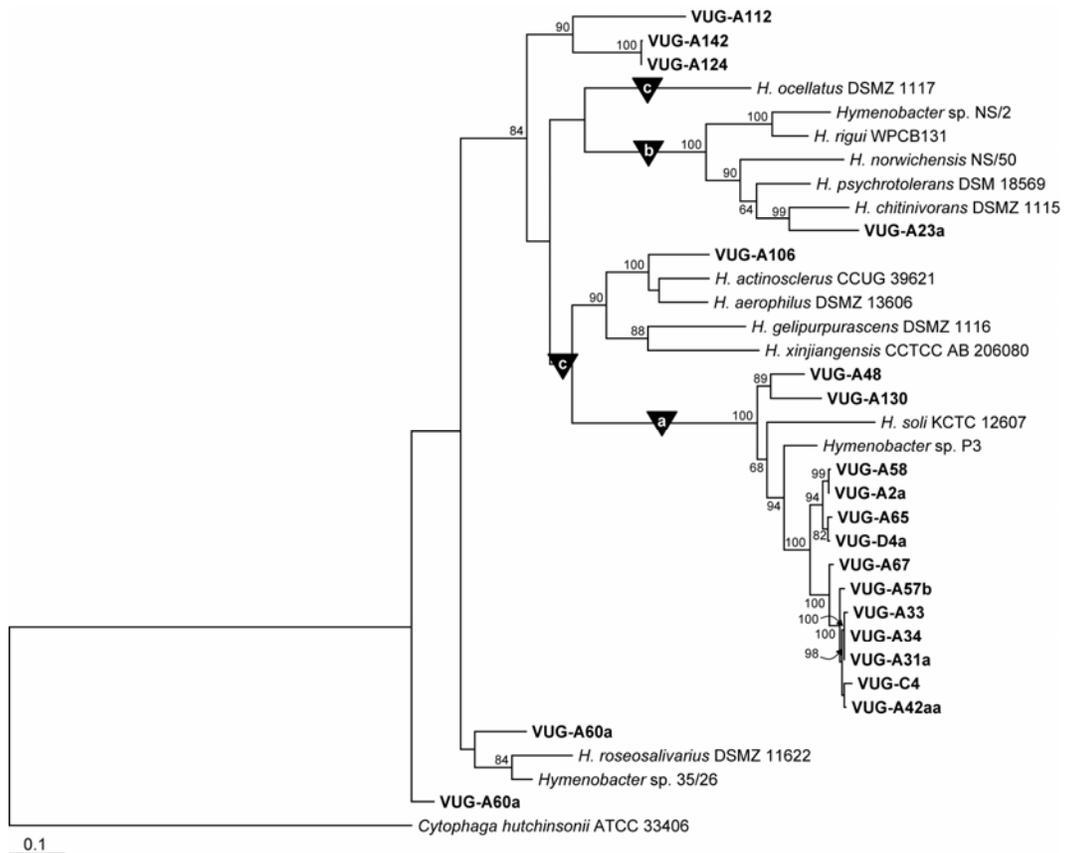


Figure 3.2. Maximum-likelihood phylogeny of partial (≈ 1200 nt) *gyrB* nucleotide sequences. *C. hutchinsonii* was included as an outgroup; no *gyrB* sequences from *Adhaeribacter*, *Effluviibacter* or *Pontibacter* were available for analysis. Only bootstrap values ≥ 60 are shown, and the scale bar represents 10% sequence divergence. Phylogenetically conserved insertions are shown by triangles on top of the insertion-containing branch; different letters indicate unique insertions. Those insertions present in only one sequence are not shown, and all were excluded from the alignment used to generate the phylogenetic tree.

3.3.2. Phenotypic characterization of *Hymenobacter* species and related VUG strains

To complement the genotypic characterization of *Hymenobacter* and related VUG strains discussed above, all strains were assayed for a wide variety of growth phenotypes including temperature, pH, oxygen depletion and salt tolerance, growth on rich media, oxidase and catalase reactions (Table 3.2) and colony and cellular morphology (Table 3.3). *Hymenobacter* and related strains are Gram negative by both staining and the KOH test and are oxidase positive. Most strains grow at both 4°C and 18°C, with some, particularly previously described species and related VUG isolates, also able to grow at 28°C. These latter strains were also more likely to produce catalase. In contrast to previously reported results (Reichenbach 1992, Buczolits et al. 2006), no strains were able to grow at 37°C; the reason for this discrepancy is unknown. Most strains poorly tolerated NaCl and rich media but could commence growth at high pH values. At extreme pH values many strains grew only poorly or changed colony morphology, e.g., enhanced capsule production. Whether observation of growth, therefore, truly reflects differential pH preference or differential protection from extreme pH values is uncertain. No strains were able to grow anaerobically, and the ability to grow microaerobically was variable.

Most *Hymenobacter* and related VUG strains were morphologically similar, existing as single bacilli or slight vibrios with an approximate size of 1.5-2 µm x 0.5-0.7 µm. Note that these measurements are approximations due to differing degrees of curvature between cells. Several strains (VUG-C4, *H. gelipurpurascens*, *H. rigui*, NS/2) also clustered along their long axes as palisades and some (VUG-A48, VUG-A112) were noticeably longer than their relatives. Whereas most VUG *Hymenobacter*-like strains formed small, smooth, shiny colonies, several reference strains exhibited much larger colony sizes, especially emphasized by their ability to spread into uninoculated regions of the plate. This characteristic may be due to sub-optimal growth conditions used for these strains e.g., due to sub-optimal pH (see above).

Table 3.2. Phenotypic characterization of *Hymenobacter* and reference VUG strains. The order in which strains are presented is the same as that in Figure 3.1. Except where otherwise indicated, all tests were conducted using the standard growth conditions.

Strain	Gram reaction	KOH test	Oxidase test	Catalase test	Growth permissive temperatures (°C)	Relative growth rate	Microaerobic growth	Anaerobic growth	Maximum % NaCl tolerated	Growth-permissive pH range	Growth on Luria-Bertani Agar	Growth on TSA
VUG-A130	-	+	+	+	4, 10, 18, 28	Slow	+	-	< 0.5	5-11	-	-
VUG-D4a	-	+	+	-	4, 10, 18	Slow	+	-	0.5 (w) ^d	6-12	-	-
VUG-A42aa	-	+	+	-	4, 10, 18	Slow	+	-	< 0.5	6-12	-	-
VUG-C4	-	+	+	-	4, 10, 18	Slow	+	-	0.5 (w)	6-12	-	-
VUG-A67	-	+	+	-	4, 10, 18	Slow	-	-	< 0.5	6-10	-	-
VUG-A57b	-	+	+	-	4, 10, 18	Slow	+	-	< 0.5	5-12	-	-
VUG-A33	-	+	+	-	4, 10, 18	Slow	+	-	0.5	6-11	-	-
VUG-A34	-	+	+	-	4, 10, 18	Slow	-	-	< 0.5	6-11	-	-
VUG-A31a	-	+	+	-	4, 10, 18	Slow	-	-	0.5 (w)	6-12	-	-
VUG-A48	-	+	+	-	4, 10, 18	Slow	+	-	0.5	5-11	-	-
VUG-A2a	-	+	+	-	4, 10, 18	Slow	+	-	< 0.5	5-11	-	-
VUG-A58	-	+	+	-	4, 10, 18	Slow	+	-	< 0.5	6-11	-	-
VUG-A65	-	+	+	-	4, 10, 18	Slow	+	-	< 0.5	6-11	-	-
<i>H. soli</i>	ND ^e	ND	ND	ND	4, 10, 18, 28	Fast	ND ^f	-	0.5	5-11	-	-
<i>Hymenobacter</i> sp. P3	-	+	+	+	4, 10, 18	Slow	-	-	< 0.5	5-11	-	-
VUG-A23a	-	+	+	-	4, 10, 18	Slow	-	-	< 0.5	6-11	-	-
VUG-A124	-	+	+	-	4, 10, 18	Slow	-	-	< 0.5	6-10	-	-
VUG-A142	-	+	+	-	4, 10, 18	Slow	-	-	< 0.5	7-10	-	-
<i>H. norwichensis</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	1	5-12	-	+

Strain	Gram reaction	KOH test	Oxidase test	Catalase test	Growth permissive temperatures (°C)	Relative growth rate	Microaerobic growth	Anaerobic growth	Maximum % NaCl tolerated	Growth-permissive pH range	Growth on Luria-Bertani Agar	Growth on TSA
VUG-A112	-	+	+	-	4 (w), 10, 18	Slow	-	-	< 0.5	6-11	-	-
<i>H. chitinivorans</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	1 (w)	5-12	w	-
<i>Hymenobacter</i> sp. 35/26	-	+	+	+	4, 10, 18, 28 (w)	Slow	-	-	0.5 (w)	7-11	-	-
VUG-A60a	-	+	+	+	4 (w), 10, 18, 28 (w)	Slow	-	-	1	7-9	-	-
VUG-A141a	-	+	+	+	4, 10, 18, 28	Slow	-	-	2 (w)	6-12	-	w
<i>H. roseosalivarius</i>	-	+	+	+	10, 18 (w)	Slow	-	-	< 0.5	6-12	-	-
<i>H. actinosclerus</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	2	5-12	+	-
VUG-A106	-	+	+	-	4, 10, 18, 28	Slow	-	-	3 (w)	6-12	+	+
<i>H. aerophilus</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	2	5-12	+	+
<i>H. xinjiangensis</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	1	5-11	+	+
<i>H. psychrotolerans</i>	ND	ND	ND	ND	4, 10, 18, 28	Fast	ND ^f	-	1 (w)	5-11	-	-
<i>H. gelipurpurascens</i>	-	+	+	+	4, 10, 18, 28	Fast	+	-	1	5-12	-	w
<i>H. rigui</i>	-	+	+	-	4, 10, 18, 28	Fast	+	-	1	5-12	-	+
<i>Hymenobacter</i> sp. NS/2	-	+	+	w	4, 10, 18, 28	Fast	+	-	1	5-12	w	-
<i>H. ocellatus</i>	-	+	+	+	10, 18, 28	Fast ^g	+	-	1	5-12	-	-

^aTemperatures tested: 4, 10, 18, 28 and 37°C

^bConcentrations tested: 0.5, 1, 2, 3, and 4% (w/v)

^cValues reflect the ability of each strain to commence growth at pH 3-12. The degree to which the cell alters the pH of the medium during growth is unknown.

^d(w): Weak growth

^eND: not done; ^fmanufacture of the Anaerocult C system used for these tests was discontinued before these strains could be assayed.

^gSlow and fast growth required approximately 7 and 3 days for robust growth to appear, respectively.

Table 3.3. Colony and cell morphology of *Hymenobacter* and related VUG strains grown using standard growth conditions. The order in which strains are presented is the same as that in Figure 3.1.

Strain	Cell morphology						Colony morphology						
	Morphology	Arrangement	Length (μm)	Width (μm)	Number of cells counted	Inclusions	Form	Elevation	Margin	Surface	Transparency	Diameter (mm)	Spreading
VUG-A130	Rods, slight vibrio	Singles	1.73 \pm 0.37	0.61 \pm 0.15	41	+	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.25-0.5	-
VUG-D4a	Rods, slight vibrio	Singles, pairs	1.63 \pm 0.41	0.68 \pm 0.12	31	-	Circular	Convex	Even	Shiny	Transparent	0.5-1.25	-
VUG-A42aa	Rods	Singles	1.71 \pm 0.40	0.63 \pm 0.12	60	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.5-0.75	-
VUG-C4	Rods, slight vibrio	Singles, pairs, palisades clusters	1.59 \pm 0.32	0.48 \pm 0.08	38	-	Circular	Flat	Even	Shiny	Transparent	0.5-1	-
VUG-A67	Rods, slight vibrio	Singles	1.54 \pm 0.29	0.72 \pm 0.16	33	+	Circular	Flat	Even	Shiny	Opaque, transparent edge	0.25-0.75	-
VUG-A57b	Rods	Singles	1.43 \pm 0.35	0.89 \pm 0.14	29	-	Circular	Flat	Even	Smooth	Opaque, transparent edge	0.25	-
VUG-A33	Rods, slight vibrio	Singles	2.12 \pm 0.64	0.73 \pm 0.12	48	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.5	-
VUG-A34	Rods, longer forms	Singles	1.91 \pm 1.40	0.79 \pm 0.15	29	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.25-0.75	-
VUG-A31a	Rods, slight vibrio	Singles	1.34 \pm 0.25	0.57 \pm 0.08	52	+	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.5-1	-
VUG-A48	Vibrio	Singles	4.23 \pm 1.22	1.04 \pm 0.16	15	-	Circular	Raised	Even	Dry	Opaque	0.25-0.5	-

Strain	Cellular morphology						Colony morphology						
	Morphology	Arrangement	Length (µm)	Width (µm)	Number of cells counted	Inclusions	Form	Elevation	Margin	Surface	Transparency	Diameter (mm)	Spreading
VUG-A2a	Rods	Singles	1.35 ± 0.27	0.63 ± 0.08	34	-	Circular	Flat	Even	Shiny	Opaque, transparent edge	0.25-1.25	-
VUG-A58	Rods, slight vibrio	Singles	1.39 ± 0.24	0.61 ± 0.11	46	-	Punctiform	Convex	Even	Shiny	Opaque, transparent edge	0.1-0.25	-
VUG-A65	Rods, slight vibrio	Singles	1.75 ± 0.85	0.57 ± 0.10	36	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.5	-
<i>Hymenobacter</i> sp. P3	Rounded rods	Singles	1.53 ± 0.29	0.69 ± 0.11	34	-	Circular	Convex	Even	Shiny	Transparent	0.1	-
VUG-A23a	Rods, slight vibrio	Singles	1.92 ± 0.43	0.60 ± 0.11	22	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.5-1.25	-
VUG-A124	Rods, some bent, slight vibrio	Singles	2.32 ± 0.98	0.60 ± 0.12	33	-	Punctiform	Convex	Even	Shiny	Opaque, transparent edge	0.1-0.5	-
VUG-A142	Rods, some bent, slight vibrio	Singles	1.82 ± 0.48	0.89 ± 0.14	24	-	Punctiform	Convex	Even	Shiny	Transparent	0.1-0.5	-
<i>H. norwichensis</i>	Rods	Singles	1.06 ± 0.25	0.72 ± 0.11	24	-	Irregular	Flat	Diffuse	Shiny	Opaque, two distinct transparent zones about transparent centre	1-5	+
VUG-A112	Vibrio	Singles, pairs	4.03 ± 1.95	0.65 ± 0.15	24	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.1-0.25	+
<i>H. chitinivorans</i>	Rod, slight vibrio	Singles	2.10 ± 0.40	0.54 ± 0.08	29	-	Irregular	Flat	Diffuse	Shiny	Transparent	~ 2	+
<i>Hymenobacter</i> sp. 35/26	Rods	Singles, clusters	1.53 ± 0.29	0.69 ± 0.11	22	-	Punctiform	Convex	Even	Shiny	Transparent	0.1-0.25	-
VUG-A60a	Rods, slight vibrio	Singles	1.66 ± 0.50	0.51 ± 0.09	38	+	Circular	Flat	Even	Shiny	Opaque, transparent edge	0.1	-

Strain	Cellular morphology						Colony morphology						
	Morphology	Arrangement	Length (µm)	Width (µm)	Number of cells counted	Inclusions	Form	Elevation	Margin	Surface	Transparency	Diameter (mm)	Spreading
VUG-A141a	Rods	Singles	1.82 ± 0.48	0.89 ± 0.14	24	+	Circular	Convex	Even	Shiny	Opaque, small transparent edge	0.1-0.25	-
<i>H. roseosalivarius</i>	Rods, slight vibrio	Singles	1.64 ± 0.59	0.71 ± 0.13	29	-	Circular	Convex	Even	Shiny	Transparent	0.1	-
<i>H. actinosclerus</i>	Rods	Singles	1.42 ± 0.29	0.73 ± 0.13	22	-	Circular	Flat	Even	Shiny	Transparent	1	+
VUG-A106	Rods, slight vibrio	Singles	1.47 ± 0.50	0.59 ± 0.10	31	-	Circular	Convex	Even	Shiny	Opaque, transparent edge	0.25-2	-
<i>H. aerophilus</i>	Rods, smeared ends	Singles	1.42 ± 0.29	0.73 ± 0.13	28	-	Irregular	Convex	Even	Shiny	Transparent	2-2.5	-
<i>H. xinjiangensis</i>	Rods	Singles	2.10 ± 0.56	0.83 ± 0.11	54	+	Filamentous	Convex	Filamentous	Shiny, wrinkled	Transparent	~ 15	+
<i>H. gelipurpurascens</i>	Rods	Singles, clusters in palisades	1.31 ± 0.20	0.54 ± 0.08	17	+	Irregular	Flat	Diffuse	Shiny	Transparent	~ 0.5	+
<i>H. rigui</i> WPCB131	Rods, slight vibrio	Singles, pairs, clusters in palisades	1.94 ± 0.31	0.64 ± 0.11	21	-	Irregular – three concentric zones separated by clearing	Convex	Diffuse	Shiny	Transparent	~ 1-5	+
<i>Hymenobacter</i> sp. NS/2	Rods, slight vibrio	Singles, pairs, clusters in palisades	1.67 ± 0.32	0.58 ± 0.07	23	-	Irregular	Flat	Diffuse	Shiny	Transparent	~ 1-5	+
<i>H. ocellatus</i>	Rods	Singles	1.43 ± 0.21	0.47 ± 0.08	41	-	Circular	Convex	Entire	Shiny	Opaque, two distinct transparent zones about transparent centre	2.5-7.5	ND

Despite some variability, the phenotypes (Table 3.2), cell and colony morphologies (Table 3.3) can be clustered relative to the gene phylogenies discussed above (Figures 3.1 and 3.2). The phylogenetic split between the majority of described *Hymenobacter* strains and their relatives and *H. soli*, P3 and related VUG strains is particularly well conserved. Compared to the latter, the former have greater tolerance of NaCl and rich media, higher growth temperature tolerance, greater abundance of catalase, faster relative growth rates and greater propensity to form large, spreading colonies on R2A incubated at 18°C in the dark. Minor phenotypic differences in VUG-A130, VUG-A48, P3 and *H. soli* relative to the main cluster of VUG isolates support their relationship to this cluster as outlying neighbors. Among other strains, the unusually high (for *Hymenobacter*) tolerance to salt and high nutrient concentrations of VUG-A106, *H. actinosclerus* and *H. aerophilus* is also congruent with their phylogenetic relatedness.

3.3.3. Characterization of the fatty acid composition of *Hymenobacter* species and related strains

Fatty acid compositions of several *Hymenobacter* species have been determined using a wide variety of growth conditions in previous literature reports. Because growth media and incubation temperature are both known to affect cellular fatty acid composition (Ponder et al. 2005), FAME analysis was conducted for *Hymenobacter* and related strains using the same growth conditions (Table 3.4). A caveat to this analysis is that the different growth rates of microorganisms within this collection (Table 3.2) could not be standardized among strains; due to the requirements for solid medium the effect of this variable remains unknown. Furthermore, the extent to which the 18°C growth temperature used causes physiologically similar responses among strains (e.g., optimal versus stressful growth conditions) is unknown but may differ given their different growth temperature ranges (Table 3.2).

Overall, fatty acid composition in *Hymenobacter* and related strains is quite homogeneous (Table 3.4), with the peaks eluting at 11.1, 11.3, 14.3, 14.6, 15.0

and 16.6 min prominent in all strains. However, strains which appear closely related according to gene phylogenies (Figures 3.1 and 3.2) can have particular fatty acids which vary considerably in their abundance (e.g., compare the 15.0 and 15.9 min peaks of VUG-A124 and VUG-A142 and note the variability between *H. soli*-like VUG strains; excluding VUG-A130 and VUG-A48, the percentage of the 14.3 min peak ranges from 29.1% - 53.6%; mean \pm stdev = 40.7% \pm 6.5%). This is especially true of iso- and anteiso-isomers of the same fatty acid (Table 3.5), the proportions of which adjust to maintain membrane homeostasis (Russell 2008, Zhang and Rock 2008). The extent to which apparent differences between fatty acid compositions are strain- or species-specific, or simply reflect differences in fine-scale tuning of membrane fluidity, is therefore difficult to determine where few strains have been analyzed in parallel, as is the case for most described *Hymenobacter* species. Previously described *Hymenobacter* species and related strains, however, can be robustly differentiated from the clade containing *H. soli* and the majority of novel VUG strains by a greater amount of the 14.0 and 16.6 min peaks in the former. In congruence with the 16S rRNA gene tree (Figure 3.1), the fatty acid composition of *H. ocellatus* is clearly distinct from that of all other *Hymenobacter* and related strains, containing low proportions of the 14.3 and 14.6 min peaks and exceptionally high proportions of the 13.8, 15.9 and 16.6 min peaks.

Considering the groups well supported in gene phylogenies (Figures 3.1 and 3.2), fatty acid composition supports the common grouping of most VUG strains with VUG-A48, VUG-A130 and *H. soli* as outlying neighbors, based upon lesser accumulation of the 16.6 min peak. A common relationship between

Table 3.4. (Next page) Whole cell fatty acid composition of *Hymenobacter* and related isolates. All strains were grown in triplicate using the standard growth conditions. All values are expressed as the mean percentage \pm one standard deviation, and only those present in all three replicates with a mean $>$ 1% are shown. Peaks present below this threshold but present above it in one other strain are indicated as “tr” (trace); those absent entirely are indicated using a dash (-). Strains are presented in the same order as that in Figure 3.1. See Table 3.5 for presumptive fatty acid identifications.

Retention time (min)	VUG-A130	VUG-D4a	VUG-A42aa	VUG-C4	VUG-A67	VUG-A57b	VUG-A33	VUG-A34	VUG-A31a	VUG-A48	VUG-A2a
7.5	tr	-	tr	-	tr	-	-	tr	tr	tr	tr
9.0	tr	tr	tr	2.4 ± 0.2	tr	tr	tr	1.7 ± 0.1	tr	tr	tr
9.7	tr	tr	1.2 ± 0.2	tr	1.3 ± 0.3	tr	tr	1.3 ± 0.1	1.1 ± 0.1	9.7 ± 1.2	tr
10.6	13.4 ± 0.9	1.8 ± 0.3	tr	tr	tr	1.9 ± 0.1	tr	2.4 ± 0.3	tr	2.4 ± 0.8	3.3 ± 0.2
10.7	tr	tr	tr	tr	tr	2.3 ± 0.1	tr	2.2 ± 0.2	1.0 ± 0.1	tr	tr
10.8	5.5 ± 0.4	tr	tr	tr	-	tr	tr	tr	tr	1.6 ± 0.6	tr
11.1	14.5 ± 0.3	9.8 ± 0.3	9.6 ± 0.6	9.9 ± 0.6	7.7 ± 0.1	19.5 ± 0.3	4.9 ± 1.1	15.8 ± 0.3	15.2 ± 0.3	11.2 ± 2.5	12.2 ± 0.4
11.3	21.5 ± 0.7	12.1 ± 0.2	10.1 ± 0.7	12.1 ± 0.5	10.1 ± 0.2	7.5 ± 0.1	4.4 ± 1.0	20.1 ± 0.5	15.7 ± 0.4	15.9 ± 0.4	12.2 ± 0.3
12.0	tr	tr	tr	tr	-	tr	tr	tr	tr	tr	tr
13.2	tr	tr	tr	5.8 ± 0.1	tr	tr	1.4 ± 0.2	1.4 ± 0.1	tr	tr	tr
13.6	3.9 ± 0.7	2.6 ± 0.4	tr	tr	tr	tr	3.5 ± 0.9	tr	tr	2.0 ± 0.5	1.5 ± 0.1
13.8	tr	tr	tr	1.2 ± 0.1	-	-	tr	tr	tr	-	tr
14.0	1.1 ± 0.2	tr	tr	tr	tr	tr	1.3 ± 0.2	tr	tr	tr	tr
14.3	21.6 ± 0.9	45.2 ± 2.1	49.5 ± 1.0	39.1 ± 0.8	53.6 ± 0.8	35.7 ± 0.5	37.7 ± 2.3	29.1 ± 0.3	38.3 ± 0.6	37.9 ± 2.5	39.1 ± 1.3
14.6	8.7 ± 1.9	19.0 ± 0.9	19.3 ± 0.6	20.3 ± 0.4	20.3 ± 0.6	16.9 ± 0.5	26.3 ± 1.6	15.2 ± 0.4	17.6 ± 0.1	11.9 ± 1.3	19.0 ± 0.3
14.8	-	tr	-	-	-	-	-	-	-	-	-
15.0	1.2 ± 1.1	2.0 ± 0.7	4.0 ± 1.2	2.5 ± 2.2	2.7 ± 0.8	2.3 ± 0.3	5.9 ± 2.1	1.1 ± 0.3	2.1 ± 0.3	4.2 ± 2.3	3.3 ± 0.8
15.9	tr	tr	-	tr	-	tr	tr	tr	-	tr	tr
16.3	tr	tr	tr	tr	-	-	-	tr	tr	tr	tr
16.6	6.4 ± 1.7	5.4 ± 1.0	6.2 ± 0.1	5.6 ± 0.7	3.2 ± 0.7	12.2 ± 0.1	5.6 ± 1.4	9.7 ± 0.8	9.0 ± 0.1	8.2 ± 0.5	5.1 ± 0.2
17.3	2.4 ± 0.1	tr									
17.6	tr	-	-	-	-	-	tr	-	-	tr	-
18.1	tr	tr	tr	tr	-	tr	tr	tr	tr	tr	tr
22.7	-	-	-	-	-	-	1.2 ± 0.6	-	-	-	-
23.4	tr	tr	tr	tr	tr	tr	3.8 ± 3.6	tr	tr	3.7 ± 3.6	1.7 ± 1.0
23.6	tr	2.1 ± 1.2	1.2 ± 0.8	1.2 ± 0.7	1.2 ± 1.1	1.6 ± 0.8	4.1 ± 0.8	tr	tr	tr	2.1 ± 0.7

Retention time (min)	VUG-A58	VUG-A65	<i>H. soli</i>	P3	VUG-A23a	VUG-A124	VUG-A142	<i>H. norwichensis</i>	VUG-A112	<i>H. chitinivorans</i>	35/26
7.5	tr	tr	1.5 ± 0.1	-	tr	tr	tr	-	tr	tr	-
9.0	tr	tr	2.5 ± 0.1	tr	tr	tr	tr	tr	-	tr	-
9.7	1.1 ± 0.3	tr	tr	tr	tr	tr	tr	1.2 ± 0.9	tr	tr	tr
10.6	2.4 ± 0.2	5.6 ± 0.5	13.0 ± 0.3	tr	tr	tr	tr	tr	3.0 ± 0.03	2.7 ± 0.2	-
10.7	tr	tr	2.2 ± 0.2	-	tr	1.8 ± 0.1	1.4 ± 0.6	tr	1.3 ± 0.004	2.4 ± 0.04	tr
10.8	tr	tr	2.0 ± 0.1	tr	tr	-	-	tr	-	tr	-
11.1	10.4 ± 0.2	14.6 ± 0.1	28.7 ± 1.1	2.6 ± 0.2	9.4 ± 0.2	10.9 ± 0.04	10.1 ± 2.3	13.9 ± 1.8	16.5 ± 1.0	20.7 ± 1.1	5.6 ± 0.5
11.3	12.8 ± 0.4	13.5 ± 0.04	15.2 ± 0.4	3.6 ± 0.2	3.0 ± 0.1	-	-	8.0 ± 1.0	tr	10.6 ± 0.2	tr
12.0	tr	tr	tr	tr	tr	tr	tr	tr	tr	tr	tr
13.2	tr	tr	tr	1.6 ± 0.03	2.0 ± 0.1	4.1 ± 0.08	tr	1.2 ± 0.2	-	1.4 ± 0.04	1.2 ± 0.2
13.6	3.2 ± 0.2	2.2 ± 0.1	tr	2.3 ± 0.04	2.9 ± 0.01	1.6 ± 0.04	tr	tr	1.5 ± 0.2	tr	tr
13.8	tr	tr	tr	tr	tr	1.1 ± 0.1	tr	tr	-	tr	tr
14.0	tr	tr	tr	tr	4.0 ± 0.1	3.4 ± 0.1	3.1 ± 0.5	4.0 ± 0.7	5.8 ± 0.2	4.5 ± 0.2	7.2 ± 0.3
14.3	43.3 ± 1.2	37.0 ± 0.2	17.6 ± 0.3	49.9 ± 0.3	29.5 ± 0.4	28.8 ± 0.6	29.8 ± 6.8	28.6 ± 2.8	16.3 ± 0.3	11.2 ± 0.3	30.8 ± 0.5
14.6	17.8 ± 1.0	19.4 ± 0.2	7.0 ± 0.2	27.7 ± 0.2	22.3 ± 0.1	18.5 ± 0.1	22.9 ± 0.2	15.3 ± 2.0	17.2 ± 0.3	10.3 ± 0.3	33.6 ± 2.2
14.8	tr	-	-	-	tr	-	-	-	tr	-	-
15.0	2.7 ± 1.7	2.0 ± 0.7	2.3 ± 1.3	3.2 ± 0.1	2.4 ± 0.4	1.6 ± 0.7	7.3 ± 1.9	8.3 ± 8.0	5.5 ± 0.3	2.8 ± 0.9	3.1 ± 1.0
15.9	tr	tr	tr	-	8.4 ± 0.1	9.9 ± 0.5	4.0 ± 0.5	5.5 ± 1.2	12.4 ± 0.4	11.4 ± 0.9	1.2 ± 0.1
16.3	tr	tr	tr	-	tr	tr	tr	1.1 ± 0.9	tr	1.6 ± 0.8	tr
16.6	4.3 ± 0.2	5.8 ± 0.8	6.2 ± 0.7	tr	16.1 ± 0.5	16.8 ± 0.6	17.0 ± 1.1	8.9 ± 0.5	18.6 ± 1.0	15.9 ± 1.0	9.4 ± 0.7
17.3	tr	tr	1.9 ± 0.1	-	tr	1.7 ± 0.1	1.6 ± 0.4	2.7 ± 0.1	1.8 ± 0.2	2.4 ± 0.3	3.2 ± 0.2
17.6	-	tr	-	-	tr	-	-	tr	tr	tr	tr
18.1	tr	tr	tr	tr	tr	tr	tr	tr	-	tr	1.1 ± 0.1
22.7	-	-	-	-	-	-	-	-	-	-	1.5 ± 2.2
23.4	tr	tr	tr	tr	tr	tr	1.4 ± 0.6	tr	tr	2.2 ± 0.5	1.9 ± 1.3
23.6	1.9 ± 1.4	tr	tr	tr	tr	tr	1.4 ± 0.3	1.3 ± 0.9	tr	tr	tr

Retention time (min)	VUG-A60a	VUG-A141a	<i>H. roseosativarius</i>	<i>H. actinosclerius</i>	VUG-A106	<i>H. aerophilus</i>	<i>H. xinjiangensis</i>	<i>H. psychrotolerans</i>	<i>H. gelipurpurascens</i>	<i>H. rigui</i>	NS/2	<i>H. ocellatus</i>
7.5	tr	-	tr	-	-	-	-	tr	tr	tr	tr	tr
9.0	tr	-	tr	tr	tr	tr	tr	-	tr	tr	tr	tr
9.7	tr	tr	tr	tr	tr	tr	tr	tr	tr	tr	tr	tr
10.6	tr	tr	tr	tr	tr	-	2.1 ± 0.4	1.7 ± 0.1	tr	tr	tr	tr
10.7	tr	tr	tr	tr	tr	tr	tr	tr	tr	2.6 ± 0.3	1.3 ± 0.1	1.9 ± 0.1
10.8	-	tr	-	tr	tr	1.1 ± 0.1	1.2 ± 0.2	tr	tr	1.8 ± 0.1	1.7 ± 0.1	-
11.1	10.5 ± 0.4	6.4 ± 0.2	9.8 ± 0.4	13.5 ± 0.3	7.1 ± 0.1	6.0 ± 0.2	16.5 ± 0.8	15.5 ± 0.4	11.5 ± 0.2	24.1 ± 0.7	14.3 ± 0.1	28.3 ± 0.5
11.3	1.6 ± 0.04	3.9 ± 0.1	1.1 ± 0.02	15.3 ± 1.0	10.3 ± 0.04	16.4 ± 0.7	8.0 ± 0.04	11.4 ± 0.5	10.4 ± 0.2	17.4 ± 0.6	16.6 ± 0.4	13.0 ± 0.2
12.0	tr	tr	tr	1.2 ± 0.3	tr	tr	tr	tr	tr	tr	tr	1.1 ± 0.1
13.2	1.9 ± 0.1	1.5 ± 0.004	1.8 ± 0.02	2.2 ± 0.1	2.1 ± 0.1	1.7 ± 0.05	3.9 ± 0.5	1.7 ± 0.1	4.9 ± 0.1	1.1 ± 0.02	1.6 ± 0.1	tr
13.6	2.3 ± 0.1	2.1 ± 0.1	tr	1.3 ± 0.1	1.5 ± 0.1	1.7 ± 0.1	2.0 ± 0.1	2.4 ± 0.1	1.7 ± 0.1	1.1 ± 0.1	2.3 ± 0.2	-
13.8	1.1 ± 0.1	tr	tr	1.5 ± 0.02	tr	tr	3.2 ± 0.4	tr	3.0 ± 0.1	tr	tr	13.8 ± 0.1
14.0	9.1 ± 0.7	9.1 ± 0.1	6.4 ± 1.6	1.2 ± 0.02	4.6 ± 0.3	1.2 ± 0.1	4.3 ± 0.2	3.9 ± 0.03	8.2 ± 0.1	6.8 ± 0.3	6.0 ± 0.4	tr
14.3	23.2 ± 0.5	26.2 ± 0.6	29.3 ± 0.4	29.6 ± 0.5	33.6 ± 0.7	32.3 ± 1.4	13.8 ± 1.6	23.5 ± 0.5	22.9 ± 0.3	13.2 ± 0.6	17.1 ± 0.3	3.7 ± 0.3
14.6	22.5 ± 0.3	26.5 ± 0.6	24.0 ± 0.5	12.6 ± 0.1	24.3 ± 0.8	10.0 ± 0.4	12.2 ± 1.4	14.2 ± 0.3	22.6 ± 0.3	11.9 ± 0.5	14.9 ± 0.8	3.2 ± 0.2
14.8	-	-	-	-	-	-	1.6 ± 0.4	-	-	-	-	-
15.0	2.1 ± 1.1	1.9 ± 0.9	2.6 ± 0.6	4.3 ± 0.9	3.7 ± 2.2	4.3 ± 0.7	9.3 ± 2.6	4.0 ± 0.2	2.1 ± 0.5	2.1 ± 1.2	tr	2.6 ± 1.8
15.9	2.3 ± 0.1	2.0 ± 0.1	2.9 ± 0.5	3.7 ± 0.4	1.1 ± 0.2	1.4 ± 0.03	3.9 ± 0.3	6.8 ± 1.3	3.5 ± 0.1	8.2 ± 1.9	9.7 ± 1.0	11.7 ± 0.5
16.3	tr	tr	tr	1.2 ± 0.04	tr	1.2 ± 0.5	tr	tr	tr	tr	1.4 ± 0.3	1.3 ± 0.03
16.6	18.9 ± 1.3	12.3 ± 0.6	16.0 ± 1.0	10.6 ± 0.3	9.2 ± 1.2	14.5 ± 0.4	7.3 ± 1.4	11.6 ± 0.6	6.1 ± 0.6	7.4 ± 0.03	10.9 ± 0.7	25.9 ± 0.2
17.3	4.5 ± 0.8	1.8 ± 0.2	2.1 ± 0.4	1.9 ± 0.1	17.3 ± 1.2	4.0 ± 0.4	2.6 ± 0.2	1.6 ± 0.04	1.7 ± 0.2	2.3 ± 0.3	2.2 ± 0.2	4.3 ± 0.1
17.6	tr	1.1 ± 0.1	tr	tr	tr	2.4 ± 0.2	1.4 ± 0.2	tr	tr	tr	tr	1.9 ± 0.02
18.1	tr	tr	tr	tr	tr	tr	tr	tr	tr	-	tr	tr
22.7	-	-	tr	-	-	-	-	tr	-	-	-	-
23.4	tr	1.5 ± 1.4	1.6 ± 1.1	tr	tr	tr	6.8 ± 3.2	tr	tr	tr	tr	tr
23.6	tr	3.8 ± 1.0	2.2 ± 0.9	tr	1.4 ± 1.1	1.8 ± 1.2	tr	1.6 ± 0.7	1.3 ± 1.0	tr	tr	-

VUG-A48, VUG-A130 and *H. soli*, based upon common accumulation of the 10.6 min peak, is further suggested. Related but divergent, strain P3 lacks most of the major peaks found in other genetically related strains. Amongst other clades, the higher levels of the 11.1 and 15.9 min peaks and lower levels of the 14.3 and 14.6 min peaks suggest greater similarity of VUG-A112 to *H. chitinivorans*, as suggested by the 16S rRNA gene phylogeny (Figure 3.1), than to VUG-A124 and

Table 3.5. Presumptive identifications of detected fatty acids based on comparison to the BAME standard, literature values (e.g., Buczolits et al. 2006) and/or derivation from their molecular masses. Note that the GS-MS protocol used could either not resolve or caused the degradation of the hydroxylated fatty acids in the BAME standard; their contribution to *Hymenobacter* fatty acid composition remains unknown.

Retention time (min)	Molecular Mass of FAME	Presumptive Identification	Basis for Identification
7.5	228	13:0	MS ^a
9.0	242	14:0 straight chain, iso or anteiso	MS
9.7	242	14:0 straight chain, iso or anteiso	BAME
10.6	254	15:1 iso	Literature
10.7	254	15:1 anteiso	Literature
10.8	254	15:1	MS
11.1	256	15:0 iso	BAME
11.3	256	15:0 anteiso	BAME
12.0	256	15:0	BAME
13.2	268	16:1	MS
13.6	266	16:2	MS
13.8	270	16:0 iso	BAME
14.0	266	16:2	MS
14.3	268	16:1 $\omega 5c$	BAME
14.6	268	16:1 $\omega 7c$	Literature
14.8	γ^b	γ^b	-
15.0	270	16:0	BAME
15.9	280	17:2	MS
16.3	268	16:1	MS
16.6	282	17:1	MS
17.3	284	17:0 iso	BAME
17.6	284	17:0 anteiso	Literature
18.1	290	18:1	MS
22.7	268	γ^c	BAME – 18:1 $\omega 5t^c$
23.4	292	18:0	MS
23.6	236	γ^b	MS

^aPeak identification derived exclusively from the molecular mass of the parent ion

^bUnknown

^cNote that whereas the retention time of this peak matches that of the 18:1 $\omega 5t$ FAME in the BAME standard the molecular mass of its parent ion does not.

VUG-A142 as suggested by the *gyrB* phylogeny (Figure 3.2). VUG-A60a, VUG-A141a, 35/26 and *H. roseosalivarius* all share an elevated proportion of the 14.0 min peak, although this trait is also shared with *H. gelipurpurascens*, *H. rigui* and NS/2. Clustering of these latter three microorganisms is congruent with 16S rRNA gene but not *gyrB* phylogeny. In summary, despite considerable strain-specific variability, differences between clades of *Hymenobacter* and related species can be resolved. Where the 16S rRNA gene phylogeny differs from that of *gyrB*, fatty acid composition is more congruent with the latter than the former. The few strains analyzed for currently described species, however, precludes a deep understanding of how their fatty acid composition reflects their species- rather than strain-specific traits.

3.3.4. Biolog characterization of carbon source utilization profiles of *Hymenobacter* species and related strains

Similar to fatty acid analysis, carbon source utilization has hitherto been compared amongst described *Hymenobacter* species using different methods and growth conditions, which, again, are known to impact the results obtained (Ponder et al. 2005). The Biolog GN-2 system was therefore used in the current study to compare carbon source utilization patterns among *Hymenobacter* and related strains using common growth conditions. Unfortunately, the propensity of *Hymenobacter* strains to aggregate and produce a capsule may have in some cases resulted in false positive reactions in this assay due primarily to heterologous inoculation densities and capsule catabolism, respectively. Whereas the former could be accounted for by visual inspection of the plates following incubation, the latter resulted in plate-wide reaction of some strains with the tetrazolium dye,

Table 3.6. (Next page) Biolog GN-2 carbon source utilization by selected *Hymenobacter* and related strains. All plates were inoculated using a suspension of cells giving $\approx 50\%$ transmittance. Inocula were grown according to standard conditions and results measured following 2 weeks incubation. Wells were visually examined for false positive and negative results resulting from cell clumping or reduced liquid volume due to evaporation. Strains are presented in the same order as that in Figure 3.1.

Compound	VUG-A130 ^c	VUG-A42aa	VUG-A67	VUG-A57b	VUG-A33	VUG-A34	VUG-A2a	VUG-A58	<i>H. soli</i>	P3	VUG-A23a	VUG-A124	VUG-A142	<i>H. norwichensis</i>	VUG-A112	<i>H. chitinivorans</i>	VUG-A60a	VUG-A141a	<i>H. roseosativarius</i>	<i>H. actinosclerius</i>	VUG-A106	<i>H. aerophilus</i>	<i>H. xinjiangensis</i>	<i>H. psychrotolerans</i>	<i>H. gelupurplescans</i>	<i>H. rigui</i>	NS/2	
α -Cyclodextrin	- ^b	-	w	-	-	+	-	-	+	-	-	-	-	+	-	+	+	+	w	+	+	+	+	+	+	+	+	
Dextrin	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Glycogen	w	-	w	-	-	+	-	-	+	-	-	w	-	+	-	w	-	+	+	+	+	+	-	+	+	+	+	+
Tween 40	-	w	-	-	-	-	+	+	-	+	+	w	+	-	+	-	-	-	-	w	+	w	-	w	-	w	w	w
Tween 80	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
<i>N</i> -Acetyl-D-Galactosamine	-	w	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-
<i>N</i> -Acetyl-D-Glucosamine	-	w	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-
Adonitol	-	-	w	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-
L-Arabinose	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	w
D-Arabitol	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	w	-	w	-	-
D-Cellobiose	+	-	w	-	-	+	-	-	-	-	-	-	-	w	-	-	+	-	-	-	-	-	w	-	+	-	+	+
<i>i</i> -Erythritol	w	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	w	w	-	-	+	-	+	+	-
D-Fructose	-	w	w	-	-	w	-	-	+	-	-	-	-	+	-	+	-	+	w	-	w	+	w	+	+	+	+	+
L-Fucose	-	w	w	-	-	-	-	-	-	-	-	-	-	w	-	w	-	-	-	-	-	w	-	-	-	w	w	w
D-Galactose	w	w	+	+	-	+	-	-	+	-	-	-	-	+	-	+	+	+	+	-	+	w	w	+	+	+	+	+
Gentiobiose	-	w	w	-	-	-	-	-	+	-	-	-	-	+	-	w	-	+	+	-	-	-	w	-	+	+	+	+
α -D-Glucose	-	w	+	-	-	+	-	-	+	-	-	w	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>m</i> -Inositol	w	w	w	-	-	-	-	-	-	-	-	w	-	-	-	+	-	-	-	w	-	-	-	-	w	+	+	+
α -D-Lactose	-	-	w	-	-	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	+	+	+	+
Lactulose	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	w	w
Maltose	-	w	+	+	-	+	-	-	+	-	-	w	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+
D-Mannitol	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+	w	-	-	-	-	-	-
D-Mannose	w	w	w	-	-	+	-	-	-	-	-	-	-	+	-	-	-	-	+	-	-	w	-	-	+	+	+	w
D-Melibiose	+	-	-	-	-	w	-	-	-	-	-	-	-	+	-	+	-	-	-	w	w	w	+	+	+	+	+	
β -Methyl-D-Glucose	+	-	w	-	-	w	-	-	-	-	-	-	-	+	-	-	-	-	-	w	-	w	w	-	w	w	-	

Compound	VUG-A130 ^c	VUG-A42aa	VUG-A67	VUG-A57b	VUG-A33	VUG-A34	VUG-A2a	VUG-A58	<i>H. soli</i>	P3	VUG-A23a	VUG-A124	VUG-A142	<i>H. norwichensis</i>	VUG-A112	<i>H. chitinivorans</i>	VUG-A60a	VUG-A141a	<i>H. roseosalivarius</i>	<i>H. actinoscleris</i>	VUG-A106	<i>H. aerophilus</i>	<i>H. xinjiangensis</i>	<i>H. psychrotolerans</i>	<i>H. gelipurpurascens</i>	<i>H. rigui</i>	NS/2	
D-Psicose	-	-	-	-	-	-	-	-	+	-	-	-	-	w	-	w	-	-	+	-	-	w	+	+	+	+	w	
D-Raffinose	-	w	-	-	-	w	-	-	-	-	-	-	-	+	-	+	-	-	+	-	+	-	-	+	+	+	+	
L-Rhamnose	w	-	w	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	w	-	-	-	w	
D-Sorbitol	-	w	w	-	-	+	-	-	-	-	-	-	-	-	-	w	-	-	-	-	+	-	w	-	-	w	+	
Sucrose	w	w	w	-	-	w	-	-	+	-	-	w	-	+	-	+	-	+	+	-	-	+	-	+	+	+	+	
D-Trehalose	-	w	-	-	-	w	-	-	+	-	-	w	-	+	-	+	+	+	+	-	-	w	+	+	+	+	+	
Turanose	-	w	w	-	-	w	-	-	-	-	-	-	-	+	-	w	-	-	-	+	-	w	+	-	+	+	+	
Xylitol	+	w	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	-	-	-	w	w
Pyruvate Methyl Ester	w	w	+	+	-	+	+	+	+	-	-	+	+	-	-	w	-	-	-	w	+	w	-	-	+	-	w	
Succinate Mono-Methyl Ester	-	w	w	w	-	+	-	-	-	-	-	-	-	+	-	w	-	-	-	+	+	+	-	+	+	+	+	
Acetate	+	-	-	w	w	+	-	-	+	+	-	-	-	+	-	-	-	+	+	+	+	+	+	+	+	+	+	
<i>cis</i> -Aconitate	w	-	-	-	-	-	-	-	-	-	-	-	-	w	-	w	-	-	-	-	-	-	w	-	-	-	w	
Citrate	-	w	w	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	+	w	-	-	-	-	+	+	
Formate	w	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	+	-	-	w	-	+	+	-	
D-Galactonate Lactone	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	w	-	
D-Galacturonate	w	-	w	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	+	-	-	w	+	-	+	+	+	
D-Gluconate	-	-	w	-	-	w	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	w	+	-	w	+	-	
D-Glucosamine	w	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	
D-Glucuronate	-	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	w	
α -Hydroxybutyrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
β -Hydroxybutyrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-	
γ -Hydroxybutyrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
<i>p</i> -Hydroxyphenylacetate	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Itaconate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	w	-	-	-	w	

Compound	VUG-A130 ^a	VUG-A42aa	VUG-A67	VUG-A57b	VUG-A33	VUG-A34	VUG-A2a	VUG-A58	<i>H. soli</i>	P3	VUG-A23a	VUG-A124	VUG-A142	<i>H. norwichensis</i>	VUG-A112	<i>H. chitinivorans</i>	VUG-A60a	VUG-A141a	<i>H. roseosativarius</i>	<i>H. actinosclerius</i>	VUG-A106	<i>H. aerophilus</i>	<i>H. xinjiangensis</i>	<i>H. psychrotolerans</i>	<i>H. gelipurpurascens</i>	<i>H. rigui</i>	NS/2
α-Ketobutyrate	+	w	w	+	+	+	+	+	-	-	-	+	+	+	-	-	-	+	-	-	+	w	w	-	w	+	+
α-Ketoglutarate	w	w	w	-	-	-	+	+	-	-	-	w	+	-	-	w	-	-	-	-	w	-	w	-	-	w	w
α-Ketovalerate	+	w	w	w	-	+	+	+	-	+	-	+	+	-	-	-	-	-	-	+	w	w	+	-	-	w	-
D,L-Lactate	-	w	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	w	-	-
Malonate	-	w	w	-	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	+	w	-	+	w
Propionate	w	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Quinate	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-
D-Saccharate	-	w	w	-	-	w	-	-	-	-	-	-	-	-	-	w	-	-	-	-	w	-	+	-	-	-	+
Sebacate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Succinate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	w	-	w	-
Bromosuccinate	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	+	-	-	-	w
Succinamate	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	+	-	-	-	-
Glucuronamide	w	-	w	-	-	w	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	+	-
L-Alaninamide	+	w	+	-	-	+	-	-	-	-	-	-	-	-	-	w	-	-	-	+	-	w	-	w	-	w	w
D-Alanine	+	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	w	-
L-Alanine	+	+	w	-	-	+	-	-	-	-	-	w	-	+	-	w	-	-	+	-	+	w	-	w	+	-	+
L-Alanylglycine	-	+	+	+	-	+	-	-	-	-	-	w	-	+	-	w	-	+	-	+	+	+	-	+	+	+	w
L-Asparagine	-	w	+	-	-	+	-	-	-	-	-	-	-	+	-	+	-	-	-	+	+	w	-	+	+	+	-
L-Aspartate	-	w	w	-	-	+	-	-	+	-	-	-	-	+	-	+	-	-	-	+	+	w	w	+	-	+	w
L-Glutamate	+	+	+	-	-	+	+	+	+	+	-	+	+	+	-	w	+	+	+	+	+	+	+	+	+	+	+
Glycyl-L-Aspartate	-	-	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	w	+	w	+	+	+	+	+
Glycyl-L-Glutamate	+	w	w	+	-	+	-	-	+	-	-	w	-	+	-	w	-	-	-	+	+	+	+	+	+	+	+
L-Histidine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-
Hydroxy-L-Proline	w	w	w	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	w	-	-	w	-
L-Leucine	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	w	-	-	-

Compound	VUG-A130 ^a	VUG-A42aa	VUG-A67	VUG-A57b	VUG-A33	VUG-A34	VUG-A2a	VUG-A58	<i>H. soli</i>	P3	VUG-A23a	VUG-A124	VUG-A142	<i>H. norwichensis</i>	VUG-A112	<i>H. chitinivorans</i>	VUG-A60a	VUG-A141a	<i>H. roseosalivarius</i>	<i>H. actinoscleris</i>	VUG-A106	<i>H. aerophilus</i>	<i>H. xinjiangensis</i>	<i>H. psychrotolerans</i>	<i>H. gelipurascens</i>	<i>H. rigui</i>	NS/2
L-Ornithine	+	w	-	-	-	+	+	-	-	-	-	w	-	-	-	w	-	-	-	w	w	+	+	+	+	w	w
L-Phenylalanine	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-
L-Proline	+	-	+	-	-	+	+	+	+	-	-	w	-	+	-	w	-	-	-	+	+	+	w	w	-	+	-
L-Pyroglutamine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-
D-Serine	-	w	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	w	-
L-Serine	-	-	-	w	-	+	-	-	-	-	-	-	-	-	-	w	-	-	-	w	w	w	-	-	-	-	w
L-Threonine	-	w	+	-	+	+	+	-	-	-	-	w	-	+	-	w	-	-	+	+	+	w	-	w	+	+	w
D,L-Carnitine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	-	-	-	-
γ-Aminobutyrate	+	w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-
Urocanate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-
Inosine	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	w	-	-	-
Uridine	-	w	-	-	-	+	-	-	-	-	-	w	-	+	-	+	-	-	-	+	-	w	-	w	+	-	-
Thymidine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	w	w	-
Phenylethylamine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-
Putrescine	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2-Aminoethanol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-
2,3-Butanediol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	-	w	-	-	-
Glycerol	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	+	-
D,L-α-Glycerol Phosphate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	w	-	-	-	-	-	-	w	-
α-D-Glucose-1-Phosphate	+	-	-	-	-	+	-	-	+	+	-	-	-	+	-	w	-	+	+	+	-	+	-	+	+	+	w
D-Glucose-6-Phosphate	-	-	-	-	-	w	-	-	-	-	-	-	-	+ ^c	-	w	-	-	-	-	-	-	-	w	+	w ^c	w

^aResults are shown following 1 week incubation because evaporation of the water blank by week 2 invalidated comparisons

^b(-): Negative result; (w): Weak positive result; (+): Strong positive result

^cResults for these wells are given for week 1 due to evaporation of these wells by week 2

decreasing the sensitivity of the assay and its ability to discriminate between negative and weak positive reactions and weak and strong positive reactions.

In contrast to fatty acid composition (Section 3.3.3), carbon source utilization patterns were variable between strains (Table 3.6). Only two carbon sources, dextrin and L-glutamate, were used by all (or nearly all) *Hymenobacter* and related strains (Table 3.6). Several carbon sources, most notably α -cyclodextrin, glycogen, D-galactose, maltose, D-trehalose and acetate, were used by nearly all described species, in contrast to *H. soli* and its relatives and strains VUG-A23a, VUG-A124 and VUG-A142 (Table 3.6). *H. soli* and its relatives typically utilized relatively few carbon sources, although strain-specific exceptions exist (e.g., VUG-A34; Table 3.6). Considering the groups suggested by gene phylogenies (Figures 3.1 and 3.2), *H. gelipurpurascens*, *H. rigui* and NS/2 similarly utilized more carbon sources than their phylogenetic neighbors; the opposite is true for VUG-A60a, VUG-A141a and *H. roseosalivarius* and (even more so) for VUG-A124 and VUG-A142, which utilize comparatively few carbon sources (Table 3.6). Very few positive assays were recorded for both VUG-A112 and VUG-A23a, making analysis of these strains uninformative. In summary, whereas some phylogenetically informative differences in carbon source utilization patterns exist between *Hymenobacter* and related strains, Biolog results display strain-specific heterogeneity. Again, to what extent these differences are strain- or species-specific remains unknown due to the paucity of analyzed strains both here and in the literature.

3.4. Discussion

Based upon gene phylogenies (Figures 3.1 and 3.2), phenotype (Table 3.2), morphotype (Table 3.3), fatty acid content (Table 3.4) and carbon source utilization (Table 3.6), all strains analyzed in this study belong to the genus *Hymenobacter* as currently defined (Buczolits et al. 2006). Additionally, the common production of 2'-hydroxyflexixanthin-type carotenoids by all *Hymenobacter* and related strains (Chapters 4 and 5; Klassen and Foght 2008; and data not shown) is also congruent with classification of these strains in the same

genus. Based on these results, however, divisions exist between three clades: (i) *H. soli*, P3 and related VUG strains; (ii) *H. roseosalivarius*, VUG-A60a, VUG-A141a and 35/26; and (iii) the majority of named *Hymenobacter* strains plus NS/2 and VUG-A106. Whether it is more appropriate for all of these clades to be included in the genus *Hymenobacter* or to be renamed as separate genera remains a subject for future study. The phylogenetic affiliation of *H. norwichensis*, *H. chitinivorans*, VUG-A23a, VUG-A112, VUG-A124 and VUG-A142 to one of the above groups remains ambiguous, although the carbon source utilization, phenotypic and morphological characteristics of *H. norwichensis* and *H. chitinivorans* are notably congruent with those of group (iii). Whether these strains will form unique lineages or collapse into those defined above with further study is unknown.

The lack of congruence between phylogenetic trees for the 16S rRNA and *gyrB* genes is particularly remarkable amongst the results of this study. Fatty acid composition (Table 3.4) supports a 16S rRNA gene-like rather than a *gyrB*-like phylogeny. This suggests several horizontal transfers of *gyrB*, which is surprising given its well-conserved nature and wide application as a phylogenetic marker (Watanabe et al. 2001, Santos and Ochman 2004). Whether other *Hymenobacter* reference genes display similar phylogenetic incongruencies remains unknown, in part due to the difficulties in designing primers from currently available sequences and applying those developed previously (e.g., *cpn60* and *rpoB*; data not shown). These problems are likely due to the deep divergence of *Hymenobacter* relative to better studied *Bacteroidetes* species. The genome sequencing project currently underway for *H. roseosalivarius* (<http://www.jgi.doe.gov/sequencing/statusreporter/psr.php?projectid=98064>) will help resolve this question.

Considering the current taxonomy of *Hymenobacter*, most species have been classified based upon characterization of only a few (typically one or two) strains. Those studies typically did not use reference strains analyzed in parallel but solely referred to information in different published studies. Problematically,

each study uses different methods and growth conditions, depending on the strains analyzed and the researchers performing the analyses. Additionally, this may bias the strains available for study; e.g., VUG-A124 and VUG-A142 will grow only on R2A and no other tested medium, including those upon which most other *Hymenobacter* species have been isolated and maintained (data not shown). The results shown here indicate significant differences between closely related taxa (e.g., VUG-A124 and VUG-A142 and *H. soli*-like VUG strains; Tables 3.4 and 3.6), and greater similarity between named species, particularly in fatty acid composition (Table 3.4), than recognized previously. The assignment of a species name to a previously unknown strain depends not only upon demonstrating its phylogenetic and phenotypic novelty but also that this novelty is phylogenetically conserved. Poor recognition of both these criteria has been noted previously as a widespread and misleading taxonomic practice (Christensen et al. 2001, Felis and Dellaglio 2007). The results reported here highlight the importance of proper consideration of these principles.

Naming novel species is important to provide a framework for future study, communication and regulation of the relevant taxa (Gevers et al. 2005). It is worthwhile, however, to consider the theoretical basis for doing so for the genus *Hymenobacter*, especially considering their incongruent housekeeping gene phylogenies and high levels of strain-specific carbon source utilization phenotypes. As discussed previously (Section 1.1), *Hymenobacter* are commonly detected in aerosols and therefore likely have a cosmopolitan distribution. Theory suggests that well-mixed populations with significant capacity for horizontal gene transfer have little capacity for allopatric speciation (Berg and Kurland 2002); both the aerosol mode of transport and apparently frequent horizontal transfer of the *gyrB* gene are true for *Hymenobacter*. Furthermore, the capacity for periodic re-introduction of ice-bound genotypes (such as VUG strains), which may have elsewhere become extinct, might facilitate their long-term survival in the global population (Rogers et al. 2004); the frequent detection of *Hymenobacter* in icy environments (Section 1.1) suggests the relevance of this scenario. It is therefore possible that genome evolution in *Hymenobacter* is both reticulate (i.e., web-like)

due to well-distributed horizontal gene transfer and “bi-directional”, whereby “less-evolved” genotypes can survive an evolutionary bottleneck by archival with subsequent rise to high frequencies in the global population. Both of these forces would theoretically loosen species barriers, making taxonomic classification of these organisms especially problematic.

Defining species based upon strains which have been dormant (but likely metabolically active, even if at a low maintenance level; Price and Sowers 2004) is theoretically problematic for another reason. Given the diversity of VUG genotypes isolated and the wide distribution of modern-day *Hymenobacter* strains, it seems most likely that VUG *Hymenobacter* strains are not specifically adapted to the ice from which they were isolated. This is, however, the environment through which selection has acted since their interment (likely \approx 4,000 years ago; J. Barker, personal communication), which may have resulted in phenotypic divergence from their modern relatives. For example, carbon sources in ice veins (Price 2000, Krembs and Demming 2008) likely do not match those which *Hymenobacter* more usually encounters elsewhere. Selection for survival within the ice vein, therefore, may have resulted in catabolic gene mutations not shared by the modern relatives of the archived strain, which may not be evolutionary fixed in these strains upon their re-emergence following melting. It will be especially interesting to compare the *H. soli*-like VUG strains with their modern close relatives, both to determine the prevalence of these gene loss processes and to ensure that the proper characters are designated as typical in their species assignment. These isolates represent an excellent model system for the further study of the evolutionary implications of long-term genotypic archival. A better understanding of *Hymenobacter* evolution will therefore greatly enhance our understanding of microbial evolution and biogeography more generally.

Which strains characterized here should be considered new species? Both 16S rRNA (Figure 3.1) and *gyrB* (Figure 3.2) gene phylogenies suggest a common clade comprising VUG-A2a, VUG-A31a, VUG-A33, VUG-A34, VUG-A42aa, VUG-A57b, VUG-A58, VUG-A65, VUG-A67, VUG-C4 and VUG-D4a,

related to but distinct from *H. soli*, P3, VUG-A48 and VUG-A130. Compared to all described *Hymenobacter* species, these strains produce greater amounts of the 14.0 and 16.6 FAMES (Table 3.4), typically utilize fewer carbon sources (Table 3.6), produce 2'-methoxyflexixanthin as a major carotenoid (Chapter 4, Klassen and Foght 2008 and data not shown), are less tolerant to higher concentrations of NaCl and higher growth temperatures, growth slower, typically do not produce catalases (Table 3.2) and are do not form large, spreading colonies on R2A (Table 3.3). These differences and their conservation in all aforementioned strains indicated the appropriateness of designating this taxon as a novel species. Extending the Biolog carbon source utilization data to the strains within this clade not studied here and determining the abundances of hydroxylated fatty acids in these strains will complete the present data set such that it will be sufficiently complete to establish which of these characters are conserved among this novel species; these experiments are ongoing. Whereas both 16S rRNA and *gyrB* phylogenies (Figures 3.1 and 3.2) indicate subgroups within the main clade of *H. soli*-like VUG strains, subclade memberships differ between phylogenetic markers and are not obvious from phenotypic or chemotaxonomic data (Tables 3.2-3.4 and 3.6). Much more extensive data sets (particularly genotypic) are required to identify potential cryptic species within this clade.

Whereas strain-specific characteristics exist for *H. soli*, P3, VUG-A48 and VUG-A130, the lack of closely related strains and considerable intra-strain heterogeneity of characters makes assessing their significance in designating new species difficult. Despite the few strains studied, it is likely that both VUG-A23a and the pair of VUG-A124 and VUG-A142 represent novel species based upon their differences in phenotype (Table 3.2), carbon source utilization (Table 3.6) and their genetic divergence from other taxa (Figures 3.1 and 3.2). Although the newly described *H. daecheongensis* (Xu et al. 2009) is related to these strains according to 16S rRNA gene phylogeny (Figure 3.1), the reported elevated growth temperature (30°C) and fast growth of this strain (Xu et al. 2009) suggests that it differs substantially from them. Classification of VUG-A23a and the pair of VUG-A124 and VUG-A142 as “*species proponenda*”, that is, clearly distinct

from their taxonomic relatives yet described based on characterization of a very few strains (Felis and Dellaglio 2007), is likely most appropriate in these cases. This designation would both appropriately indicate the novelty of these strains and indicate that the characteristics described here may not be representative of other (currently unknown) members of these species. The other unnamed strains analyzed here (VUG-A60a, VUG-A141a, VUG-A112, 35/26, NS/2) are relatively closely related to named taxa and, while perhaps in fact novel species, require further study with more closely related isolates and higher-resolution taxonomic methods (e.g., whole genome hybridization) for their definitive classification.

3.5. Literature cited

- Aislabie, J.M., Broady, P.A., and Saul, D.J. 2006. Culturable aerobic heterotrophic bacteria from high altitude, high latitude soil of La Gorce Mountains (86°30'S, 147°W), Antarctica. *Antarct. Sci.*, 18: 313-321.
- Baik, K.S., Seong, C.N., Moon, E.Y., Park, Y.-D., Yi, H., and Chun, J. 2006. *Hymenobacter rigui* sp. nov., isolated from wetland freshwater. *Int. J. Syst. Evol. Microbiol.*, 56: 2189-2192.
- Berg, O.G., and Kurland, C.G. 2002. Evolution of microbial genomes: sequence acquisition and loss. *Mol. Biol. Evol.*, 19: 2265-2276.
- Buczolits, S., Denner, E.B.M., Vybiral, D., Wieser, M., Kämpfer, P., and Busse, H.J. 2002. Classification of three airborne bacteria and proposal of *Hymenobacter aerophilus* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 52: 445-456.
- Buczolits, S., Denner, E.B.M., Kämpfer, P., and Busse, H.-J. 2006. Proposal of *Hymenobacter norwichensis* sp. nov., classification of '*Taxeobacter ocellatus*', '*Taxeobacter gelipurpurascens*' and '*Taxeobacter chitinivorans*' as *Hymenobacter ocellatus* sp. nov., *Hymenobacter gelipurpurascens* sp. nov. and *Hymenobacter chitinivorans* sp. nov., respectively, and emended description of the genus *Hymenobacter* Hirsch *et al.* 1999. *Int. J. Syst. Evol. Microbiol.*, 56: 2071-2078.
- Cheng, S.M., and Foght, J.M. 2007. Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.*, 59: 318-330.

- Christensen, H., Bisgaard, M., Frederiksen, W., Mutters, R., Kuhnert, P., and Olsen, J.E. 2001. Is characterization of a single isolate sufficient for valid publication of a new genus or species? Proposal to modify Recommendation 30b of the *Bacteriological Code* (1990 Revision). *Int. J. Syst. Evol. Microbiol.*, 51: 2221-2225.
- Collins, M.D., Hutson, R.A., Grant, I.R., and Patterson, M.F. 2000. Phylogenetic characterization of a novel radiation-resistant bacterium from irradiated pork: description of *Hymenobacter actinosclerus* sp. nov. *Int. J. Syst. Evol. Microbiol.*, 50: 731-734.
- De Bruijn, F.J. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain-reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Appl. Environ. Microbiol.*, 58: 2180-2187.
- Dear, S., and Staden, R. 1991. A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.*, 19: 3907-3911.
- Embley, T., and Wait, R. 1994. Structural Lipids of Eubacteria. *In* Chemical methods in prokaryotic systematics. M. Goodfellow and A.G. O'Donnell. (ed.) Wiley, Chichester, NY. pp. 121-161.
- Felis, G.E., and Dellaglio, F. 2007. On species descriptions based on a single strain: proposal to introduce the status *species proponenda* (sp. pr.). *Int. J. Syst. Evol. Microbiol.*, 57: 2185-2187.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics*, 5: 164-166.
- Foght, J., Aislabie, J., Turner, S., Brown, C.E., Ryburn, J., Saul, D.J., and Lawson, W. 2004. Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb. Ecol.*, 47: 329-340.
- Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., and Swings, J. 2005. Re-evaluating prokaryotic species. *Nat. Rev. Microbiol.*, 3: 733-739.
- Hirsch, P., Ludwig, W., Hethke, C., Sittig, M., Hoffmann, B., and Gallikowski, C.A. 1998. *Hymenobacter roseosalivarius* gen. nov., sp. nov. from continental Antarctic soils and sandstone: bacteria of the *Cytophaga/Flavobacterium/Bacteroides* line of phylogenetic descent. *Syst. Appl. Microbiol.*, 21: 374-383.

- Kim, K.-H., Im, W.-T., and Lee, S.-T. 2008. *Hymenobacter soli* sp. nov., isolated from grass soil. *Int. J. Syst. Evol. Microbiol.*, 58: 941-945.
- Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.
- Krembs, C., and Demming, J.W. 2008. The role of exopolymers in microbial adaptation to sea ice. *In Psychrophiles: from biodiversity to biotechnology*. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 247-264.
- Nedashkovskaya, O.L., Kim, S.B., Suzuki, M., Shevchenko, L.S., Lee, M.S., Lee, K.H., Park, S.M., Frolova, G.M., Oh, H.W., Bae, S.K., Park, H.-Y., and Mikhailov, V.V. 2005. *Pontibacter actiniarum* gen. nov., sp. nov., a novel member of the phylum 'Bacteroidetes', and proposal of *Reichenbachiella* gen. nov. as a replacement for the illegitimate prokaryotic generic name *Reichenbachia* Nedashkovskaya *et al.* 2003. *Int. J. Syst. Evol. Microbiol.*, 55: 2583-2588.
- Ponder, M.A., Gilmour, S.J., Bergholz, P.W., Mindock, C.A., Hollingsworth, R., Thomashow, M.F., and Tiedje, J.M. 2005. Characterization of potential stress responses in ancient Siberian permafrost psychroactive bacteria. *FEMS Microbiol. Ecol.*, 53: 103-115.
- Price, P.B. 2000. A habitat for psychrophiles in deep Antarctic ice. *Proc. Natl. Acad. Sci. U. S. A.*, 97: 1247-1251.
- Price, P.B., and Sowers, T. 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. *Proc. Natl. Acad. Sci. U. S. A.*, 101: 4631-4636.
- Reichenbach, H. 1992. *Taxeobacter*, a new genus of the *Cytophagales* with three new species. *In Advances in the taxonomy and significance of Flavobacterium, Cytophaga and related bacteria: proceedings of the 2nd international symposium on Flavobacterium, Cytophaga and related bacteria*. P.J. Jooste. (ed.) University of the Orange Free State, Bloemfontein, South Africa. pp. 182-185.
- Rickard, A.H., Stead, A.T., O'May, G.A., Lindsay, S., Banner, M., Handley, P.S., and Gilbert, P. 2005. *Adhaeribacter aquaticus* gen. nov., sp. nov., a Gram-negative isolate from a potable water biofilm. *Int. J. Syst. Evol. Microbiol.*, 55: 821-829.
- Rogers, S.O., Starmer, W.T., and Castello, J.D. 2004. Recycling of pathogenic

- microbes through survival in ice. *Med. Hypotheses*, 63: 773-777.
- Russell, N.J. 2008. Membrane components and cold sensing. *In Psychrophiles: from biodiversity to biotechnology*. R. Margesin, F. Schinner, J.-C. Marx, and C. Gerday. (ed.) Springer-Verlag, Berlin. pp. 177-190.
- Santos, S.R., and Ochman, H. 2004. Identification and phylogenetic sorting of bacterial lineages with universally conserved genes and proteins. *Environ. Microbiol.*, 6: 754-759.
- Saul, D.J., Aislabie, J.M., Brown, C.E., Harris, L., and Foght, J.M. 2005. Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. *FEMS Microbiol. Ecol.*, 53: 141-155.
- Stamatakis, A., Hoover, P., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, 57: 758-771.
- Suresh, K., Mayilraj, S., and Chakrabarti, T. 2006. *Effluviibacter roseus* gen. nov., sp. nov., isolated from muddy water, belonging to the family 'Flexibacteraceae'. *Int. J. Syst. Evol. Microbiol.*, 56: 1703-1707.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. *MEGA4*: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24: 1596-1599.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25: 4876-4882.
- Watanabe, K., Nelson, J.S., Harayama, S., and Kasai, H. 2001. ICB database: the *gyrB* database for identification and classification of bacteria. *Nucleic Acids Res.*, 29: 344-345.
- Xu, J.-L., Liu, Q.-M., Yu, H.-S., Jin, F.-X., Lee, S.-T., and Im, W.-T. 2009. *Hymenobacter daecheongensis* sp. nov., isolated from stream sediment. *Int. J. Syst. Evol. Microbiol.*, 59: 1183-1187.
- Yamamoto, S., and Harayama, S. 1995. PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. *Appl. Environ. Microbiol.*, 61: 1183-1187.

Microbiol., 61: 1104-1109.

- Zhang, G., Niu, F., Busse, H.-J., Ma, X., Liu, W., Dong, M., Feng, H., An, L., and Cheng, G. 2008a. *Hymenobacter psychrotolerans* sp. nov., isolated from the Qinghai-Tibet Plateau permafrost region. *Int. J. Syst. Evol. Microbiol.*, 58: 1215-1220.
- Zhang, J.-Y., Liu, X.-Y., and Liu, S.-J. 2009a. *Adhaeribacter terreus* sp. nov., isolated from forest soil. *Int. J. Syst. Evol. Microbiol.*, 59: 1595-1598.
- Zhang, L., Zhang, Q., Luo, X., Tang, Y., Dai, J., Li, Y., Wang, Y., Chen, G., and Fang, C. 2008b. *Pontibacter korlensis* sp. nov., isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 58: 1210-1214.
- Zhang, L., Dai, J., Tang, Y., Luo, X., Wang, Y., An, H., Fang, C., and Zhang, C. 2009b. *Hymenobacter deserti* sp. nov., isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 59: 77-82.
- Zhang, Q., Liu, C., Tang, Y., Zhou, G., Shen, P., Fang, C., and Yokota, A. 2007. *Hymenobacter xinjiangensis* sp. nov., a radiation-resistant bacterium isolated from the desert of Xinjiang, China. *Int. J. Syst. Evol. Microbiol.*, 57: 1752-1756.
- Zhang, Y.-M., and Rock, C.O. 2008. Membrane lipid homeostasis in bacteria. *Nat. Rev. Microbiol.*, 6: 222-233.
- Zhou, Y., Wang, X., Liu, H., Zhang, K.-Y., Zhang, Y.-Q., Lai, R., and Li, W.-J. 2007. *Pontibacter akesuensis* sp. nov., isolated from a desert soil in China. *Int. J. Syst. Evol. Microbiol.*, 57: 321-325.

4. Differences in Carotenoid Composition Among *Hymenobacter* and Related Strains Support a Tree-Like Model of Carotenoid Evolution¹

4.1. Introduction

Over 600 known structurally unique carotenoids are distributed throughout all major lineages of the tree of life (Schmidt et al. 1994, Britton et al. 2004). Recently, carotenoids have gained biotechnological interest as natural nutritional supplements (Sandmann et al. 1999, Fraser and Bramley 2004) and natural pigments (Mortensen 2006) with a projected market in 2010 exceeding one billion US dollars (cited in Del Campo et al. 2007). To this end, carotenoid biosynthesis has been well studied in many microorganisms (Sieiro et al. 2003, Cheng 2006), particularly regarding the creation of recombinant biosynthetic pathways leading to the production of novel pigments (Sandmann et al. 1999, Umeno et al. 2005) with biotechnologically interesting properties such as improved antioxidant activity (Albrecht et al. 2000, Nishida et al. 2005). These approaches require the identification of novel biosynthetic enzymes having expanded substrate ranges and activities, the discovery of which requires a clear understanding of carotenoid diversity and distribution. Most high-resolution structural studies are limited to the carotenoids present in single representative strains. Whether this approach captures all structural diversity is unknown because systematic, high resolution studies of carotenoid composition in taxonomically related strains are lacking. The current understanding of the evolution of carotenoid metabolism (and that of other secondary metabolites) suggests a “tree-like” model (Umeno et al. 2005) containing a highly conserved, core biosynthetic pathway with terminal “branches” that are freer to evolve. The degree of evolutionary plasticity exhibited by these terminal biosynthetic branches remains unclear due to insufficient study.

¹A version of this chapter has been published. Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.

This work describes the carotenoids produced by bacteria of the genus *Hymenobacter* (*Flexibacteraceae*; *Bacteroidetes*) including several novel strains isolated from Victoria Upper Glacier, Antarctica (VUG; Chapter 2). An undefined “*Taxeobacter*” (now *Hymenobacter*; Buczolits et al. 2006) strain has been previously reported to contain 2'-hydroxyflexixanthin and 3-deoxy-2'-hydroxyflexixanthin as major carotenoids (Bircher and Pfander 1997), although this identification is considered to be poorly supported (Britton et al. 2004). To obtain sufficient structural resolution, HPLC was used to separate and compare the carotenoids present in this collection of *Hymenobacter* and related strains. Because of the previously reported carotenoid structures and taxonomic relatedness of these strains, it was hypothesized that systematic examination of their carotenoids would indicate the extent of carotenoid evolution within these related genera and species.

4.2. Materials and methods

4.2.1. Bacterial strains and growth conditions

All *Hymenobacter* reference strains were purchased from the DSMZ and CCUG culture collections, except for *H. norwichensis*, *H. rigui* and *Hymenobacter* sp. str. NS/2 which were kindly provided by Hans-Jürgen Busse (University of Vienna). VUG *Hymenobacter*-like strains were isolated by direct plating of aseptically melted glacier ice onto R2A agar (Difco) and incubation at 4°C, 10°C or room temperature (approximately 20°C) in the dark with subsequent restreaking to purity (Chapter 2). Because most strains did not grow consistently in liquid culture or at room temperature, cultures were maintained on chilled R2A plates incubated at 10°C for 4 weeks. Except where indicated, growth experiments were conducted at 18°C in the dark for 7 d, using 1 week-old cultures incubated at 10°C as inocula to ensure active growth.

4.2.2. 16S rRNA gene sequencing and phylogenetic analysis

Genomic DNA was extracted from R2A plate-grown cells using a bead-beating and chemical lysis method described previously (Foght et al. 2004) and

near full length 16S rRNA gene sequences (*Escherichia coli* positions 8-1509 (Brosius et al. 1981)) were amplified by PCR using primers PB36F and PB38R (Foght et al. 2004; see Section 3.2.1. for sequences). PCR products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Instruments; ABI) and an ABI 3700 DNA sequencer (ABI) using PB36F, PB38R and internal primers (Cheng and Foght 2007). Sequences were assembled using the PREGAP v1.5 and GAP4 v4.10 programs of the Staden package (Dear and Staden 1991), checked for chimeras using PINTAIL (Ashelford et al. 2005) and submitted to GenBank. To construct bootstrapped trees, 16S rRNA genes of VUG and reference *Hymenobacter* strains were aligned using CLUSTAL_X (Thompson et al. 1997) and trimmed using the SEQRET program of the EMBOSS bioinformatics suite (Rice et al. 2000) to eliminate primer sequences and ensure the comparison of sequences of equal lengths (nucleotides 38 to 1486, *E. coli* numbering (Brosius et al. 1981)). Using PHYLIP version 3.6a3 (Felsenstein 1989) 100 phylogenetic trees were created using the Kimura 2-parameter method (Kimura 1980) and a consensus neighbor-joining tree was constructed using superimposed branch lengths generated using the method of Fitch and Margoliash (1967). The 16S rRNA gene sequence of *E. coli* (GenBank accession number J01859) was used as an outgroup.

4.2.3. Carotenoid extraction and analysis

To extract carotenoids, plate-grown *Hymenobacter* cultures were added to 2 mL of HPLC-grade methanol (Fisher Scientific) and extracted at 65°C for 5 min. Cell debris was pelleted at low speed in a clinical centrifuge, the supernatant removed and the pellet extraction repeated using fresh methanol. Pooled supernatants were dried under a stream of nitrogen and redissolved in a small volume of methanol by gentle heating.

UV-Vis absorption spectra of crude methanolic extracts were determined using an Ultraspec 3000 spectrophotometer (Pharmacia Biotech). Wavelengths from 200 nm to 800 nm were recorded at 0.5 nm intervals.

HPLC analyses of concentrated methanolic extracts were conducted using an Agilent 1100 series HPLC (Agilent) equipped with a 125x4 mm LiChrospher 100 RP-18 column, 5 μ m particle size (Agilent). Carotenoids were eluted at a flow rate of 1.5 mL/min by a 10 min linear gradient of 100% 80:20 HPLC-grade methanol:MilliQ-filtered water to 100% 80:20 HPLC-grade methanol:HPLC-grade ethyl acetate (BDH) followed by 8 min of isocratic 80:20 methanol:ethyl acetate. All solvents were degassed by vacuum filtration through a 0.45 μ m pore diameter HVLP filter (Millipore) prior to use. Carotenoids were detected at 470 nm using the online photodiode array (PDA) detector with a 600 nm reference wavelength.

Peak areas generated by the HPLC auto-integrator were used for carotenoid quantification because all detected carotenoids shared similar absorption maxima (Table 4.2) and their absorption coefficients were unknown. For data analysis, peak areas were exported to an EXCEL spreadsheet and peaks with retention times less than 1.2 min or areas less than 100 milli absorbance units (mAu) \cdot sec were excluded to eliminate injection artifacts and background noise, respectively. Peaks with widths greater than 0.22 min were also excluded except in cases where a distinct peak was obviously apparent from visual inspection of the HPLC chromatogram. In time course assays, certain peaks below the 100 mAu \cdot sec threshold were included for uniformity based on their presence in more concentrated parallel samples. Peak areas of *cis*-carotenoids, as determined by their distinct UV spectra (Britton et al. 2004), were added to those of their all *trans*-isomers to further simplify analysis. The identities of these *cis*-peaks were confirmed by the homogeneous matrix-assisted laser desorption/ionization (MALDI) mass spectra of mixtures containing both *cis*- and *trans*-isomers (J. L. Klassen, unpublished results).

4.2.4. Carotenoid purification and mass spectrometry

To isolate large volumes of carotenoids, cell biomass from 25-30 replicate R2A plates of isolates VUG-A42aa and VUG-A141a were separately extracted twice in 100 mL of HPLC-grade methanol at 65°C for 5 min and gravity filtered

through Whatman #1 filter paper (Whatman), which was finally washed twice with 25 mL of methanol. Pooled filtrate was concentrated to near dryness by rotary evaporation, the carotenoid extract removed and the flask washed several times with small volumes of methanol which was added to the concentrate and dried under a nitrogen stream until visual precipitation occurred. After warming to room temperature, carotenoid precipitates were dissolved by adding a minimal volume of methanol.

Individual carotenoids were purified by repeated preparative HPLC using the conditions described above or slight modifications thereof. Pooled fractions were concentrated by rotary evaporation and/or drying under nitrogen gas, redissolved in methanol and repurified as described above until no further contaminants were visible using the PDA detector. High-resolution mass spectra were determined by electrospray ionization (ESI) as sodiated derivatives using a Mariner Biospectrometry Workstation (ABI) in positive ionization mode (carotenoids 5 and 6) or by high resolution MALDI in a *trans*-2-[3-{4-*tert*-butylphenyl}-2-methyl-2-propenylidene]malononitrile matrix using a Bruker Daltonics 9.4T Apex-Qe FTICR mass spectrometer (Bruker Daltonics) in positive ionization mode (carotenoids 3, 4, 7, 8, and 9).

4.2.5. Nucleotide sequence accession numbers

The sequences obtained in this study were deposited in GenBank under accession numbers EU155008 to EU155017 and EU159489.

4.3. Results

4.3.1. Phylogenetic analysis of the genus *Hymenobacter* and related VUG strains

An updated phylogenetic tree of the genus *Hymenobacter* (Figure 4.1) was constructed using the nearly full length 16S rRNA gene sequences of all currently described *Hymenobacter* species (October 2007) plus 10 novel VUG strains. Isolates VUG-A60a and VUG-A141a were closely related to *H. roseosalivarius*

and isolates VUG-A23a, VUG-A142 and VUG-A124 were loosely affiliated with *H. chitinivorans*. Isolates VUG-A33, VUG-A34, VUG-A42aa, VUG-A67 and VUG-C4 formed a cluster (hereafter “novel VUG clade”) distinct from all other strains and may comprise a novel genus based on <95% 16S rRNA gene sequence similarity (Stackebrandt and Goebel 1994) to all other described *Hymenobacter* species.

4.3.2. Carotenoid content in *Hymenobacter* and related strains

One feature common to all *Hymenobacter* and VUG strains is their bright red-pink pigmentation. Analyses of crude methanolic extracts of *Hymenobacter* and VUG strains by UV-Vis spectroscopy revealed the common presence of a

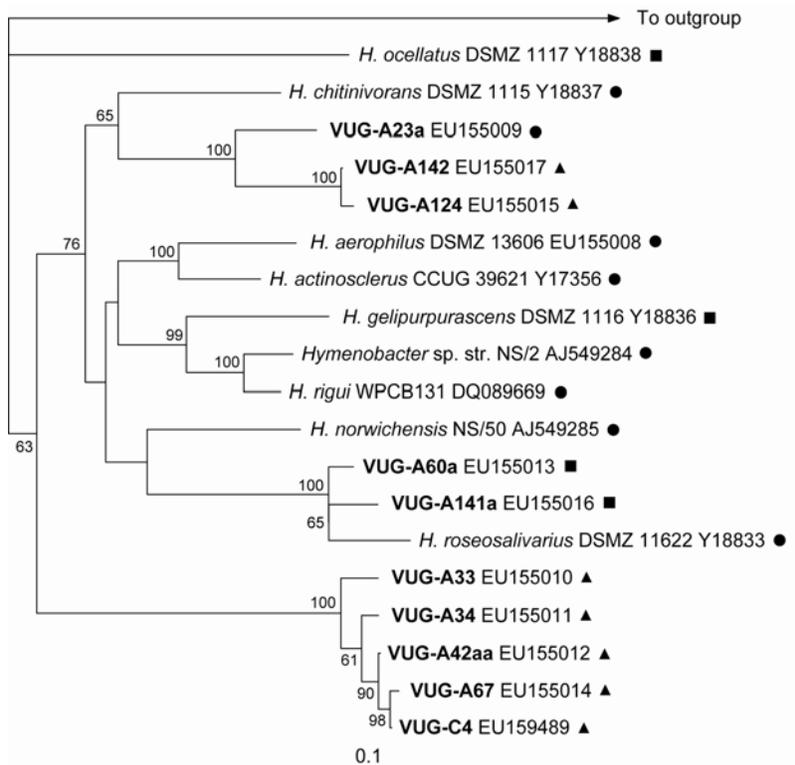


Figure 4.1. Near full-length 16S rRNA gene phylogeny of cultured *Hymenobacter* and *Hymenobacter*-like VUG strains (in boldface) constructed by neighbor-joining. Only bootstrap values >50 are shown. The scale bar shown represents a 10% difference in nucleotide composition. The 16S rRNA gene sequence of *E. coli* was used as an outgroup. GenBank accession numbers are indicated to the right of the sequence name. The major carotenoid content of strains is indicated as follows: circles, carotenoid 5 only; triangles, carotenoid 5 and 6; squares, carotenoids 3, 4 and 5.

single broad peak with an absorbance maximum at 500 nm and a shoulder at 517 nm (results not shown). In contrast, HPLC analyses revealed a total of nine unique carotenoids, the distribution of which differed between strains (Table 4.1, Figure 4.1). Stereoisomers of carotenoids 5-7 (bracketed in Figure 4.2) were detected and resolved by this HPLC method and identified by their characteristic UV absorption spectra (Schmidt et al. 1994, Britton et al. 2004). No stereoisomers of any other carotenoids were observed.

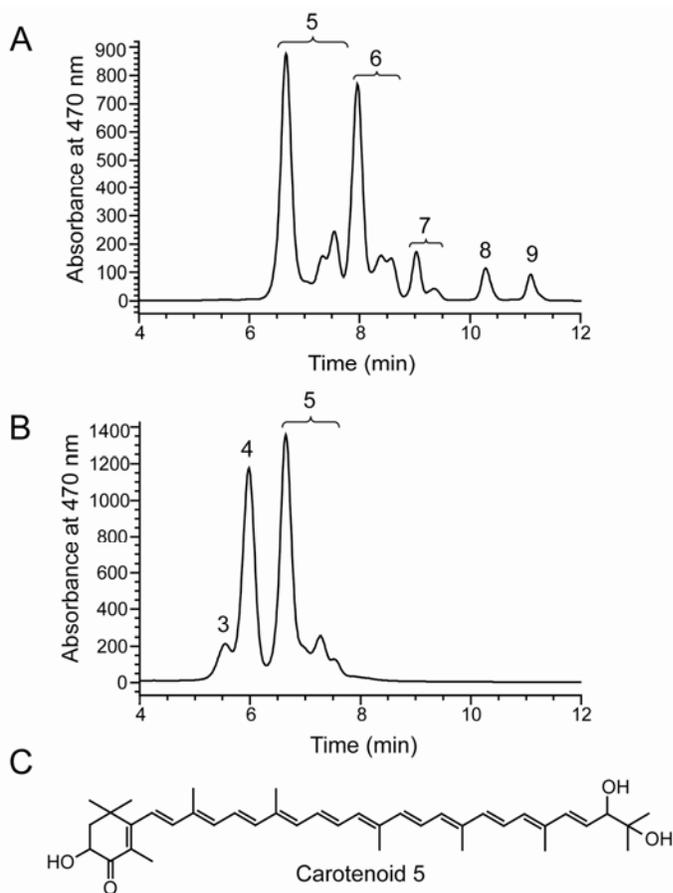


Figure 4.2. Representative HPLC chromatograms for (A) VUG-A42aa and (B) VUG-A141a. Carotenoids were detected using the HPLC PDA detector at 470 nm and numbered in accordance with Tables 4.1 and 4.2. Carotenoids 1 and 2 were present at concentrations below detection level in this sample and are therefore not labeled. Brackets indicate stereoisomer groups as determined by the presence of *cis*-peaks in their UV-absorption spectra. An abbreviated timescale showing only 4-12 min is shown because no carotenoids were detected outside this range. Panel (C) shows the proposed structure of the *trans*-isomer of carotenoid 5, 2'-hydroxyflexixanthin.

Table 4.1. Percent abundance of carotenoids present in VUG and reference *Hymenobacter* strains grown on R2A plates for 1 week at 18°C in the dark. Values represent the means of three independent replicates and their standard deviations. Strains are presented in the same order as in Figure 4.1, and carotenoid abundances in boldface correspond to the groupings in the figure.

Strain	Percent total carotenoid peak area								
	1	2	3	4	5	6	7	8	9
<i>H. ocellatus</i> DSMZ 1117	- ^a	-	11 ± 0.7	9.3 ± 0.4	80 ± 0.8	tr	tr	-	-
<i>H. chitinivorans</i> DSMZ 1116	-	-	-	-	100 ± 0	-	-	-	-
VUG-A23a	-	-	-	tr	95 ± 0.7	3.7 ± 0.4	-	tr	-
VUG-A142	-	-	-	-	16 ± 0.6	84 ± 0.6	-	-	tr
VUG-A124	-	-	-	-	15 ± 2	84 ± 2	-	-	tr
<i>H. aerophilus</i> DSMZ 13606	-	-	-	-	93 ± 0.7	2.9 ± 0.7	-	1.5 ± 0.2	2.2 ± 0.1
<i>H. actinosclerus</i> CCUG 39621	-	-	-	-	99 ± 0.01	1.1 ± 0.01	-	-	-
<i>H. gelipurpurascens</i> DSMZ 1116	-	-	8.8 ± 0.1	8.2 ± 0.1	83 ± 0.2	-	-	-	-
<i>Hymenobacter</i> str. NS/2	-	-	-	tr	100 ± 0.3	-	-	-	-
<i>H. rigui</i> WPCB131	-	-	-	-	100 ± 0.2	-	-	-	-
<i>H. norwichensis</i> NS/50	-	-	-	tr	98 ± 0.1	1.1 ± 0.2	-	-	-
VUG-A60a	2.7 ± 0.6	4.0 ± 0.1	14 ± 0.5	23 ± 1	49 ± 2	tr	-	-	-
VUG-A141a	tr ^b	2.0 ± 0.1	9.6 ± 0.2	30 ± 1	56 ± 2	1.1 ± 0.03	-	-	-
<i>H. roseosalivarius</i> DSMZ 11622	-	-	-	tr	99 ± 0.2	tr	-	-	-
VUG-A33	-	-	-	-	62 ± 2	37 ± 2	tr	tr	tr
VUG-A34	-	-	-	-	79 ± 2	15 ± 2	tr	3.6 ± 0.3	2.1 ± 0.2
VUG-A42aa	-	-	-	-	49 ± 2	40 ± 0.9	5.1 ± 0.6	3.3 ± 0.5	2.8 ± 0.3
VUG-A67	-	-	-	-	27 ± 0.3	68 ± 1	2.9 ± 0.6	tr	1.7 ± 0.1
VUG-C4	-	-	-	-	86 ± 0.3	6.9 ± 0.2	-	6.8 ± 0.3	-

^a-: Below detection level

^btr: Trace amounts (<1% of the total carotenoids) detected

Carotenoid 5 was present in all *Hymenobacter* and VUG strains and in most cases was the major pigment observed (Table 4.1). In isolates VUG-A142 and VUG-A124, carotenoid 6 formed the major pigment with carotenoid 5 present in smaller amounts (Table 4.1). Carotenoid 6 was also present in members of the novel VUG clade (Table 4.1) despite their lack of close phylogenetic relationship with strains VUG-A124 and VUG-A142 (Figure 4.1). Carotenoids 7-9 were present in *H. aerophilus* and strains of the novel VUG clade (Table 4.1). Carotenoids 1-4 were detected in only two strains, VUG-A60a and VUG-A141a, whereas the phylogenetically close *H. roseosalivarius* (Figure 4.1) contained solely carotenoid 5 (Table 4.1). Carotenoids 3 and 4, but not carotenoids 1 and 2, were also present in *H. ocellatus* and *H. gelipurpurascens* (Table 4.1).

4.3.3. Chemical characterization of carotenoids

In order to chemically characterize all detected carotenoids from VUG-A141a and VUG-A42aa, which together possessed all carotenoids 1-9, UV-Vis absorption spectra of each carotenoid were obtained using the HPLC online PDA detector (Table 4.2). Additionally, carotenoids 3-9 were isolated by preparative HPLC and characterized by high-resolution mass spectrometry (Table 4.2). The UV-Vis absorption spectra of carotenoids 1 and 2 featured one broad peak at 478 nm consistent with a ketolated photochrome containing 11-12 double bonds (Schmidt et al. 1994, Britton et al. 2004). Insufficient amounts of these carotenoids were recovered for their mass spectra to be determined. Carotenoids 3-7 exhibited identical UV-Vis absorption spectra with a single peak centered at 480-486 nm and a shoulder at 502 nm suggesting a common ketolated photochrome backbone containing 12 double bonds (Schmidt et al. 1994, Britton et al. 2004). The chemical formula inferred from the high-resolution mass spectrum of the sodiated ion of carotenoid 5, $C_{40}H_{54}O_4$, is consistent with that of 2'-hydroxyflexixanthin (Figure 4.2C), the previously determined major carotenoid of a "*Taxeobacter*" (now *Hymenobacter*; Buczolits et al. 2006) strain (Bircher and Pfander 1997). The inferred chemical formulae of carotenoids 3 ($C_{46}H_{64}O_9$), 4 ($C_{45}H_{62}O_8$), and 6 ($C_{41}H_{56}O_4$, derived from a sodiated ion) are consistent with

hexosyl, pentosyl and methyl derivatives of carotenoid 5, respectively. None of these carotenoids have been reported previously. The inferred chemical formula of carotenoid 7, $C_{41}H_{56}O_3$, is consistent with that of carotenoid 6 lacking a single hydroxyl group. Although three structural isomers are possible (methyl-3-deoxy-2'-hydroxyflexixanthin, methyl-flexixanthin and methyl-1',2'-dihydro-2,2'-dihydroxy-3',4'-didehydro-4-keto- γ -carotene; Figure 4.3A), methyl-3-deoxy-2'-hydroxyflexixanthin may be the most plausible given the prior report (Bircher and Pfander 1997) of 3-deoxy-2'-hydroxyflexixanthin as a minor carotenoid present in

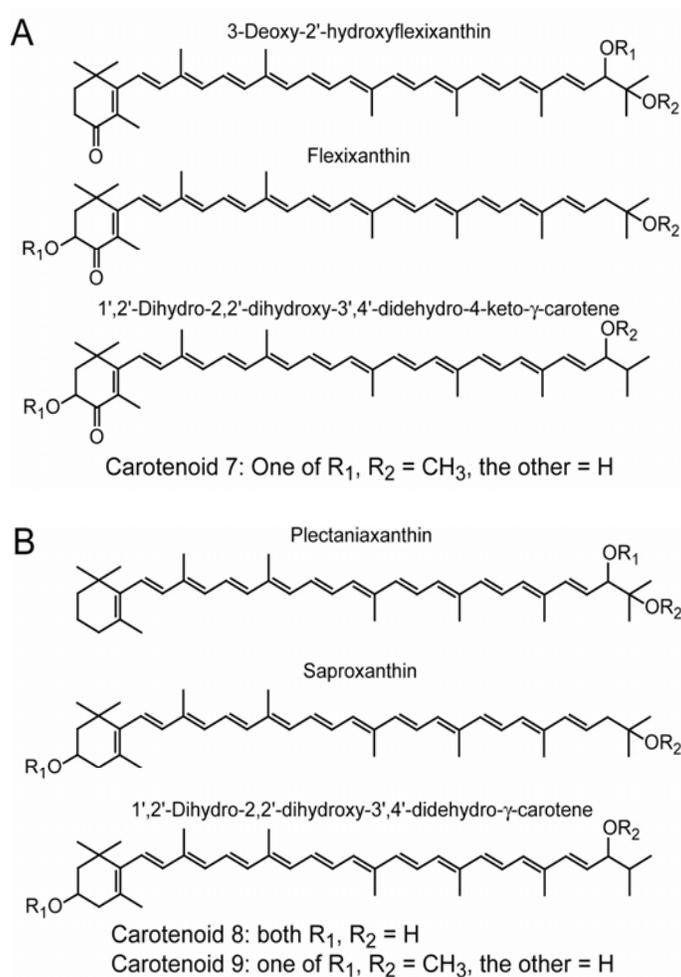


Figure 4.3. Possible chemical structures of (A) carotenoid 7 and (B) carotenoids 8 and 9. The location of the methyl group may occur at either oxygen in the structures of carotenoids 7 and 9.

Table 4.2. UV-Vis absorption maxima, high-resolution MALDI molecular weights, deduced chemical formulae, and presumptive identifications of carotenoids purified from strains VUG-A141a (carotenoids 1-5) and VUG-A42aa (carotenoids 6-9). The high-resolution ESI molecular weights of carotenoids 5 and 6 are given as sodiated derivatives with their non-sodiated low-resolution molecular weights in brackets.

Carotenoid	UV-Vis maxima, nm	Mass ions, <i>m/z</i>	Deduced Chemical Formula	Presumptive Identity
1	478	nd ^a	nd	Unknown
2	478	nd	nd	Unknown
3	484, 502, sh ^b	760.45449	C ₄₆ H ₆₄ O ₉	Hexosyl-2'-hydroxyflexixanthin
4	480, 502, sh	730.44392	C ₄₅ H ₆₂ O ₈	Pentosyl-2'-hydroxyflexixanthin
5	486, 502, sh	621.39143 (598.4)	C ₄₀ H ₅₄ O ₄ Na C ₄₀ H ₅₄ O ₄	2'-Hydroxyflexixanthin
6	484, 502, sh	635.40708 (612.4)	C ₄₁ H ₅₆ O ₄ Na C ₄₁ H ₅₆ O ₄	Methyl-2'-hydroxyflexixanthin
7	480, 502, sh	596.42240	C ₄₁ H ₅₆ O ₃	Methyl-3-deoxy-2'-hydroxyflexixanthin, methyl-flexixanthin or methyl-1',2'-dihydro-2,2'-dihydroxy-3',4'-didehydro-4-keto- γ -carotene
8	446, 472, 502	568.42748	C ₄₀ H ₅₆ O ₂	Plectanixanthin, saproxanthin or 1',2'-dihydro-2,2'-dihydroxy-3',4'-didehydro- γ -carotene
9	446, 470, 500	582.44313	C ₄₁ H ₅₈ O ₂	Methyl-plectanixanthin, methyl-saproxanthin or methyl-1',2'-dihydro-2,2'-dihydroxy-3',4'-didehydro- γ -carotene

^and: not determined due to low abundance (Table 4.1)

^bsh: shoulder peak

the previously studied “*Taxeobacter*” strain; the latter compound was not detected in the current study. Carotenoids 8 and 9 exhibited UV-Vis absorption spectra with maxima at 446, 470-472 and 500-502 nm, suggesting a common non-ketolated photochrome containing 12 double bonds (Schmidt et al. 1994, Britton et al. 2004). The mass spectra of carotenoids 8 (C₄₀H₅₆O₂) and 9 (C₄₁H₅₈O₂) are consistent with plectanixanthin, saporixanthin or 1',2'-dihydro-2,2'-dihydroxy-3',4'-didehydro- γ -carotene (Figure 4.3B) and their methyl derivatives, respectively. Plectanixanthin and methyl-plectanixanthin are the most likely, based on the structures described previously (Bircher and Pfander 1997) and, assuming a shared biosynthetic pathway.

4.3.4. The effect of culture age on carotenoid composition in *Hymenobacter* and related strains

Because of my inability to culture the VUG strains reliably in liquid media, I was unable to standardize carotenoid production based on optical density or growth phase. Previous studies of other genera have determined that carotenoid content is maximal and biosynthesis most complete during stationary phase (Bhosale and Bernstein 2004, Veiga-Crespo et al. 2005). All *Hymenobacter* and VUG strains were therefore incubated on plates under the standard conditions of 1 week at 18°C in the dark because these incubation conditions uniformly resulted in visually robust growth presumed to correspond to maximal cell density. To determine the effect of these growth conditions on the absolute quantification of carotenoid content in the *Hymenobacter* and VUG strains, I examined the carotenoid composition of parallel cultures grown on R2A plates and harvested at different times. Some strains such as *H. gelipurpurascens* (Figure 4.4A) and VUG-A124 (data not shown) maintained a constant carotenoid composition throughout incubation for 26 d. Others, such as VUG-A141a (Figure 4.4B), *H. aerophilus* and VUG-A33 (data not shown), required 7-9 d for carotenoid composition to stabilize. In these cases, presumptive glycosylated or methylated carotenoids accumulated with concomitant reduction of their precursor carotenoid 5 (Figure 4.4B).

The standard culture conditions used in these experiments were chosen without prior quantification of the effect of culture age on carotenoid composition. Although in hindsight I recognize that this could have compromised quantification of the carotenoid content in the strains (Table 4.1), the results demonstrate that whereas absolute quantification of carotenoid composition may not be precise without standardization of growth phase (Figure 4.4B), this method is capable of identifying the major carotenoids present and discriminating between the carotenoid compositions of related organisms (Table 4.1). Furthermore, absolute quantification of carotenoid mixtures using coupled HPLC

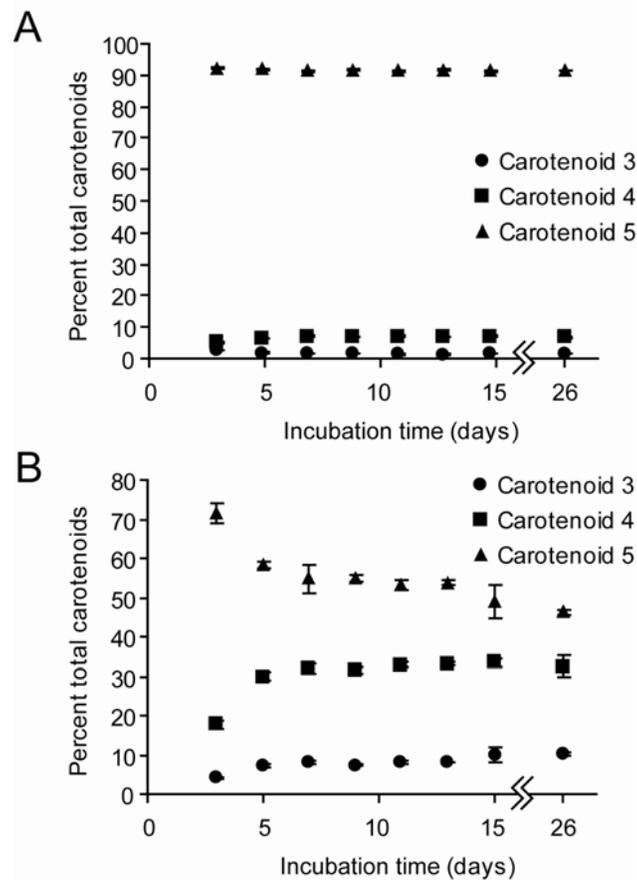


Figure 4.4. Relative carotenoid composition of (A) *H. gelipurpurascens* and (B) VUG-A141a during incubation for 26 d on parallel R2A plates incubated in the dark at 18°C. Percentages are determined from on-line PDA detection of absorption at 470 nm relative to 600 nm as quantified using the HPLC auto-integrator. Values are the means of three independent replicates and error bars (where visible) show 1 standard deviation.

and UV-Vis absorption spectroscopy is technically difficult due to variability of their absorbance maxima and extinction coefficients. I am satisfied that the method applied here is sufficient to accomplish the purpose of this study, namely, to systematically identify the carotenoids present in a related group of organisms for comparison with their 16S rRNA gene phylogeny.

4.4. Discussion

Carotenoid diversity is important because of both its biotechnological potential (Sandmann et al. 1999) and its role in understanding the evolution of secondary metabolism (Umeno et al. 2005). Ascertaining the extent of carotenoid diversity from the literature is problematic due to the focus of most high-resolution studies on a single strain and/or the utilization of methods having insufficient resolution to differentiate among related compounds. For example, the microorganisms in the *Bacteroidetes* division for which high resolution carotenoid structures have been determined (Bircher and Pfander 1997, Lutnaes et al. 2002, Lutnaes et al. 2004, Shindo et al. 2007) are all taxonomically distantly related. From these data it cannot be deduced if the carotenoids described are representative of the taxa studied. In contrast, when taxonomically close microorganisms have been compared (e.g., Bernardet et al. 2002 and references therein) the methods used could only characterize pigments as carotenoids or, at best, flexirubrin-like. These methods are clearly insufficient to accurately analyze carotenoid distribution and diversity in these microorganisms. In contrast, the current study used an HPLC-based method to detect nine different carotenoids, including some previously unreported in the literature, and to estimate their relative abundance in a taxonomically related group of 19 strains. Comparison of the carotenoid distribution in these strains with their 16S rRNA gene phylogeny (Figure 4.1) indicates that carotenoid composition varies among genera (i.e., the novel VUG clade vs. all other described *Hymenobacter* spp.) and between some species (e.g., *H. gelupurpurascens* vs. *H. rigui*). Furthermore, the erratic phylogenetic distribution of certain carotenoids (e.g., carotenoids 3 and 4)

strongly suggests differences in the evolution of their cognate biosynthetic pathways relative to their phylogenetic neighbors.

In nearly all organisms, carotenoid biosynthesis proceeds through a conserved central pathway leading from geranylgeranyl pyrophosphate to lycopene (Sieiro et al. 2003). At this point, the pathway branches by further cyclization, ketolation, hydroxylation, esterification, desaturation or epoxidation reactions (Sieiro et al. 2003, Cheng 2006); the differential presence or regulation of the responsible enzymes results in the formation of distinct carotenoids in different taxa. As terminal “branches” in a tree-like model of carotenoid evolution (Umeno et al. 2005) these later biosynthetic steps are the most evolutionarily plastic due to their modulation (as opposed to gain or loss) of carotenoid structure and, therefore, function.

Based on the carotenoids identified in this study, I infer a biosynthetic pathway in *Hymenobacter* leading from geranylgeranyl pyrophosphate to lycopene, typical of most other bacteria (Sieiro et al. 2003). Lycopene is subsequently cyclized, ketolated and hydroxylated at one end of the molecule and hydroxylated and desaturated at the other, in accordance with other related studies (Teramoto et al. 2003, Teramoto et al. 2004, Tao et al. 2006). It is unclear from the current results whether carotenoids 7-9 represent biosynthetic intermediates of this pathway or the products of a separate, parallel one (e.g., Giraud et al. 2004). The inferred terminal biosynthetic step in the *Hymenobacter* carotenoid biosynthetic pathway is glycosylation and methylation leading to the production of carotenoids 3 and 4 or 6, 7 and 9, respectively. The erratic distribution of these carotenoids (Table 4.1) clearly indicates differences in the distribution or regulation of the cognate enzymes as a consequence of differential gene gain, gene loss or evolution of regulatory mechanisms in these organisms. This inference is in accordance with the increased evolutionary plasticity of these terminal biosynthetic “branches”.

From a taxonomic perspective, these results support the application of carotenoid composition to discriminate between organisms of different genera, as

typically applied in polyphasic taxonomic studies (Vandamme et al. 1996). In some cases carotenoid composition can differ between species (Table 4.1); this property must be properly supported by other taxonomic and/or genetic tests to achieve specific identification. The results of this study do, however, support the increased application of HPLC-based methods (or others with similar structural resolution) in the systematic study of carotenoid distribution such as that compiled by Bernardet et al. for the *Flavobacteriaceae* (Bernardet et al. 2002). This will allow the identification of novel carotenoids, the cognate biosynthetic enzymes of which may be applicable in the design of recombinant biosynthesis pathways (Sandmann et al. 1999, Umeno et al. 2005). Of the 213 papers published in 2006 in the International Journal of Systematic and Evolutionary Microbiology reporting pigmented strains, only two used HPLC to precisely determine the pigments present. This suggests that increased adoption of these methods by the systematics community will greatly enhance our currently limited knowledge of carotenoid diversity and distribution.

4.5. Literature cited

- Albrecht, M., Takaichi, S., Steiger, S., Wang, Z.-Y., and Sandmann, G. 2000. Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*. *Nat. Biotechnol.*, 18: 843-846.
- Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., and Weightman, A.J. 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl. Environ. Microbiol.*, 71: 7724-7736.
- Bernardet, J.F., Nakagawa, Y., and Holmes, B. 2002. Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int. J. Syst. Evol. Microbiol.*, 52: 1049-1070.
- Bhosale, P., and Bernstein, P.S. 2004. β -Carotene production by *Flavobacterium multivorum* in the presence of inorganic salts and urea. *J. Ind. Microbiol. Biotechnol.*, 31: 565-571.
- Bircher, C., and Pfander, H. 1997. Synthesis of carotenoids in the gliding bacteria *Taxeobacter*: (all-*E*,2'*R*)-3-deoxy-2'-hydroxyflexixanthin. *Helv. Chim. Acta*, 80: 832-837.
- Britton, G., Liaaen-Jensen, S., and Pfander, H. 2004. Carotenoids handbook Birkhäuser Verlag, Basal, Switzerland.

- Brosius, J., Dull, T.J., Sleeter, D.D., and Noller, H.F. 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J. Mol. Biol.*, 148: 107-127.
- Cheng, Q. 2006. Structural diversity and functional novelty of new carotenoid biosynthesis genes. *J. Ind. Microbiol. Biotechnol.*, 33: 552-559.
- Cheng, S.M., and Foght, J.M. 2007. Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.*, 59: 318-330.
- Dear, S., and Staden, R. 1991. A sequence assembly and editing program for efficient management of large projects. *Nucleic Acids Res.*, 19: 3907-3911.
- Del Campo, J.A., García-González, M., and Guerrero, M.G. 2007. Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. *Appl. Microbiol. Biotechnol.*, 74: 1163-1174.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics*, 5: 164-166.
- Fitch, W.M., and Margoliash, E. 1967. The construction of phylogenetic trees. *Science*, 155: 279-284.
- Foght, J., Aislabie, J., Turner, S., Brown, C.E., Ryburn, J., Saul, D.J., and Lawson, W. 2004. Culturable bacteria in subglacial sediments and ice from two Southern Hemisphere glaciers. *Microb. Ecol.*, 47: 329-340.
- Fraser, P.D., and Bramley, P.M. 2004. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.*, 43: 228-265.
- Giraud, E., Hannibal, L., Fardoux, J., Jaubert, M., Jourand, P., Dreyfus, B., Sturgis, J.N., and Verméglio, A. 2004. Two distinct *crt* gene clusters for two different functional classes of carotenoid in *Bradyrhizobium*. *J. Biol. Chem.*, 279: 15076-15083.
- Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16: 111-120.
- Lutnaes, B.F., Oren, A., and Liaaen-Jensen, S. 2002. New C₄₀-carotenoid acyl glycoside as principal carotenoid in *Salinibacter ruber*, an extremely halophilic eubacterium. *J. Nat. Prod.*, 65: 1340-1343.
- Lutnaes, B.F., Strand, A., Pétursdóttir, S.K., and Liaaen-Jensen, S. 2004. Carotenoids of thermophilic bacteria - *Rhodothermus marinus* from submarine Icelandic hot springs. *Biochem. Syst. Ecol.*, 32: 455-468.
- Mortensen, A. 2006. Carotenoids and other pigments as natural colorants. *Pure Appl. Chem.*, 78: 1477-1491.
- Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., Komemushi, S., Miki, W., and Misawa, N. 2005. Elucidation of a

- carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'- β -hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanthophylls. *Appl. Environ. Microbiol.*, 71: 4286-4296.
- Rice, P., Longden, I., and Bleasby, A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, 16: 276-277.
- Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, O., and Böger, P. 1999. The biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. *Trends Biotechnol.*, 17: 233-237.
- Schmidt, K., Connor, A., and Britton, G. 1994. Analysis of pigments: carotenoids and related polyenes. *In* Chemical methods in prokaryotic systematics. M. Goodfellow and A.G. O'Donnell. (ed.) John Wiley & Sons, Chichester. pp. 403-461.
- Shindo, K., Kikuta, K., Suzuki, A., Katsuta, A., Kasai, H., Yasumoto-Hirose, M., Matsuo, Y., Misawa, N., and Takaichi, S. 2007. Rare carotenoids, (3R)-saproxanthin and (3R,2'S)-myxol, isolated from novel marine bacteria (*Flavobacteriaceae*) and their antioxidative activities. *Appl. Microbiol. Biotechnol.*, 74: 1350-1357.
- Sieiro, C., Poza, M., de Miguel, T., and Villa, T.G. 2003. Genetic basis of microbial carotenogenesis. *Int. Microbiol.*, 6: 11-16.
- Stackebrandt, E., and Goebel, B.M. 1994. A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Bacteriol.*, 44: 846-849.
- Tao, L., Yao, H., Kasai, H., Misawa, N., and Cheng, Q. 2006. A carotenoid synthesis gene cluster from *Algoriphagus* sp. KK10202C with a novel fusion-type lycopene β -cyclase gene. *Mol. Genet. Genomics*, 276: 79-86.
- Teramoto, M., Rählert, N., Misawa, N., and Sandmann, G. 2004. 1-Hydroxy monocyclic carotenoid 3,4-dehydrogenase from a marine bacterium that produces myxol. *FEBS Lett.*, 570: 184-188.
- Teramoto, M., Takaichi, S., Inomata, Y., Ikenaga, H., and Misawa, N. 2003. Structural and functional analysis of a lycopene β -monocyclase gene isolated from a unique marine bacterium that produces myxol. *FEBS Lett.*, 545: 120-126.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25: 4876-4882.
- Umeno, D., Tobias, A.V., and Arnold, F.H. 2005. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol. Mol. Biol. Rev.*, 69: 51-78.

- Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K., and Swings, J. 1996.
Polyphasic taxonomy, a consensus approach to bacterial systematics.
Microb. Rev., 60: 407-438.
- Veiga-Crespo, P., Blasco, L., dos Santos, F.R., Poza, M., and Villa, T.G. 2005.
Influence of culture conditions of *Gordonia jacobaea* MV-26 on
canthaxanthin production. Int. Microbiol., 8: 55-58.

5. 1'-Xylosyl and 2'-Methyl derivatives of 2'-Hydroxyflexixanthin are Major Carotenoid Pigments of *Hymenobacter*²

5.1. Introduction

Bacteria of the genus *Hymenobacter* (*Flexibacteraceae*, *Bacteroidetes*) are often isolated from dry environments subject to intense oxidative stress (see Section 1.1). Accordingly, all isolated *Hymenobacter* strains are colored bright red-pink due to the presence of carotenoids, natural pigments with notable biotechnological application as natural colorants (Mortensen 2006) and nutritional supplements (Sandmann et al. 1999, Fraser and Bramley 2004). Carotenoid antioxidant function is well established, especially in *Deinococcus* (Tian et al. 2007, Zhang et al. 2007, Tian et al. 2008); this genus has been repeatedly co-detected with *Hymenobacter* (Rainey et al. 2005, Saul et al. 2005, Fredrickson et al. 2008). Carotenoids likely play an important antioxidative role in both genera.

Ten *Hymenobacter*-related strains were previously isolated from Victoria Upper Glacier, Antarctica glacial ice (Chapters 2-4, Klassen and Foght 2008). In these strains and an additional nine reference *Hymenobacter* species, I detected and isolated seven chemically distinct carotenoids using high-performance liquid chromatography (HPLC), with the exact carotenoid composition varying among strains (Chapters 3 and 4, Klassen and Foght 2008). Based on in-line HPLC UV-Vis spectra, high-resolution mass spectrometry (MS) and comparison to previous synthetic work (Bircher and Pfander 1997) claiming to represent the major carotenoids of an unidentified "*Taxeobacter*" (now *Hymenobacter*; Buczolits et al. 2006) strain, these carotenoids were proposed to be 2'-hydroxyflexixanthin, its pentosyl-, hexosyl- and methyl- derivatives and related non-ketolated precursors. Because the previous synthetic work (Bircher and Pfander 1997) did not present the isolation and characterization of 2'-hydroxyflexixanthin from its natural source, the identification of this pigment as the major carotenoid in "*Taxeobacter*"

²A version of this chapter has been published. Klassen, J. L., McKay, R. and Foght J. M. 2009. 1'-Xylosyl and 2'-Methyl Derivatives of 2'-Hydroxyflexixanthin and Major Carotenoids Pigments of *Hymenobacter*. *Tetrahedron Lett.* 50:2656-2660

(now *Hymenobacter*) is considered unsubstantiated (Britton et al. 2004). Here I report full characterization of the four most abundant carotenoids isolated from *Hymenobacter*-like strains VUG-A42aa and VUG-A141a using ^1H -NMR, circular dichroism (CD) and UV-Vis spectroscopy and previously generated high-resolution MS data (Chapter 4, Klassen and Foght 2008).

5.2. Materials and methods

5.2.1. General experimental procedures

UV-Vis absorption spectra of all purified carotenoids were determined in HPLC-grade methanol and HPLC-grade chloroform (Fisher) using an Ultrospec 3100 pro spectrometer (Biochrom). Absorption at wavelengths between 200 and 800 nm was recorded at 1 nm intervals. CD spectra of all purified carotenoids were determined at room temperature in 95% ethanol (Fisher) using an Olis DSM 17 spectrometer. Polarization was recorded at 1 nm intervals between 195 and 600 nm.

Both one-dimensional ^1H and two-dimensional $^1\text{H}, ^1\text{H}$ total correlation spectroscopy (TOCSY; Levitt et al. 1982, Bax and Davis 1985, Kupce and Freeman 1993) was conducted by Dr. Ryan McKay at the National High Field NMR Center (NANUC), Edmonton, Alberta. NMR spectra were collected at 800 MHz (Oxford 18.8 T) on a Varian Inova console controlled by a Sun Blade (Solaris) host computer. Samples were dissolved in 600 μL of methylene chloride (D_2 – 99.9%; Cambridge Isotope Laboratories Inc.), placed in 5 mm Wilmad 535-PP-9 constricted tubes and stoppered. Flame sealing of these tubes often caused abrupt sample degradation. Samples were either prepared just prior to data acquisition or stored in the dark at -20°C until run at 25°C . The TOCSY pulse sequence used was the standard BioPack (Varian Inc.) version 2008-08-19 and experimental details are provided in Table 5.1. All experiments were conducted in a linear-uniform fashion with sweep widths of 11990 and 8000 Hz for the directly and indirectly detected dimensions, respectively. All experiments were processed and analyzed using Varian software VNMRJ 1.1D native Macintosh on a G5 PowerPC with OS X 10.5x. Each spectrum was referenced to the major

CD₂Cl₂ peak (defined as δ5.31). When processing the data a Pi/3 shifted cosine squared apodization was applied prior to zero filling the acquired data to a total data size of twice the acquired data.

5.2.2. Bacterial material and isolation procedures

The isolation and phylogenetic position of strains VUG-A141a and VUG-A42aa within the genus *Hymenobacter* have been described elsewhere (Chapters 2-4, Klassen and Foght 2008). To generate biomass for carotenoid extraction, both strains were grown on R2A agar plates (Difco) in the dark at 18°C for 1 week because growth in liquid media is poor or absent (Chapters 3 and 4, Klassen and Foght 2008). For each strain, biomass harvested from 45 replicate plates was extracted twice using 100 mL of 80:20 methanol-acetone (HPLC grade; Fisher) at 65°C for 5 min and gravity filtering the extract each time through Whatman #1 filter paper which was finally washed twice with 25 mL 80:20 methanol-acetone. For each sample the resulting filtrate was pooled, dried by rotary evaporation and dissolved in a minimal volume of methanol. Carotenoids were purified by preparative HPLC as described previously (Chapter 4, Klassen and Foght 2008) until only one peak was detected by the in-line UV-Vis photo diode array (PDA) detector at 478 nm. In-line UV-Vis spectroscopy suggested that each collected peak contained nearly entirely the all-*trans* isomer, with *cis*-carotenoids excluded from analysis for simplicity.

Table 5.1. Experimental Parameters for 2D-¹H, ¹H-TOCSY NMR Experiments.

Carotenoid	Recycle Delay (sec)	Scans per Increment	Pulse Width (μs)	Complex points (direct, indirect)	Spinlock duration and field strength	Pulse sequence ^a	Total Time (h)
4	2	16	10.25	2048,128	38 μs at 9000 Hz	MF	2.5
5	1	64	8.375	4096, 256	44 μs at 5000 Hz	No presat	13
6	3	32	9.25	2048,128	44 μs at 8000 Hz	SF	7.5
7	2	16	10.25	2048,128	38 μs at 9000 Hz	MF	2.5
8	3	32	10.5	4096,128	44 μs at 8000 Hz	No presat	7.5

^aSF refers to single frequency saturation of the residual water resonance; MF refers to multi-frequency saturation of both the residual water and non-deuterated (i.e., CDCl₂H) solvents; No presat indicates that no saturation pulses were used.

5.3. Results and discussion

High resolution MS and in-line HPLC UV-Vis spectroscopy previously suggested that the dominant carotenoid in all surveyed *Hymenobacter* strains (carotenoid 5; Chapter 4, Klassen and Foght 2008) is a monocyclic ketocarotenoid containing 12 double bonds and having a molecular formula of $C_{40}H_{54}O_4$. UV-Vis spectroscopy of purified carotenoid 5 shows a broad peak with a central absorbance maximum at 494 nm and minimal spectral fine structure (Supplemental Figure C1), consistent with this proposed structure. The 1H -NMR spectra of carotenoid 5 (Table 5.2), Supplemental Figures C2 and C3) are consistent with those previously reported for 2'-hydroxyflexixanthin (Andrewes et al. 1984), with important diagnostic peaks at δ 1.76 (H-2; doublet of doublets split by H-3 and OH-3; see Figure 5.1 for numbering scheme), δ 4.28 (H-3; doublet of doublets split by H-2_{ax} and H-2_{eq}) and δ 3.96 (H-2'; linked in the 2D-TOCSY to H-3' and H-4'); the latter shift clearly suggests hydroxylation at the 2'-position by comparison with the NMR structure of flexixanthin (Andrewes et al. 1984). The CD spectrum of carotenoid 5 (Figure 5.2) clearly indicates *S* stereochemistry at the 2'-hydroxyl position by comparison to previously published data observed for any of the 2'-methoxy carotenoids identified in this study (see below). I therefore hesitate to apply the additivity hypothesis for carotenoids (Bartlett et al. 1969) to the 3-hydroxyl group of carotenoid 5 and consider its stereochemistry undetermined. Based on this evidence, I identify carotenoid 5 as 2'*S*-hydroxyflexixanthin (Table 5.2), as previously reported (albeit without full substantiation; Britton et al. 2004) from a *Taxeobacter* (now *Hymenobacter*) species (Chapter 4, Klassen and Foght 2008).

Previous analyses (Chapter 4, Klassen and Foght 2008) suggested the presence of pentosyl (carotenoid 4; molecular formula: $C_{45}H_{62}O_8$) and hexosyl (carotenoid 3; molecular formula: $C_{46}H_{64}O_9$) derivatives of carotenoid 5 in strain VUG-A141a. The latter was present only in small amounts and was not analyzed further. Both the UV-Vis and 1H -NMR spectra (Supplemental Figures C1, C4

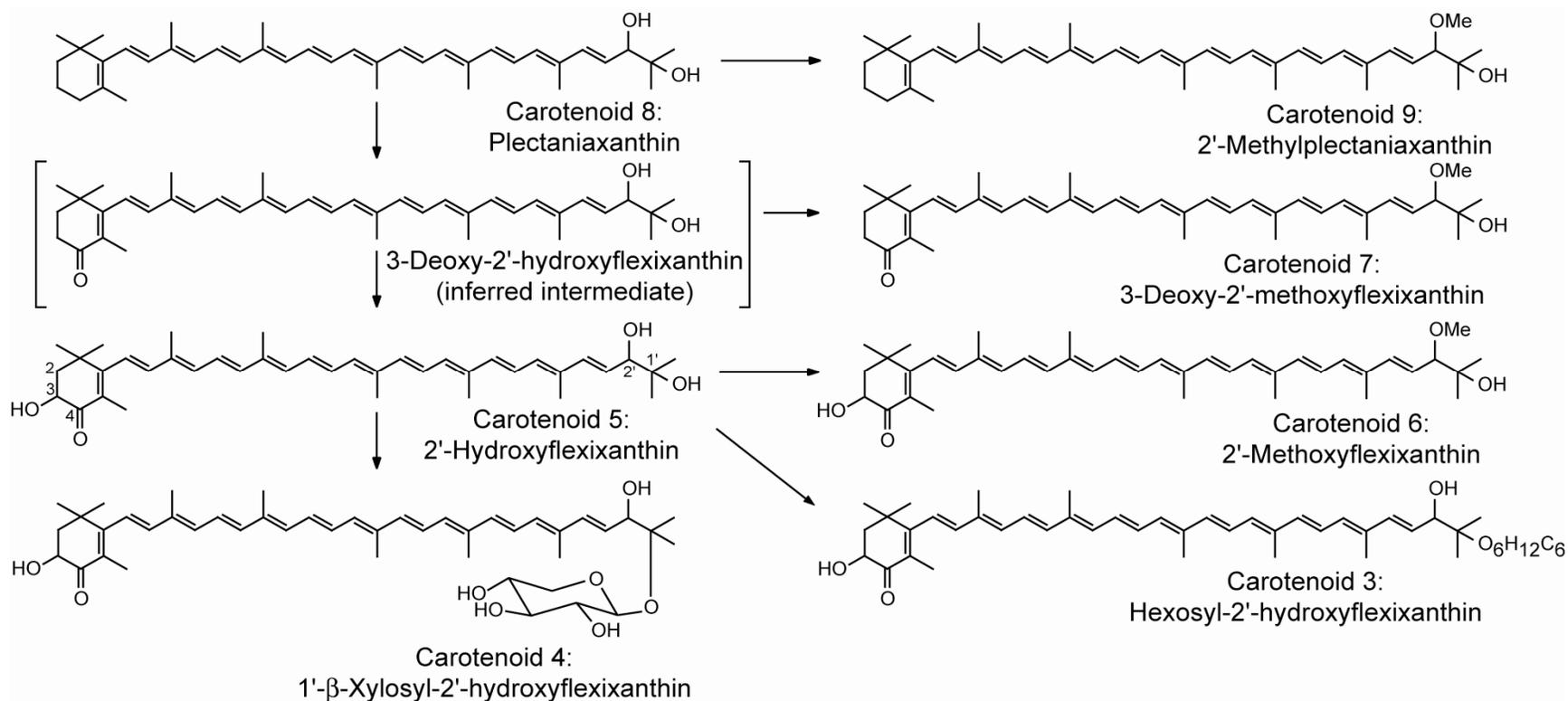


Figure 5.1. The major carotenoids in *Hymenobacter* and their proposed biosynthetic relationships. Compounds described in this study are numbered as in Chapter 4 and Klassen and Foght (2008). 3-Deoxy-2'-hydroxyflexixanthin, an inferred intermediate compound not detected in this study, is indicated in square brackets. Carbon numbers mentioned in the text are indicated for carotenoid 5.

Table 5.2. ¹H-NMR spectral assignments in CD₂Cl₂, referenced to the major solvent peak defined as δ5.31, for carotenoids 3 and 4, isolated from strain VUG-A141a, and 5 and 6, isolated from strain VUG-A42aa, determined using 1D-¹H-NMR and 2D-¹H-TOCSY spectra (Supplemental, Figures C2-C8).

Proton Location	Carotenoid							
	3		4		5		6	
	δ (ppm)	Multiplicity and <i>J</i> (Hz)	δ (ppm)	Multiplicity and <i>J</i> (Hz)	δ (ppm)	Multiplicity and <i>J</i> (Hz)	δ (ppm)	Multiplicity and <i>J</i> (Hz)
H-2	1.76	dd, 6.5, 6.5	1.75	dd, 6.6, 6.6	1.76	dd, 6.2, 6.2	1.84	t, 6.9
H-3	4.28	dd, 6.2, 13.9	4.19	dd, 5.9, 14.0	4.28	dd, 9.7, 20.4	2.45	t, 9.1
H-7	6.11	d, 10.3	6.19	d, 11.3	6.21	d, 17.2	6.21	d, 11.5
H-8	6.32	m	6.44	d, 16.2	6.52	d, 16.2	6.37	d, 17.2
H-10	6.20	d, 10.7	6.23	d, 15.0	6.33	d, 17.9	6.26	d, 15.9
H-11	6.62	t, 15.7	6.62	dd, 11.4, 15.0	6.63	t, 13.8	6.63	t, 15.8
H-12	6.39	d, 16.6	6.39	d, 9.9	6.40	d, 17.2	6.40	d, 19.6
H-14	6.29	m	6.26	m	6.30	m	6.30	m
H-15	6.95	m	6.94	m	6.94	m	6.94	m
H-15'	6.99	m	6.99	dd, 9.2, 15.8	6.98	m	6.98	m
H-14'	6.35	m	6.24	m	6.35	m	6.33	m
H-12'	6.45	d, 12.8	6.45	d, 14.8	6.45	d, 17.7	6.44	d, 15.3
H-11'	6.68	t, 10.75	6.67	m	6.68	t, 15.1	6.68	t, 16.4
H-10'	6.30	d, 9.28	6.32	m	6.29	m	6.36	m
H-8'	6.30	d, 8.42	6.30	m	6.26	d, 16.8	6.28	d, 13.9
H-7'	6.84	dd, 11.7,	6.84	m	6.85	dd, 11.5,	6.84	dd, 11.5,
H-6'	6.15	m	6.11	d, 12.1	6.16	d, 10.9	6.12	m
H-4'	6.37	d, 16.71	6.37	d, 16.3	6.33	d, 11.9	6.33	m
H-3'	5.71	t, 7.73	5.64	dd, 7.0, 15.8	5.53	t, 14.0	5.53	dd, 8.5, 16.5
H-2'	3.96	m	4.08	d, 7.0	3.41	m	3.40	d, 14.61
H-16	1.13	s	1.18	s	1.10	s	1.10	s
H-17	1.15	s	1.19	s	1.12	s	1.11	s
H-18	1.91	s	1.90	s	1.91	s	1.93	s
H-19	1.98	s	1.99	s	1.98	s	1.98	s
H-20	2.00	s	1.98	s	2.00	s	1.99	s
H-20'	2.00	s	1.98	s	2.00	s	2.00	s
H-19'	1.98	s	1.98	s	1.98	s	1.93	s
H-18'	1.92	s	1.90	s	1.93	s	1.25	s
H-17'	1.30	s	1.25	s	1.30	s	1.18	s
H-16'	1.19	s	1.23	s	1.18	s	2.11	s
2-OH	2.11	m	2.11	m	2.11	m	-	-
1'-OH	2.11	m	-	-	2.11	m	2.11	m
2'-OH	2.11	m	2.11	m	-	-	-	-
2'-Methyl	-	-	-	-	3.28	m	3.41	d, 5.57
Sugar:								
H-1''	-	-	4.53	d, 7.36	-	-	-	-
H-2''	-	-	3.30	dd, 7.36, 8.72	-	-	-	-
H-3''	-	-	3.48	dd, 8.69, 8.69	-	-	-	-
H-4''	-	-	3.67	ddd, 5.34,	-	-	-	-
H-5''a	-	-	3.29	dd, 9.92,	-	-	-	-
H-5''b	-	-	3.99	dd, 5.18,	-	-	-	-

and C5) of carotenoid 4 are nearly identical to those of carotenoid 5, excepting $^1\text{H-NMR}$ peaks for carotenoid 4 corresponding to a β -linked xylose (compare Table 5.2 with $^1\text{H-NMR}$ assignments for β -glucosylated and acyl- β -glucosylated monocyclic carotenoids; Lutnaes et al. 2002, Lutnaes et al. 2004; xylose has the same stereochemistry as glucose, excepting the C6 substituent; see also Shindo et al. 2008). Because TOCSY cross-correlations were present for both 2- and 2'-hydroxyl protons, xylosylation is presumed to be at the 1'-hydroxyl group, consistent with the position of glycosylation in other flavobacterial monocyclic carotenoids (Lutnaes et al. 2002, Lutnaes et al. 2004). The CD spectrum of carotenoid 4 (not shown) showed only a weak signal and was not comparable to that of carotenoid 5 or similar previously published spectra (Bartlett et al. 1969, Rønneberg et al. 1985), likely due to the presence of the glycosyl moiety. The stereochemistry of the 2'-hydroxyl group of carotenoid 5 therefore remains undetermined. I identify carotenoid 4 as 1'- β -xylosyl-2'-hydroxyflexixanthin (Table 5.2). This is the second report of a xylosylated carotenoid in the literature (Shindo et al. 2008) and the first in which the xylose moiety is not otherwise modified.

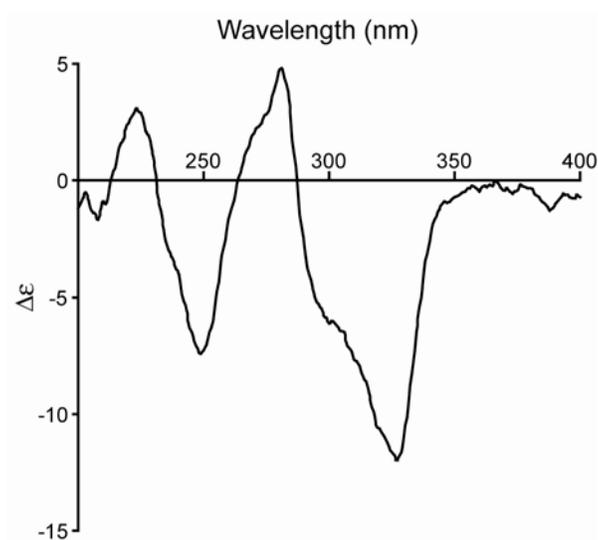


Figure 5.2. CD spectrum (200-400 nm) in ethanol at room temperature of carotenoid 5 isolated from strain VUG-A141a.

Strain VUG-A42aa possess two 12-conjugated double bond-containing ketocarotenoids in addition to 2'-hydroxyflexixanthin: carotenoid 6 (Chapter 4, Klassen and Foght 2008), with a molecular formula of $C_{41}H_{56}O_4$, presumed to be methyl-2'-hydroxyflexixanthin; and carotenoid 7 (Chapter 4, Klassen and Foght 2008), with a molecular formula of $C_{41}H_{56}O_3$, chemically similar to carotenoid 6 but lacking a hydroxyl group. In both cases 1H -NMR spectra (Table 5.2, Supplemental Figures C6-C8) suggest methylation at the 2'-hydroxyl group, based on the lack of interactions in the TOCSY spectrum between protons H-2' and H-3' and the hydroxyl hydrogen (δ 2.11; see also Bircher and Pfander 1997). This is consistent with the lack of CD signal for either carotenoid 6 or 7 (data not shown), presumably from masking or abolishment of stereochemistry at this position due to methylation. The presence of a peak at δ 2.45 instead of δ 4.28 (as with carotenoid 5 and carotenoid 6) in the 1H -NMR spectrum of carotenoid 7 also indicated the absence of a 2-hydroxyl group. The novel methyl-carotenoids carotenoid 6 and carotenoid 7 are therefore identified as 2'-methoxyflexixanthin and 3-deoxy-2'-methoxyflexixanthin, respectively (Table 5.2).

Previously determined UV-Vis spectra and molecular formulae inferred from mass spectrometry (Chapter 4, Klassen and Foght 2008) suggested the presence of two non-ketolated carotenoids in strain VUG-A42aa, one of which (carotenoid 9; molecular formula: $C_{41}H_{58}O_2$; Chapter 4, Klassen and Foght 2008) was presumed to be a methyl derivative of the other (carotenoid 8; molecular formula: $C_{40}H_{56}O_2$; Chapter 4, Klassen and Foght 2008). Carotenoid 9 was present only in trace amounts and was not analyzed further. Unfortunately, replicate extracts of carotenoid 8 yielded a 1H -NMR spectrum consistently contaminated by an unknown glycosyl moiety (Supplemental Figures C9 and C10). Clearly present in this spectrum, however, were H-2', H-3' and H-4' signals similar to those determined for carotenoid 5 and TOCSY cross-correlated peaks at δ 1.31, δ 1.59 and δ 2.30 (annotated as H-4, H-3 and H-2, respectively). The similarity of these signals to those determined by others (Madhour et al. 2005) and the lack of a 2'-hydroxyflexixanthin-like H-3 peak suggest that carotenoid 8 is

most likely plectanixanthin; carotenoid 9 is most likely 2'-methylplectanixanthin by analogy to carotenoids 6 and 7 (Table 5.2).

The structural similarity of the carotenoids identified in this study suggests a metabolic relationship between them, as indicated in (Figure 5.1). The accumulation of carotenoid 8 in relatively small amounts in strain VUG-A42aa (Chapter 4, Klassen and Foght 2008) suggests that it is a biosynthetic precursor for carotenoid 5, which is likely in turn a precursor of carotenoid 6 in VUG-A42aa and of carotenoids 3 and 4 in VUG-A141a. The reactions leading to carotenoids similar to carotenoid 8 have been at least partially elucidated in other Bacteroidetes (Teramoto et al. 2003, Teramoto et al. 2004, Tao et al. 2006), Cyanobacteria (Takaichi and Mochimaru 2007, Maresca et al. 2008) and *Deinococcus radiodurans* (Tian et al. 2007, Zhang et al. 2007, Tian et al. 2008). Carotenoid 7 is likely a methylation product of the presumed intermediate between carotenoids 8 and 5, 3-deoxy-2'-hydroxyflexixanthin. It is most parsimonious to suppose the presence of a single methyltransferase having broad specificity leading to carotenoids 6, 7 and 9, although my data cannot preclude the existence of multiple dedicated enzymes. Further determination of this pathway will be assisted by the genome sequence for the 2'-hydroxyflexixanthin-producing *Hymenobacter roseosalivarius*, undergoing sequencing at the Joint Genome Institute (<http://www.jgi.doe.gov/sequencing/statusreporter/psr.php?projectid=98064>, accessed February 2009).

Carotenoid production is widespread in the *Flavobacteria* and *Sphingobacteria* (which includes the genus *Hymenobacter*), with carotenoid biosynthetic gene homologs present in all currently available genome sequences excepting the insect symbiont *Candidatus* 'Sulcia muelleri' (Integrated Microbial Genome database version 2.7 [last updated December 2008]; <http://img.jgi.doe.gov/cgi-bin/pub/main.cgi>). Many, but not all, *Flavobacteria* and *Sphingobacteria* produce highly oxygenated monocyclic carotenoids (Yokoyama and Miki 1995, Lutnaes et al. 2002, Lutnaes et al. 2004, Tao et al.

2006, Shindo et al. 2007), some of which possess antioxidant activity surpassing that of oxygenated bicyclic carotenoids such as zeaxanthin (Shindo et al. 2007). Although the exact reasons for the current extent of flavobacterial and sphingobacterial structural diversity remains unknown, possibilities may include fine-scale tuning of electrochemical activity or membrane fluidity. The presence of these potent antioxidants in *Hymenobacter* very likely facilitates their survival in the high-radiation environments in which they have most commonly been detected.

5.4. Chemical characteristics of *Hymenobacter* carotenoids

2'-S-Hydroxyflexixanthin ((2'S)3,1',2'-trihydroxy- β , ψ -caroten-4-one)

(Carotenoid 5): Major carotenoid of *Hymenobacter* str. VUG-A141a; UV-Vis (MeOH): λ_{\max} 452 (sh), 479, 508 nm (%III/II = 0), see Supplemental Figure C1; UV-Vis (CHCl₃): λ_{\max} 464 (sh), 492, 520 (sh) (%III/II = 0); CD: see Figure 5.2; ¹H-NMR (CD₂Cl₂, 800MHz): see Table 5.2 and Supplemental Figures C2 and C3. High resolution MALDI (sodiated from Chapter 4, Klassen and Foght 2008): *m/z* 621.39161 (C₄₀H₅₄O₄Na).

1'-Xylosyl-2'-hydroxyflexixanthin (1'- β -D-xylosyl-3,2'-dihydroxy- β , ψ -caroten-

4-one) (Carotenoid 4): Major carotenoid of *Hymenobacter* str. VUG-A141a; UV-Vis (MeOH): λ_{\max} 452 (sh), 479, 502.5 nm (%III/II = 5.1), see Supplemental Figure C1; UV-Vis (CHCl₃): λ_{\max} 464 (sh), 494, 522 (sh) (%III/II = 0); ¹H-NMR (CD₂Cl₂, 800MHz): see Table 5.2 and Supplemental Figures C4 and C5. High resolution MALDI (from Chapter 4, Klassen and Foght 2008): *m/z* 730.44392 (C₄₅H₆₂O₈).

Hexosyl-2'-hydroxyflexixanthin (hexosyl-3,1',2'-trihydroxy- β , ψ -caroten-4-

one) (Carotenoid 3): Minor carotenoid of *Hymenobacter* str. VUG-A141a, see Chapter 4, Klassen and Foght (2008) for characterization.

2'-Methoxyflexixanthin (2'-methoxy-3,1'-dihydroxy- β , ψ -caroten-4-one)

(Carotenoid 6): Major carotenoid of *Hymenobacter* str. VUG-A42aa; UV-Vis (MeOH): λ_{\max} 453 (sh), 478, 504 nm (%III/II = 9.1), see Supplemental Figure C1;

UV-Vis (CHCl₃): λ_{\max} 464 (sh), 492.5, 522 (sh) (%III/II = 0); ¹H-NMR (CD₂Cl₂, 800MHz): see Table 5.2 and Supplemental Figure C6. The 1D-¹H-NMR spectrum was unfortunately overwritten during data analysis and is therefore not shown. High resolution MALDI (sodiated; from Chapter 4, Klassen and Foght 2008): m/z 635.40708 (C₄₁H₅₆O₄Na).

3-Deoxy-2'-methoxyflexixanthin (2'-methoxy-1'-hydroxy- β,ψ -caroten-4-one) (Carotenoid 7): Major carotenoid of *Hymenobacter* str. VUG-A42aa; UV-Vis (MeOH): λ_{\max} 452 (sh), 479, 508 nm (%III/II = 0), see Supplemental Figure C1; UV-Vis (CHCl₃): λ_{\max} 464 (sh), 492, 520 (sh) (%III/II = 0); ¹H-NMR (CD₂Cl₂, 800MHz): see Table 5.2 and Supplemental Figures C7 and C8. High resolution MALDI (from Chapter 4, Klassen and Foght 2008): m/z 596.42240 (C₄₁H₅₆O₃).

Plectanixanthin (1',2'-dihydroxyl- β,ψ -carotene) (Carotenoid 8): Minor carotenoid of *Hymenobacter* str. VUG-A42aa, see Chapter 4, Klassen and Foght (2008) for chemical characterization. Isolated as an impure preparation. ¹H-NMR (CD₂Cl₂, 800MHz): see Supplemental Figures C9 and C10 and highlighted signals in the text.

2'-Methylplectanixanthin (2'-methoxy-1'-hydroxy- β,ψ -carotene) (Carotenoid 9): Minor carotenoid of *Hymenobacter* str. VUG-A42aa, see Chapter 4, Klassen and Foght (2008) for characterization.

5.5. Literature cited

- Andrewes, A.G., Foss, P., Borch, G., and Liaaen-Jensen, S. 1984. Bacterial carotenoids. 50. On the structures of (3*S*)-flexixanthin and (3*S*,2'*S*)-2'-hydroxyflexixanthin. *Acta Chem. Scand.*, 38: 337-339.
- Bartlett, L., Klyne, W., Mose, W.P., Scopes, P.M., Galasko, G., Mallams, A.K., Weedon, B.C.L., Szabolcs, J., and Tóth, G. 1969. Optical rotatory dispersion of carotenoids. *J. Chem. Soc. C*: 2527-2544.
- Bax, A., and Davis, D.G. 1985. MLEV-17-based two-dimensional homonuclear magnetization transfer spectroscopy. *J. Magn. Reson.*, 65: 355-360.
- Bircher, C., and Pfander, H. 1997. Synthesis of carotenoids in the gliding bacteria *Taxeobacter*: (all-*E*,2'*R*)-3-deoxy-2'-hydroxyflexixanthin. *Helv. Chim. Acta*, 80: 832-837.

- Britton, G., Liaaen-Jensen, S., and Pfander, H. 2004. Carotenoids handbook Birkhäuser Verlag, Basal, Switzerland.
- Fraser, P.D., and Bramley, P.M. 2004. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.*, 43: 228-265.
- Fredrickson, J.K., Li, S.-m.W., Gaidamakova, E.K., Matrosova, V.Y., Zhai, M., Sulloway, H.M., Scholten, J.C., Brown, M.G., Balkwill, D.L., and Daly, M.J. 2008. Protein oxidation: key to bacterial desiccation resistance? *ISME J.*, 2: 393-403.
- Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.
- Kupce, E., and Freeman, R. 1993. Techniques for multisite excitation. *J. Magn. Reson. A*, 105: 234-238.
- Levitt, M.H., Freeman, R., and Frenkiel, T. 1982. Broadband heteronuclear decoupling. *J. Magn. Reson.*, 47: 328-330.
- Lutnaes, B.F., Oren, A., and Liaaen-Jensen, S. 2002. New C₄₀-carotenoid acyl glycoside as principal carotenoid in *Salinibacter ruber*, an extremely halophilic eubacterium. *J. Nat. Prod.*, 65: 1340-1343.
- Lutnaes, B.F., Strand, A., Pétursdóttir, S.K., and Liaaen-Jensen, S. 2004. Carotenoids of thermophilic bacteria - *Rhodothermus marinus* from submarine Icelandic hot springs. *Biochem. Syst. Ecol.*, 32: 455-468.
- Madhour, A., Anke, H., Mucci, A., Davoli, P., and Weber, R.W.S. 2005. Biosynthesis of the xanthophyll plectanixanthin as a stress response in the red yeast *Dioszegia* (Tremellales, Heterobasidiomycetes, Fungi). *Phytochemistry*, 66: 2617-2626.
- Maresca, J.A., Graham, J.E., and Bryant, D.A. 2008. The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. *Photosynth. Res.*, 97: 121-140.
- Mortensen, A. 2006. Carotenoids and other pigments as natural colorants. *Pure Appl. Chem.*, 78: 1477-1491.
- Rainey, F.A., Ray, K., Ferreira, M., Gatz, B.Z., Nobre, M.F., Bagaley, D., Rash, B.A., Park, M.-J., Earl, A.M., Shank, N.C., Small, A.M., Henk, M.C., Battista, J.R., Kämpfer, P., and da Costa, M.S. 2005. Extensive diversity of ionizing-radiation-resistant bacteria recovered from Sonoran Desert soil and description of nine new species of the genus *Deinococcus* obtained from a single soil sample. *Appl. Environ. Microbiol.*, 71: 5225-5235.
- Rønneberg, H., Andrewes, A.G., Borch, G., Berger, R., and Liaaen-Jensen, S. 1985. CD correlation of C-2' substituted monocyclic carotenoids. *Phytochemistry*, 24: 309-319.
- Sandmann, G., Albrecht, M., Schnurr, G., Knörzer, O., and Böger, P. 1999. The

- biotechnological potential and design of novel carotenoids by gene combination in *Escherichia coli*. Trends Biotechnol., 17: 233-237.
- Saul, D.J., Aislabie, J.M., Brown, C.E., Harris, L., and Foght, J.M. 2005. Hydrocarbon contamination changes the bacterial diversity of soil from around Scott Base, Antarctica. FEMS Microbiol. Ecol., 53: 141-155.
- Shindo, K., Kikuta, K., Suzuki, A., Katsuta, A., Kasai, H., Yasumoto-Hirose, M., Matsuo, Y., Misawa, N., and Takaichi, S. 2007. Rare carotenoids, (3*R*)-saproxanthin and (3*R*,2'*S*)-myxol, isolated from novel marine bacteria (*Flavobacteriaceae*) and their antioxidative activities. Appl. Microbiol. Biotechnol., 74: 1350-1357.
- Shindo, K., Asagi, E., Sano, A., Hotta, E., Minemura, N., Mikami, K., Tamesada, E., Misawa, N., and Moaka, T. 2008. Diapolycopene acid xylosyl esters A, B, and C, novel antioxidative glyco-C₃₀-carotenoic acids produced by a new marine bacterium *Rubritalea squalenifaciens*. J. Antibiot., 61: 185-191.
- Takaichi, S., and Mochimaru, M. 2007. Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. Cell. Mol. Life Sci., 64: 2607-2619.
- Tao, L., Yao, H., Kasai, H., Misawa, N., and Cheng, Q. 2006. A carotenoid synthesis gene cluster from *Algoriphagus* sp. KK10202C with a novel fusion-type lycopene β -cyclase gene. Mol. Genet. Genomics, 276: 79-86.
- Teramoto, M., Rahlert, N., Misawa, N., and Sandmann, G. 2004. 1-Hydroxy monocyclic carotenoid 3,4-dehydrogenase from a marine bacterium that produces myxol. FEBS Lett., 570: 184-188.
- Teramoto, M., Takaichi, S., Inomata, Y., Ikenaga, H., and Misawa, N. 2003. Structural and functional analysis of a lycopene β -monocyclase gene isolated from a unique marine bacterium that produces myxol. FEBS Lett., 545: 120-126.
- Tian, B., Xu, Z., Sun, Z., Lin, J., and Hua, Y. 2007. Evaluation of the antioxidant effects of carotenoids from *Deinococcus radiodurans* through targeted mutagenesis, chemiluminescence, and DNA damage analyses. Biochim. Biophys. Acta, 1770: 902-911.
- Tian, B., Sun, Z., Xu, Z., Shen, S., Wang, H., and Hua, Y. 2008. Carotenoid 3',4'-desaturase is involved in carotenoid biosynthesis in the radioresistant bacterium *Deinococcus radiodurans*. Microbiology, 154: 3697-3706.
- Yokoyama, A., and Miki, W. 1995. Isolation of myxol from a marine bacterium *Flavobacterium* sp. associated with a marine sponge. Fish. Sci., 61: 684-686.
- Zhang, L., Yang, Q., Luo, X., Fang, C., Zhang, Q., and Tang, Y. 2007. Knockout of *crtB* or *crtI* gene blocks the carotenoid biosynthetic pathway in *Deinococcus radiodurans* R1 and influences its resistance to oxidative

DNA-damaging agents due to change of free radicals scavenging ability.
Arch. Microbiol., 188: 411-419.

6. Microbial Carotenoid Diversity and Lineage-Specific Evolutionary Mechanisms are Revealed by Comparative Genomics: Implications for the Rational Study of Pathway Diversity

6.1. Introduction

6.1.1. General introduction and study rationale

Carotenoids comprise a large secondary metabolite family of over 600 isoprenoid compounds produced by most plants and many microorganisms (Britton et al. 2004). Depending on the length of their conjugated double bond chain and the nature of their substituents, carotenoids most often absorb light in the 300-600 nm range, appearing yellow, orange or red (Britton 1995). This extended polyene chain also promotes extensive electrochemical activity (Britton 1995). Carotenoid function is perhaps best understood in the photosynthetic light-harvesting complex, where carotenoids dissipate excess energy and radicals from excited oxygen and (bacterio)chlorophyll molecules, physically structure the photosynthetic reaction center and act as accessory light-harvesting pigments (Frank and Cogdell 1996, Fraser et al. 2001, Frank and Brudvig 2004). Furthermore, in all organisms carotenoids may function as antioxidants and promote oxidative stress resistance (Tian et al. 2007, Zhang et al. 2007), even acting as a virulence factor in *Staphylococcus aureus* by promoting resistance to neutrophil oxidative burst (Liu et al. 2005). Carotenoids are also precursors for many apocarotenoids (i.e., cleaved carotenoids) such as retinal, the cofactor for the photoactive rhodopsin protein found in many microorganisms (Sharma et al. 2006, Fuhrman et al. 2008) and functionally similar light-sensing proteins in vertebrates, which use as a cofactor the apocarotenoid retinol (Vitamin A; Spudich et al. 2000). At least one rhodopsin (xanthorhodopsin) also interacts directly with antennae carotenoids (Lanyi and Balashov 2008). Other apocarotenoids include plant hormones, fungal pheromones and antifungal compounds (Auldridge et al. 2006). In all organisms, membrane fluidity and proton permeability may also be modulated by carotenoids, depending on

carotenoid structure and concentration (Gruszecki and Strzalka 2005, Kupisz et al. 2008); these latter functions remain poorly studied, especially *in vivo*.

Carotenoids are biotechnologically high-value compounds with an annual market estimated to exceed one billion US dollars by 2010 (cited by Del Campo et al. 2007). Applications include natural pigments (Mortensen 2006) and nutraceuticals based on the potential of carotenoids to decrease the risk of several human diseases (Fraser and Bramley 2004, Krinsky and Johnson 2005, Rao and Rao 2007). This biotechnological interest has prompted extensive research into both natural (Del Campo et al. 2007) and recombinant carotenoid production, particularly in microbes (Das et al. 2007). As part of the latter approach, carotenoids are a model system (Umeno et al. 2005) to study recombinant biosynthetic pathway engineering (Schmidt-Dannert 2000, Sandmann 2002b, Wang et al. 2007), by which novel compounds are produced by combining genes from multiple organisms in a heterologous host. This approach has resulted in novel carotenoids with enhanced biotechnologically relevant properties such as antioxidative activity (Albrecht et al. 2000, Nishida et al. 2005). Despite underlying pathway engineering initiatives, however, microbial carotenoid biosynthetic and structural diversity and distribution have been significantly underestimated due to utilization of methods lacking either taxonomic breadth or structural resolution (Klassen and Foght 2008). In this study microbial carotenoid diversity and distribution are estimated using comparative genomics, bypassing the labor-intensive requirement of detailed phenotypic study across taxonomically diverse organisms.

The goal of this study was to determine the diversity of microbial carotenoid biosynthetic genes using comparative genomics for comparison with known structural and biosynthetic diversity. Unlike several recent studies and annotation approaches that rely almost exclusively on pair-wise BLAST similarity to single seed sequences, this work conservatively identifies carotenoid biosynthetic homologs in genomic databases based on phylogenetic congruence with biochemically- or genetically-confirmed seed sequences and logical

reconstruction of metabolic pathways. In this way, a high false-positive homolog detection rate is avoided, contrasting with poorly-constrained studies attempting to detect all possible homologs based upon homology to a few functionally confirmed sequences (as discussed for microbial rhodopsins; Fuhrman et al. 2008). Furthermore, the inherent parsability of the recovered sequence data facilitates pathway-level evolutionary analyses to determine the factors underlying the generation of carotenoid diversity and distribution. Evaluation of diversity through this lens allows projection of the potential for functional novelty across phylogenetic space. In this way, the most promising avenues for further study and biotechnological exploitation are suggested based on homology, or lack thereof, to high-confidence data and the nature of lineage-specific diversity-generating evolutionary mechanisms.

6.1.2. Known microbial carotenoid biosynthetic pathways

Biochemically or genetically demonstrated carotenoid biosynthetic protein types and their corresponding protein and nucleotide sequences were determined from previous literature reports (Table 6.1 and Supplemental Table D1; Sieiro et al. 2003, Cheng 2006, Takaichi and Mochimaru 2007, Maresca et al. 2008a, Tanaka et al. 2008). For reviews of carotenoid chemistry and nomenclature see Britton (Britton 1995, Britton et al. 2004). Carbon numbers mentioned in the text are indicated for lycopene and β -carotene in Figure 6.1. From this review, a map of known carotenoid biosynthetic pathways (Figure 6.1) was constructed, summarized below, which formed a scaffold for homolog annotation. Due to the paucity of data, apocarotenoid biosynthesis (except those leading to neurosporoxanthin) is not considered here, and for simplicity only bacterial enzyme nomenclature is used; see Table 6.1 for synonyms.

Carotenoid biosynthesis has been described as “tree-like”

Figure 6.1. (next page) Known carotenoid biosynthetic pathways. For simplicity, most pathway intermediates are not shown. Apocarotenoids, except neurosporoxanthin, have been excluded. Non-homologous, functionally equivalent enzymes are indicated by a slash; for alternative names of homologous sequences see Table 6.1. Carbon numbers referred to in the text are indicated for lycopene and β -carotene.

Table 6.1. Known microbial carotenoid biosynthetic proteins used for *in silico* carotenoid biosynthetic pathway reconstruction, their synonyms and biochemical functions.

Protein Name	Synonyms ^a	E.C. Number ^b	Biochemical Function
Ald	-	-	4,4'-Diapolycopene-4,4'-dial oxidase
CAO-2	-	-	3',4'-Didehydrolycopene or Torulene (3',4'-didehydro- γ -carotene) 3',4'-oxidase
CHYB	-	1.14.13.-	β -Carotene 3,(3')-(di)hydroxylase
CHYE	-	1.14.13.-	ϵ -Carotene 3,(3')-(di)hydroxylase
CrtA	-	1.-.-.-	1'-Methoxy linear xanthophyll 2'-ketolase or -hydroxylase
CrtB	PSY, CarA ^c , CarP ^c , AL-2	2.5.1.32	Phytoene synthase
CrtC	-	-	Lycopene 1-hydroxylase, γ -carotene 1'-hydroxylase
CrtD	-	1.14.99.-	1,2-Dihydrolycopene-3,4-desaturase, 1',2'-dihydro- γ -carotene-3',4'-desaturase
CrtEb	LitC ^d	-	Lycopene prenyl transferase
CrtF	-	2.1.1.-	Linear xanthophyll methyltransferase
CrtG	-	-	β -Carotene 2,(2')-(di)hydroxylase
CrtH	CRTISO	5.-.-.-	7,9,7',9'- <i>cis</i> -Lycopene isomerase
CrtI	CarB, AL-1	1.14.99.-	Phytoene desaturase
CrtL	CrtLb, LYCB	1.14.-.-	Lycopene β -mono- or β -bicyclase
CrtLe	LYCE	1.14.-.-	Lycopene ϵ -bicyclase
CrtM	-	2.5.1.-	4,4'-Diapophytoene synthase
CrtN	-	1.14.99.-	4,4'-Diapophytoene desaturase
CrtNb	CrtP	1.-.-.-	4,4'-Diaponeurosporene or 4,4'-Diapolycopene oxidase
CrtO	-	-	β -Carotene 4,(4')-(di)ketolase
"CrtOat"	-	2.3.1.-	4,4'-Diaponeurosporen-4-oic acid glycosyl transferase
CrtP	PDS	1.14.99.-	Phytoene desaturase
CrtQ	ZDS	1.14.99.30	ζ -Carotene desaturase
"CrtQgt"	-	2.4.1	4,4'-Diaponeurosporen-4-oic acid glycoside acyl transferase
CrtR	-	1.14.13.-	β -Carotene 3,(3')-(di)hydroxylase
CrtU	-	-	β -Carotene ϕ -desaturase
CrtW	BKT	-	β -Carotene 4,(4')-(di)ketolase
CrtX	-	-	Zeaxanthin glycosyl transferase
CrtY	CrtYm	1.14.-.-	Lycopene β -mono- or β -bicyclase
CrtYcd ^e	CarR ^c	-	Lycopene β -mono- or β -bicyclase
CrtYef	-	-	Flavuxanthin (acyclic C ₅₀ carotenoid) ϵ -bicyclase
CrtZ	-	1.14.13.-	β -Carotene 3,(3')-(di)hydroxylase
CruA	-	-	Lycopene β -bicyclase
CruB	-	-	γ -Carotene β -monocyclase
CruC	-	-	1'-Hydroxylchlorobactene glycosyl transferase
CruD	-	-	1'-Hydroxylchlorobactene glycoside acyl transferase
CruE	-	-	β -Carotene χ -desaturase
CruF	-	-	γ -Carotene 1'-hydroxylase
CruG	-	-	Myxol or ketomyxol glycotransferase
CruH	-	-	Renierapurpurin (χ,χ -carotene) dicarboxylase
CruP	-	-	Lycopene β -monocyclase
CYP175A1	-	-	β -Carotene 3,(3')-(di)hydroxylase
LitAB ^d	-	-	Flavuxanthin (acyclic C ₅₀ carotenoid) β -bicyclase
ORF10	-	-	Isorenieratene (ϕ,ϕ -carotene) 3,(3')-(di)hydroxylase
VDE	-	1.10.99.3	Violaxanthin (diepoxyzeaxanthin) de-epoxidase
YLO-1	-	-	Apo-4'-lycopenal or β -apo-4'-carotenal oxidase
ZEP	-	1.14.13.90	Zeaxanthin epoxidase

^aSynonyms beginning with “AL” refer to *Neurospora crassa*; synonyms beginning with “Car” refer to most other fungi; CrtP in *Staphylococcus aureus* is synonymous with CrtNb; all synonyms in block capital letters refer to photosynthetic eukaryotes

^bE. C. numbers (where available) were obtained from [Kyoto Encyclopaedia of Genes and Genomes](#) pathway map00906, last updated February 24, 2009

^cCarAP and CarPR exist as heterodimers

^dLitBC exists in *Dietzia* sp. CQ4 as a heterodimer, although separated homologs exist in other organisms

^eIn some, but not all, organisms CrtYcd exists as a heterodimer

(Umeno et al. 2005), whereby all pathways share a common conserved “trunk” from which branch a diversity of sub-pathways. In nearly all carotenogenic organisms, phytoene is produced by the phytoene synthase CrtB and is subsequently desaturated by the phytoene desaturase CrtI to form lycopene or, more rarely, neurosporene. In the Cyanobacteria, Chlorobi and photosynthetic eukaryotes the latter step is accomplished by three separate enzymes: the desaturases CrtP and CrtQ and the isomerase CrtH. An analogous trunk is present in C₃₀ carotenoid biosynthesis, where 4,4'-diapophytoene is produced by the 4,4'-diapophytoene synthase CrtM and is desaturated by the 4,4'-diapophytoene desaturase CrtN. From this point, a myriad of biosynthetic branches occur. β- and ε-cyclic end groups can be formed at one or both ends of a linear carotenoid and further desaturated to form aromatic φ- and χ- end groups. Each Greek letter indicates a different end group arrangement; β- and ε and φ- and χ-end groups differ from each other by the location of their double bonds and methyl groups, respectively (Britton 1995, Britton et al. 2004). In C₅₀ carotenoid biosynthesis, cyclization may occur following prenylation at the linear ψ-end group 2- position. Xanthophylls (i.e., oxygenated carotenoids) can be formed by many different transformations, most often 4- ketolation and 2- and 3-hydroxylation of β- and, in the latter case, ε-end groups, or 2-ketolation and 1- and 2-hydroxylation of φ-end groups. Formation of aldehydes and carboxylic acids is also possible, along with a host of less-studied modifications including glycosylation and acylation. Compounding this diversity is the potential for asymmetry between end groups. Different branches of the metabolic “tree” (Figure 6.1) are formed by the

differential presence of carotenoid biosynthetic enzymes. Inference of the carotenoids most likely produced by a phylogenetic lineage is therefore possible by identifying the phylogenetic distribution of their cognate biosynthetic enzyme homologs. These inferences underpin the analyses of microbial carotenoid diversity and lineage-specific evolution described below.

6.2. Materials and methods

6.2.1. Dataset construction

Carotenoid biosynthetic enzymes with known function were identified from the literature (see Supplemental Table D1) and their corresponding amino acid sequences retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/>). Biosynthetic function of an enzyme was considered demonstrated if (in order of confidence): (i) it had been confirmed by *in vitro* biochemical studies; (ii) its recombinant expression in a non-carotenogenic host resulted in an appropriate anabolic reaction or (iii) *in vivo* mutation of its cognate gene resulted in a loss of function. In the later case, functional assignments were subsequently confirmed by phylogenetic placement of these sequences with related homologs of known function, due to the possibility of polar mutations eliciting misleading phenotypes. In a few cases, amino acid sequences for proteins of confirmed function were unidentifiable due to missing GenBank accession numbers or genomic gene identifiers; because alternative close homologs were available these sequences were omitted from the initial seed database.

Initial, non-bootstrapped phylogenetic trees for each protein type in the initial seed database were constructed, and representatives from each resulting phylogenetic cluster were used to iteratively search the Integrated Microbial Genome (IMG) database version 2.4 (Markowitz et al. 2007), last updated December 2007, using BLASTp (Altschul et al. 1990). For each protein type, all BLAST hits with an expectation value $< 1 \times 10^{-20}$ were exported along with their corresponding nucleotide sequences. In several instances paralogous proteins were recovered from the same organism. To eliminate obviously spurious and paralogous sequences, non-bootstrapped phylogenetic analyses were conducted to

determine to which, if any, carotenoid biosynthesis enzyme family the recovered sequences belonged. Sequences were binned based primarily upon phylogenetic clustering with those of demonstrated functions from the initial seed database, either in obvious clades or adjacent to them in accordance with the taxonomy of their originating organisms. Genes were also binned based upon the construction of logical carotenoid biosynthetic pathways, based upon both currently described carotenoid biosynthetic pathways and known chemical structures (Figure 6.1, Supplemental Table D1). In all cases sequence assignments were made conservatively, i.e., sequences were removed if there was no clear reason for their inclusion, favoring a lower rate of false-positive assignment at the expense of a higher false-negative assignment rate.

Because the IMG database is updated only intermittently, representative sequences for each protein type retrieved from the IMG database were used as inputs for PSI-BLAST (Altschul et al. 1997) searches against the NCBI reference protein sequence database. Non-genome derived sequences present in the NCBI non-redundant database were excluded because their organismal identities typically lacked corroborating evidence. Three PSI-BLAST iterations were conducted with an expectation value threshold set such that all previously identified sequences were recovered. Sequences obtained by this approach were compared to those from the IMG and initial seed databases using non-bootstrapped phylogenetic trees. Those unique to the NCBI database and which clustered internal to previously recovered IMG and seed sequences were retained. In cases where a particular sequence was absent from the biosynthetic pathway inferred from our sorted sequence database, the corresponding genome was specifically queried for that homolog using BLAST. Where multiple closely related strains (i.e., nearly 100% protein sequence identity for all protein types) were recovered, only one sequence was retained as a representative (Supplemental Table D1). Although in most cases seed sequences (i.e., those recovered from the literature) were used in preference to genomic data, occasionally a genome-sequenced strain was chosen as representative due to the greater number of putative carotenoid biosynthesis enzyme sequences present

(Supplemental Table D1). Because the IMG and NCBI databases contained few algal genomes, the genome database sites for *Cyanidioschyzon merolae* (<http://merolae.biol.s.u-tokyo.ac.jp/>), *Galdieria sulphuraria* (<http://genomics.msu.edu/galdieria/>), *Phaeodactylum tricornutum* (<http://genome.jgi-sf.org/Phatr2/Phatr2.home.html/>) and *Thalassiosira pseudonana* (<http://genomeportal.jgi-psf.org/Thaps3/Thaps3.home.html/>) were individually BLAST searched using previously identified algal and cyanobacterial sequences. The genome sequence of *Heliobacterium modesticaldum* Ice1 was also individually queried using C30 carotenoid biosynthesis protein sequences after its inclusion in the NCBI database.

In addition to whole-genome sequence data, carotenoid biosynthetic protein sequences from uncultured microorganisms present on large-insert fosmid clones from oceanic surface waters of Monterey Bay and the North Pacific Subtropical Gyre (McCarren and DeLong 2007) were included to better represent natural proteorhodopsin diversity. Only fosmids containing a putative full carotenoid biosynthetic pathway leading to rhodopsin and a clear phylogenetic identity were included in the dataset to best facilitate pathway reconstruction, limiting usage of more abundant but fragmented metagenomic data. The presence of rhodopsin genes in the analyzed genome sequences was determined by searching the GenBank refseq database using three sequential PSI-BLAST iterations with a 1×10^{-5} expectation value cut-off. Searches were conducted using rhodopsins from *Halobacterium salinarium*, *Nostoc* sp. PCC 7120 and *Pelagibacter ubique* HTCC1062 (GenBank accession numbers 0501217A, NP_487205 and AAZ21446, respectively) as seed sequences. Sequences below this threshold were compared phylogenetically without bootstrapping to exclude sequences outlying those with previously demonstrated function, those from the included metagenomic study (McCarren and DeLong 2007) or organisms lacking appropriate carotenoid biosynthetic enzyme homologs.

To phylogenetically scaffold carotenoid biosynthetic pathways derived from genome sequences and the literature, 16S rRNA gene sequences were obtained using either BLAST searches against each individual genome or directly

from the NCBI database. The 16S rRNA gene was chosen primarily because it is most routinely used for organism identification and many partial sequences, therefore, were available for organisms for which complete genome sequences were unavailable.

6.2.2. Phylogenetic methods

All sequences were aligned using CLUSTALW v.2.0.5 (Thompson et al. 1994) or CLUSTALX v.1.83 (Thompson et al. 1997). Alignments were examined visually and obviously aberrant sequences (e.g., those from incomplete draft genome sequences) were omitted. Extreme 5' and 3' sequence ends, which were often of uneven length and poorly aligned, were excluded, as were indels present in only one sequence. Other lineage-specific indels were included to maximize the phylogenetic signal for intra-clade phylogenies, even at the expense of resolution at deeper nodes. All conclusions discussed in the text are supported by separate analyses using reduced datasets in which all indels were removed (data not shown). Heterodimeric sequences, where present, were trimmed such that only a single domain was included (Supplemental Table D2). When occurring separately, CrtYcd and CrtYef heterodimer sequences were fused to match their monomeric homologs and to maximize the phylogenetic signal.

Phylogenetic analyses were conducted primarily using RAxML v.7.0.4 (Stamatakis 2006, Stamatakis et al. 2008) as implemented through the CIPRES web portal (<http://www.phylo.org/>). In all cases the Jones-Taylor-Thornton (JTT) substitution matrix was used, the proportion of invariant sites estimated automatically and the best scoring tree used for visualization. Preliminary RAxML experiments using other substitution matrices (BLOSUM62, DAYHOFF and WAG) gave equivalent results, albeit with slightly lower median bootstrap values (data not shown). Nucleotide trees were also created using RAxML according to the default parameters, again using the best tree and estimating the proportion of invariant sites. Further experiments using parsimony (PROTPARS, one jumble per replicate) and distance (PROTDIST, Dayhoff PAM matrix and NEIGHBOR, neighbor joining method) tree construction methods implemented in

PHYLIP v.3.66, 3.67 or 3.68 (Felsenstein 1989) also yielded congruent results. Between methods, nodes were often non-equivalent due to differential placement of poorly-supported and often deep-branching sequences between methods; parsimony and distance results are therefore not shown for simplicity but are available upon request. Most trees were rooted only to their midpoint using RETREE (PHYLIP). In preliminary experiments, trees rooted using basal-branching outgroup sequences were consistently rooted within the same clade in multiple analyses, but with an unclear intra-clade rooting pattern (data not shown). In these experiments, outgroup sequences were selected from a neighboring COG family showing homology over the entire sequence length, as determined using the NCIB Conserved Domain Database (Marchler-Bauer et al. 2007). Where one COG family included several carotenoid biosynthesis enzyme sequences, the same outgroup sequence was used for consistency. As this work focused on intra-clade relationships, midpoint-rooted trees were used to avoid the intra-clade phylogenetic distortions caused by uncertainly placed roots; relevant observations from rooted trees are indicated in the text.

6.2.3. Statistical methods

Synonymous and non-synonymous substitution rates were calculated separately using the Nei-Gojobori method with the Jukes-Cantor correction for same-site mutations, as implemented in MEGA v.4.0 (Tamura et al. 2007) and the d_n/d_s calculated in EXCEL for all pair-wise comparisons with $d_s < 1.5$ (to account for mutational saturation) and $d_n > 0.01$ (to ensure a sufficient number of informative substitutions), similar to cutoffs used elsewhere (Novichkov et al. 2009). For this analysis nucleotide sequences were aligned as translated amino acid sequences to conserve codon groupings using MEGA. Two-tailed P values were calculated in SPSS v14.0 using the Mann-Whitney U test by comparing all elevated d_n/d_s pair-wise comparisons for a particular carotenoid biosynthetic gene type and phylogenetic lineage (bolded in Supplemental Figure D10) to those not elevated, excluding values generated by pair-wise comparison of two sequences with elevated d_n/d_s ratios. To identify putative recombination events, third codon-

position, ungapped nucleotide sequence alignments from each cluster were created using MEGA, and maximum-likelihood trees were created using the HKY+gamma substitution matrix implemented in PAUP* v.4.0 (Sinauer Associates, Inc. Publishers, Sunderland Massachusetts). Evolutionary rate heterogeneity (Worobey 2001) was determined using 1000 bootstrap replications for each tree using PIST v.1.0 (<http://evolve.zoo.ox.ac.uksoftware.html?id=PIST/>).

6.3. Results

6.3.1. Distribution of microbial carotenoid biosynthetic pathways

The carotenoid biosynthetic enzyme sequence database created in this study (Supplemental Table D1) was created using two approaches: BLASTp searching of the IMG database using multiple, diverse, seed sequences of known function; and iterative PSI-BLAST searching of the GenBank refseq database using single seed sequences, both followed by phylogenetic analyses and *in silico* pathway construction. Importantly, this dataset was constructed conservatively; i.e., homologs, even with high BLAST expectation scores, were excluded if they were phylogenetically distant from sequences with corroborative biochemical or genetic evidence, as were sequences which did not form a logical biosynthetic pathway. The distribution and evolution of each major pathway type will be discussed separately below.

Carotenoid biosynthesis is widely distributed in bacteria and archaea (Figure 6.2), occurring in 29% of the genomes present in the IMG database v.2.4 comprising 12 of 20 phyla. Due to the few eukaryotic strains present in the IMG database at the time of writing, fungi and algae were not included in this analysis. Note that several apparently non-carotenogenic phyla are represented by very few genome sequences and that some of these phyla are known to contain carotenogenic members (e.g., Verrucomicrobia and photosynthetic Acidobacteria; Bryant et al. 2007, Shindo et al. 2008). As expected, phyla containing photosynthetic members, including α -, β -, and γ -Proteobacteria, Cyanobacteria, Chloroflexi, Chlorobi and the Heliobacteria (Firmicutes) all possess carotenoid

biosynthetic protein homolog-containing strains. Many non-photosynthetic Proteobacteria are also carotenogenic. Surprisingly, several non-photosynthetic classes and phyla, notably Flavobacteria, Sphingobacteria, Actinobacteria and *Deinococcus-Thermus* were predominantly composed of carotenogenic strains. This broad distribution, sample bias notwithstanding, likely indicates the important and underappreciated ecophysiological function of carotenoids in these taxa.

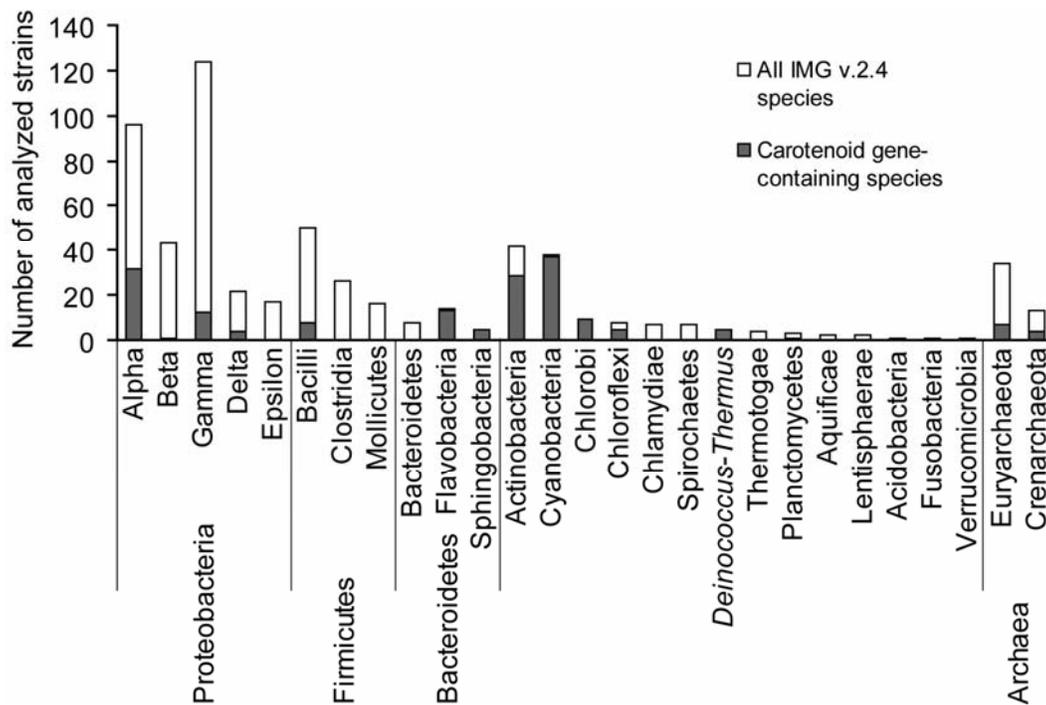


Figure 6.2. Distribution of carotenoid biosynthetic pathways (as inferred from Supplemental Table D1) in genome sequences of the IMG database, version 2.4. Representative strains were selected for each species except Cyanobacteria, for which species designations are not comparable to other bacteria. Because incomplete genomes were included this analysis represents an underestimate.

6.3.2. Broad evolutionary patterns are revealed by core carotenoid biosynthetic protein phylogenies

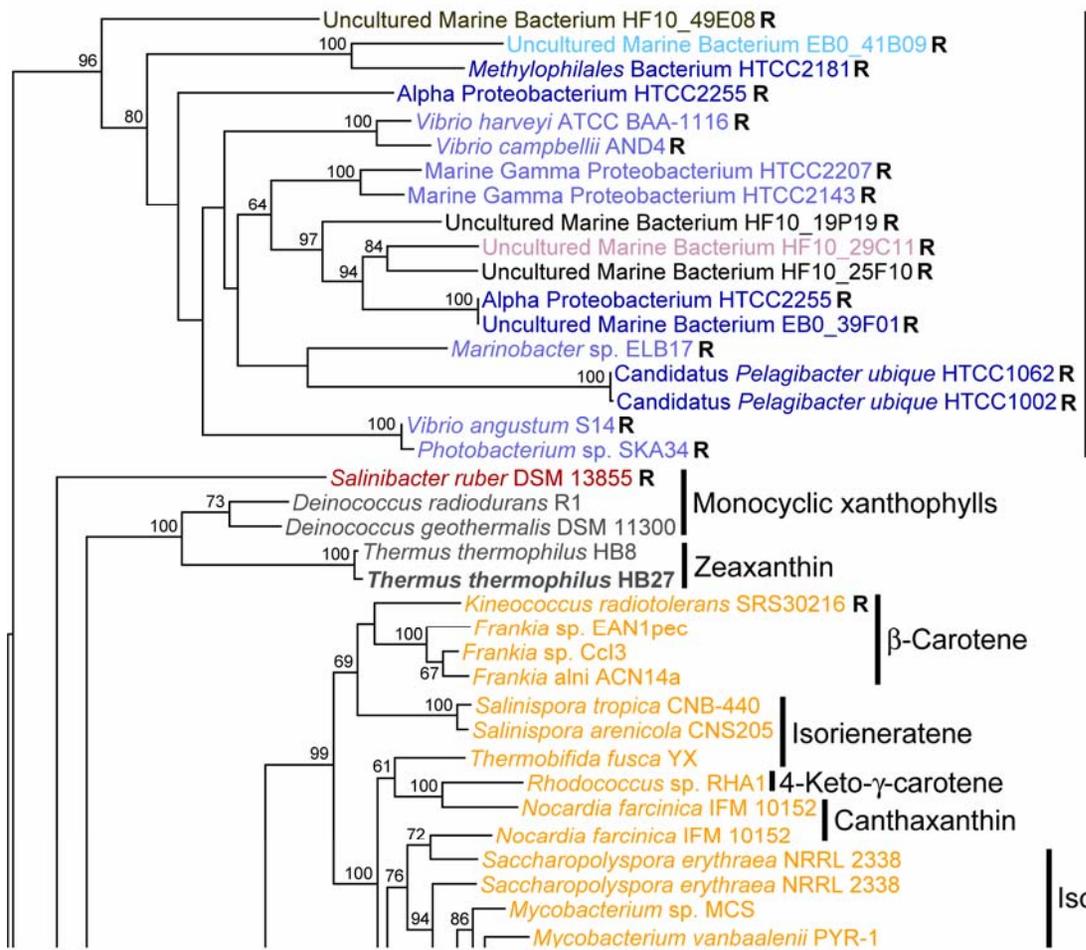
Phylogenetic analyses of carotenoid biosynthetic protein sequences allow the determination of carotenoid evolutionary history. Although previous evolutionary studies exist (e.g., Sandmann 2002a, Phadwal 2005) they: (i) model evolution based upon biochemical characteristics without fully considering sequence distribution or pathway context; (ii) use limited datasets; (iii) use misidentified organisms to represent entire phyla (especially *Paracoccus zeaxanthinifaciens* misidentified as *Flavobacterium* sp.) or (iv) use less rigorous phylogenetic methods than those employed here. By predominantly using reference genomic sequence databases containing a wide diversity of taxa identified using numerous taxonomic markers, the search and analysis strategies used here avoid these pitfalls.

Because of the tree-like nature of the carotenoid biosynthetic pathway (Umeno et al. 2005; Figure 6.1), delineation of carotenoid biosynthetic protein lineages is most easily visualized using phylogenies of the core proteins: CrtB and its homolog CrtM, CrtI, CrtY and CrtYcd (Figures 6.3 and 6.4, Supplemental Figures D1 and D2). Indeed, only CrtB/M sequences are conserved in all carotenogenic taxa, with carotenoid biosynthesis in Firmicutes and the Chlorobi-Cyanobacteria-photosynthetic microbial eukaryote lineage diverging after this step. As expected from their divergent pathway content, both of these carotenoid biosynthesis lineages form coherent clades in the CrtB/M tree, although poor resolution was observed for the deepest Chlorobi node (Figure 6.3). In other analyses using representative sequences and omitting gapped amino acid positions, however, the node between Chlorobi and Chloroflexi is robustly supported (data not shown). This is intriguingly reminiscent of phylogenies of (bacterio)chlorophyll biosynthetic protein phylogenies (Xiong et al. 2000). Unfortunately, the short length of the CrtB protein restricts the phylogenetic signal that it contains such that the other deep nodes in this tree remain poorly resolved, limiting its utility to infer the evolution of photosynthesis (e.g. between

purple bacteria, Cyanobacteria and the pair of Chlorobi and Chloroflexi) more generally.

Two other major carotenoid biosynthetic lineages are clearly evident from core carotenoid biosynthetic protein phylogenies (Figures 6.3 and 6.4, Supplemental Figures D1 and D2), one containing nearly all proteobacterial sequences and one containing sequences from Actinobacteria, Bacteroidetes and Archaea. These divisions were preserved in rooted trees (data not shown). The former can be further divided into three groups: linear xanthophyll-producing Proteobacteria, proteorhodopsin-containing organisms and β -bicyclic xanthophyll-producing α - and γ -Proteobacteria. In the latter, the CrtB/M phylogeny and rooted CrtI phylogenies (data not shown), suggest that sequences from isorenieratene-producing Actinobacteria and their phylogenetic neighbors (hereafter the C₄₀ lineage) predate those from C₅₀ carotenoid-producing Actinobacteria (hereafter the C₅₀ lineage), Bacteroidetes and Archaea. The branching order of these latter taxa is not conserved between CrtB and CrtI trees, suggesting multiple ancient horizontal transfer events. The overall phylogenetic cohesion of most clades (i.e., each clade contains sequences from only one taxonomic group) suggests that carotenoid biosynthesis evolved soon after the divergence of the major bacterial lineages. This inference most simply explains both the wide distribution of carotenoid biosynthesis in many taxonomic lineages and also its absence in others (Figure 6.2). Several taxa, namely δ -Proteobacteria, Chloroflexi, fungi and *Deinococcus-Thermus* do not unequivocally group with

Figure 6.3. (Next 7 Pages) Phylogenetic tree of CrtB and CrtM protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Genomes containing a rhodopsin homolog are indicated by an “R” and sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree is rooted to its midpoint to maximize the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given. Due to its extreme branch length, the sequence from *Aspergillus niger*, while homologous to all other sequences, was excluded.



β-Carotene

- Taxa:
- Actinobacteria
 - Archaea
 - Bacteroidetes
 - Planctomycetes
 - α-Proteobacteria
 - β-Proteobacteria
 - γ-Proteobacteria
 - Thermus/Deinococcus
 - Unclassified

Monocyclic xanthophylls

Zeaxanthin

β-Carotene

Isorieneratene

14-Keto-γ-carotene

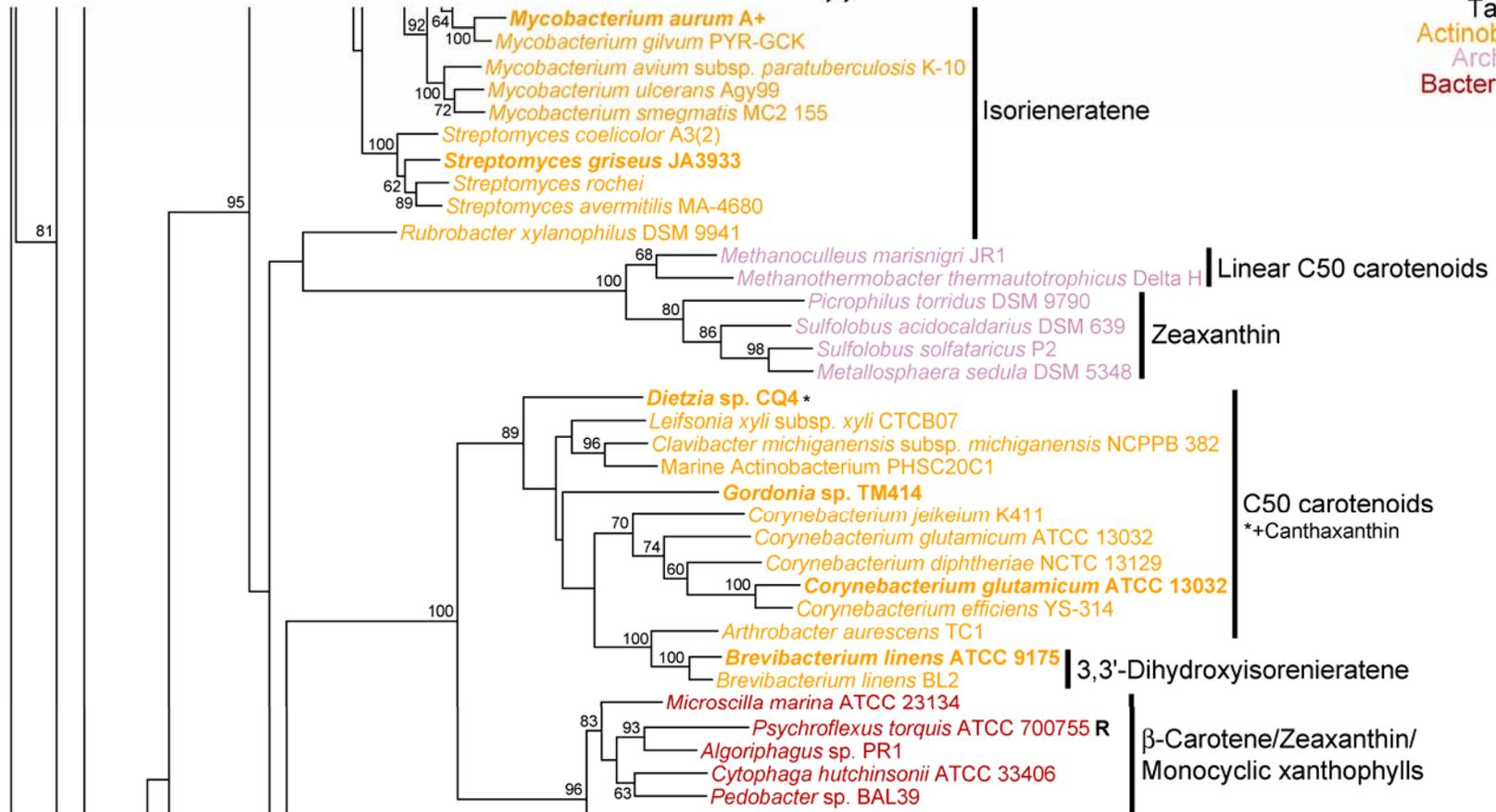
Canthaxanthin

Isorieneratene

(Continued on next page)

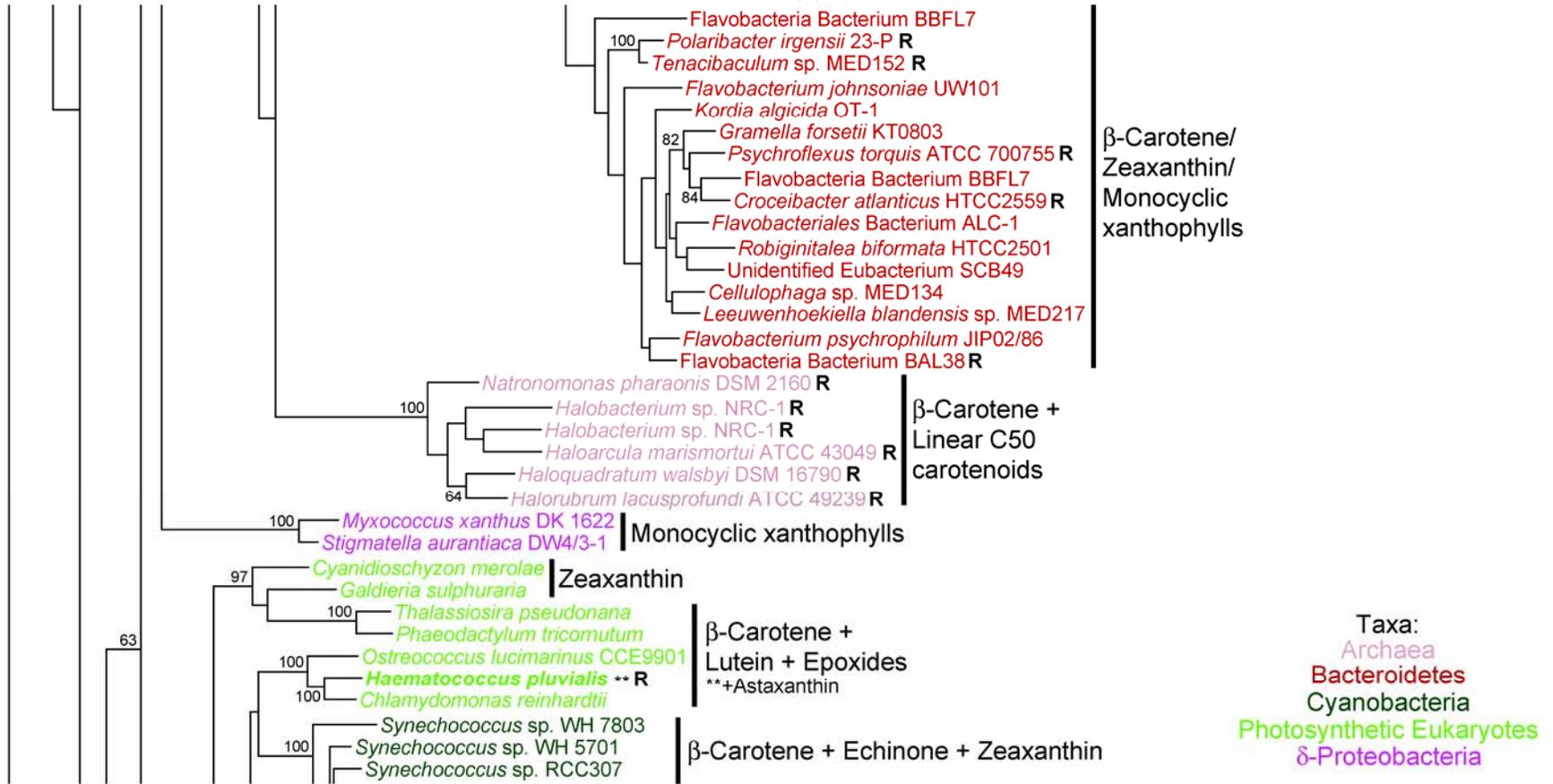
(Continued from previous page)

Taxa:
Actinobacteria
Archaea
Bacteroidetes



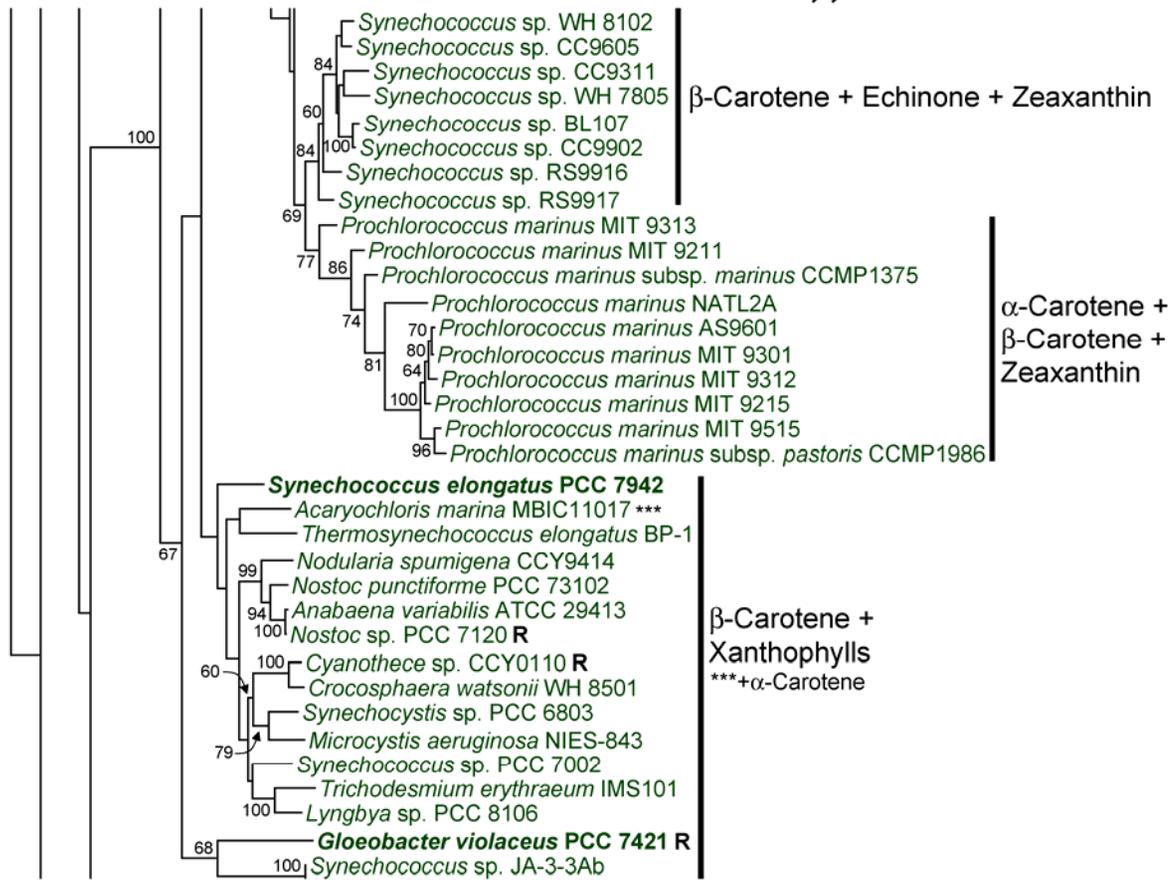
(Continued on next page)

(Continued from previous page)



(Continued on next page)

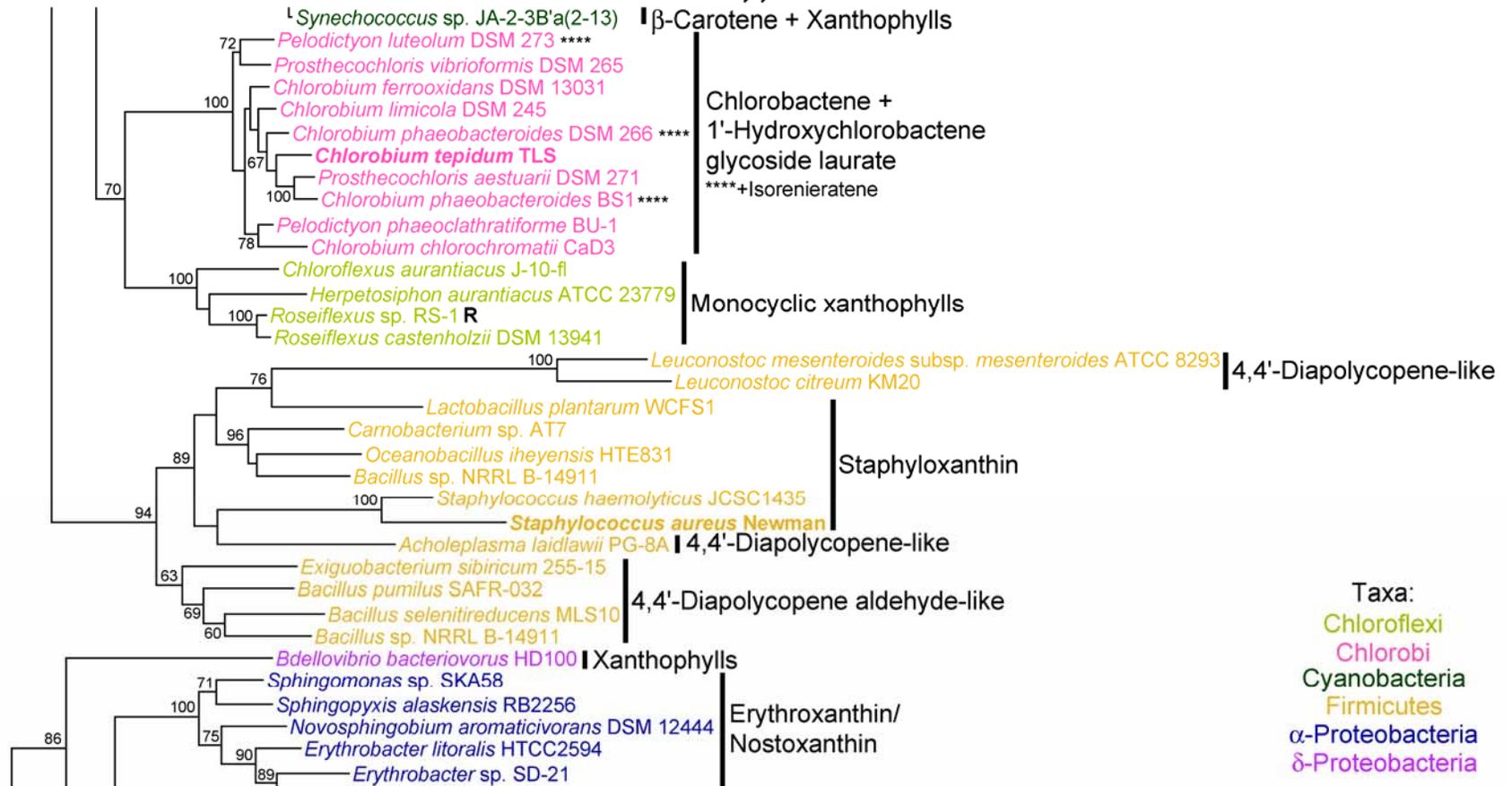
(Continued from previous page)



Taxon:
Cyanobacteria

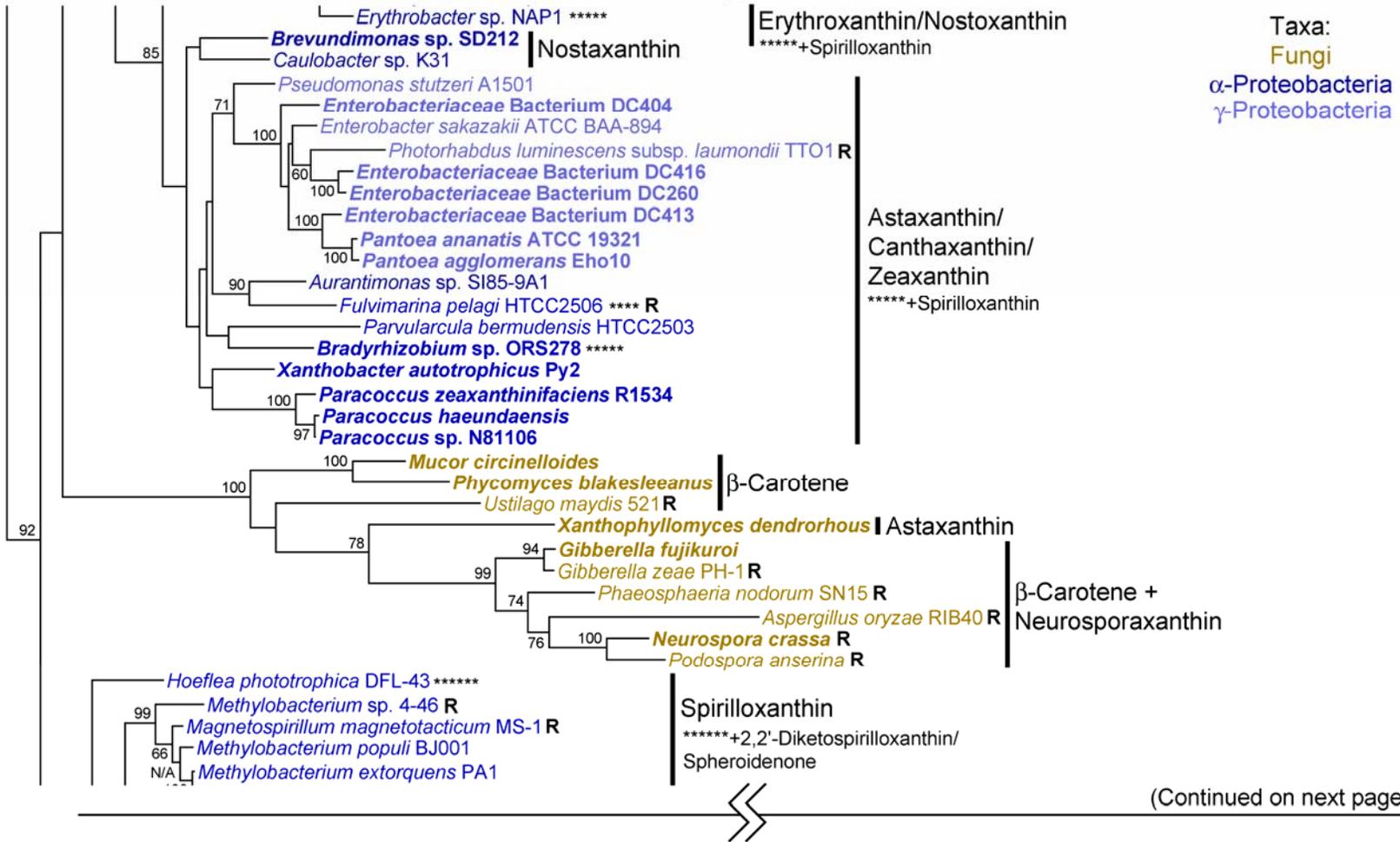
(Continued on next page)

(Continued from previous page)



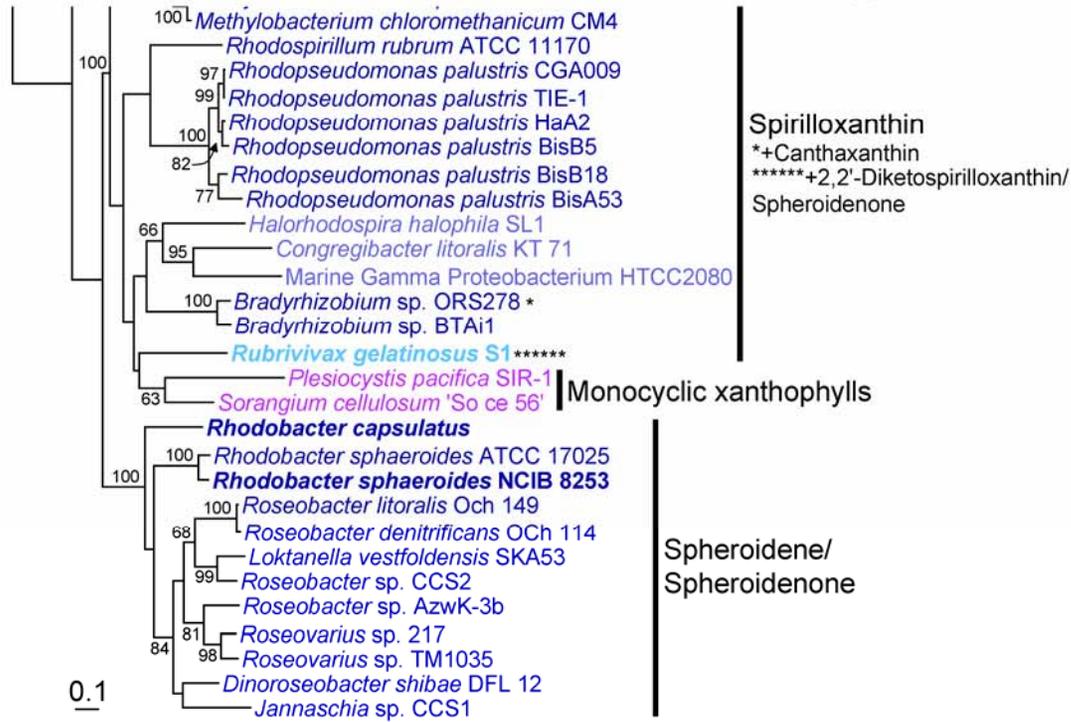
(Continued on next page)

(Continued from next page)



(Continued on next page)

(Continued from previous page)



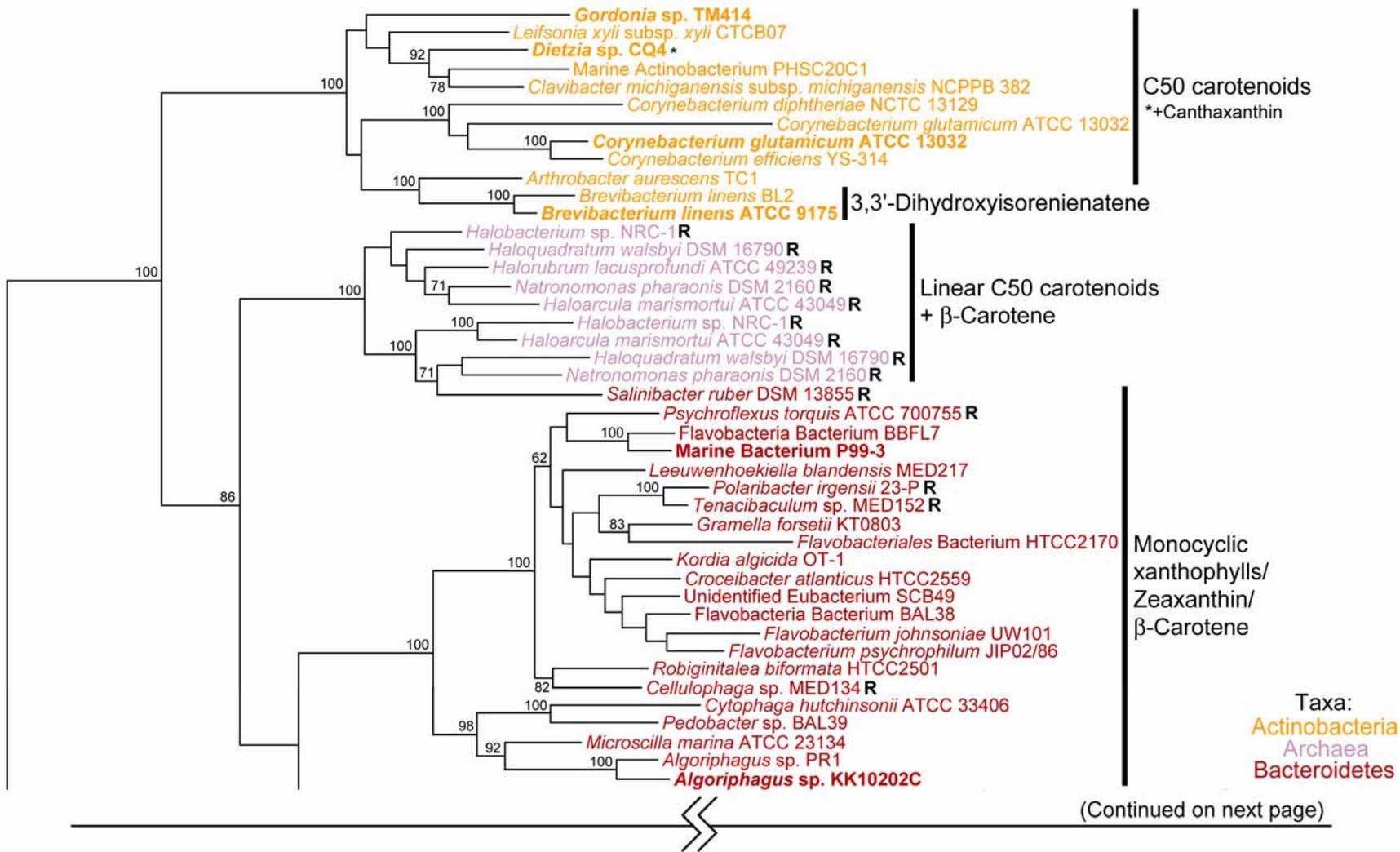
Taxa:
 α -Proteobacteria
 β -Proteobacteria
 γ -Proteobacteria
 δ -Proteobacteria

one of the major identified lineages. This may be due to extreme sequence divergence rendering single protein trees non-informative, analytical difficulties (e.g., from uncertainties in splitting heterodimeric fungal CrtBYcd sequences) or evolution via horizontal gene transfer, as particularly evinced from the presence of multiple lycopene cyclase types in Chloroflexi (CrtL, CrtY and CruA) and *Deinococcus-Thermus* (CrtL and CrtYcd; Supplemental Table D1, Supplemental Figures D1-D4).

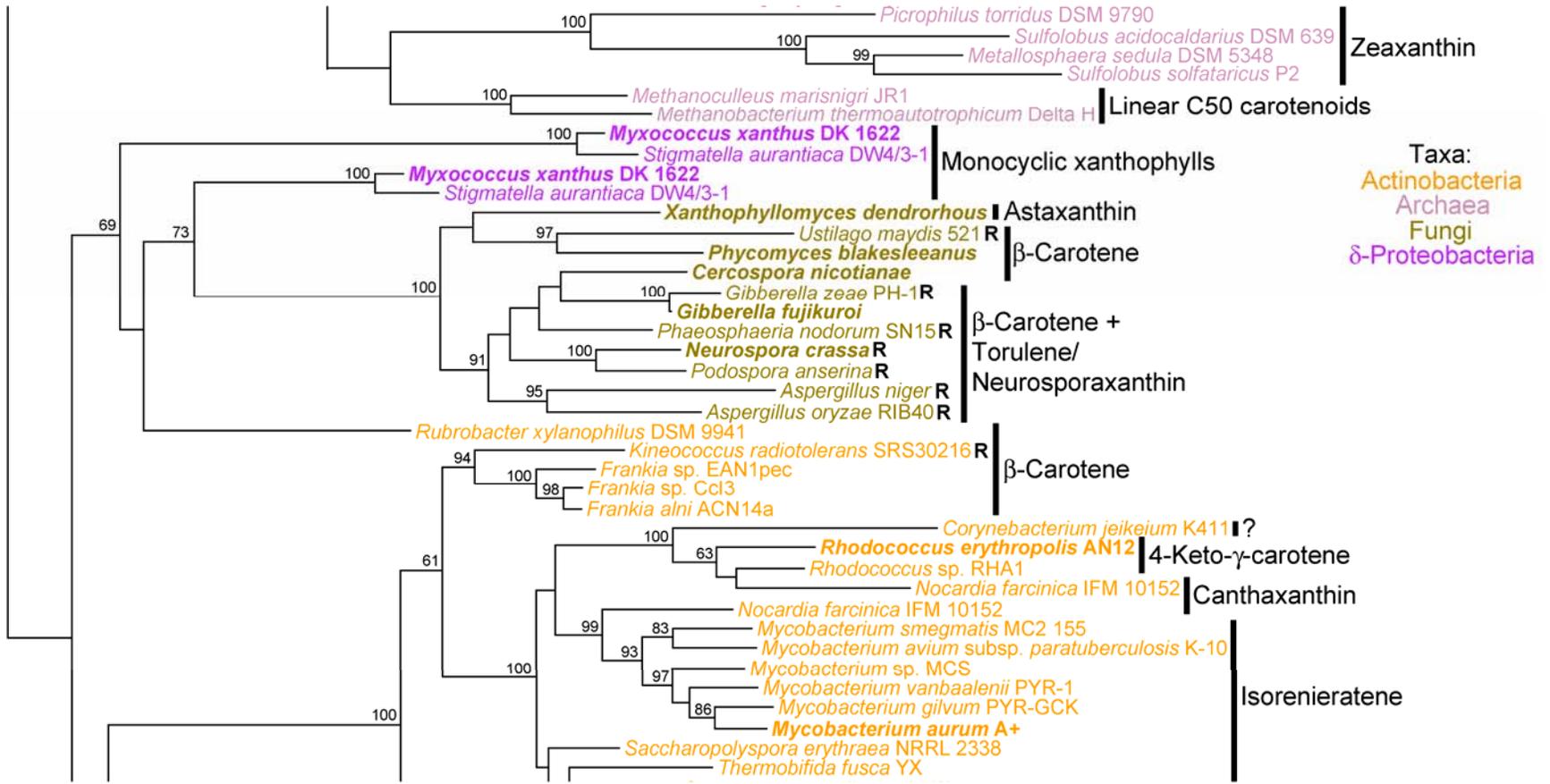
6.3.3. Carotenoid biosynthesis in Proteobacteria that produce β -bicyclic xanthophylls

β -Carotene is produced by the sequential activity of CrtB, CrtI and CrtY, and can be subsequently hydroxylated by CrtG and CrtZ and ketolated by CrtW to form a myriad of β -bicyclic xanthophylls including nostoxanthin, zeaxanthin, canthaxanthin and astaxanthin (Figure 1, Table 1). As is evident from the phylogenies of these proteins (Figures 6.3 and 6.4, Supplemental Figures D1, D5 and D6), α - and γ -Proteobacteria that produce β -bicyclic xanthophylls form a sub-clade in the Proteobacteria carotenoid biosynthetic lineage. The congruency of the CrtZ phylogeny with those of CrtB, CrtI and CrtY and its wide distribution within this lineage suggests ancestral zeaxanthin production within the α -Proteobacteria which was transferred horizontally into the Enterobacteriaceae (γ -Proteobacteria). Canthaxanthin, astaxanthin and nostoxanthin-like carotenoids (additionally requiring CrtG and/or CrtW) have a much more limited and heterogeneous phylogenetic dispersal suggesting later lineage-specific

Figure 6.4. (Next 5 Pages) Phylogenetic tree of CrtI protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Genomes containing a rhodopsin homolog are indicated by an “R” and sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximize the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



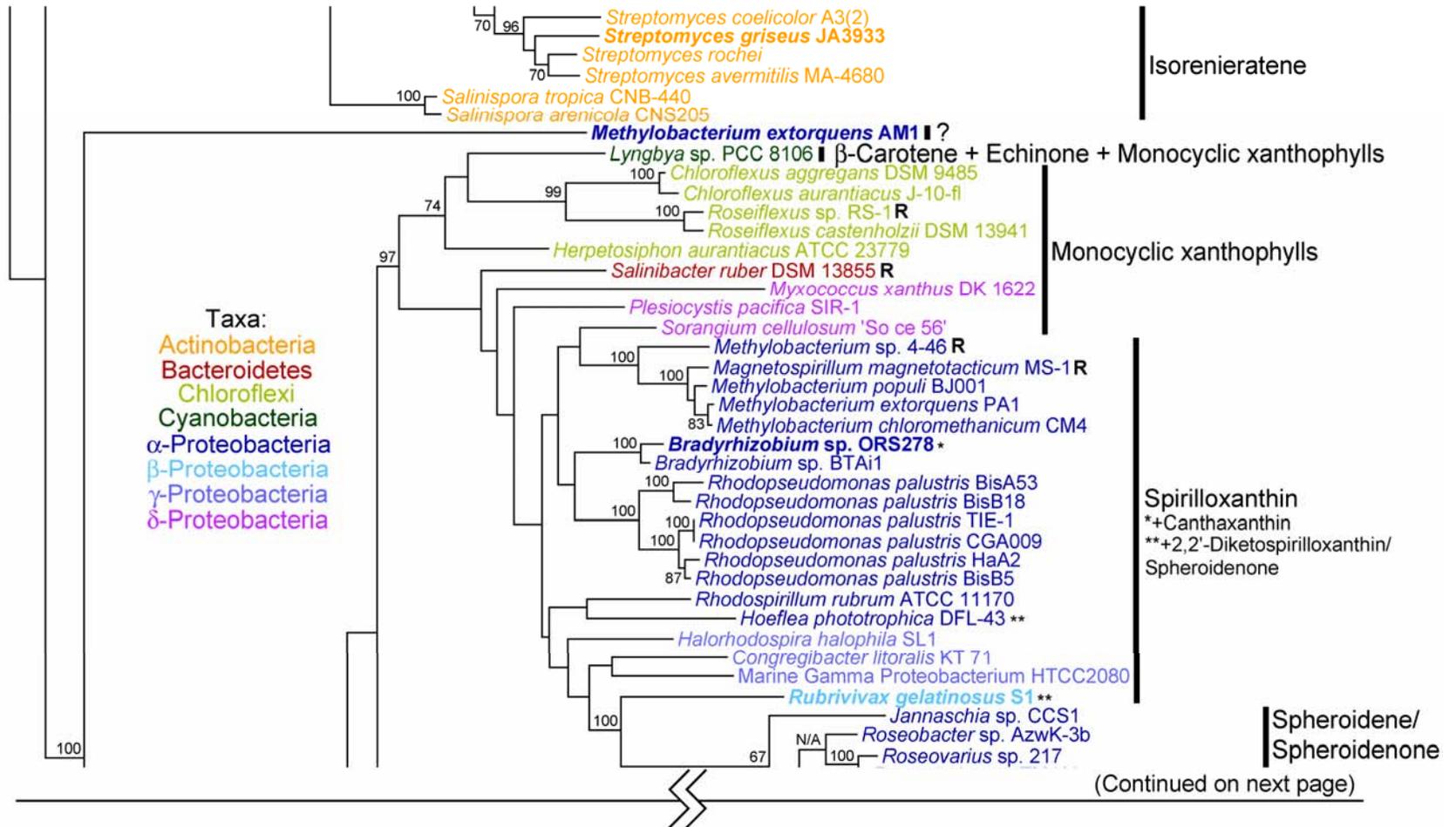
(Continued from previous page)



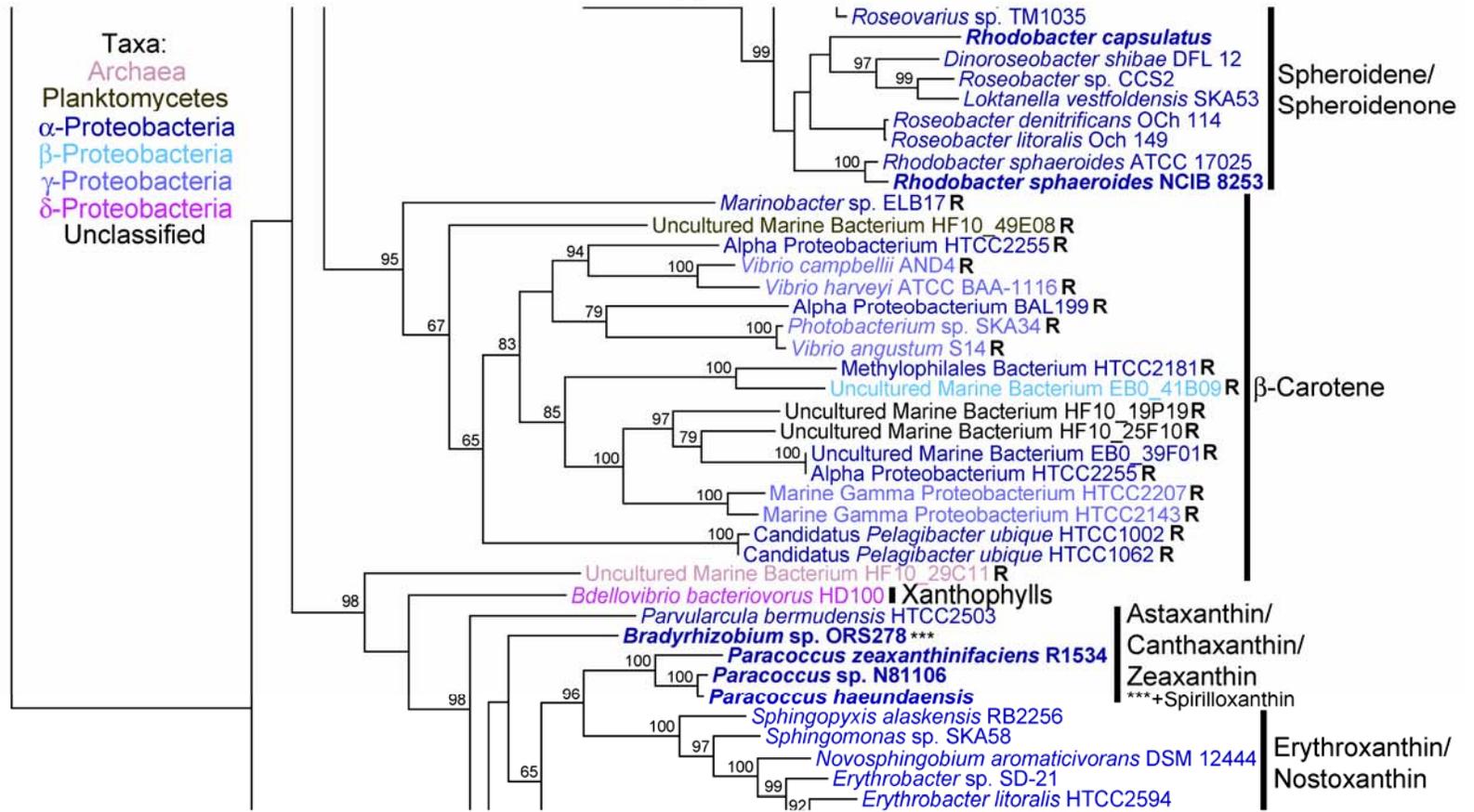
Taxa:
Actinobacteria
Archaea
Fungi
δ-Proteobacteria

(Continued on next page)

(Continued from previous page)

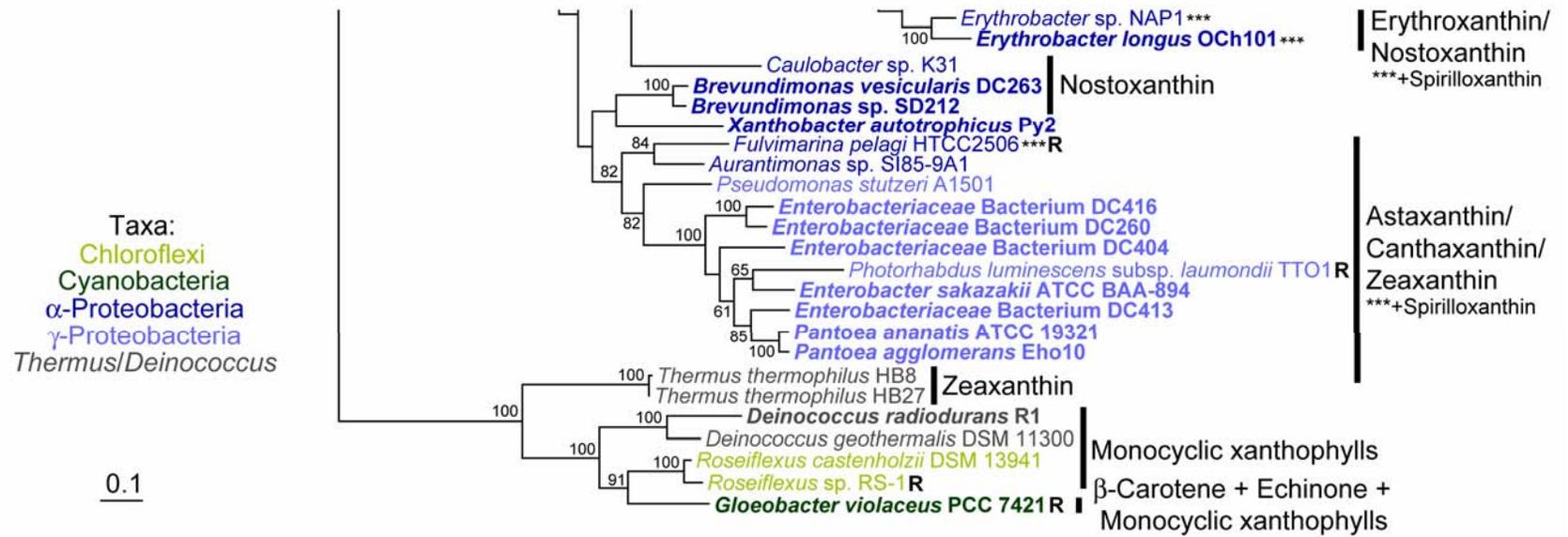


(Continued from previous page)



(Continued on next page)

(Continued from previous page)



acquisition, in some cases coupled to loss of CrtZ. Indeed, the CrtW phylogeny (Supplemental Figure D6) clearly indicates horizontal transfer between α -Proteobacteria e.g., into the *Erythrobacter* lineage leading to the more recent innovation of erythrooxanthin production. In summary, two evolutionary forces, lateral transfer of CrtW and CrtG and differential loss of CrtZ, resulted in heterogeneously distributed carotenoid biosynthetic pathways in phylogenetically related β -bicyclic xanthophyll-producing α -Proteobacteria. Additionally, the combinatorial nature of these pathways can lead to even greater structural diversity via the production of chemically asymmetrical intermediates. Accumulation of these intermediates depends on fine-scale sequence variation and differential protein expression, both currently undetectable using comparative genomics.

6.3.4. Carotenoid production in Proteobacteria that produce linear xanthophylls

Purple Proteobacteria produce linear carotenoids of the spheroiden(on)e and spirilloxanthin series' (Supplemental Table D1) that interact with the light harvesting complex in these microorganisms. Because the detailed evolution of this carotenoid biosynthetic sub-pathway is discussed in Chapter 7, only broad-scale patterns of diversity are highlighted here. Co-clustering of linear xanthophyll biosynthetic protein sequences (Figures 6.3 and 6.4, Supplemental Figure D7) reflects a common pathway for structural divergence, likely constrained by interactions of these carotenoids with the proteobacterial photosynthetic reaction center. CrtA, however, is present in *Rubrivivax gelatinosus* S1 and *Hoeflea phototrophica* DFL-43. Because CrtA is otherwise restricted to spheroidenone-producing bacteria, this pattern distribution strongly suggests its horizontal transfer. Recent horizontal transfer in this biosynthetic lineage is otherwise rare, although more ancient transfers have occurred involving all carotenoid biosynthetic pathway genes as part of a photosynthetic gene cluster (Igarashi et al. 2001). Production of carotenoids of either the spheroidene or spirilloxanthin pathways, which differ primarily by the length of their conjugated

double bond chain, is determined by the substrate specificities of specific pathway members, especially CrtC and CrtI (Steiger et al. 2003, Stickforth and Sandmann 2007), and perhaps also by the rate of metabolic flux (Stickforth and Sandmann 2007). *R. gelatinosus* S1 produces carotenoids of both the spheroidene and spirilloxanthin series (Takaichi 1999), suggesting the potential for pathway plasticity. This is conceptually similar to the production of asymmetrical carotenoids by the β -bicyclic xanthophyll-producing α -Proteobacteria whereby carotenoid diversity is mediated by fine-scale amino acid sequence variation rather than gene presence or absence. Carotenoid biosynthesis in Proteobacteria producing linear xanthophylls occurs primarily through vertical inheritance, likely resulting from selection imposed by physical interaction with the photosynthetic reaction center. Other options, however, are also possible, such as acquisition by horizontal transfer of CrtA and combinatorial biosynthesis due to fine scale changes in enzyme substrate specificity.

6.3.5. Carotenoids as precursors for retinal biosynthesis

Rhodopsin homologs are present in many carotenogenic taxa (Figures 6.3 and 6.4, Supplemental Figures D1 and D2, Supplemental Table D1) as befits the function of β -carotene as a retinal precursor. Strikingly, CrtB, CrtI and CrtY sequences from most proteorhodopsin-containing organisms form one coherent clade branching from the main Proteobacteria lineage (Figures 6.3 and 6.4, Supplemental Figure D1), indicating their coevolution with proteorhodopsins, apparently constrained from further biosynthetic diversification by their role in retinal production. Extensive horizontal transfer of these sequences is suggested by co-clustering of α -, β - and γ -Proteobacteria, Euryarchaeote and Planctomycete sequences (Figures 6.3 and 6.4). Similar results have been reported previously, including the heterologous phylogenetic placement of different pathway enzymes (McCarren and DeLong 2007) which, while present in these trees (e.g., clone HF10_29C11), is relatively rare (Figures 6.3 and 6.4; Supplemental Figure D1). Carotenoid biosynthetic protein sequences and rhodopsins also co-occur in all carotenogenic fungi and haloarchaea, suggesting their long evolutionary coupling

in these lineages, separate from the proteobacterial lineage and perhaps also each other. Carotenoid biosynthetic gene duplication apparently occurred early in the history of the haloarchaeal lineage (Figure 6.4), likely to accommodate retinal production; similar duplications have also occurred in the Bacteroidetes (Figure 6.4).

Comparison of the phylogenies presented here (Figures 6.3 and 6.4, Supplemental Figures D1 and D2) to those generated by Sharma et al. (2006) for rhodopsins similarly reveals two rhodopsin-producing lineages, one containing haloarchaeal and fungal rhodopsins and the other proteorhodopsins. Interestingly, the rhodopsins phylogenetically most closely related to the proteorhodopsins (Sharma et al. 2006) are those which have most obviously undergone horizontal transfer according to phylogenies of carotenoid biosynthetic proteins, namely, those from *Nostoc* sp. PCC 7120, *Gloeobacter violaceus*, *Kineococcus radiotolerans*, *Rubrobacter xylanophilus* and the Bacteroidetes (Figures 6.3 and 6.4). In these cases, a lack of co-clustering between rhodopsins and carotenoid biosynthetic genes suggests that retinal production evolved by co-opting a preexisting carotenoid biosynthetic pathway. The proteorhodopsin progenitor therefore likely underwent numerous horizontal transfers as a single gene before its linkage with a specific carotenoid biosynthetic lineage. Following this event, both the proteorhodopsin progenitor its associated carotenoid biosynthetic proteins were transferred as a gene cluster, constraining carotenoid biosynthesis in this lineage from further diversification due retinal production. Co-evolution of rhodopsins and carotenoid biosynthetic proteins also occurred in the fungi and archaea, although with greater carotenoid diversification, perhaps accommodated in part by gene duplication.

6.3.6. Carotenoid biosynthesis in the Bacteroidetes

Aside from β -carotene production as a substrate for retinal biosynthesis, several different carotenoid biosynthetic pathways exist within the phylum Bacteroidetes leading to the production of zeaxanthin and a variety of monocyclic xanthophylls (Figure 6.5A, Supplemental Table D1). Monophyletic clustering of

CrtB, CrtI, CrtY(cd) and CrtZ sequences, except for *Salinibacter ruber* (Figures 6.3 and 6.4, Supplemental Figures D1, D2 and D5), suggests the evolution of carotenoid biosynthesis in the phylum Bacteroidetes from a common ancestor that likely produced zeaxanthin. Subsequent divergence of the class Bacteroidetes from the classes Flavobacteria and Sphingobacteria would account for the lack of carotenogenesis in the former taxon (Figure 6.2). As discussed above, this biosynthetic lineage evolved from ancient horizontal transfer of the core pathway genes from Actinobacteria and perhaps also Archaea (Figures 6.3 and 6.4, Supplemental Figure D2). Extensive horizontal gene transfer is evident from the incongruence between 16S rRNA gene phylogeny and the distribution of both carotenoid (Figure 6.5A) and lycopene cyclase type (Supplemental Figures D1 and D2).

Biosynthesis of monocyclic xanthophylls in the phylum Bacteroidetes likely originated from a zeaxanthin-producing pathway by acquisition of enzymes acting on an acyclic ψ -end group such as CrtD and CrtA. Because both mono- and bicyclases from this lineage have been reported (Teramoto et al. 2003, Tao et al. 2006), it is unclear whether γ -carotene preceded or followed the acquisition of CrtD and CrtA. The unique presence of CrtW in *Algoriphagus* suggests the more recent evolution of flexixanthin production in this lineage involving horizontal transfer with *Myxococcus xanthus* (Supplemental Figure D6). Unlike all other Bacteroidetes carotenoid biosynthesis protein sequences, those from *S. ruber* cluster with a variety of phylogenetic neighbors suggesting a polyglot origin for this pathway, in accordance with extensive horizontal gene transfer into this organism (Monogodin et al. 2005). In summary, the phylum Bacteroidetes produces a wide diversity of carotenoids, primarily due to horizontal transfer of carotenoid biosynthetic enzymes both into and within this lineage. As with the β -bicyclic xanthophyll-producing α -Proteobacteria, the combinatorial nature of the β -monocyclic xanthophyll pathway (Figure 6.1) combined with evidence of horizontal transfer within this lineage suggests a high potential for structural diversity of β -monocyclic xanthophylls within this phylum.

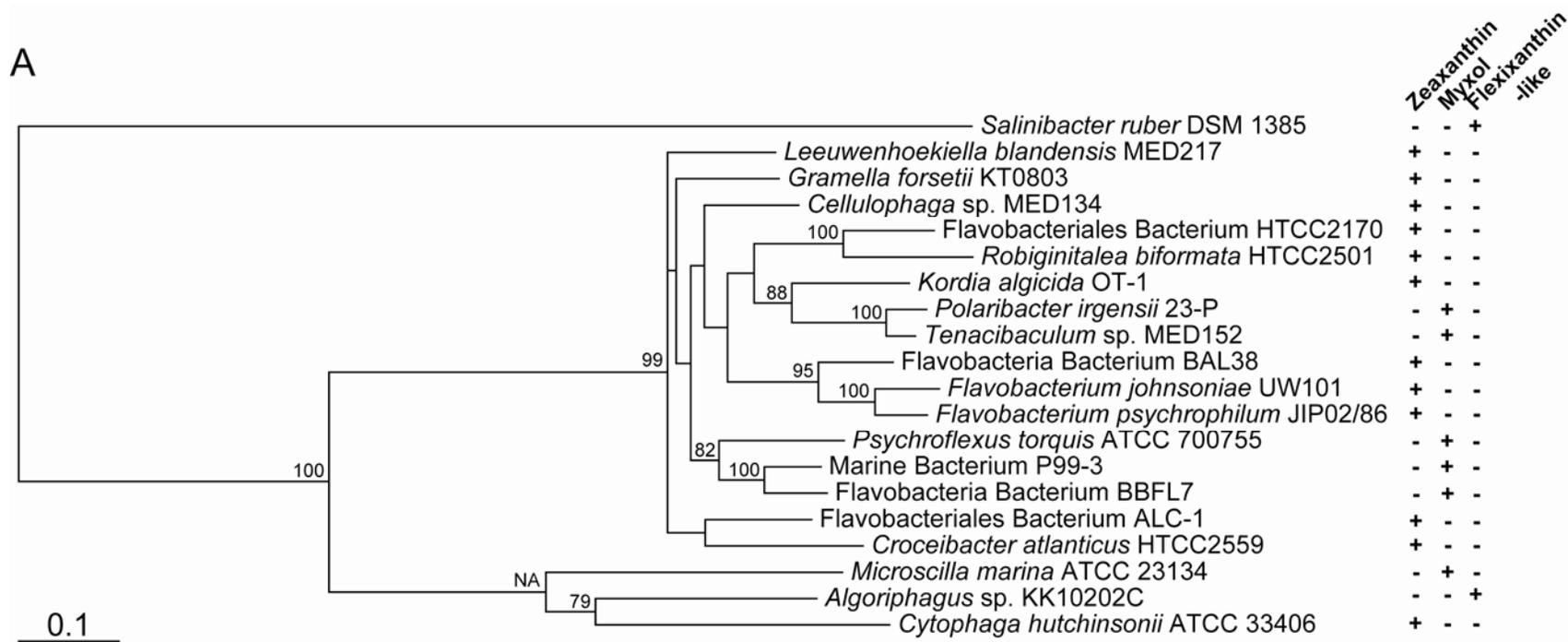
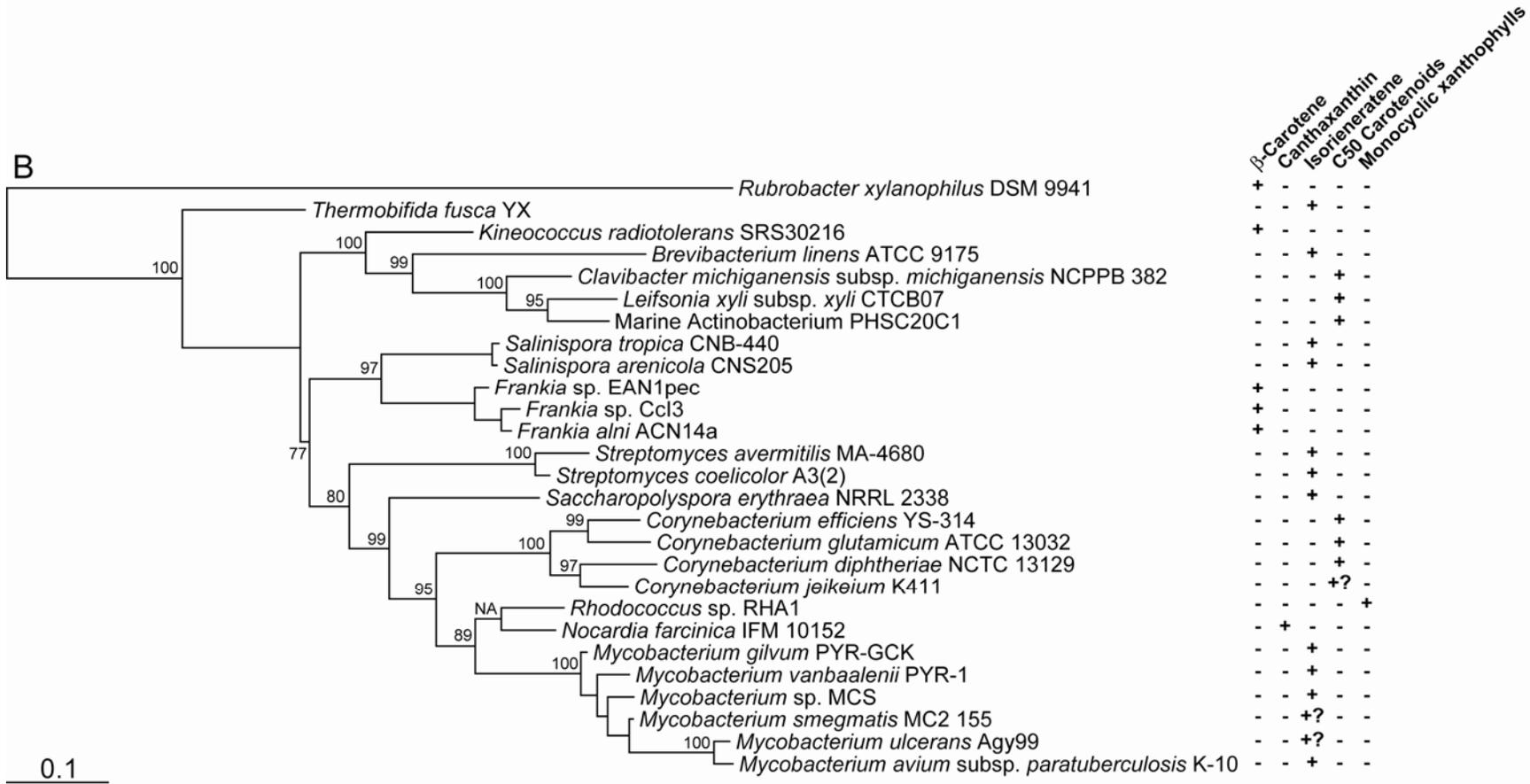
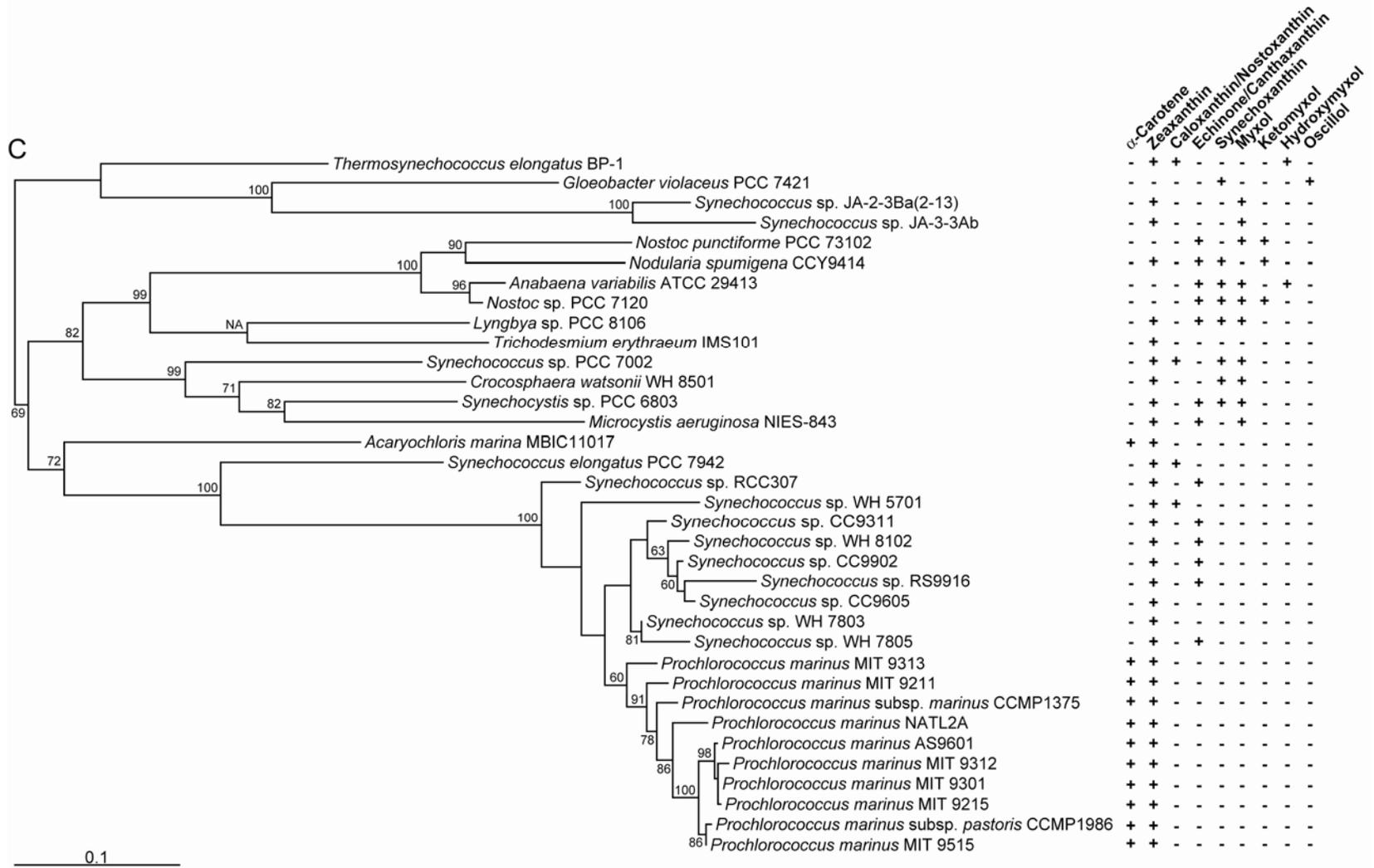


Figure 6.5. Phylogenetic trees constructed from nearly full-length 16S rRNA genes from carotenoid-producing members of (A) the Bacteroidetes; (B) Actinobacteria and (C) Cyanobacteria constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Carotenoids produced, inferred from biosynthetic pathway reconstructions (Supplemental Table D1), are shown to the right of each tree. All trees are rooted to their midpoint, and the scale bar represents 10% sequence divergence. “NA” indicates the ML basal node for which no bootstrap value was given, and “?” indicates carotenoids inferred for pathways for which carotenoid biosynthetic genes have been lost.





6.3.7. Carotenoid biosynthesis in the Actinobacteria

As in the Bacteroidetes, multiple carotenoid biosynthetic pathways exist within the Actinobacteria. The existence of two major actinobacterial CrtB and CrtI clades (Figures 6.3 and 6.4) suggests either independent *de novo* early pathway evolution or ancient horizontal transfer and divergence. Carotenoid composition differs between lineages (Figures 6.3, 6.4 and 6.5B), with the C₅₀ lineage containing the enzymes CrtEb and CrtYef (the latter unique to this lineage and perhaps functionally evolved *de novo*) leading to C.p.450 and decaprenoxanthin, and the C₄₀ lineage containing CrtU for the synthesis of isorenieratene. Horizontal gene transfer has occurred between the C₄₀ and C₅₀ lineages (Figure 6.5B), as exemplified by isorenieratene (C₄₀) production in *Brevundimonas* (C₅₀ lineage) due to the presumed displacement of CrtEb and CrtYef by a C₄₀ lineage CrtU, and the transfer of CrtL into *Dietzia* sp. CQ4 (C₅₀ lineage; Supplemental Figure D3) enabling canthaxanthin production; in this case the C₅₀ carotenoid C.p.450 is still produced. Whether β -carotene production in *Kineococcus radiotolerans* and *Frankia* species and canthaxanthin and 4-keto- γ -carotene production in the *Nocardia-Rhodococcus* lineage (Figure 6.4), Supplemental Figures D1 and D2) is due to CrtU loss or predates its acquisition by the C₄₀ lineage remains unclear due to poor bootstrap support for rooted CrtU trees (data not shown). Mutation leading to incomplete biosynthetic pathways has also occurred in the Actinobacteria, (e.g., *Mycobacterium ulcerans* Agy99, *M. smegmatis* MC2 155, *Corynebacterium diphtheriae* NCTC 13129 and *C. jeikeium* K411), likely as a result of ongoing genome reduction during evolution into a new pathogenic niche presumably not requiring the biosynthesis of isorenieratene (e.g., Stinear et al. 2007). Surprisingly, there exist in the C₄₀ lineage three different classes of lycopene cyclase: CrtL, CrtY and CrtYcd. The heterologous phylogenetic distribution of these proteins again suggests horizontal gene transfer with concomitant loss of the ancestral cyclase. In summary, horizontal gene transfer, coupled in most cases with concomitant gene loss, and possibly the *de novo* evolution of novel biochemical functions (e.g., CrtYef) within the

Actinobacteria has resulted in a diverse array of carotenoid biosynthetic pathway types. However, these pathways are not as heterologously distributed or combinatorially evolved as in the Bacteroidetes and β -bicyclic xanthophyll-producing α -Proteobacteria.

6.3.8. C30 carotenoid biosynthesis

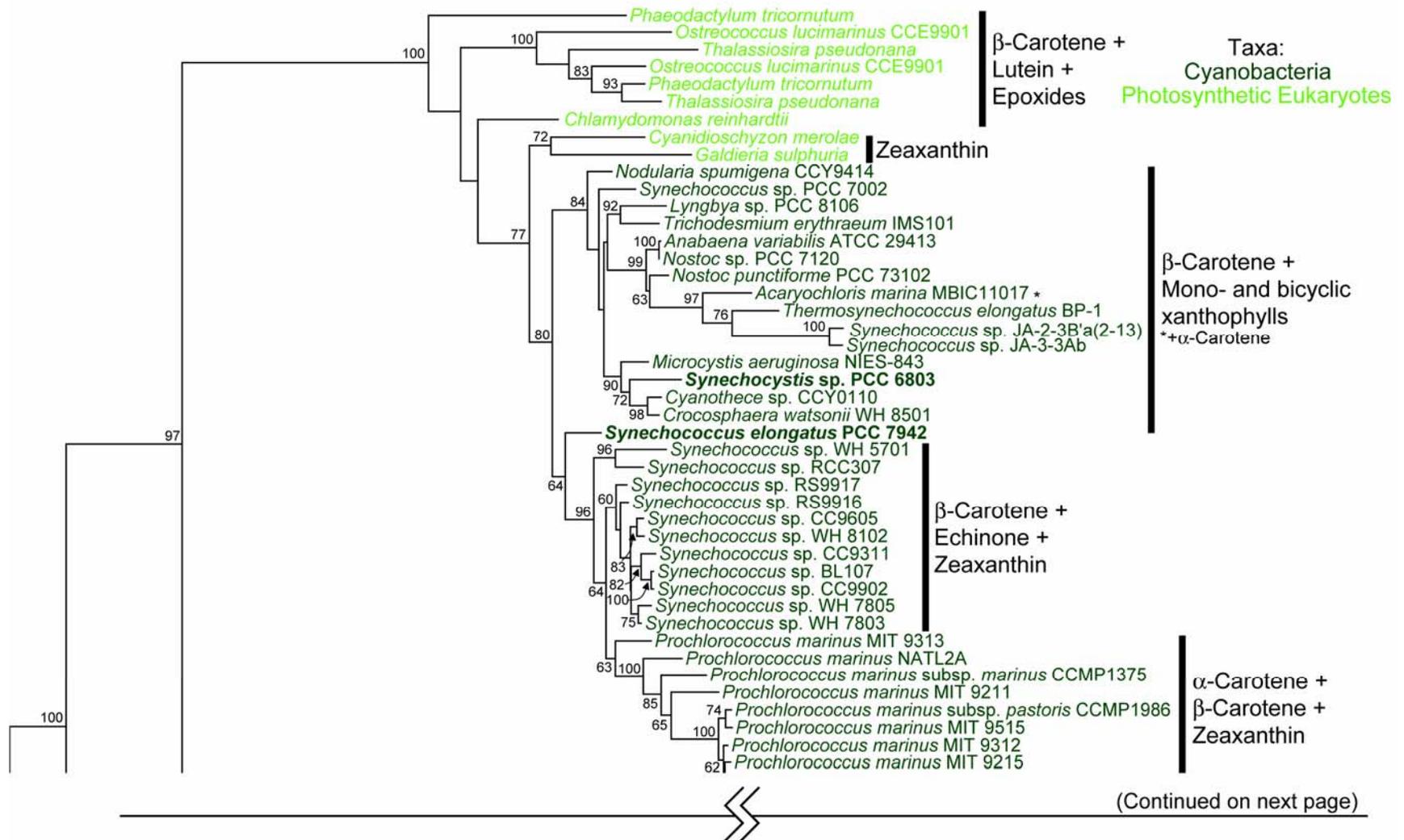
C30 carotenoid biosynthetic protein homologs are present in many *Bacilli* and certain γ -Proteobacteria and Planctomycetes. A C30 carotenoid has also been reported from Verrucomicrobia (Shindo et al. 2008), although no carotenoid biosynthetic protein sequences are available for analysis here. The CrtB/M phylogeny (Figure 6.3) suggests that C30 carotenoid biosynthesis is distinct from all other carotenoid biosynthetic protein phylogenetic lineages; indeed, in rooted trees it is consistently closest to the root, albeit with poor intra-clade resolution (data not shown). *Methylomonas* sp. 16a, *Rhodopirellula baltica* SH 1, *Gemmata obscuriglobus* UQM 2246 and *Heliobacterium modesticaldum* Ice1 all lack close homologs of CrtM, suggesting at least one novel type of 4,4'-diapophytoene synthase in these organisms. Accurate assignment of C30 carotenoid biosynthetic pathway function is rendered difficult by the ability of homologous proteins to catalyze different reactions in different organisms e.g., *Staphylococcus aureus* vs. *Methylomonas* sp. 16a (Figure 6.1, Supplemental Table D1). Some microorganisms, including *Oceanobacillus iheyensis* HTE831, *Carnobacterium* sp. AT7 and *Bacillus* sp. NRRL B-14911 presumably possess an entire biosynthetic pathway leading to staphyloxanthin, the major carotenoid of *S. aureus*. Interestingly, none of the former are pathogenic, implying a broader ecophysiological role for this pigment than exclusively as a virulence factor (Liu et al. 2005). Whether C30 carotenoid biosynthetic pathways in other microorganisms are typically short or involve currently unidentified proteins remains a matter for future study.

6.3.9. Carotenoid biosynthesis in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage

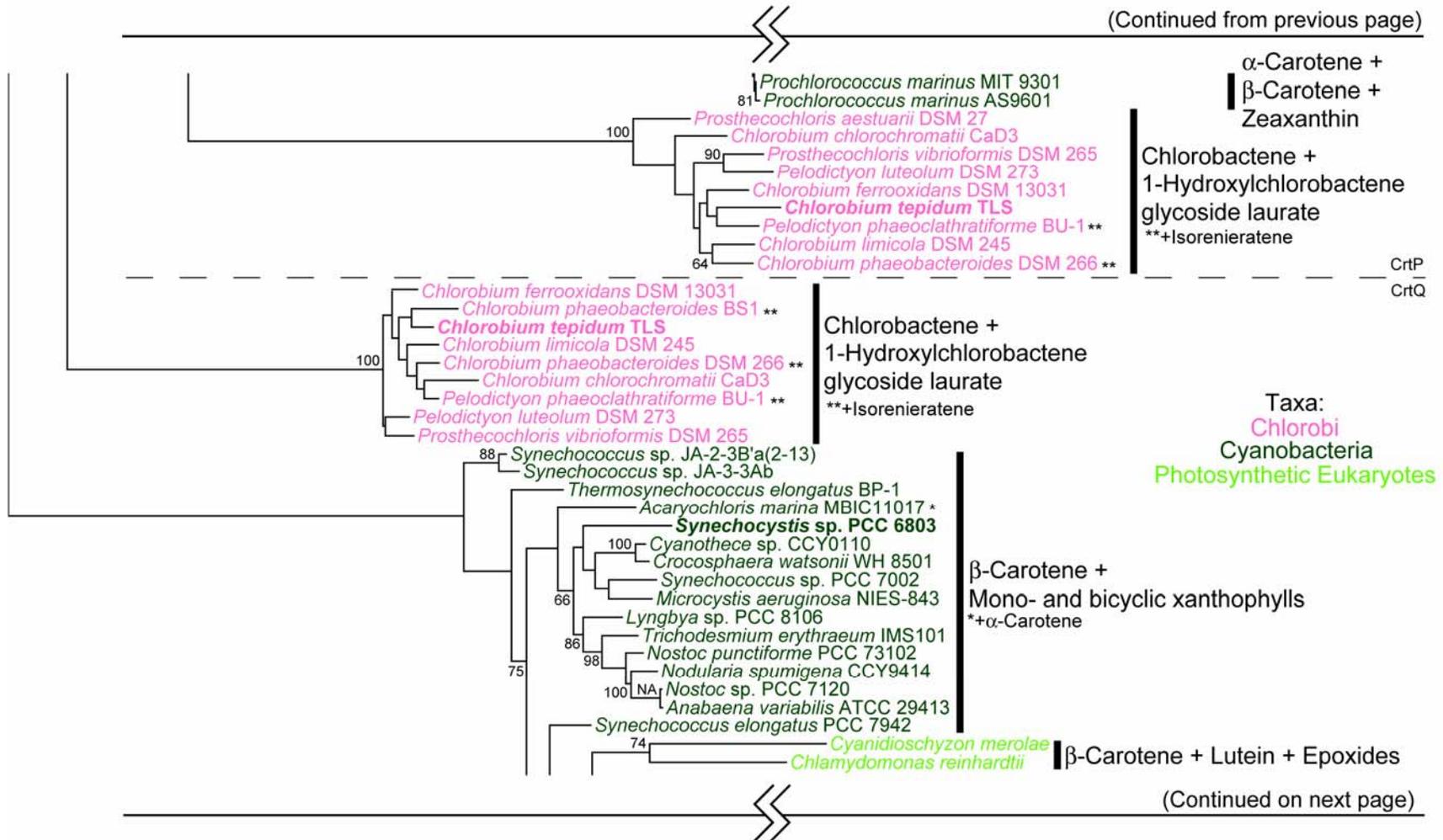
Carotenoid biosynthesis in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage features both monocyclic (Chlorobi) and bicyclic (photosynthetic eukaryotes) branches, even within the same organism (some Cyanobacteria). As discussed above, the combinatorial nature of these sub-pathways leads to an extensive range of possible structures, which is further compounded by extensive *de novo* sub-pathway evolution. Examples include unique sub-pathways leading to α -carotene via CrtLe, synechoxanthin via CruE and CruH, 1'-hydroxychlorobactene glycoside laurate via CruC and CruD and isorenieratene via CruB. While some pathways (especially those leading to α -carotene and canthaxanthin and/or echinone) have a discrete distribution, all Cyanobacteria other than *Prochlorococcus* and related *Synechococcus* (hereafter “other Cyanobacteria”), despite forming a phylogenetically coherent clade, additionally produce many different, heterologously distributed xanthophylls (Figure 6.5C).

Horizontal transfer of cyanobacterial carotenoid biosynthetic proteins is only obviously apparent at deeper nodes (e.g., CrtW, Supplemental Figure D6), and may involve transfer from non-photosynthetic organisms (e.g., CrtG, CrtL, CrtO, CrtW); outgroup sequences for these clades could either not be determined (CrtL) or generated poorly supported basal nodes (all other proteins; data not shown). The well supported basal branching of red algal CrtB sequences (99% bootstrap support; Figure 6.3) relative to the green algae and Cyanobacteria

Figure 6.6. (Next 3 Pages) Phylogenetic tree of the paralogs CrtP and CrtQ (PDS and ZDS in eukaryotes) protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated function are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. All trees are rooted to their midpoint, and the scale bar represents 10% sequence divergence. NA indicates the ML basal node for which no bootstrap value was given.



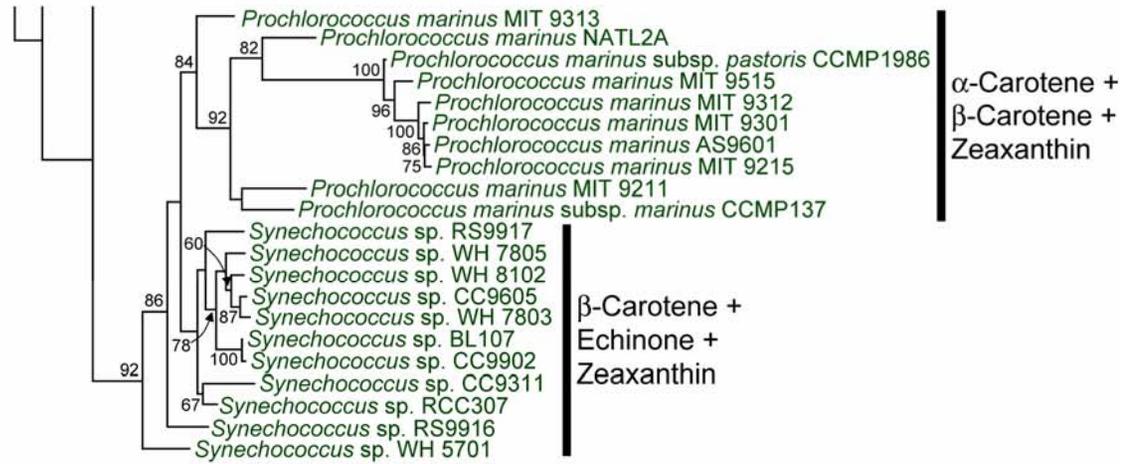
(Continued from previous page)



(Continued on next page)

(Continued from previous page)

Taxon:
Cyanobacteria



0.1

s

(branching order less well resolved; Figure 6.3) in particularly interesting, and implies the transfer of carotenoid biosynthesis from the (pro-)chloroplastic lineage into that of modern Cyanobacteria. This feature is not observed in other gene trees due to the presence of other confounding evolutionary events in the photosynthetic eukaryotes, such as the lack of well-supported CrtH homologs and CrtP and CrtQ loss (Figure 6.6; see also Frommolt et al. 2008). Although intriguing, support for this observation requires further detailed study outside of the scope of this research.

More recent horizontal transfer events, although detectable (see *Nodularia spumigena* CCY9414 CrtP, Figure 6.6) and impossible to rule out at poorly resolved nodes, appear rarely in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage, as determined from the general reproducibility of all core and many peripheral carotenoid biosynthesis protein phylogenies (Figure 6.6, supplemental Figures D6 and D8). Based on this reproducibility, the combinatorial nature of monocyclic xanthophyll biosynthesis and the phylogenetic cohesion of the “other Cyanobacteria”, the most parsimonious hypothesis is that differential gene loss plays a dominant role in generating monocyclic xanthophyll diversity in the “other Cyanobacteria”, with recent horizontal transfer playing a comparatively minor role. The greater evolutionary freedom of monocyclic xanthophyll biosynthesis may be in part due to the lack of selection pressure caused by direct interaction with the photosynthetic reaction center, in contrast to other carotenoids such as β -carotene (Jordan et al. 2001, Loll et al. 2005).

Also remarkable in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage is the presence of five types of lycopene cyclase: CruA, a lycopene β -bicyclase; CruB, a γ -carotene β -monocyclase; CruP, a lycopene β -mono- and bicyclase; CrtLb, a lycopene β -bicyclase; and CrtLe, a lycopene and γ -carotene ϵ -bicyclase. CruA, CruB and CruP are paralogs, as are CrtLb and CrtLe. CruA/B and CrtLb/Le paralog evolution has occurred relatively recently and is responsible for the unique production of isorenieratene and α -carotene in brown Chlorobi and *Prochlorococcus*, respectively (Supplemental Figures D3 and D4).

The duplication of CrtLb to form CrtLe is independent of that in the eukaryotic lineage producing the biochemically equivalent cyclases LYCB and LYCE (Supplemental Figure D3). Divergence of CruA from CruP likely occurred in the more distant past as suggested by the deep separation of these clades (Supplemental Figure D4), possibly to accommodate the production of monocyclic xanthophylls in the “other Cyanobacteria”. Because of its presence in all Chlorobi and “other Cyanobacteria”, it may be hypothesized that CruA-type lycopene cyclases are ancestral in the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage, with CrtL-type cyclases evolving in non-photosynthetic bacteria and subsequently transferred into the *Prochlorococcus-Synechococcus* lineage with concomitant CruA displacement. Unfortunately, attempts to reliably root the CrtL tree to test this hypothesis were unsuccessful due to a lack of suitable outgroups identifiable using the NCBI Conserved Domain Database (data not shown).

In summary, the Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage contains a wide variety of mono- and bi-cyclic carotenoids mostly distributed in a lineage-specific manner. Heterologous distribution of monocyclic xanthophylls in the “other Cyanobacteria” is most parsimoniously explained by differential gene loss. Early pathway assembly most likely involved (i) *de novo* evolution of multiple biosynthetic enzyme types; (ii) cyclase gene duplication and divergent evolution of paralogs and (iii) possibly horizontal transfer, otherwise notably rare in this lineage, of some genes from non-photosynthetic organisms.

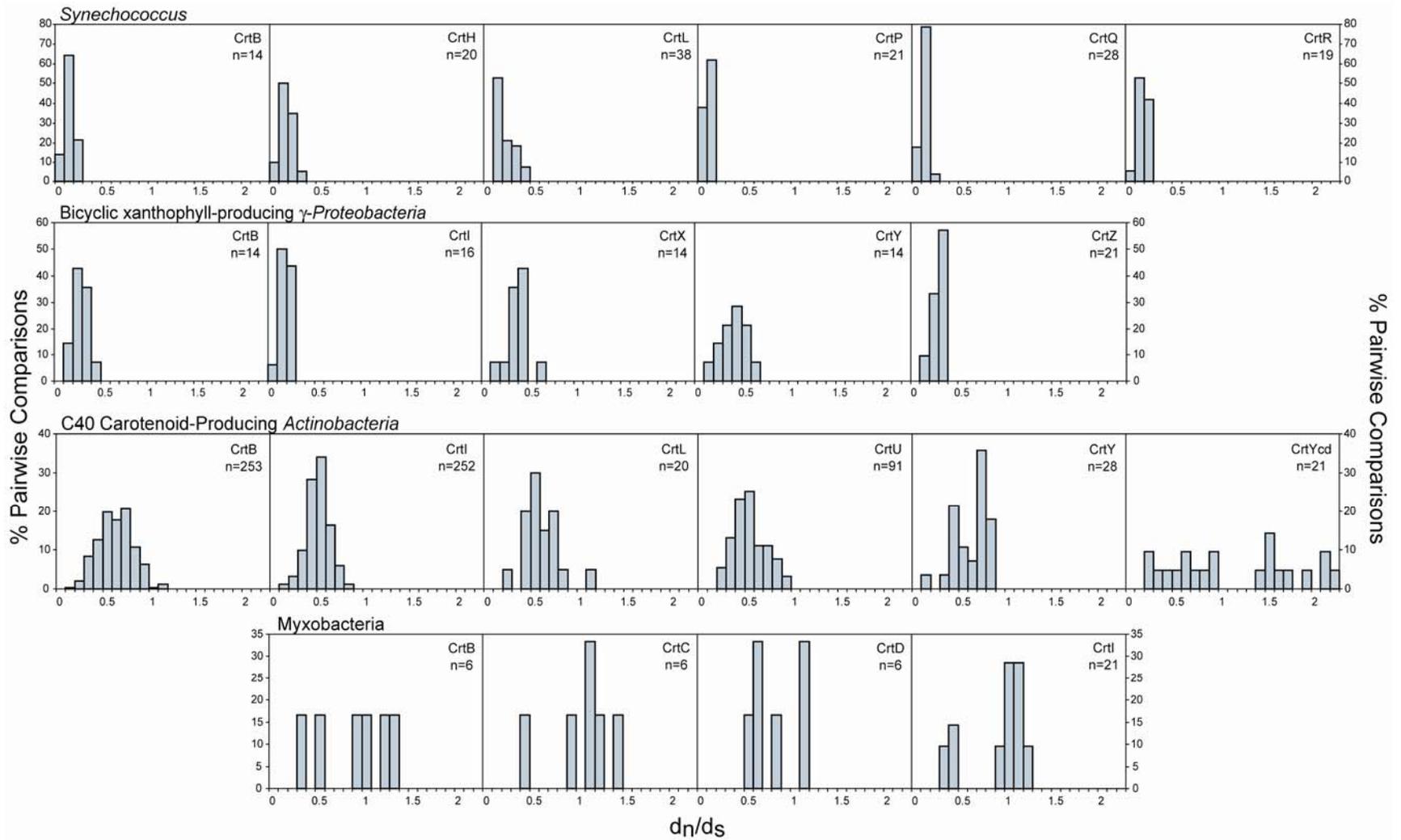
6.3.10. Evolutionary selection in carotenoid biosynthesis pathways

Positive evolutionary selection may increase carotenoid biosynthetic protein diversity by selecting for altered protein functions leading to evolutionarily advantageous phenotypes, especially following gene duplication or horizontal transfer. This phenomenon is detectable as an elevated non-synonymous/synonymous nucleotide substitution ratio (d_n/d_s ; Hurst 2002). Genes for each protein type and carotenoid biosynthetic lineage were compared in a pair-

wise manner, considering only d_n values > 0.1 to ensure sufficient sequence variation and d_s values < 1.5 to account for mutational saturation due to divergence (i.e., resulting from back mutations; Hurst 2002), in general agreement with cutoffs used elsewhere (Novichkov et al. 2009). These cutoff levels, while eliminating obviously aberrant comparisons, also resulted in rejection (due to d_s values > 1.5) of most comparisons of cyanobacterial sequences, many of which are obviously only minimally divergent (Figure 6.6, Supplemental Figure D9). Sequences comparisons within these groups also showed low d_n values, suggesting minimal positive selection operating on these genes. As indicated previously, the evolution of carotenoid biosynthesis in the purple bacteria is analyzed in greater detail in Chapter 7 and is therefore considered only briefly here.

To determine potentially lineage-specific evolutionary mechanisms in carotenoid biosynthesis, pair-wise d_n/d_s comparisons were binned by rounding to one decimal place and the frequency of each value plotted (Figure 6.7, Supplemental Figure D9). Positive selection upon sequences within these datasets was inferred if the resulting distribution was bimodal (as opposed to unimodal if selection was approximately uniform among the sequences analyzed) with a portion centered about a value of one or greater. Upon detection, the original pair-wise matrices were examined to determine the sequence(s) that might be responsible for the elevated values (Supplemental Figure D10) which were compared statistically with non-elevated values from the same lineage (Table 6.2). This approach was chosen over other, more statistically informative analyses such as codeml (Yang 2007) due to its better accommodation of

Figure 6.7. (Next page) Distributions of pair-wise d_n/d_s values, rounded to one decimal place, for *Synechococcus*, bicyclic xanthophyll-producing γ -Proteobacteria, C_{40} carotenoid-producing Actinobacteria and myxobacteria. Values are expressed as a percentage of the total number of comparisons (n) for each sequence cluster protein. Only values with $d_n > 0.01$ and $d_s < 1.5$ were included; note that these cut-offs underestimate values at the lower range of the distributions shown, especially for *Synechococcus*. Results for other taxa are shown in Supplemental Figure D9.



divergent sequences and lesser requirements for computational resources required by the large datasets analyzed in this study. Using this method, d_n/d_s values > 1 were detected for *Mycobacterium aurum* A+ and *Frankia alni* ACN14a *crtYcd*, *Dietzia* sp. CQ4 *crtYef* and carotenoid biosynthetic gene sequences for *Stigmatella aurantiaca* DW4/3-1 and *Myxococcus xanthus* DK 1622 (Table 6.2, and Supplemental Figures D9 and D10). Elevated d_n/d_s ratios, therefore, can occur either for specific genes (e.g., in the Actinobacteria) or for entire pathways and phylogenetic lineages (e.g., in *Myxococcus* and *Stigmatella*). Intriguingly, *M. xanthus* contains two CrtI proteins responsible for separate desaturations (Iniesta et al. 2007), possibly a result of recent divergence due to positive selection. While the rationale for positive selection within the myxobacteria is unclear, it is consistent with the large genome size and abundance of gene duplications

Table 6.2. Inferred positive selection on carotenoid biosynthetic genes as determined from elevated d_n/d_s for specific sequences compared to their phylogenetic neighbors.

Sequences	d_n/d_s (mean \pm standard deviation) ^a	Number of pair-wise comparisons (elevated / non-elevated)	Mann-Whitney U Test Versus Other Sequences From the Same Lineage
<i>Mycobacterium aurum</i> A+ and <i>Frankia alni</i> ACN14a <i>crtYcd</i>	1.75 \pm 0.30	10 / 10	Two-tailed $P = 0.000$ $Z = -3.780$
<i>Dietzia</i> sp. CQ4 <i>crtYef</i>	1.05 \pm 0.08	3 / 5	Two-tailed $P = 0.025$ $Z = -2.236$
<i>Myxococcus xanthus</i> DK 1622 and <i>Stigmatella aurantiaca</i> DW4/3-1 <i>crtB</i>	1.11 \pm 0.17	4 / 1	_b
<i>Myxococcus xanthus</i> DK 1622 and <i>Stigmatella aurantiaca</i> DW4/3-1 <i>crtC</i>	1.20 \pm 0.15	4 / 1	-
<i>Myxococcus xanthus</i> DK 1622 and <i>Stigmatella aurantiaca</i> DW4/3-1 <i>crtD</i>	1.05 \pm 0.03	4 / 1	-
<i>Myxococcus xanthus</i> DK 1622 and <i>Stigmatella aurantiaca</i> DW4/3-1 <i>crtI</i>	1.05 \pm 0.08	16 / 3	Two-tailed $P = 0.007$ $Z = -2.683$

^aPair-wise comparisons between sequences with elevated d_n/d_s ratios were aberrantly low and excluded from this calculation; see Supplemental Figure D10.

^bToo few sequences available for statistical comparison.

reported for these microorganisms (Goldman et al. 2006).

Aside from the evidence of positive selection highlighted above, differences between the overall d_n/d_s ratios over the entire pathway between phylogenetic groups were also detected, albeit with the caveats concerning the d_s cutoffs used and methodological accommodations for the highly divergent sequences analyzed. Considering all carotenoid biosynthetic pathway genes together, d_n/d_s ratios were lowest in Cyanobacteria (d_n/d_s centered about $\approx 0.1-0.2$; Figure 6.7 and Supplemental Figure D9), followed by the spheroidenone-producing Proteobacteria, bicyclic xanthophyll-producing γ -Proteobacteria, Sphingomonadales and Bacteroidetes (d_n/d_s centered about $\approx 0.2-0.3$; Figure 6.7 and Supplemental Figure D9) and spirilloxanthin-producing Proteobacteria, bicyclic xanthophyll-producing α -Proteobacteria, proteorhodopsin-producing bacteria, *Deinococcus-Thermus*, Haloarchaea, Firmicutes and C_{40} and C_{50} carotenoid-producing Actinobacteria (d_n/d_s centered about $\approx 0.4-0.5$; Figure 6.7, Supplemental Figure D9). Although not considered in greater detail here, the differences in selection operative on the carotenoid biosynthetic pathways of different phylogenetic lineages is clearly a topic for future study. Interestingly, differences between d_n/d_s ratios for different pathway steps were not apparent, in contrast to the biosynthetic genes for the plant pigment anthocyanin (Rausher et al. 1999, Lu and Rausher 2003). Whether this is a general feature resulting from the metabolic pathway topology of carotenoid biosynthesis might also benefit from future study.

6.3.11. Recombination in carotenoid biosynthetic pathway evolution

One striking feature of all phylogenetic trees analyzed in this study was the poor bootstrap support for the Chlorobi and Bacteroidetes lineages. A similar result reported by others was attributed to low levels of phylogenetically informative sequence positions despite long branch lengths (Maresca et al. 2008b). Although bootstrap values were improved in maximum likelihood phylogenies considering only Chlorobi sequences, this was not true of Bacteroidetes CrtB, CrtI and CrtZ trees (data not shown). Interestingly, a recent

study identified *Flavobacterium psychrophilum* as having the highest recombination rate of all tested microorganisms (Vos and Didelot 2009). To determine the impact of recombination on the evolution of carotenoid biosynthetic pathways, the heterogeneous rate test (Worobey 2001) was applied to the same sequence groups used for the d_n/d_s calculation. In nearly all cases the ratio of two-state parsimony-informative to all polymorphic sites (q) was <0.35 (average $q = 0.24$) with low associated P values (data not shown). This indicates that homologous recombination was not detected by this method, and therefore likely plays only a minor role in microbial carotenoid biosynthetic pathway evolution.

6.4. Discussion

To rationally select a model system for study and exploitation, the diversity of homologous systems must be characterized to ensure the representativeness of proposed model. Typically, diversity is assessed in three ways. First, and perhaps most commonly, a system well studied in one organism, generally due to historical precedent, is reviewed, compared and contrasted, most often anecdotally, with similar analyses conducted in other organisms. Problematically, this approach arbitrarily assumes, typically without corroborating evidence, that the best-studied system is sufficiently representative of both the taxon studied and natural diversity as a whole to form a baseline for further comparison. Secondly, diversity may be studied by direct comparison, as exemplified in taxonomic studies where phenotypes are considered amongst related strains. Unfortunately, the data generated by these studies are: (a) often at a level of a chemical resolution too low to be informative (as was argued previously for carotenoids; Klassen and Foght 2008); (b) biased by tests developed for well studied and fast-growing model or medically-important organisms; and (c) without integration into models of cellular physiology. Thirdly, organisms and their encoded pathways can be compared by computing gene and protein similarity to those with demonstrated function. This approach has the potential for high-quality, in-depth annotation with a phylogenetic breadth exceeding that possible with other methods. Unfortunately, false-positive annotations are

frequent and easily propagated in subsequent genome sequencing projects e.g., as discussed by Feist et al. (2009). This occurs because of the statistical and pairwise nature of the most common BLAST-based annotation tools, which by themselves present no reliable cutoff below which functional assignment is confident. Additionally, depending on their evolutionary history, even closely related proteins may vary in function. This becomes particularly problematic where few genetically and biochemically validated homologs exist (e.g., rhodopsins; Fuhrman et al. 2008) or in pathways containing many paralogs (e.g., carotenoids). Thus, although harboring vast potential, genome annotations are only truly informative when coupled with corroborating data from the scientific literature and phylogenetic and evolutionary analyses.

In order to maximize the utility of sequence homology to predict natural diversity, this study employed a conservative search strategy limited by: (a) phylogenetic relation of genome-derived sequences to those of demonstrated function; and (b) plausible inclusion of sequence products in a reaction pathway. Emphasizing phylogenetic proximity of sequences lacking functional validation to those with demonstrated functions maximizes the probability of their shared function. Assuming that genetic drift is the major driver of sequence divergence, the probability of shared biochemical function of two sequences co-varies both with their phylogenetic proximity and that of the organisms from which they are derived. Deviations from this probabilistic distribution may result from other evolutionary processes such as horizontal gene transfer or positive selection. Application of this theoretical framework within the constraints of a pathway model, as exemplified here, further reduces the false-negative rate by assuming that homologs share a function only if the phenotype (i.e., the pathway end-product and its structural relatives) is conserved.

In addition to overall phylogeny and evolutionary history, this study inferred four distinct, lineage-specific evolutionary mechanisms of microbial carotenoid evolution (summarized in Table 6.3). First, lateral gene transfer is a particularly important diversification mechanism in some lineages (e.g.,

Table 6.3. Summary of lineage-specific evolutionary mechanisms identified in this study.

Phylogenetic Lineage	Inferred Evolutionary Mechanism			
	Horizontal transfer	Differential <i>de novo</i> evolution and/or gene loss	Coevolution with other proteins	Mutational selection pressure
β -Bicyclic xanthophyll-producing Proteobacteria	Extensive	Rare, CrtZ loss only	None	Possibly slightly elevated overall
Linear xanthophyll-producing Proteobacteria	Rare, sometimes with entire photosynthetic gene cluster	Rare, some variance in enzyme substrate specificity	Extensive, with photosynthetic reaction center	Possibly slightly elevated overall
Proteorhodopsin-producing microbes	Extensive, with proteorhodopsin	None	Extensive, with proteorhodopsin	Possibly slightly elevated overall
Bacteroidetes	Extensive	None	None	None
Actinobacteria	Some, phylogenetically relatively well constrained	Some, especially in the C ₅₀ lineage	None	Possibly slightly elevated overall, some cyclases highly elevated
C30 Carotenoid-producing bacteria	Some, few data	Some, few data	None	Possibly slightly elevated overall
Cyanobacteria-Chlorobi-photosynthetic microbial eukaryote lineage	Possibly at deeper nodes, otherwise rare	Extensive <i>de novo</i> evolution, sometimes via duplication, differential loss in the “other <i>Cyanobacteria</i> ”	Possibly with photosynthetic reaction center	Likely decreased overall
Myxobacteria	Unresolved	None	None	<i>Myxococcus</i> and <i>Stigmatella</i> highly elevated

Actinobacteria and Bacteroidetes) but not others (e.g., Cyanobacteria, at least at relatively shallow nodes). Horizontal gene transfer is particularly important at deep phylogenetic nodes, where newly-evolved biochemical functions were shuffled among evolving phylogenetic lineages such as Actinobacteria, Bacteroidetes, and Archaea, leading to modern patterns of carotenoid diversity. This early period of extensive pathway-constructing horizontal gene transfer has been suggested by others (Woese 1998), is consistent with the inferred early evolution of biochemical functionality (Caetano-Anollés et al. 2007) and is congruent with the “complexity hypothesis” (Jain et al. 1999) whereby propensity for horizontal transfer is inversely related to physiological network connectivity. Secondly, in some lineages (notably, monocyclic xanthophyll-producing “other Cyanobacteria”) carotenoid distribution is best explained by *de novo* biosynthetic evolution, often coupled to differential gene loss, whereby modern organisms contain only a subset of their ancestral genes. In some lineages *de novo* biosynthetic evolution involves gene duplication and divergence. Thirdly, some lineages, notably in the myxobacteria, promote evolution through positive selection. This may, at least in part, be responsible for the biochemically distinct CrtI paralogs present in *Myxococcus xanthus* (Iniesta et al. 2007). Positive selection was also noted for specific reactions in other taxa, perhaps indicating functional divergence; these genes are good candidates for further study and may indirectly reflect other evolutionary processes, e.g., possible ϵ -cyclase *crtYef* evolution in *Dietzia* sp. CQ4 following compensatory acquisition of β -cyclase CrtL. Whether the smaller differences in d_n/d_s between lineages have evolutionary significance remains unclear. Fourthly, co-evolution between carotenoid biosynthesis enzymes and other structures, most notably proteorhodopsins and the photosynthetic reaction center, results in constraints on structural divergence but can potentially access new selective forces to facilitate horizontal transfer as part of a larger evolutionary unit, such as with proteorhodopsins (McCarren and DeLong 2007). Finally, although intra-clade homologous recombination was undetectable in this dataset, its existence is not definitively rejected, particularly for poorly-resolved Bacteroidetes sequences.

Previous analysis based upon carotenoid biosynthetic pathway enzyme diversity has suggested a “tree-like” model of carotenoid evolution (Umeno et al. 2005), whereby early biosynthetic steps of carotenoid biosynthesis (the “root”) are highly conserved in all biosynthetic pathway types whereas later steps (the “braches”) exhibit greater divergence leading linearly to structural diversity. As expected from the tree-like model, in this analysis core biosynthetic genes are more conserved than terminal pathway steps. However, many of the enzymes which are intermediately conserved (e.g., carotenoid cyclases, CrtW, CrtZ) are present within most of the major carotenoid biosynthetic lineages but are not absolutely conserved. This suggests that these enzymes have been subjected to non-vertical evolutionary mechanisms such as differential gene loss and/or gain and horizontal gene transfer. This pattern is remarkably congruent with those present at a much lower phylogenetic scale within the genus *Hymenobacter* (Chapter 5; Klassen and Foght 2008). Considering these data, the model which emerges for carotenoid biosynthetic evolution features a gradient of evolutionary plasticity, greatest for proteins at biosynthetic pathway termini and least at the pathway root; this pattern is typical of metabolic pathway evolution more generally (Jain et al. 1999, Wellner et al. 2007). Carotenoid biosynthetic pathway evolution is therefore much more web-like than tree-like, with proteins evolving in one pathway impacting the evolution of another due to horizontal transfer. Additionally, entire carotenoid biosynthetic pathways have been transferred between different organisms (e.g., proteorhodopsin-producing microbes, linear xanthophyll-producing Proteobacteria, Actinobacteria). Whereas diversification always occurs within the selective confines placed upon a host organism, at a larger phylogenetic scale organismal and biosynthetic pathway evolution can be independent processes.

Understanding the diversity of evolutionary mechanisms acting on a metabolic pathway and their distribution can direct detailed biochemical and genetic studies towards those taxa most likely to possess novel physiological diversity. This approach is especially important in bioprospecting for pathways having biotechnological relevance, such as carotenoid biosynthesis. For

example, novel carotenoids may be most likely discovered in lineages in which horizontal gene transfer and *de novo* biosynthetic evolution are prevalent, compared to those lineages in which evolution is constrained by co-evolution with other proteins. Additionally, evidence for positive selection may identify enzymes having altered specificity, which are likely excellent candidates for further study.

6.5. Conclusions

Previous attempts to identify the evolutionary history and diversity of carotenoid biosynthesis in its totality have been restricted to those proteins which have been studied genetically or biochemically (Sandmann 2002a, Phadwal 2005). In contrast, the present comparative genomics analysis identified several major phylogenetic lineages of carotenoid biosynthetic proteins absent or unrecognized in previous syntheses and for the first time rigorously defined the relationships between them whenever possible. This previously lacking phylogenetic framework provides an alternative view of carotenoid diversity to those based on structural and/or biosynthetic information. Testing the previously developed hypothesis of tree-like carotenoid biosynthetic evolution (Umeno et al. 2005) against this novel phylogenetic framework suggested that, rather than a linear progression of increasing diversity from root to tip building on previous innovations, the evolution of carotenoid biosynthesis is much more web-like with an increasing propensity for non-vertical modes of evolution increasing towards pathway termini. Furthermore, the degree to and mechanisms by which pathway diversification occurred was variable between phylogenetic lineages, constrained in some by interactions with other biochemical structures and fostered in others by more frequent occurrences of positive selection, horizontal gene transfer and differential gene gain and/or loss. Understanding the taxonomic distribution of these processes may particularly enhance the efficiency of bioprospecting for novel carotenoids by predicting which taxa are most likely to contain novel structural diversity due to their evolutionary tendencies towards pathway diversification.

6.6. Literature cited

- Albrecht, M., Takaichi, S., Steiger, S., Wang, Z.-Y., and Sandmann, G. 2000. Novel hydroxycarotenoids with improved antioxidative properties produced by gene combination in *Escherichia coli*. *Nat. Biotechnol.*, 18: 843-846.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.*, 215: 403-410.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, 25: 3389-3402.
- Auldridge, M.E., McCarty, D.R., and Klee, H.J. 2006. Plant carotenoid cleavage oxygenases and their apocarotenoid products. *Curr. Opin. Plant Biol.*, 9: 315-321.
- Britton, G. 1995. Structure and properties of carotenoids in relation to function. *FASEB J.*, 9: 1551-1558.
- Britton, G., Liaaen-Jensen, S., and Pfander, H. 2004. *Carotenoids handbook* Birkhäuser Verlag, Basel, Switzerland.
- Bryant, D.A., Garcia Costas, A.M., Maresca, J.A., Chew, A.G.M., Klatt, C.G., Bateson, M.M., Tallon, L.J., Hostetler, J., Nelson, W.C., Heidelberg, J.F., and Ward, D.M. 2007. *Candidatus Chloracidobacterium thermophilum*: an aerobic phototrophic acidobacterium. *Science*, 317: 523-526.
- Caetano-Anollés, G., Shin Kim, H., and Mittenthal, J.E. 2007. The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture. *Proc. Nat. Acad. Sci. USA*, 104: 9358-9363.
- Cheng, Q. 2006. Structural diversity and functional novelty of new carotenoid biosynthesis genes. *J. Ind. Microbiol. Biotechnol.*, 33: 552-559.
- Das, A., Yoon, S.-H., Lee, S.-H., Kim, J.-Y., Oh, D.-K., and Kim, S.-W. 2007. An update on microbial carotenoid production: application of recent metabolic engineering tools. *Appl. Microbiol. Biotechnol.*, 77: 505-512.
- Del Campo, J.A., García-González, M., and Guerrero, M.G. 2007. Outdoor cultivation of microalgae for carotenoid production: current state and perspectives. *Appl. Microbiol. Biotechnol.*, 74: 1163-1174.
- Feist, A.M., Herrgård, M.J., Thiele, I., Reed, J.L., and Palsson, B.Ø. 2009.

- Reconstruction of biochemical networks in microorganisms. *Nat. Rev. Microbiol.*, 7: 129-143.
- Felsenstein, J. 1989. PHYLIP: phylogeny inference package (version 3.2). *Cladistics*, 5: 164-166.
- Frank, H.A., and Cogdell, R.J. 1996. Carotenoids in photosynthesis. *Photochem. Photobiol.*, 63: 257-264.
- Frank, H.A., and Brudvig, G.W. 2004. Redox functions of carotenoids in photosynthesis. *Biochemistry*, 43: 8607-8615.
- Fraser, N.J., Hashimoto, H., and Cogdell, R.J. 2001. Carotenoids and bacterial photosynthesis: the story so far... *Photosynth. Res.*, 70: 249-256.
- Fraser, P.D., and Bramley, P.M. 2004. The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.*, 43: 228-265.
- Frommolt, R., Werner, S., Paulsen, H., Goss, R., Wilhelm, C., Zauner, S., Maier, U.G., Grossman, A.R., Bhattacharya, D., and Lohr, M. 2008. Ancient recruitment by chromists of green algal genes encoding enzymes for carotenoid biosynthesis. *Mol. Biol. Evol.*, 25: 2653-2667.
- Fuhrman, J.A., Schwalbach, M.S., and Stingl, U. 2008. Proteorhodopsins: an array of physiological roles? *Nat. Rev. Microbiol.*, 6: 488-494.
- Goldman, B.S., Nierman, W.C., Kaiser, D., Slater, S.C., Durkin, A.S., Eisen, J.A., Ronning, C.M., Barbazuk, W.B., Blanchard, M., Field, C., Halling, C., Hinkle, G., Iartchuk, O., Kim, H.S., Mackenzie, C., Madupu, R., Miller, N., Shvartsbeyn, A., Sullivan, S.A., Vaudin, M., Wiegand, R., and Kaplan, H.B. 2006. Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Nat. Acad. Sci. USA*, 103: 15200-15205.
- Gruszecki, W.I., and Strzalka, K. 2005. Carotenoids as modulators of lipid membrane physical properties. *Biochim. Biophys. Acta*, 1740: 108-115.
- Hurst, L.D. 2002. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.*, 18: 486-487.
- Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., and Nagashima, K.V.P. 2001. Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.*, 52: 333-341.
- Iniesta, A.A., Cervantes, M., and Murillo, F.J. 2007. Cooperation of two carotene desaturases in the production of lycopene in *Myxococcus xanthus*. *FEBS*

J., 274: 4306-4314.

- Jain, R., Rivera, M.C., and Lake, J.A. 1999. Horizontal gene transfer among genomes: the complexity hypothesis. *Proc. Nat. Acad. Sci. USA*, 96: 3801-3806.
- Jordan, P., Fromme, P., Witt, H.T., Klukas, O., Saenger, W., and Krauß, N. 2001. Three-dimensional structure of cyanobacterial photosystem I at 2.5 Å resolution. *Nature*, 411: 909-917.
- Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.
- Krinsky, N.I., and Johnson, E.J. 2005. Carotenoid actions and their relation to health and disease. *Mol. Aspects Med.*, 26: 459-516.
- Kupisz, K., Sujak, A., Patyra, M., Trebacz, K., and Gruszecki, W.I. 2008. Can membrane-bound carotenoid pigment zeaxanthin carry out a transmembrane proton transfer? *Biochim. Biophys. Acta*, 1778: 2334-2340.
- Lanyi, J.K., and Balashov, S.P. 2008. Xanthorhodopsin: a bacteriorhodopsin-like proton pump with a carotenoid antenna. *Biochim. Biophys. Acta*, 1777: 684-688.
- Liu, G.Y., Essex, A., Buchanan, J.T., Datta, V., Hoffman, H.M., Bastian, J.F., Fierer, J., and Nizet, V. 2005. *Staphylococcus aureus* golden pigment impairs neutrophil killing and promotes virulence through its antioxidant activity. *J. Exp. Med.*, 202: 209-215.
- Loll, B., Kern, J., Saenger, W., Zouni, A., and Biesiadka, J. 2005. Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. *Nature*, 438: 1040-1044.
- Lu, Y., and Rausher, M.D. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol. Biol. Evol.*, 20: 1844-1853.
- Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Krylov, D., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Lu, S., Marchler, G.H., Mullokandov, M., Song, J.S., Thanki, N., Yamashita, R.A., Yin, J.J., Zhang, D., and Bryant, S.H. 2007. CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.*, 35: D237-D240.

- Maresca, J.A., Graham, J.E., and Bryant, D.A. 2008a. The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. *Photosynth. Res.*, 97: 121-140.
- Maresca, J.A., Romberger, S.P., and Bryant, D.A. 2008b. Isorenieratene biosynthesis in green sulfur bacteria requires the cooperative actions of two carotenoid cyclases. *J. Bacteriol.*, 190: 6384-6391.
- Markowitz, V.M., Szeto, E., Palaniappan, K., Grechkin, Y., Chu, K., Chen, I.-M.A., Dubchak, I., Anderson, I., Lykidis, A., Mavromatis, K., Ivanova, N.N., and Kyrpides, N.C. 2007. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res.*, 36: D528-D533.
- McCarren, J., and DeLong, E.F. 2007. Proteorhodopsin photosystem gene clusters exhibit co-evolutionary trends and shared ancestry among diverse marine microbial phyla. *Environ. Microbiol.*, 9: 846-858.
- Monogodin, E.F., Nelson, K.E., Daugherty, S., DeBoy, R.T., Wister, J., Khouri, H., Weidman, J., Walsh, D.A., Papke, R.T., Sanchez Perez, G., Sharma, A.K., Nesbø, C.L., MacLeod, D., Baptiste, E., Doolittle, W.F., Charlebois, R.L., Legault, B., and Rodriguez-Valera, F. 2005. The genome of *Salinibacter ruber*: convergence and gene exchange among hyperhalophilic bacteria and archaea. *Proc. Nat. Acad. Sci. USA*, 102: 18147-18152.
- Mortensen, A. 2006. Carotenoids and other pigments as natural colorants. *Pure Appl. Chem.*, 78: 1477-1491.
- Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., Komemushi, S., Miki, W., and Misawa, N. 2005. Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'- β -hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanthophylls. *Appl. Environ. Microbiol.*, 71: 4286-4296.
- Novichkov, P.S., Wolf, Y.I., Dubchak, I., and Koonin, E.V. 2009. Trends in prokaryotic evolution revealed by comparison of closely related bacterial and archaeal genomes. *J. Bacteriol.*, 191: 65-73.
- Phadwal, K. 2005. Carotenoid biosynthetic pathway: molecular phylogenies and evolutionary behavior of *crt* genes in eubacteria. *Gene*, 345: 35-43.
- Rao, A.V., and Rao, L.G. 2007. Carotenoids and human health. *Pharmacol. Res.*, 55: 207-216.

- Rausher, M.D., Miller, R.E., and Tiffin, P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.*, 16: 266-274.
- Sandmann, G. 2002a. Molecular evolution of carotenoid biosynthesis from bacteria to plants. *Physiol. Plant.*, 116: 431-440.
- Sandmann, G. 2002b. Combinatorial biosynthesis of carotenoids in a heterologous host: a powerful approach for the biosynthesis of novel structures. *ChemBioChem*, 3: 629-635.
- Schmidt-Dannert, C. 2000. Engineering novel carotenoids in microorganisms. *Curr. Opin. Biotechnol.*, 11: 255-261.
- Sharma, A.K., Spudich, J.L., and Doolittle, W.F. 2006. Microbial rhodopsins: functional versatility and genetic mobility. *Trends Microbiol.*, 14: 463-469.
- Shindo, K., Asagi, E., Sano, A., Hotta, E., Minemura, N., Mikami, K., Tamesada, E., Misawa, N., and Moaka, T. 2008. Diapolycopene acid xylosyl esters A, B, and C, novel antioxidative glyco-C₃₀-carotenoic acids produced by a new marine bacterium *Rubritalea squalenifaciens*. *J. Antibiot.*, 61: 185-191.
- Sieiro, C., Poza, M., de Miguel, T., and Villa, T.G. 2003. Genetic basis of microbial carotenogenesis. *Int. Microbiol.*, 6: 11-16.
- Spudich, J.L., Yang, C.-S., Jung, K.-H., and Spudich, E.N. 2000. Retinylidene proteins: structures and functions from archaea to humans. *Annu. Rev. Cell Dev. Biol.*, 16: 365-392.
- Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22: 2688-2690.
- Stamatakis, A., Hoover, P., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, 57: 758-771.
- Steiger, S., Mazet, A., and Sandmann, G. 2003. Heterologous expression, purification, and enzymatic characterization of the acyclic carotenoid 1,2-hydratase from *Rubrivivax gelatinosus*. *Arch. Biochem. Biophys.*, 414: 51-58.
- Stickforth, P., and Sandmann, G. 2007. Kinetic variations determine the product pattern of phytoene desaturase from *Rubrivivax gelatinosus*. *Arch. Biochem. Biophys.*, 461: 235-241.

- Stinear, T.P., Seemann, T., Pidot, S., Frigui, W., Reysset, G., Garnier, T., Meurice, G., Simon, D., Bouchier, C., Ma, L., Tichit, M., Porter, J.L., Ryan, J., Johnson, P.D.R., Davis, J.K., Jenkin, G.A., Small, P.L.C., Jones, L.M., Tekaiia, F., Laval, R., Daffé, M., Parkhill, J., and Cole, S.T. 2007. Reductive evolution and niche adaptation inferred from the genome of *Mycobacterium ulcerans*, the causative agent of Buruli ulcer. *Genome Res.*, 17: 192-200.
- Takaichi, S. 1999. Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. *In* The photochemistry of carotenoids. H.A. Frank, A.J. Young, G. Britton, and R.J. Cogdell. (ed.) Kluwer Academic Publishers, New York, NY. pp. 39-69.
- Takaichi, S., and Mochimaru, M. 2007. Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. *Cell. Mol. Life Sci.*, 64: 2607-2619.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. *MEGA4*: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24: 1596-1599.
- Tanaka, Y., Sasaki, N., and Ohmiya, A. 2008. Biosynthesis of plant pigments: anthocyanins, betalins and carotenoids. *Plant J.*, 54: 733-749.
- Tao, L., Yao, H., Kasai, H., Misawa, N., and Cheng, Q. 2006. A carotenoid synthesis gene cluster from *Algoriphagus* sp. KK10202C with a novel fusion-type lycopene β -cyclase gene. *Mol. Genet. Genomics*, 276: 79-86.
- Teramoto, M., Takaichi, S., Inomata, Y., Ikenaga, H., and Misawa, N. 2003. Structural and functional analysis of a lycopene β -monocyclase gene isolated from a unique marine bacterium that produces myxol. *FEBS Lett.*, 545: 120-126.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, 22: 4673-4680.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25: 4876-4882.
- Tian, B., Xu, Z., Sun, Z., Lin, J., and Hua, Y. 2007. Evaluation of the antioxidant effects of carotenoids from *Deinococcus radiodurans* through targeted

- mutagenesis, chemiluminescence, and DNA damage analyses. *Biochim. Biophys. Acta*, 1770: 902-911.
- Umeno, D., Tobias, A.V., and Arnold, F.H. 2005. Diversifying carotenoid biosynthetic pathways by directed evolution. *Microbiol. Mol. Biol. Rev.*, 69: 51-78.
- Vos, M., and Didelot, X. 2009. A comparison of homologous recombination rates in bacteria and archaea. *ISME J.*, 3: 199-208.
- Wang, F., Jiang, J.G., and Chen, Q. 2007. Progress on molecular breeding and metabolic engineering of biosynthesis pathways of C₃₀, C₃₅, C₄₀, C₄₅, C₅₀ carotenoids. *Biotechnol. Adv.*, 25: 211-222.
- Wellner, A., Lurie, M.N., and Gophna, U. 2007. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. *Genome Biol.*, 8: R156.
- Woese, C.R. 1998. The universal ancestor. *Proc. Nat. Acad. Sci. USA*, 95: 6854-6859.
- Worobey, M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. *Mol. Biol. Evol.*, 18: 1425-1434.
- Xiong, J., Fischer, W.M., Inoue, K., Nakahara, M., and Bauer, C.E. 2000. Molecular evidence for the early evolution of photosynthesis. *Science*, 289: 1724-1730.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, 24: 1586-1591.
- Zhang, L., Yang, Q., Luo, X., Fang, C., Zhang, Q., and Tang, Y. 2007. Knockout of *crtB* or *crtI* gene blocks the carotenoid biosynthetic pathway in *Deinococcus radiodurans* R1 and influences its resistance to oxidative DNA-damaging agents due to change of free radicals scavenging ability. *Arch. Microbiol.*, 188: 411-419.

7. Pathway Evolution by Horizontal Transfer and Positive Selection is Accommodated by Relaxed Negative Selection Upon Upstream Pathway Genes in Purple Bacterial Carotenoid Biosynthesis³

7.1. Introduction

Biochemical pathway evolution has been extensively examined, particularly regarding mechanisms by which novel functions can be generated, diversified and maintained (Caetano-Anollés et al. 2009, Fani and Fondi 2009). Best studied in this regard is the role of gene duplication followed by divergence, resulting in paralog families which, despite sharing a common evolutionary ancestor, possess different functions (Conant and Wolfe 2008). In clonally reproducing organisms such as bacteria and archaea this type of diversification is further compounded by horizontal gene transfer (Gogarten and Townsend 2005), whereby a divergent ortholog from one organism is introduced into the metabolic network of another thereby becoming a “xenolog” (Koonin et al. 2001). Horizontal gene transfer between distantly related organisms is especially diversifying due to the likelihood of altering the genome structure or biochemical and regulatory networks of the recipient, in contrast to recombination between close relatives which may promote genetic cohesion (Lawrence and Retchless 2009).

Selection controls phenotypic diversity as a function of evolutionary fitness. Three scenarios can be detected from patterns of nucleotide substitutions (Hurst 2002): (i) positive selection, by which advantageous functionally divergent mutants are further optimized by increased mutational sampling of phenotypic space; (ii) negative (purifying) selection, by which deleterious mutations are purged; and (iii) neutral mutation, in which mutations accumulate that do not affect the selected phenotype, resulting in genetic drift. Considering horizontal transfer, selection will favor fixation of a horizontally transferred gene if it is

³A version of this chapter has been accepted for publication. 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.* Accepted September 25, 2009.

advantageous and disfavor it when either the gene product or the alterations that it causes in the host network are deleterious. Successful horizontal gene transfer resulting in gene fixation is the result of *net* evolutionary benefit towards the host, both due to the horizontally transferred gene itself and minimal sub-optimal alteration of the host metabolic and genetic networks into which it is integrating. Genetic parasites such as plasmids, transposons and integrated phages are exceptional in directly promoting their own retention.

In this study, the evolution of carotenoid biosynthesis by horizontal gene transfer and selection is evaluated in the purple bacteria (anoxygenic phototrophic *Proteobacteria* which can use reduced sulfur compounds as electron sources) for which carotenoid pathway diversity has been studied previously (Section 6.3.4). Whereas phylogenetically most purple bacteria belong the α -*Proteobacteria*, some belonging to the β - and γ -*Proteobacteria* have evolved by horizontal transfer of α -proteobacterial photosynthetic super-operons, which include carotenoid biosynthetic genes (Nagashima et al. 1997, Igarashi et al. 2001). Carotenoid biosynthesis in the purple bacteria (Figure 7.1) begins with the condensation of two molecules of geranylgeranyl pyrophosphate by the phytoene synthase CrtB, forming phytoene. The phytoene desaturase CrtI then desaturates phytoene either three or four times producing neurosporene or lycopene, respectively. Both of these intermediates are subsequently hydroxylated at the 1-position by the hydroxylase CrtC, desaturated at the 3- and 4-positions by the CrtI homolog CrtD, methylated at the 1-hydroxyl group by the methyltransferase CrtF and, in spheroidenone and 2,2'-diketospirilloxanthin-producing organisms, ketolated at the 2-position by the ketolase CrtA (Figure 7.1). Considerable sub-pathway diversity also exists because of the potential for asymmetry between carotenoid ends.

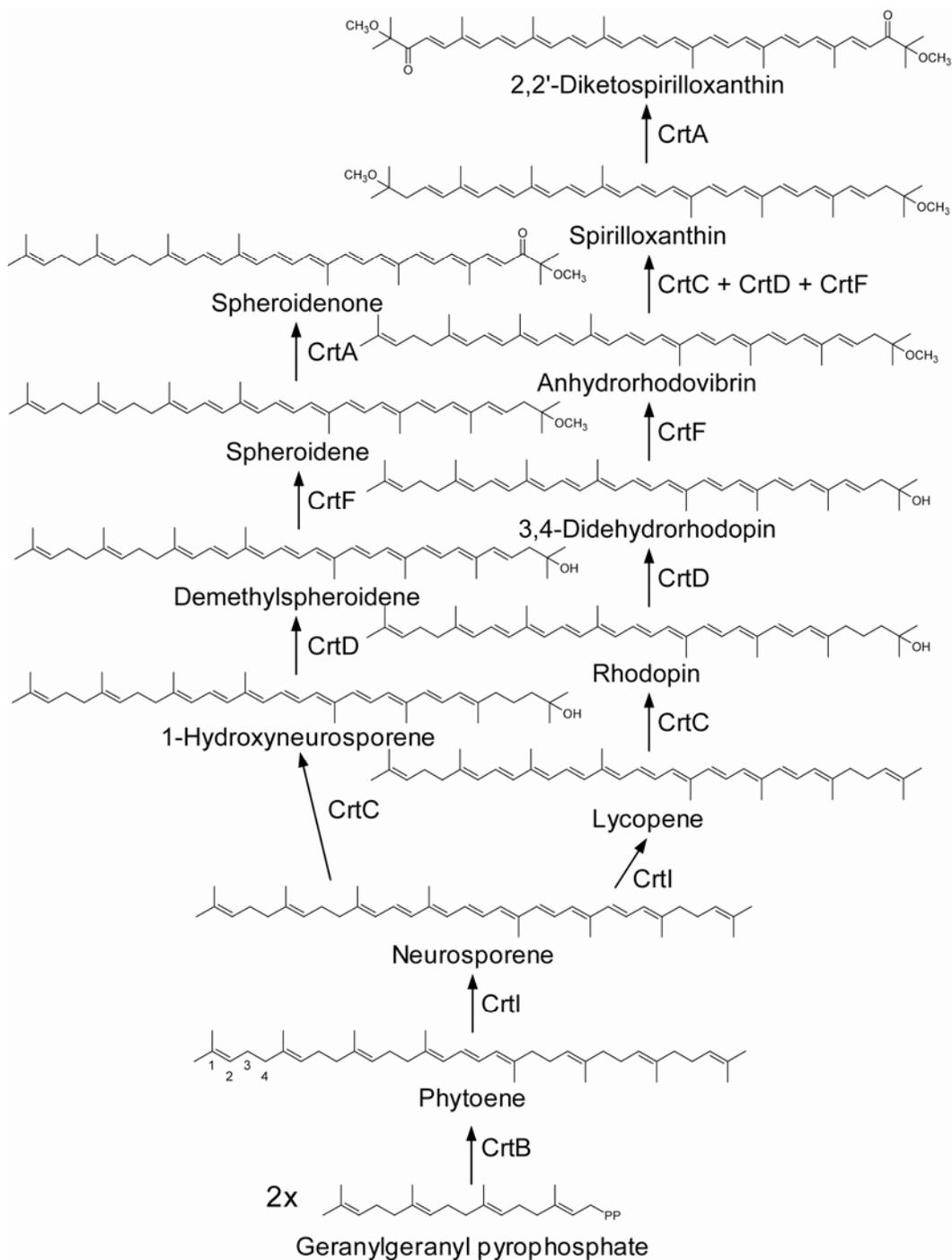


Figure 7.1. The carotenoid biosynthetic pathway in purple bacteria. For simplicity not all sub-pathways are shown. Carbon numbers for the ψ end-group are shown for phytoene.

Carotenoid biosynthesis has been well studied biochemically and genetically in the spheroidene-producing *Rhodobacter* (hereafter *Rb.*) *capsulatus* (Armstrong et al. 1989, Raisig et al. 1996, Badenhop et al. 2003, Gerjets et al. 2009) and *Rb. sphaeroides* (Lang et al. 1994, Lang et al. 1995, Albrecht et al. 1997) and the spirilloxanthin-producing *Bradyrhizobium* sp. ORS278 (Giraud et al. 2004), *Thiocapsa roseopersicina* (Kovács et al. 2003) and *Rubrivivax* (hereafter *Rv.*) *gelatinosus* (Ouchane et al. 1997a, Ouchane et al. 1997b, Steiger et al. 2000, Harada et al. 2001a, Harada et al. 2001b, Pinta et al. 2003, Steiger et al. 2003, Stickforth and Sandmann 2007, Gerjets et al. 2009). This latter microorganism produces 2,2'-diketospirilloxanthin (lycopene-derived), spheroidenone (neurosporene-derived) and their precursors using the same enzymes. Pathway utilization in *Rv. gelatinosus* is determined primarily by the substrate specificities of CrtC, CrtD and CrtI and the rate of metabolic flux (Steiger et al. 2000, Steiger et al. 2003, Stickforth and Sandmann 2007). These extensive biochemical and genetic studies provide a solid framework for sequence-based evolutionary analyses, making this pathway a valuable model to study biochemical pathway evolution.

7.2. Materials and methods

All sequences were obtained from GenBank and the Integrated Microbial Genome database as described previously (Section 6.2.2). Nucleotide sequences were aligned as translated proteins using CLUSTALW (Thompson et al. 1994) as implemented in MEGA 4.0 (Tamura et al. 2007) and indel regions were removed. Maximum likelihood protein phylogenetic trees were generated using the JTT substitution matrix and estimated proportions of invariable sites and bootstrap replicates using RAxML (Stamatakis et al. 2008) as implemented through the CIPRES web portal (<http://www.phylo.org/portal/Home.do>). The phylogenetic trees obtained were congruent with those presented previously (Chapter 6), including those generated using parsimony and neighbor-joining methods.

To determine the effect of selection on purple bacterial carotenoid biosynthetic genes, non-synonymous (d_n) and synonymous (d_s) mutation rates

were calculated in MEGA 4.0 (Tamura et al. 2007) using the Nei-Gojobori (Nei and Gojobori 1986) with Jukes-Cantor correction, Li-Wu-Lou (Li et al. 1985), Pamilo-Bianchi-Li (Li 1993, Pamilo and Bianchi 1993) and Kumar (Nei and Kumar 2000) methods and the above described sequence alignments, excluding all gapped positions and the truncated *crtB* and *crtI* sequences from the draft *Magnetospirillum magnetotacticum* MS-1 genome. Pair-wise estimations of d_n and d_s were used in preference to tree-based methods (such as PAML; Yang 2007) due to poor bootstrap support (typically $< 60\%$) at the deepest nodes in the phylogenies of most analyzed genes (Supplemental Figure E1). Pair-wise d_n and d_s estimations for *Rv. gelatinosus*, *H. phototrophica* and all other spirilloxanthin-producers were separately aggregated excluding pair-wise comparisons between *Rv. gelatinosus* and *H. phototrophica* and compared using SPSS v15.0. Only pair-wise d_n and d_s estimations yielding values of $d_n \geq 0.01$ and $d_s \leq 1.5$ were considered, so as to allow for sufficient mutational signal and to avoid the effect of back mutations which would artificially increase d_s with increased sequence divergence, respectively (Hurst 2002). Statistical comparison of all values or those with $d_s \leq 2$ did not substantially alter the conclusions drawn here (data not shown). Note that these cutoffs minimize but do not eliminate the possibility of back mutation. However, because elevated d_s estimations due to back mutation result in the overestimation of negative selection, the identification of positive selection and relaxed negative selection is more conservative; the detection of these processes reported here is therefore not invalidated by this effect.

7.3. Results

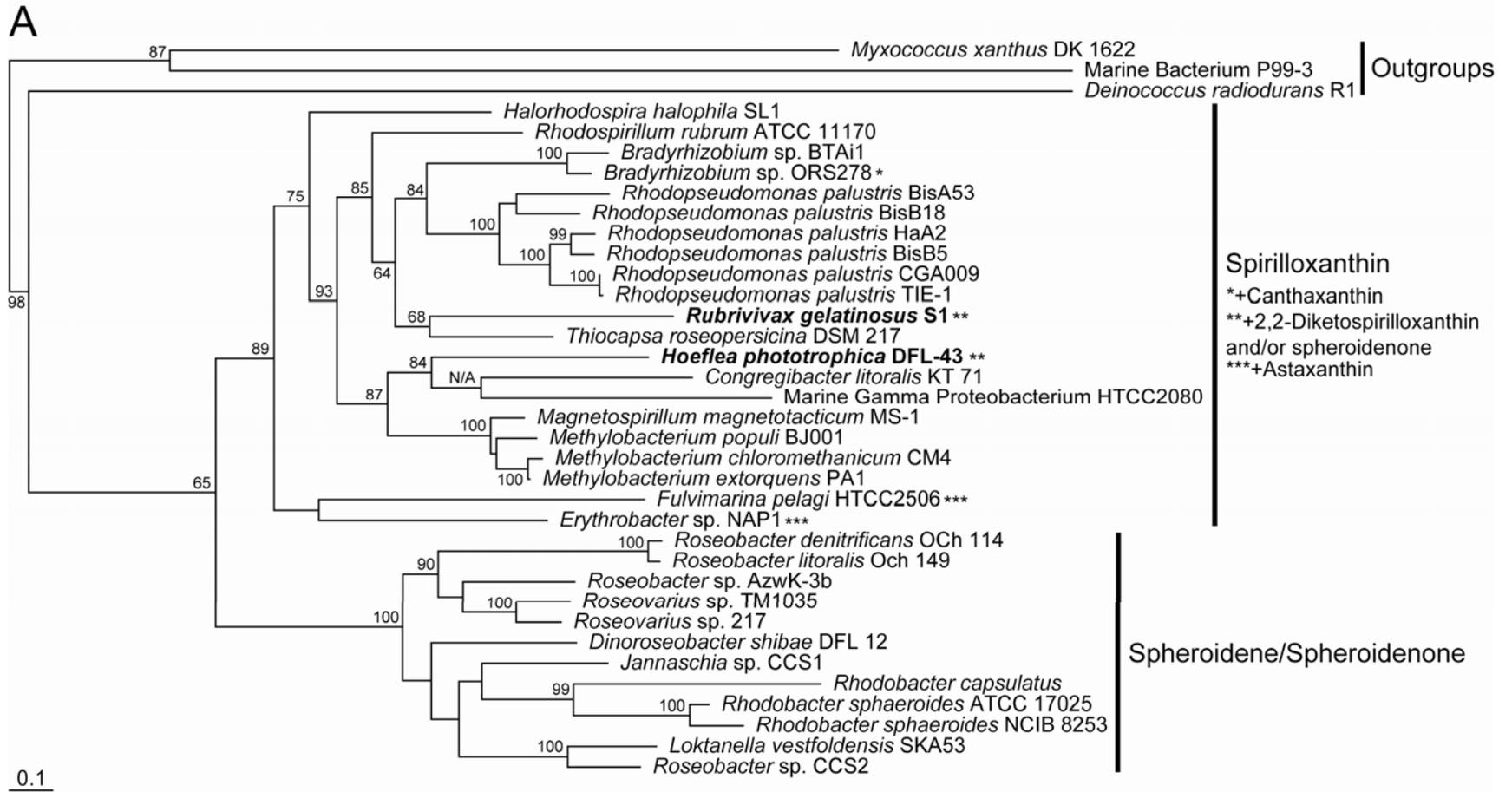
7.3.1. Identification of carotenoid biosynthetic pathway distribution in the purple bacteria and horizontally transferred carotenoid biosynthesis genes

Phylogenetic analyses of purple bacteria carotenoid biosynthetic protein sequences and their inferred and known pathway products (Figure 7.2, Supplemental Figure E1) indicate two major phylogenetic lineages, one producing spheroidenone and the other spirilloxanthin (see also Section 6.3.4). Note that in

all cases precursor carotenoids also accumulate; pathways are discussed here in terms of their end-products for simplicity. Spheroidenone production resulted from the acquisition of the ketolase CrtA in the phylogenetic lineage containing *Rhodobacter* and the aerobic anoxygenic photosynthetic bacteria. The lack of spirilloxanthin production (Takaichi 1999) and fixation of *crtA* in all studied organisms from this lineage suggests that spheroidenone production may confer to them a selective advantage relative to that of spirilloxanthin.

Two exceptions to the above phylogenetic scheme are *Rv. gelatinosus*, which produces both spheroidenone and 2,2'-diketospirilloxanthin (Takaichi 1999), and *H. phototrophica*, which presumptively produces spheroidenone, as inferred from UV-vis spectroscopy of whole cell extracts (Biebl et al. 2006). Carotenoid biosynthetic proteins from both of these organisms cluster most strongly with spirilloxanthin-producing purple bacteria (Figure 7.2). Whereas some phylogenies for other carotenoid biosynthetic proteins indicate branching of *H. phototrophica* at the base of the spheroidenone-producing clade with poor bootstrap support (Supplemental Figure E1), CrtA phylogenies and patterns of selection suggest the recent evolution of spheroidenone biosynthesis in this microorganism (see below). The lack of 2,2'-diketospirilloxanthin and/or spheroidenone production in the phylogenetic neighbors of *Rv. gelatinosus* and *H. phototrophica* suggests that these pathways evolved relatively recently and uniquely in these phylogenetic lineages, independent of and likely subsequent to the acquisition of *crtA* by the major spheroidenone-producing lineage.

In some instances, carotenoid biosynthetic pathway distribution in the purple bacteria differs from that described above e.g., the production of astaxanthin and canthaxanthin by *Erythrobacter* sp. NAP1 (Kobližek et al. 2003) and *Bradyrhizobium* sp. ORS278 (Takaichi 1999), respectively. Production of these carotenoids involves separate gene clusters (Hannibal et al. 2000, Giraud et al. 2004) and therefore different evolutionary histories independent of those producing spirilloxanthin and are not considered further here.



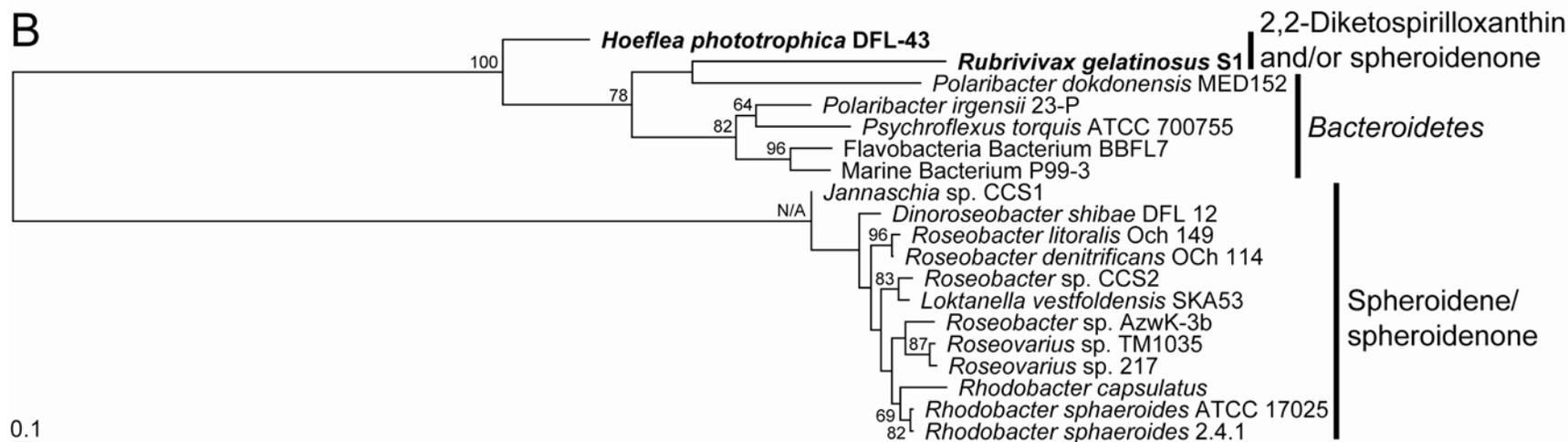


Figure 7.2. (See also previous page) RAxML phylogenetic tree of (A) CrtD and (B) CrtA amino acid sequences in purple bacteria and (for CrtA) the *Bacteroidetes*. Bootstrap values are expressed as a percentage with only those ≥ 60 are shown. The scale bar indicates 10% sequence divergence. Both trees were rooted to their midpoints, and outgroups for (A) were selected based on previous analyses (Chapter 6). Carotenoids produced for each strain in (A) were inferred from the published literature and previous phylogenetic analyses (Section 6.3.4); exceptions to broad groupings are indicated using asterisks. *Rv. gelatinosus* and *H. phototrophica* are bolded for clarity. See Supplemental Figure E1 for phylogenies of other carotenoid biosynthesis proteins in purple bacteria.

Phylogenetic analysis of CrtA protein sequences (Figure 7.2B) clearly indicates horizontal transfer from the *Bacteroidetes* into *Rv. gelatinosus*. The only characterized *Bacteroidetes crtA* gene encodes a 2'-hydroxylase (Rählert et al. 2009), and carotenoids with 2'-hydroxylated but not 2'-ketolated ψ -end groups are widespread in this phylum (Chapter 5; Klassen et al. 2009 and references within). It is most parsimonious, therefore, to assume that all *Bacteroidetes*-like CrtA enzymes function as hydroxylases, not ketolases as in the purple bacteria, and that this was the ancestral function of the sequence transferred into *Rv. gelatinosus*. Because the *H. phototrophica* CrtA sequence also clusters with the *Bacteroidetes* (Figure 7.2B) and phylogenies of other *H. phototrophica* carotenoid biosynthetic genes suggest a closer relationship to spirilloxanthin- versus spheroidenone-producing microorganisms (Figure 7.2A, Supplemental Figure E1), it is most reasonable to conclude that the *H. phototrophica crtA* gene was also horizontally transferred. Because *H. phototrophica* and *Rv. gelatinosus* are not phylogenetic neighbors for any analyzed protein, horizontal transfer of *crtA* presumably occurred independently in each lineage. The origin and direction of horizontal transfer of *crtA* are more obscure for *H. phototrophica*; this analysis cannot determine whether *crtA* ancestrally functioned as a ketolase from which CrtA-type hydroxylases arose in the *Bacteroidetes* or will cluster internally to other *Bacteroidetes* CrtA hydroxylase sequences when they become available, suggesting ancestral hydroxylase function.

7.3.2. Horizontally transferred *crtA* genes in *Rv. gelatinosus* and *H. phototrophica* are under positive selection

According to phylogenetic analyses, production of spheroidenone by horizontally transferred *crtA* genes, at least one of which most likely functioned ancestrally as a hydroxylase instead of a ketolase, is evolutionary favorable. The rates of non-synonymous and synonymous codon substitutions (d_n and d_s , respectively) were therefore estimated for these genes and their homologs by pairwise estimation using four different estimation methods (Figure 7.3, Supplemental Figure E2; trend lines are shown for d_n/d_s ratios of 1 and 0.2 to

highlight those values indicating positive and strong negative selection, respectively). Unfortunately, the high levels of divergence among sequences restricted the number of calculable d_n and d_s values, particularly for *Bacteroidetes crtA* genes. Those *Bacteroidetes* pair-wise comparisons which could be determined, however, indicated strong negative selection ($d_n/d_s \approx 0.2$) except for *Polaribacter dokdonensis* for which $d_n/d_s \approx 0.5$; this latter value may indicate

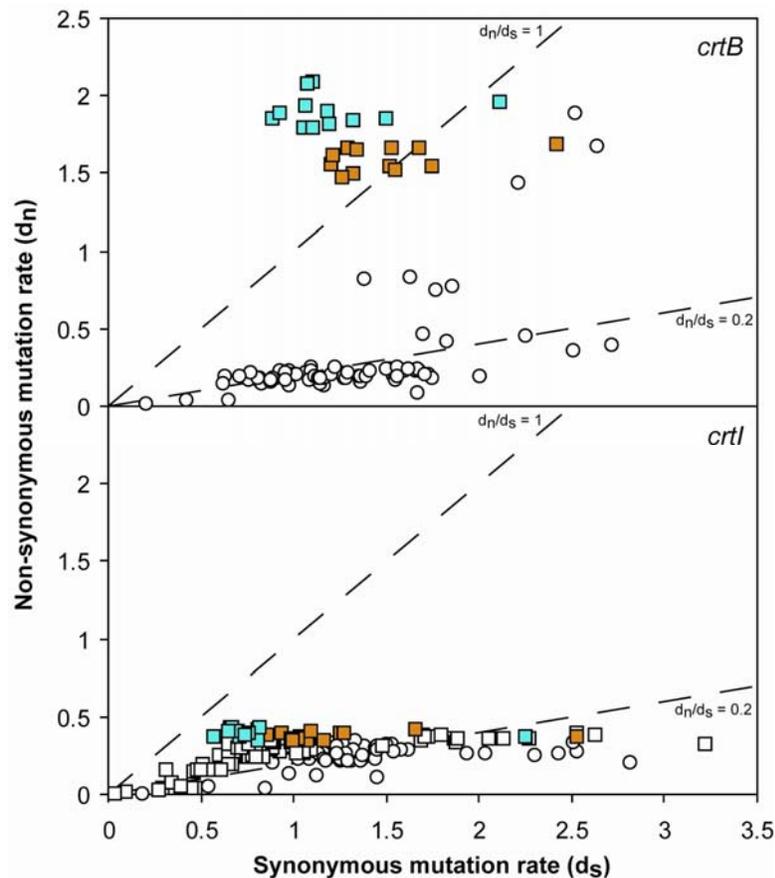


Figure 7.3. Plots of pair-wise estimated d_n and d_s substitution rates for primarily spheroidenone-producing bacteria (circles) and primarily spirilloxanthin-producing bacteria (squares). Values from comparisons involving *H. phototrophica* and *Rv. gelatinosus* are indicated in orange and blue, respectively. Also shown are trend lines for d_n/d_s ratios of 1 and 0.2. Plots for *crtA* and *crtI* were constructed using the Nei-Gojobori method with the Jukes-Cantor correction; for plots generated using this method for *crtB*, *crtC*, *crtD* and *crtF* and those using other algorithms see Supplemental Figure E2. See also summary statistics in Table 7.1.

Table 7.1. Mean d_n/d_s substitution ratios with associated standard deviations and comparison to spirilloxanthin-producing bacteria (excluding *Rv. gelatinosus* and *H. phototrophica*), or to spheroidenone-producing bacteria for *crtA*, using the Mann-Whitney U test. Only pair-wise estimations with $d_n \geq 0.01$ and $d_s \leq 1.5$ were considered. Means \pm one standard deviation and *P* values < 0.1 are bolded for clarity.

Strain	Statistical model			
	Nei-Gojobori with Jukes-Cantor correction	Li-Wu-Luo	Pamilo-Bianchi-Li	Kumar
<i>crtA</i>				
<i>H. phototrophica</i>	1.23 \pm 0.08 , $n^a = 6$, <i>P</i> < 0.001 , <i>Z</i> = -3.987	1.11 \pm 0.13 , $n = 2$, <i>P</i> = 0.017 , <i>Z</i> = -2.386	1.30 , $n = 1$, <i>P</i> = 0.089 , <i>Z</i> = -1.701	1.11 \pm 0.07 , $n = 6$, <i>P</i> < 0.001 , <i>Z</i> = -4.043
<i>Rv. gelatinosus</i>	1.75 \pm 0.23 , $n = 10$, <i>P</i> < 0.001 , <i>Z</i> = -4.984	1.50 , $n = 1$, <i>P</i> = 0.089 , <i>Z</i> = -1.702	No values within cutoff range	1.64 \pm 0.21 , $n = 2$, <i>P</i> = 0.016 , <i>Z</i> = -2.398
All other strains	0.18 \pm 0.08 , $n = 53$	0.17 \pm 0.05 , $n = 56$	0.17 \pm 0.04 , $n = 55$	0.20 \pm 0.09 , $n = 69$
<i>crtB</i>				
<i>H. phototrophica</i>	0.48 \pm 0.07 , $n = 14$, <i>P</i> = 0.006 , <i>Z</i> = -2.771	0.41 \pm 0.05 , $n = 14$, <i>P</i> = 0.005 , <i>Z</i> = -2.822	0.40 \pm 0.06 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -3.674	0.46 \pm 0.06 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -4.061
<i>Rv. gelatinosus</i>	0.60 \pm 0.10 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -5.076	0.51 \pm 0.09 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -4.822	0.45 \pm 0.10 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -4.218	0.50 \pm 0.09 , $n = 14$, <i>P</i> < 0.001 , <i>Z</i> = -4.409
All other spirilloxanthin-producers	0.39 \pm 0.13 , $n = 92$	0.33 \pm 0.11 , $n = 84$	0.31 \pm 0.10 , $n = 78$	0.35 \pm 0.11 , $n = 94$
All other spheroidene-producers	0.23 \pm 0.07 , $n = 52$, <i>P</i> < 0.001 , <i>Z</i> = -6.942	0.22 \pm 0.06 , $n = 47$, <i>P</i> < 0.001 , <i>Z</i> = -6.104	0.21 \pm 0.05 , $n = 43$, <i>P</i> < 0.001 , <i>Z</i> = -5.757	0.23 \pm 0.06 , $n = 61$, <i>P</i> < 0.001 , <i>Z</i> = -7.190
<i>crtC</i>				
<i>H. phototrophica</i>	0.51 \pm 0.06 , $n = 17$, <i>P</i> = 0.684, <i>Z</i> = -0.407	0.47 \pm 0.05 , $n = 14$, <i>P</i> = 1.000, <i>Z</i> = -0.000	0.46 \pm 0.04 , $n = 14$, <i>P</i> = 0.339, <i>Z</i> = -0.956	0.53 \pm 0.06 , $n = 17$, <i>P</i> = 0.075 , <i>Z</i> = -1.780
<i>Rv. gelatinosus</i>	0.77 \pm 0.14 , $n = 18$, <i>P</i> < 0.001 , <i>Z</i> = -5.575	0.69 \pm 0.13 , $n = 18$, <i>P</i> < 0.001 , <i>Z</i> = -5.163	0.64 \pm 0.09 , $n = 16$, <i>P</i> < 0.001 , <i>Z</i> = -5.344	0.71 \pm 0.12 , $n = 18$, <i>P</i> < 0.001 , <i>Z</i> = -5.509
All other spirilloxanthin-producers	0.52 \pm 0.13 , $n = 145$	0.47 \pm 0.12 , $n = 121$	0.44 \pm 0.11 , $n = 108$	0.50 \pm 0.12 , $n = 143$
All other spheroidene-producers	0.31 \pm 0.13 , $n = 53$, <i>P</i> < 0.001 , <i>Z</i> = -8.287	0.28 \pm 0.11 , $n = 50$, <i>P</i> < 0.001 , <i>Z</i> = -7.906	0.27 \pm 0.10 , $n = 44$, <i>P</i> < 0.001 , <i>Z</i> = -7.402	0.30 \pm 0.10 , $n = 60$, <i>P</i> < 0.001 , <i>Z</i> = -9.012

Table 7.1 continued.

<i>crtD</i>				
<i>H. phototrophica</i>	0.40 ± 0.04 , n = 17, <i>P</i> = 0.975, <i>Z</i> = -0.032	0.39 ± 0.03 , n = 14, <i>P</i> = 0.192, <i>Z</i> = -1.305	0.38 ± 0.03 , n = 14, <i>P</i> = 0.026 , <i>Z</i> = -2.226	0.38 ± 0.04 , n = 17, <i>P</i> = 0.017 , <i>Z</i> = -2.393
<i>Rv. gelatinosus</i>	0.55 ± 0.09 , n = 18, <i>P</i> < 0.001 , <i>Z</i> = -5.278	0.49 ± 0.08 , n = 17, <i>P</i> < 0.001 , <i>Z</i> = -5.133	0.44 ± 0.07 , n = 17, <i>P</i> < 0.001 , <i>Z</i> = -4.831	0.48 ± 0.08 , n = 18, <i>P</i> < 0.001 , <i>Z</i> = -4.593
All other spirilloxanthin-producers	0.39 ± 0.11 , n = 137	0.35 ± 0.09 , n = 131	0.33 ± 0.08 , n = 120	0.37 ± 0.10 , n = 141
All other spheroidene-producers	0.32 ± 0.08 , n = 54, <i>P</i> < 0.001 , <i>Z</i> = -4.508	0.31 ± 0.07 , n = 46, <i>P</i> = 0.001 , <i>Z</i> = -3.418	0.29 ± 0.06 , n = 43, <i>P</i> = 0.003 , <i>Z</i> = -2.922	0.33 ± 0.07 , n = 62, <i>P</i> = 0.001 , <i>Z</i> = -3.453
<i>crtF</i>				
<i>H. phototrophica</i>	0.52 ± 0.06 , n = 15, <i>P</i> < 0.001 , <i>Z</i> = -4.535	0.46 ± 0.05 , n = 11, <i>P</i> < 0.001 , <i>Z</i> = -3.977	0.47 ± 0.04 , n = 8, <i>P</i> < 0.001 , <i>Z</i> = -3.951	0.50 ± 0.07 , n = 13, <i>P</i> < 0.001 , <i>Z</i> = -4.082
<i>Rv. gelatinosus</i>	0.51 ± 0.05 , n = 16, <i>P</i> < 0.001 , <i>Z</i> = -4.510	0.44 ± 0.04 , n = 16, <i>P</i> < 0.001 , <i>Z</i> = -4.011	0.40 ± 0.03 , n = 14, <i>P</i> = 0.007 , <i>Z</i> = -2.684	0.44 ± 0.04 , n = 16, <i>P</i> = 0.013 , <i>Z</i> = -2.474
All other spirilloxanthin-producers	0.41 ± 0.09 , n = 117	0.36 ± 0.08 , n = 100	0.35 ± 0.09 , n = 77	0.39 ± 0.09 , n = 115
All other spheroidene-producers	0.34 ± 0.13 , n = 36, <i>P</i> < 0.001 , <i>Z</i> = -4.138	0.32 ± 0.12 , n = 28, <i>P</i> = 0.013 , <i>Z</i> = -2.487	0.30 ± 0.10 , n = 24, <i>P</i> < 0.001 , <i>Z</i> = -2.465	0.32 ± 0.10 , n = 37, <i>P</i> < 0.001 , <i>Z</i> = -3.879
<i>crtI</i>				
<i>H. phototrophica</i>	0.36 ± 0.06 , n = 14, <i>P</i> = 0.682, <i>Z</i> = -0.410	0.32 ± 0.05 , n = 14, <i>P</i> = 0.885, <i>Z</i> = -0.144	0.31 ± 0.05 , n = 13, <i>P</i> = 0.533, <i>Z</i> = -0.624	0.35 ± 0.05 , n = 14, <i>P</i> = 0.268, <i>Z</i> = -1.107
<i>Rv. gelatinosus</i>	0.56 ± 0.07 , n = 14, <i>P</i> < 0.001 , <i>Z</i> = -5.914	0.47 ± 0.05 , n = 14, <i>P</i> < 0.001 , <i>Z</i> = -5.793	0.41 ± 0.04 , n = 14, <i>P</i> < 0.001 , <i>Z</i> = -5.618	0.46 ± 0.04 , n = 14, <i>P</i> < 0.001 , <i>Z</i> = -5.707
All other spirilloxanthin-producers	0.33 ± 0.09 , n = 90	0.30 ± 0.08 , n = 89	0.28 ± 0.07 , n = 91	0.32 ± 0.08 , n = 94
All other spheroidene-producers	0.21 ± 0.06 , n = 53, <i>P</i> < 0.001 , <i>Z</i> = -7.193	0.20 ± 0.06 , n = 46, <i>P</i> < 0.001 , <i>Z</i> = -6.815	0.20 ± 0.06 , n = 39, <i>P</i> < 0.001 , <i>Z</i> = -6.206	0.21 ± 0.06 , n = 59, <i>P</i> < 0.001 , <i>Z</i> = -7.444

^aNumber of pair-wise d_n/d_s estimations with $d_n \geq 0.01$ and $d_s \leq 1.5$

reduced negative selection in this lineage. This is congruent with the strong negative selection observed for other *Bacteroidetes* carotenoid biosynthetic genes (Section 6.3.10). By comparison, *Rv. gelatinosus* and *H. phototrophica crtA* genes are clearly under positive selection ($d_n/d_s > 1$; Table 7.1 and Figure 7.3). This observation is congruent with the proposed functional transition of CrtA from a hydroxylase to a ketolase in both *Rv. gelatinosus* and *H. phototrophica*, perhaps aided by relaxed negative selection while still in the *Bacteroidetes*. Unfortunately d_n and d_s could not be estimated in MEGA between *P. dokdonensis* and either *Rv. gelatinosus* or *H. phototrophica*. More related sequences are therefore required to determine the extent to which selection in *Bacteroidetes* contributed to the evolution of ketolase function in the CrtA proteins transferred into *Rv. gelatinosus* and *H. phototrophica*.

7.3.3. Other carotenoid biosynthetic genes in *Rv. gelatinosus* and *H. phototrophica* are under relaxed negative selection

In addition to *crtA*, all carotenoid biosynthetic protein-encoding genes from *Rv. gelatinosus* had elevated d_n/d_s ratios compared to all other primarily spirilloxanthin-producing organisms (excluding *H. phototrophica*), exhibiting a difference in d_n/d_s ratio ≈ 0.15 (Mann-Whitney U Test, $Z \geq -2.474$, $P \leq 0.013$; Table 7.1 and Figure 7.1). Similarly elevated d_n/d_s ratios were also clearly determined for the *H. phototrophica crtB* and *crtF* genes (Mann-Whitney U Test, $Z \geq -2.771$, $P \leq 0.006$ Table 1), and more equivocally (i.e., not under all models tested) for the *crtC* and *crtD* genes (Table 7.1). This lack of unequivocal statistical support may reflect the greater divergence of *H. phototrophica* sequences from the other spirilloxanthin-producers analyzed, resulting in falsely inflated d_s estimations due to back mutation (Figure 7.1). Visual inspection of the matrices generated did not reveal similarly elevated d_n/d_s ratios across all genes for any other analyzed strain (data not shown). In addition to carotenoid biosynthetic genes from *Rv. gelatinosus* and *H. phototrophica*, all genes from spheroidenone-producing microorganisms had lower d_n/d_s ratios compared with all other primarily spirilloxanthin-producing microorganisms (Mann-Whitney U

Test, $Z \geq -2.487$, $P \leq 0.013$ for all genes and algorithms; see Table 7.1 and compare circles to squares in Figure 7.3). The degree to which this is due to greater divergence amongst spirilloxanthin-producing versus spheroidenone-producing microorganisms (resulting in overestimated d_s values, minimized but not eliminated by the cutoffs used for analysis) is unknown.

7.3.4. Two selection regimes exist for purple bacteria carotenoid biosynthetic genes

In addition to strain-specific differences noted above, selection also operates differentially upon purple bacteria carotenoid biosynthetic genes responsible for different pathway steps (Table 7.1 and Table 7.2). Particularly notable are the low mean d_n/d_s values for *crtA* (excluding genes from *Rv. gelatinosus* and *H. phototrophica*) relative to all other pathway genes (Table 7.1), further suggesting a selective advantage conferred by spheroidenone production. Comparisons between *crtB*, *crtC*, *crtD*, *crtF* and *crtI* genes from spheroidenone-producing microorganisms (Table 7.2) indicate that mean d_n/d_s values for each gene are not equivalent (Kruskal-Wallis H test, $X^2 \geq 54.850$, $df = 4$, $P < 0.001$), although

Table 7.2. Comparisons of differences between mean d_n/d_s values of carotenoid biosynthetic pathway genes from spheroidenone-producing bacteria (excluding *Rv. gelatinosus* and *H. phototrophica*). Comparisons between multiple genes and gene pairs were conducted using the Kruskal-Wallis H and Mann-Whitney U tests, respectively. Only pair-wise d_n/d_s ratios with $d_n \geq 0.01$ and $d_s \leq 1.5$ were included.

Genes compared	Statistical model			
	Nei-Gojobori with Jukes-Cantor correction	Li-Wu-Luo	Pamilo-Bianchi-Li	Kumar
<i>crtB</i> , <i>crtC</i> , <i>crtD</i> , <i>crtF</i> and <i>crtI</i>	$P < 0.001$, $\chi^2 = 69.802$, $df^a = 4$	$P < 0.001$, $\chi^2 = 62.728$, $df = 4$	$P < 0.001$, $\chi^2 = 54.850$, $df = 4$	$P < 0.001$, $\chi^2 = 95.950$, $df = 4$
<i>crtC</i> , <i>crtD</i> and <i>crtF</i>	$P = 0.302$, $\chi^2 = 2.393$, $df = 2$	$P = 0.123$, $\chi^2 = 4.187$, $df = 2$	$P = 0.273$, $\chi^2 = 2.594$, $df = 2$	$P = 0.131$, $\chi^2 = 4.070$, $df = 2$
<i>crtB</i> and <i>crtI</i>	$P = 0.063$, $Z = -1.859$	$P = 0.105$, $Z = -1.621$	$P = 0.271$, $Z = -1.100$	$P = 0.073$, $Z = -1.793$

^aDegrees of freedom

those for *crtC*, *crtD*, *crtF* and *crtI* (Kruskal-Wallis H test, $\chi^2 \leq 4.187$, $df = 2$, $P \geq 0.123$) and *crtB* and *crtI* (Mann-Whitney U test, Z ranges from -1.100 to -1.859, P ranges from 0.271 to 0.063) are indistinguishable, or at least very similar in the latter case. These results indicate two selection regimes acting upon the carotenoid biosynthetic genes of spheroidenone-producing bacteria, with negative selection upon *crtB* and *crtI* being more intense than that acting upon *crtC*, *crtD* and *crtF* by difference in d_n/d_s ratios ≈ 0.1 . Mean d_n/d_s ratios for genes from spirilloxanthin-producing organisms (excluding *Rv. gelatinosus* and *H. phototrophica*) exhibited the same trends as those producing spheroidenone, with *crtB* and *crtI* having the lower mean d_n/d_s values than *crtC*, *crtD*, *crtF* and *crtI*. Greater variability between means, however, obscured statistically significant differences between groups (see Table 7.1 for means; $P < 0.001$ for all between-gene tests; data not shown). Two selection regimes likely also operate on these genes, but for some genes the between-gene statistical comparisons are confounded by particular strains having d_n/d_s ratios different from the overall mean.

7.4. Discussion

Carotenoid biosynthesis in the purple bacteria produces compounds of the spheroidene or spirilloxanthin series by utilization of the common enzymes CrtB, CrtI, CrtC, CrtD and CrtF. These carotenoids differ primarily in the degree of conjugation of their double bond chains (10 in spheroidene versus 13 in spirilloxanthin), based on their differential utilization of neurosporene or lycopene as hydrocarbon precursors. Strong negative selection on the *crtA* gene, the exclusion of spirilloxanthin production in spheroidenone-producing microorganisms and three independent gene acquisition events all suggest that spheroidenone conveys a selective advantage relative to spirilloxanthin (Figure 7.2, Table 7.1, Supplemental Figure E1). What drives evolutionary selection towards production of spheroidene is unclear. One plausible hypothesis is that ketolation in purple bacteria is inherently advantageous (e.g., in structuring the photosynthetic reaction center or modulating membrane fluidity and permeability)

but that the blue shift caused by diketolation of spirilloxanthin (Schmidt et al. 1994) reduces these organisms' ability to harvest light in an ecologically relevant spectral region (Stomp et al. 2007). Further biochemical studies, perhaps using reaction centers with unnatural carotenoids, may clarify this point.

Detailed biochemical comparison of carotenoid biosynthetic proteins from the spheroidene-producing *Rb. capsulatus* with their homologs from the spheroidene- and spirilloxanthin-producing *Rv. gelatinosus* suggests that production of spheroidene versus spirilloxanthin is governed in part by the lack of lycopene production in *Rb. capsulatus* (Raisig et al. 1996, Harada et al. 2001a). For both microorganisms, lycopene production increases with increased expression of CrtI and limited production of its precursor phytoene, whereby the lower affinity of CrtI towards neurosporene versus phytoene is counteracted by the relative lack of latter substrate (Stickforth and Sandmann 2007). In contrast, upstream pathway enzymes from both *Rb. capsulatus* and *Rv. gelatinosus* utilize a wide range of substrates from either the spheroidene or spirilloxanthin pathway, sometimes with minimal differences in specificity (Steiger et al. 2000, Pinta et al. 2003, Steiger et al. 2003, Gerjets et al. 2009). Additionally, Stickforth and Sandmann (2007) observed that the substrate specificity of CrtI from *Rv. gelatinosus* could be altered by single nucleotide mutations within the *crtI* gene. Intriguingly, the mutant with the greatest increase in lycopene production relative to neurosporene had concomitantly increased yields of both CrtI and the plasmid from which it was expressed. Lycopene production in this case is most likely facilitated by mutations affecting CrtI expression and/or stability and not preeminently its catalytic parameters. These results particularly highlight the potential for single, non-synonymous mutations to substantially alter the *in vivo* substrate specificity of a metabolic pathway, and therefore its downstream products, and especially highlight the potential for carotenoid biosynthetic pathway evolution in *Rv. gelatinosus*.

In this study, elevated d_n/d_s ratios were observed for *crtA* genes horizontally transferred into *Rv. gelatinosus* and *H. phototrophica* (Figures 7.2

and 7.3, Table 7.1), indicating positive selection for CrtA function as a ketolase (versus as a hydroxylase in the *Bacteroidetes* lineage; Rahlert et al. 2009). Elevated d_n/d_s ratios were also observed for all other *Rv. gelatinosus* and some other *H. phototrophica* carotenoid biosynthetic proteins relative to other spirilloxanthin-producers (Table 7.1). There exist five possible explanations for these latter results. First, positive selection at certain amino acid sites within each encoded protein might facilitate the shift from spirilloxanthin- to spheroidenone-producing pathways, the effects of which would be masked by the gene-wide averages analyzed here. However, the reported broad substrate specificities for upstream carotenoid biosynthetic pathway enzymes in *Rv. gelatinosus* and *Rb. capsulatus* (Steiger et al. 2000, Pinta et al. 2003, Steiger et al. 2003, Gerjets et al. 2009) and the homogeneous elevation of mean d_n/d_s values (including the upstream-functioning gene *crtB*) relative to all other spirilloxanthin-producing bacteria by ≈ 0.15 argue against this type of selection (Table 7.1). Secondly, d_n/d_s values of carotenoid biosynthesis genes other than *crtA* in *Rv. gelatinosus* and *H. phototrophica* might be elevated due to their genetic hitch-hiking with adjacent, positively selected *crtA* genes, whereby the selective advantages conveyed by variations in *crtA* outweigh the disadvantages of slightly deleterious mutations in other carotenoid biosynthetic genes. Similar results have been reported elsewhere for some γ -*Proteobacteria* (Shapiro and Alm 2008) but the prevalence of this phenomenon has not been well studied. Third, horizontal gene transfer of the entire photosynthetic gene cluster into *Rv. gelatinosus* and *H. phototrophica* might result in mutation rates for all carotenoid biosynthetic genes (excepting the positively-selected *crtA*) being due to selection for optimal integration into the new host. Supporting this explanation, taxonomic neighbors of *H. phototrophica* are non-photosynthetic (Biebl et al. 2006) and horizontal transfer of the photosynthetic gene cluster from the α -*Proteobacteria* into *Rv. gelatinosus* has been well demonstrated (Nagashima et al. 1997, Igarashi et al. 2001). However, carotenoid biosynthetic genes from the three γ -*Proteobacteria* analyzed here (*Congregibacter litoralis* KT 71, *Halorhodospira halophila* SL1 and marine γ -proteobacterium HTCC2080), which have also likely undergone horizontal

transfer, did not result in similarly elevated d_n/d_s ratios for their carotenoid biosynthesis genes (data not shown). Determining the effect of horizontal transfer of the photosynthetic gene cluster on selection rates therefore awaits better dating and phylogenetic resolution of the relevant transfer events. Fourthly, recombination might yield a pattern of mutations similar to that observed here, especially considering that the present methods used yield whole-gene averages of synonymous and non-synonymous mutations. However, application of the PIST algorithm (Worobey 2001) did not identify recombination in this dataset (data not shown) and visual observation of the aligned sequences did not suggest any obvious break-points, arguing against recombination causing the mutational signatures detected here. Finally, selection for the recovery of chromosome and genetic network structures perturbed by the integration of the horizontally transferred *crtA* genes might cause elevated d_n/d_s ratios for adjacent genes, in this case the other carotenoid biosynthetic genes. Horizontally transferred genes are known to have particularly low network connectivity (Wellner et al. 2007), suggesting that extensive perturbation of a host's biochemical network by a horizontally transferred gene is disadvantageous. Other potentially negative effects on host fitness resulting from horizontally transferred genes, including alteration of the host's transcriptional profile (Warren et al. 2008), expression-dependent toxicity of the transferred gene(s) (Sorek et al. 2007) and excessive metabolic load resulting from transcription of foreign genes (Wagner 2005) are also known, and might result in selection against factors other than the phenotype of the transferred gene itself. The physical effects of integration of a horizontally transferred gene in the host genome are poorly studied, and might negatively effect adjacent loci by disrupting operon structure, the distance to regulatory elements such as promoters and activator- and repressor-binding sites and chromosomal structure, including supercoiling, chromosomal condensation and regulation due to factors such as H-NS (Dorman 2007). Whereas negative selection against these gene-level stresses might, at least in some circumstances, outweigh that against slightly deleterious protein mutations and result in an

elevated number of non-synonymous mutations, this hypothesis remains to be specifically tested.

Regardless of its cause, theory suggests that the increased neutral drift caused by relaxed selection upon *Rv. gelatinosus* and *H. phototrophica* carotenoid biosynthesis allowed an increased sampling of phenotypic space by increasing mutational robustness (i.e., tolerance to slightly deleterious mutations; Bershtein and Tawfik 2008). It is therefore possible that mutations arose, perhaps most significantly in *crtI* considering the low number of mutations apparently needed to alter its substrate specificity (Stickforth and Sandmann 2007), affecting catalytic efficiency, enzyme stability or expression to increase neurosporene production and activate the spheroidenone pathway. This shift may have been further enhanced by the selective advantage of spheroidenone production versus that of 2,2'-diketospirilloxanthin. Negative selection upon *crtB* and *crtI* is evident (Table 7.2), highlighting these genes as modulators of pathway specificity. In contrast, negative selection is lessened for *crtC*, *crtD*, and *crtF* relative to *crtB* and *crtI* (Table 7.2), likely promoting the greater catalytic promiscuity of their cognate proteins (Steiger et al. 2000, Pinta et al. 2003, Steiger et al. 2003). This greater variability likely contributed to the success of increased neurosporene production following acquisition of CrtA by allowing production of the selectively advantageous spheroidenone and not just its upstream precursors. In analogous analyses, it has been suggested that the selection upon plant anthocyanin biosynthetic pathway members is relaxed for the terminal pathway protein that is not obliged to provide precursors for alternative pathway branches (Rausher et al. 1999, Lu and Rausher 2003, Rausher et al. 2008). The results presented here also suggest strongest negative selection upon those genes primarily responsible for pathway specificity, but, in contrast to the anthocyanin pathway, suggests multiple genes under relaxed negative selection which do not form the biosynthetic pathway terminus (Tables 7.1 and 7.2).

Current perspectives on the evolution of biosynthetic pathways (reviewed by Caetano-Anollés et al. 2009, Fani and Fondi 2009) suggests that functional

evolution is most commonly due to selection acting upon promiscuous protein functions (Yčas 1974, Jensen 1976, Kacser and Beeby 1984). Particularly interesting in this regard is the emerging “avant-garde view” of proteins (Tokuriki and Tawfik 2009) whereby conformers (i.e., protein folding variants) have variable structures in equilibria and perform promiscuous reactions. Evolutionary sampling of side-product reactions might effectively occur via mutational drift, thereby generating mutants that are selectively neutral regarding the primary phenotype but which differentially perform various promiscuous reactions (Bershtein and Tawfik 2008). As applied to the evolution of purple bacteria carotenoid biosynthesis, accommodation of the horizontally transferred *crtA* in *Rv. gelatinosus* and *H. phototrophica* (Table 7.2) may have been promoted prior to its acquisition by relaxed selection on *crtC*, *crtD* and *crtF* (Table 7.2), and the ability of CrtI to readily switch its central catalytic activity to produce neurosporene, thereby counteracting the apparently negative consequences of 2,2'-diketospirilloxanthin production. These effects were particularly enhanced in *Rv. gelatinosus* and *H. phototrophica* by the increased number of mutations available to be sampled following relaxation of negative selection upon carotenoid biosynthesis genes other than *crtA* (Figure 7.3 and Table 7.2). Biosynthetic pathway evolution, therefore, depends not only on the development of new biochemical functions but also upon the ability of the receiving pathway to integrate them. This may rely upon preexisting catalytic diversity in the pathway resulting from the topology of the recipient metabolic network (i.e., differences in selection between pathway steps). Alternatively or in conjunction, accommodation of a xenologous gene might occur by reduced selection upon the metabolic pathway specificity of the recipient host, perhaps caused by integration of the xenologous gene itself.

7.5. Literature cited

Albrecht, M., Ruther, A., and Sandmann, G. 1997. Purification and biochemical characterization of a hydroxyneurosporene desaturase involved in the biosynthetic pathway of the carotenoid spheroidene in *Rhodobacter sphaeroides*. J. Bacteriol., 179: 7462-7467.

- Armstrong, G.A., Alberti, M., Leach, F., and Hearst, J.E. 1989. Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. *Mol. Gen. Genet.*, 216: 254-268.
- Badenhop, F., Steiger, S., Sandmann, M., and Sandmann, G. 2003. Expression and biochemical characterization of the 1-HO-carotenoid methylase CrtF from *Rhodobacter capsulatus*. *FEMS Microbiol. Lett.*, 222: 237-242.
- Bershtein, S., and Tawfik, D.S. 2008. Advances in laboratory evolution of enzymes. *Curr. Opin. Chem. Biol.*, 12: 151-158.
- Biebl, H., Tindall, B.J., Pukall, R., Lünsdorf, H., Allgaier, M., and Wagner-Döbler, I. 2006. *Hoeflea phototrophica* sp. nov., a novel marine aerobic alphaproteobacterium that forms bacteriochlorophyll *a*. *Int. J. Syst. Evol. Microbiol.*, 56: 821-826.
- Caetano-Anollés, G., Yafremava, L.S., Gee, H., Caetano-Anollés, D., Kim, H.S., and Mittenthal, J.E. 2009. The origin and evolution of modern metabolism. *Int. J. Biochem. Cell B.*, 41: 285-297.
- Conant, G.C., and Wolfe, K.H. 2008. Turing a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.*, 9: 938-950.
- Dorman, C.J. 2007. H-NS, the genome sentinel. *Nat. Rev. Microbiol.*, 5: 157-161.
- Fani, R., and Fondi, M. 2009. Origin and evolution of metabolic pathways. *Phys. Life Rev.*, 6: 23-52.
- Gerjets, T., Steiger, S., and Sandmann, G. 2009. Catalytic properties of the expressed acyclic carotenoid 2-ketolases from *Rhodobacter capsulatus* and *Rubrivivax gelatinosus*. *Biochim. Biophys. Acta*, 1791: 125-131.
- Giraud, E., Hannibal, L., Fardoux, J., Jaubert, M., Jourand, P., Dreyfus, B., Sturgis, J.N., and Verméglio, A. 2004. Two distinct *crt* gene clusters for two different functional classes of carotenoid in *Bradyrhizobium*. *J. Biol. Chem.*, 279: 15076-15083.
- Gogarten, J.P., and Townsend, J.P. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.*, 3: 679-687.
- Hannibal, L., Lorquin, J., D'Ortoli, N.A., Garcia, N., Chaintreuil, C., Masson-Boivin, C., Dreyfus, B., and Giraud, E. 2000. Isolation and characterization of canthaxanthin biosynthesis genes from the photosynthetic bacterium *Bradyrhizobium* sp. strain ORS278. *J. Bacteriol.*,

182: 3850-3853.

- Harada, J., Nagashima, K.V.P., Takaichi, S., Misawa, N., Matsuura, K., and Shimada, K. 2001a. Phytoene desaturase, CrtI, of the purple photosynthetic bacterium, *Rubrivivax gelatinosus*, produces both neurosporene and lycopene. *Plant Cell Physiol.*, 42: 1112-1118.
- Harada, J., Takaichi, S., Nagashima, K.V.P., Matsuura, K., and Shimada, K. 2001b. Functional analysis of spheroidene mono-oxygenase, CrtA, of the purple photosynthetic bacterium, *Rubrivivax gelatinosus*. In 12th International Congress on Photosynthesis(ed.). Brisbane, Australia. CSIRO, pp. S2-025.
- Hurst, L.D. 2002. The *Ka/Ks* ratio: diagnosing the form of sequence evolution. *Trends Genet.*, 18: 486-487.
- Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K., and Nagashima, K.V.P. 2001. Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.*, 52: 333-341.
- Jensen, R.A. 1976. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.*, 30: 409-425.
- Kacser, H., and Beeby, R. 1984. Evolution of catalytic proteins. *J. Mol. Evol.*, 20: 38-51.
- Klassen, J.L., McKay, R., and Foght, J.M. 2009. 2'-Methyl and 1'-xylosyl derivatives of 2'-hydroxyflexixanthin are major carotenoids of *Hymenobacter* species. *Tetrahedron Lett.*, 50: 2656-2660.
- Koblížek, M., Bějá, O., Bidigare, R.R., Christensen, S., Benetiz-Nelson, B., Vetriani, C., Kolber, M.K., Falkowski, P.G., and Kolber, Z.S. 2003. Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol.*, 180: 327-338.
- Koonin, E.V., Makarova, K.S., and Aravind, L. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu. Rev. Microbiol.*, 55: 709-742.
- Kovács, Á.T., Rákhely, G., and Kovács, K.L. 2003. Genes involved in the biosynthesis of photosynthetic pigments in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Appl. Environ. Microbiol.*, 69: 3093-3102.
- Lang, H.P., Cogdell, R.J., Gardiner, A.T., and Hunter, C.N. 1994. Early steps in

- carotenoid biosynthesis: sequences and transcriptional analysis of the *crtI* and *crtB* genes of *Rhodobacter sphaeroides* and overexpression and reactivation of *crtI* in *Escherichia coli* and *R. sphaeroides*. *J. Bacteriol.*, 176: 3859-3869.
- Lang, H.P., Cogdell, R.J., Takaichi, S., and Hunter, C.N. 1995. Complete DNA sequence, specific Tn5 insertion map, and gene assignment of the carotenoid biosynthesis pathway of *Rhodobacter sphaeroides*. *J. Bacteriol.*, 177: 2064-2073.
- Lawrence, J.G., and Retchless, A.C. 2009. The interplay of homologous recombination and horizontal gene transfer in bacterial speciation. *In* Horizontal Gene Transfer. M.B. Gogarten, J.P. Gogarten, and L. Olendzenski. (ed.) Humana Press, New York, NY. pp. 29-53.
- Li, W.-H., Wu, C.-I., and Lou, C.-C. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.*, 2: 150-174.
- Li, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol.*, 36: 96-99.
- Lu, Y., and Rausher, M.D. 2003. Evolutionary rate variation in anthocyanin pathway genes. *Mol. Biol. Evol.*, 20: 1844-1853.
- Nagashima, K.V.P., Hiraishi, A., Shimida, K., and Matsuura, K. 1997. Horizontal transfer of genes coding for the photosynthetic reaction centers of purple bacteria. *J. Mol. Evol.*, 45: 131-136.
- Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, 3: 418-426.
- Nei, M., and Kumar, S. 2000. Molecular evolution and phylogenetics Oxford University Press, New York.
- Ouchane, S., Picaud, M., Vernotte, C., and Astier, C. 1997a. Photooxidative stress stimulates illegitimate recombination and mutability in carotenoid-less mutants of *Rubrivivax gelatinosus*. *The EMBO Journal*, 16: 4777-4787.
- Ouchane, S., Picaud, M., Vernotte, C., Reiss-Husson, F., and Astier, C. 1997b. Pleiotropic effects of *puf* interposon mutagenesis on carotenoid biosynthesis in *Rubrivivax gelatinosus*. *J. Biol. Chem.*, 272: 1670-1676.
- Pamilo, P., and Bianchi, N.O. 1993. Evolution of the *Zfx* and *Zfy* genes: rates and

- interdependence between the genes. *Mol. Biol. Evol.*, 10: 271-281.
- Pinta, V., Ouchane, S., Picaud, M., Takaichi, S., Astier, C., and Reiss-Husson, F. 2003. Characterization of unusual hydroxy- and ketocarotenoids in *Rubrivivax gelatinosus*: involvement of enzyme CrtF or CrtA. *Arch. Microbiol.*, 179: 354-362.
- Rählert, N., Fraser, P.D., and Sandmann, G. 2009. A *crtA*-related gene from *Flavobacterium* P99-3 encodes a novel carotenoid 2-hydroxylase involved in myxol biosynthesis. *FEBS Lett.*, 583: 1605-1610.
- Raisig, A., Bartley, G., Scolnik, P., and Sandmann, G. 1996. Purification in an active state and properties of the 3-step phytoene desaturase from *Rhodobacter capsulatus* overexpressed in *Escherichia coli*. *J. Biochem. (Tokyo)*. 119: 559-564.
- Rausher, M.D., Miller, R.E., and Tiffin, P. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. *Mol. Biol. Evol.*, 16: 266-274.
- Rausher, M.D., Lu, Y., and Meyer, K. 2008. Variation in constraint versus positive selection as an explanation for evolutionary rate variation among anthocyanin genes. *J. Mol. Evol.*, 67: 137-144.
- Schmidt, K., Connor, A., and Britton, G. 1994. Analysis of pigments: carotenoids and related polyenes. *In* Chemical methods in prokaryotic systematics. M. Goodfellow and A.G. O'Donnell. (ed.) John Wiley & Sons, Chichester. pp. 403-461.
- Shapiro, B.J., and Alm, E.J. 2008. Comparing patterns of natural selection across species using selective signatures. *PLoS Genet.*, 4: e23.
- Sorek, R., Zhu, Y., Creevey, C.J., Francino, M.P., Bork, P., and Rubin, E.M. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science*, 318: 1449-1452.
- Stamatakis, A., Hoover, P., and Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst. Biol.*, 57: 758-771.
- Steiger, S., Astier, C., and Sandmann, G. 2000. Substrate specificity of the expressed carotenoid 3,4-desaturase from *Rubrivivax gelatinosus* reveals the detailed reaction sequence to spheroidene and spirilloxanthin. *Biochem. J.*, 349: 635-640.
- Steiger, S., Mazet, A., and Sandmann, G. 2003. Heterologous expression, purification, and enzymatic characterization of the acyclic carotenoid 1,2-

- hydratase from *Rubrivivax gelatinosus*. Arch. Biochem. Biophys., 414: 51-58.
- Stickforth, P., and Sandmann, G. 2007. Kinetic variations determine the product pattern of phytoene desaturase from *Rubrivivax gelatinosus*. Arch. Biochem. Biophys., 461: 235-241.
- Stomp, M., Huisman, J., Stal, L.J., and Matthijs, H.C.P. 2007. Colorful niches of phototrophic microorganisms shaped by vibrations of the water molecule. ISME J., 1: 271-282.
- Takaichi, S. 1999. Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In The photochemistry of carotenoids. H.A. Frank, A.J. Young, G. Britton, and R.J. Cogdell. (ed.) Kluwer Academic Publishers, New York, NY. pp. 39-69.
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. 2007. *MEGA4*: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol., 24: 1596-1599.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., 22: 4673-4680.
- Tokuriki, N., and Tawfik, D.S. 2009. Protein dynamism and evolvability. Science, 324: 203-207.
- Wagner, A. 2005. Periodic extinctions of transposable elements in bacterial lineages: evidence from intragenomic variation in multiple genomes. Mol. Biol. Evol., 23: 723-733.
- Warren, R.L., Freeman, J.D., Levesque, R.C., Smailus, D.E., Flibotte, S., and Holt, R.A. 2008. Transcription of foreign DNA in *Escherichia coli*. Genome Res., 18: 1798-1805.
- Wellner, A., Lurie, M.N., and Gophna, U. 2007. Complexity, connectivity, and duplicability as barriers to lateral gene transfer. Genome Biol., 8: R156.
- Worobey, M. 2001. A novel approach to detecting and measuring recombination: new insights into evolution in viruses, bacteria, and mitochondria. Mol. Biol. Evol., 18: 1425-1434.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol., 24: 1586-1591.

Yčas, M. 1974. On earlier states of the biochemical system. *J. Theor. Biol.*, 44:
145-160.

8. Synthesis and Conclusions

It is now well known that glaciers host a wide variety and abundance of microbial life (Hodson et al. 2008). Despite the low numbers of microbes recovered in this current study, this is also true of cold-based Victoria Upper Glacier, Antarctica (VUG; Chapter 2). The highest numbers of cultured bacteria were recovered from immediately above the glacial ice-basal ice boundary, in congruence with results determined previously using fluorescence spectroscopy (Barker et al. 2006). Also remarkable was the preponderance of *Hymenobacter*-like isolates recovered from VUG (Chapter 3). The bright pink pigmentation of these isolates, resulting from carotenoids of previously unknown structure (Chapter 5), highlights the potential of glaciers to harbor microbes with significant biotechnological application (e.g., Simon et al. 2009).

Based upon their polyphasic characterization alongside nearly all described *Hymenobacter* species, many of the VUG isolates recovered in this study are taxonomically novel (Chapter 3). However, these studies also revealed discrepancies between phylogenies of different reference genes, extensive strain-specific variation for many studied phenotypes and, for some phenotypes, significantly less discrimination between described species than was realized previously (Chapter 3). Similar results were obtained for carotenoid distribution (Chapter 4), which together suggests that non-vertical evolutionary processes such as horizontal gene transfer, gene gain and gene loss occur extensively within this genus. These results call into question the nature of “species” within the *Hymenobacter* lineage (Chapter 3) and emphasize the challenge of determining a meaningful taxonomy for these microorganisms. Whereas the current studies were hampered by the divergence of *Hymenobacter* from better-studied organisms (e.g., by causing poor amplification using “universal” PCR primers for *rpoB* and *cpn60*), the application of whole-genome sequencing to these microorganisms, currently underway at the Joint Genome Institute, will greatly facilitate in-depth analyses. Despite these limitations, the current study has identified at least one novel species and likely several others. The classification

of these latter taxa will benefit greatly from the future isolation of other closely related strains to better define the phenotypic and chemotaxonomic characteristics typical of these taxa and more precisely demarcate plausible species boundaries between them.

Hymenobacter carotenoids were remarkable for their novel structures (Chapter 5) and the significant diversity of carotenoid content between species (Chapter 4). This contrasts significantly with the prevailing view of carotenoid diversity which emphasizes genus-level similarities (Vandamme et al. 1996). Extending this analysis further, comparative genomics showed that the extent of carotenoid diversity ranged widely between phylogenetic lineages and varied according to lineage-specific evolutionary mechanisms (Chapter 6). In the Bacteroidetes, comparative genomics, like the chemical analysis of *Hymenobacter* carotenoids (Chapter 4), suggested that horizontal gene transfer is a prevalent mechanism of carotenoid biosynthetic pathway evolution. These analyses particularly highlight the potential of the Bacteroidetes to contain novel carotenoid structural and biosynthetic diversity and suggest the usefulness of future study of these organisms.

Comparative genomics similarly identified many gaps in the current knowledge of carotenoids, both indicating understudied phyla and those for which carotenoid biosynthesis may occur according to currently unknown mechanisms (Chapter 6). These gaps occurred even among relatively well-studied taxa such as the purple bacteria, for which detailed evolutionary analyses of carotenoid biosynthetic evolution revealed multiple, previously unrecognized, instances of pathway evolution from spirilloxanthin production to that of spheroidenone (Chapter 7). Surprisingly, these transitions involved both horizontal gene transfer and changes in mutational selection on both the horizontally transferred gene and those upstream of it in the carotenoid biosynthetic pathway. Coupled with the extensive preexisting biochemical data, this analysis suggests that purple bacterial carotenoid biosynthesis is an excellent model system to understand microbial biochemical pathway evolution, particularly as it involves interactions between

horizontal gene transfer and mutational selection. The widespread character of these mutational forces suggests that this analysis (Chapter 7) might be especially relevant towards understanding microbial evolution on a large scale.

Overall, this thesis indicates both the usefulness of traditional microbiological techniques and the promise of modern, genomics-based approaches for studying microbial diversity. The plating techniques employed to isolate novel *Hymenobacter* strains from VUG were unsophisticated, but effective in isolating new “species” (Chapters 2 and 3) with novel, biotechnologically interesting properties (Chapters 4 and 5). However, it is well-known from many other environments, including glaciers (Cheng and Foght 2007), that culture-based techniques isolate only a subset of the true diversity present and are biased towards those microorganisms which are already known (Hugenholtz et al. 1998, DeLong and Pace 2001). Evolutionary analysis of genome sequence data offers a complementary means to identify which evolutionary lineages are most likely to contain unknown diversity (Chapter 6). Evolutionary biology is therefore an applied science, whereby the likelihood of identifying novelty is directly dependent on the evolutionary history of a gene or protein. This is particularly highlighted by the diversification evident in the carotenoid biosynthetic pathways of purple bacteria containing signatures of positive selection, an indicator of functional diversification (Chapter 7). Unraveling similar selective landscapes in other organisms and their effects on pathway evolution will undoubtedly lead to new insights into the mechanisms by which pathways diverge to accommodate and exploit new ecological niches, both in nature and in synthetic biology. Undoubtedly, traditional isolation studies are important and often effective means of accessing microbial diversity; their biases, however, leave in question their representativeness and completeness. In contrast, evolutionary hypotheses offer a directed approach to studying microbial diversity which will more efficiently direct its discovery and exploitation. We may not know the diversity we will find, but at least we can predict where to look.

8.1. Literature Cited

- Barker, J.D., Sharp, M.J., Fitzsimons, S.J., and Turner, R.J. 2006. Abundance and dynamics of dissolved organic carbon in glacier systems. *Arct. Antarct. Alp. Res.*, 38: 163-172.
- Cheng, S.M., and Foght, J.M. 2007. Cultivation-independent and -dependent characterization of Bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.*, 59: 318-330.
- DeLong, E.F., and Pace, N.R. 2001. Environmental diversity of Bacteria and Archaea. *Syst. Biol.*, 50: 470-478.
- Hodson, A., Anesio, A.M., Tranter, M., Fountain, A., Osborn, M., Priscu, J., Laybourn-Parry, J., and Sattler, B. 2008. Glacial ecosystems. *Ecol. Monogr.*, 78: 41-67.
- Hugenholtz, P., Goebel, B.M., and Pace, N.R. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.*, 180: 4765-4774.
- Simon, C., Herath, J., Rockstroh, S., and Daniel, R. 2009. Rapid identification of genes encoding DNA polymerases by function-based screening of metagenomic libraries derived from glacial ice. *Appl. Environ. Microbiol.*, 75: 2964-2968.
- Vandamme, P., Pot, B., Gillis, M., De Vos, P., Kersters, K., and Swings, J. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. *Microb. Rev.*, 60: 407-438.

Appendix A – Supplemental Data for Chapter 1

- Supplemental Table A1: Otherwise undescribed *Hymenobacter*-like isolates with 16S rRNA gene sequences in GenBank as of March 2009
- Supplemental Table A2: Cloned *Hymenobacter*-like 16S rRNA gene sequences in GenBank as of March 2009

Supplemental Table A1. Unnamed *Hymenobacter*-like isolates with 16S rRNA sequences in GenBank as of March 2009. Sequences were determined by BLAST searching the GenBank nr database using a known *Hymenobacter* species and recovery of *Hymenobacter*-like sequences using the phylogenetic tree view of the results. Search parameters were adjusted such that sequences clustering distant to *Hymenobacter* were also recovered, facilitating complete recovery of all known *Hymenobacter*-like sequences in the nr database at the time of analysis.

Strain	Source	Reference	Genbank
IFAM AA-688	Antarctic McMurdo Dry Valley soil	(Hirsch et al. 1998)	Y18834
VUG-A23a, VUG-A34, VUG-A42aa, VUG-C4, VUG-A67, VUG-A33, VUG-A141a, VUG-A60a, VUG-A124, VUG-A142	Victoria Upper Glacier glacial ice	(Klassen and Foght 2008)	EU155009 – EU155017, EU159489
VUG-A112, VUG-A130, VUG-A2a, VUG-A57b, VUG-A31a, VUG-A48, VUG-A58, VUG-A65	Victoria Upper Glacier glacial ice	Unpublished	-
<i>Hymenobacter</i> sp. BSw20462	Arctic Sea Water	Unpublished	EF639389
<i>Hymenobacter</i> sp. R2A-W5	Chinese soil	Unpublished	FJ627043
<i>Deinococcus</i> sp. 6A2	USA radionuclide contaminated soil	(Fredrickson et al. 2008)	EU029121
<i>Hymenobacter</i> sp. GIC34	Greenland glacial ice core	(Miteva et al. 2004)	AY439245
Groundwater biofilm bacterium U2	Groundwater	Unpublished	FJ204433
<i>Bacteroidetes</i> bacterium P3	Antarctica: La Gorce Mountains	(Aislable et al. 2006a)	DQ351728
<i>Cytophagales</i> str. S23328	?	(Mitsui et al. 1997)	D84607
<i>Hymenobacter</i> sp. MJ1	Cloghoge river biofilm	Unpublished	AF449431
<i>Hymenobacter</i> sp. 1004, <i>Hymenobacter</i> sp. 1018	Glacier No.1 of Tianshan Mountain	Unpublished	EF423320, EF423328
<i>Hymenobacter</i> sp. 1N-12	Commercial airline cabin air	(Osman et al. 2008)	EU379243
<i>Hymenobacter</i> sp. NS/2	UK museum aerosol	(Buczolits et al. 2006)	AJ549284
<i>Hymenobacter</i> sp. 01WB02.1-58	Hard water rivulet Germany:Harz Mountain, Westerhoefer Bach	Unpublished	FM161368
<i>Hymenobacter rigui</i> strain 214	Water Purifiers in Elementary Schools Located in Gunsan Area Korea	Unpublished	EU730947
<i>Hymenobacter</i> sp. XTM003	Tibet soil	Unpublished	EU382214
<i>Hymenobacter</i> sp. SAFR-023	Spacecraft assembly facility	Unpublished	AY167836
<i>Hymenobacter</i> sp. zf-IRIt3	Ice cores, East Rongbuk Glacier, Mt. Qomolangma (Everest)	(Zhang et al. 2006)	DQ223662
<i>Hymenobacter</i> sp. 21/4, H sp 35 26	soil Antarctica: Victoria Land	(Aislable et al. 2006b)	DQ365992

Strain	Source	Reference	Genbank
Antarctic bacterium R-7572	Antarctica:McMurdo Dry Valleys, Lake Fryxell mat	(Van Trappen et al. 2002)	AJ440980
<i>Hymenobacter</i> sp. Dae14	Sediment from stream near Daechung Dam Korea	Unpublished	EU370958

Clone	Source	Reference	Genbank
100M1_G10	Peru: Cordillera Vilcanota Range, Puca Glacier, elev. ~5000 m newly deglaciated soil 100m from glacier terminus replicate 1	(Nemergut et al. 2007)	DQ514033
0M2_G2, 0M2_H7	Peru: Cordillera Vilcanota Range, Puca Glacier, elev. ~5000 m newly deglaciated soil 0m from glacier terminus replicate 2	(Nemergut et al. 2007)	DQ513915, DQ513917
Kuy-SL-50, KuyT-IWPB-31, Kuy-SL-27, Kuy-SL-98, Kuy-SL-45, KuyT-ice-18, KuyT-IWPB-97, KuyT-IWPB-100, Kuy-SL-11, KuyT-IWPS-83	Surface snow in the Kuytun Glacier 51	Unpublished	EU263711, EU263737, EU263697, EU263705, EU263713, EU263756, EU263729, EU263733, EU263709, EU263741
zd5-30	Zadang glacier, Tibet snow	Unpublished	EU527165
PIC-D12	Puruogangri glacial ice core	Unpublished	AM232823
GU-clone-3, GU-clone-4, GU-clone-1	Alaska, Gulkana Glacier surface	Unpublished	AB464936, AB464937, AB464934
FA04C02, FA01H05, FM872922	Finland floor dust	Unpublished	FM872619, FM872540, FB04F09
G6-64, G6-88, G6-26	China: Tibet, Tanggula mountains, Mount Geladandong at 5720m snow pit	Unpublished	EU153039, EU153041, EU153035
SH2B-3C, clone 22-4E	coal	Unpublished	EU073784
ANTLV2_E05	4.8 m, Lake Vida ice cover, McMurdo Dry Valleys, Antarctica	(Mosier et al. 2007)	DQ521518
clone 22-4E	Oil sands tailings pond	Unpublished	EF420219
ARKCRY-50, ARKCRY-14, ARKCRY2, ARKCRY-118, ARKCRY-39, ARKCRY-20	Arctic sea ice cryoconite	Unpublished	AF468338, AF468341, AY198110, AF468337, AF468342, AF46834
ARKMP-14	Melt pond on Arctic sea ice floe	Unpublished	AF468332
StLO87	Germany: northeast, Lake Stechlin area fagus leaf litter	Unpublished	EU218968
V51_9a6, V70_2B5	Varnish covered rocks USA: Black Canyon, Luis Lopez Mining District, SW of Socorro, New Mexico	Unpublished	FJ595620, FJ595653

Clone	Source	Reference	Genbank
NV57_1d3	Rocks with no varnish USA: Black Canyon, Luis Lopez Mining District, SW of Socorro, New Mexico	Unpublished	FJ595542,
BF0002C006, BF0001C019	Finland floor dust	(Rintala et al. 2008)	AM697309, AM697178
F15_5A_FL	Eastern Mediterranean atmosphere	(Polymenakou et al. 2008)	EF683042
FCPN403	Grassland soil, USA: northern California, Angelo Coast Range Reserve	Unpublished	EF516581
nbt108h01	Antecubital fossa (inner elbow) skin	(Grice et al. 2008)	EU539976
EHFS1_S11c	ESTEC HYDRA facility clean room	Unpublished	EU071507

Literature cited

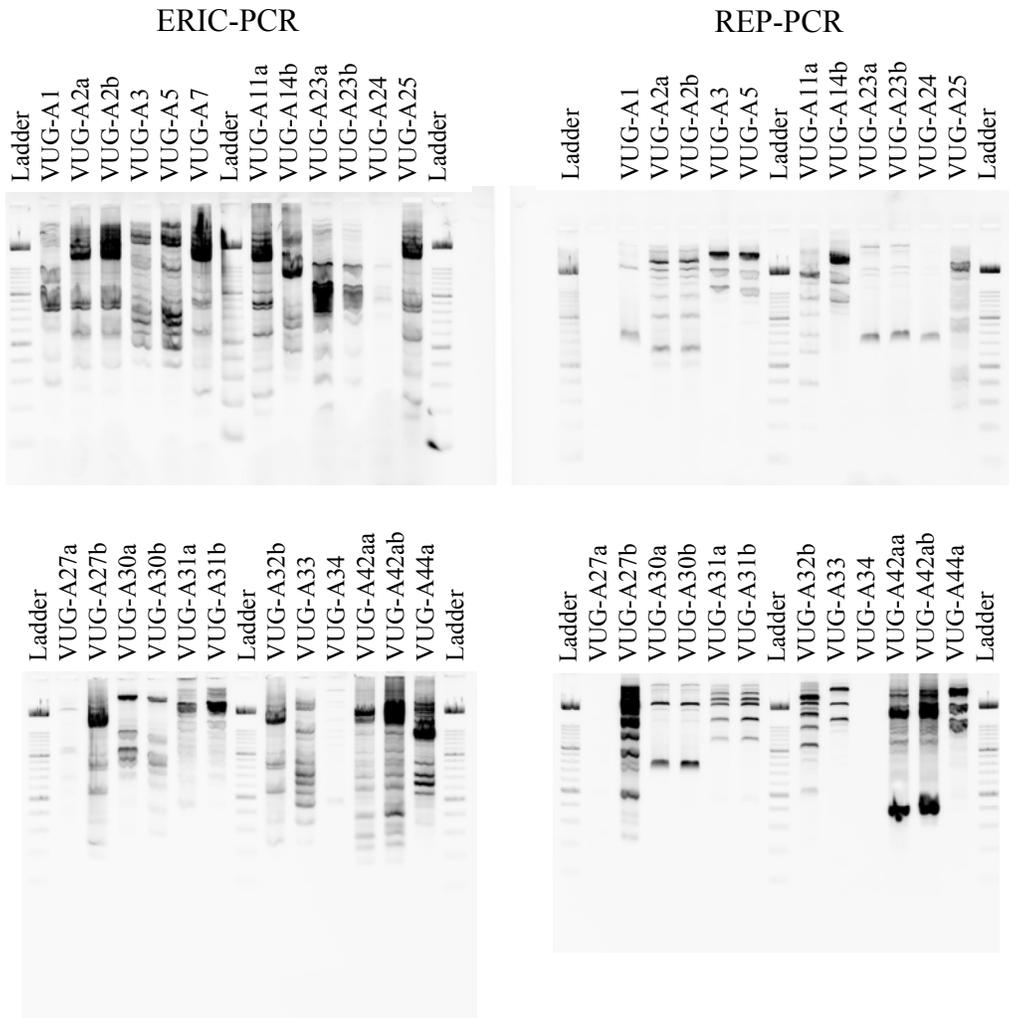
- Aislabie, J.M., Broady, P.A., and Saul, D.J. 2006a. Culturable aerobic heterotrophic bacteria from high altitude, high latitude soil of La Gorce Mountains (86°30'S, 147°W), Antarctica. *Antarct. Sci.*, 18: 313-321.
- Aislabie, J.M., Chhour, K.-L., Saul, D.J., Miyauchi, S., Ayton, J., Paetzold, R.F., and Balks, M.R. 2006b. Dominant bacteria in soils of Marble Point and Wright Valley, Victoria Land, Antarctica. *Soil Biol. Biochem.*, 38: 3041-3056.
- Bidle, K.D., Lee, S., Marchant, D.R., and Falkowski, P.G. 2007. Fossil genes and microbes in the oldest ice on Earth. *Proc. Nat. Acad. Sci. USA*, 104: 13455-13460.
- Buczolits, S., Denner, E.B.M., Kämpfer, P., and Busse, H.-J. 2006. Proposal of *Hymenobacter norwichensis* sp. nov., classification of '*Taxeobacter ocellatus*', '*Taxeobacter gelipurpurascens*' and '*Taxeobacter chitinivorans*' as *Hymenobacter ocellatus* sp. nov., *Hymenobacter gelipurpurascens* sp. nov. and *Hymenobacter chitinivorans* sp. nov., respectively, and emended description of the genus *Hymenobacter* Hirsch *et al.* 1999. *Int. J. Syst. Evol. Microbiol.*, 56: 2071-2078.
- Costello, E.K., Halloy, S.R.P., Reed, S.C., Sowell, P., and Schmidt, S.K. 2009. Fumarole-supported islands of biodiversity within a hyperarid, high-elevation landscape on Socoma Volcano, Puna de Atacama, Andes. *Appl. Environ. Microbiol.*, 75: 735-747.
- Fredrickson, J.K., Li, S.-m.W., Gaidamakova, E.K., Matrosova, V.Y., Zhai, M., Sulloway, H.M., Scholten, J.C., Brown, M.G., Balkwill, D.L., and Daly, M.J. 2008. Protein oxidation: key to bacterial desiccation resistance? *ISME J.*, 2: 393-403.
- Gao, Z., Tseng, C.-h., Pei, Z., and Blaser, M.J. 2007. Molecular analysis of human forearm superficial skin bacterial biota. *Proc. Nat. Acad. Sci. USA*, 104: 2927-2932.
- Grice, E.A., Kong, H.H., Renaud, G., Young, A.C., NISC Comparative Sequencing Program, Bouffard, G.G., Blakesley, R.W., Wolfsberg, T.G., Turner, M.L., and Segre, J.A. 2008. A diversity profile of the human skin microbiota. *Genome Res.*, 18: 1043-1050.
- Hirsch, P., Ludwig, W., Hethke, C., Sittig, M., Hoffmann, B., and Gallikowski, C.A. 1998. *Hymenobacter roseosalivarius* gen. nov., sp. nov. from continental Antarctic soils and sandstone: bacteria of the *Cytophaga/Flavobacterium/Bacteroides* line of phylogenetic descent. *Syst. Appl. Microbiol.*, 21: 374-383.
- Kent, A.D., Jones, S.E., Lauster, G.H., Graham, J.M., Newton, R.J., and McMahon, K.D. 2006. Experimental manipulations of microbial food web interactions in a humic lake: shifting biological drivers of bacterial

- community structure. *Environ. Microbiol.*, 8: 1448-1459.
- Klassen, J.L., and Foght, J.M. 2008. Differences in carotenoid composition among *Hymenobacter* and related strains support a tree-like model of carotenoid evolution. *Appl. Environ. Microbiol.*, 74: 2016-2022.
- Miteva, V.I., Sheridan, P.P., and Brenchley, J.E. 2004. Phylogenetic and physiological diversity of microorganisms isolated from a deep Greenland glacier ice core. *Appl. Environ. Microbiol.*, 70: 202-213.
- Mitsui, H., Gorlach, K., Lee, H.-j., Hattori, R., and Hattori, T. 1997. Incubation time and media requirements of culturable bacteria from different phylogenetic groups. *J. Microbiol. Methods*, 30: 103-110.
- Mosier, A.C., Murray, A.E., and Fritsen, C.H. 2007. Microbiota within the perennial ice cover of Lake Vida, Antarctica. *FEMS Microbiol. Ecol.*, 59: 274-288.
- Nemergut, D.R., Anderson, S.P., Cleveland, C.C., Martin, A.P., Miller, A.E., Seimon, A., and Schmidt, S.K. 2007. Microbial community succession in an unvegetated, recently deglaciated soil. *Microb. Ecol.*, 53: 110-122.
- Osman, S., La Duc, M.T., Dekas, A., Newcombe, D., and Venkateswaran, K. 2008. Microbial burden and diversity of commercial airline cabin air during short and long durations of travel. *ISME J.*, 2: 482-497.
- O'Sullivan, L.A., Weightman, A.J., and Fry, J.C. 2002. New degenerate *Cytophaga-Flexibacter-Bacteroides*-specific 16S ribosomal DNA-targeted oligonucleotide probes reveal high bacterial diversity in River Taff epilithon. *Appl. Environ. Microbiol.*, 68: 201-210.
- Polymenakou, P.N., Mandalakis, M., Stephanou, E.G., and Tselepidis, A. 2008. Particle size distribution of airborne microorganisms and pathogens during an intense African dust event in the Eastern Mediterranean. *Environ. Health Perspect.*, 116: 292-296.
- Rintala, H., Pitkäranta, M., Toivola, M., Paulin, L., and Nevalainen, A. 2008. Diversity and seasonal dynamics of bacterial community in indoor environment. *BMC Microbiology*, 8: 56.
- Van Trappen, S., Mergaert, J., Van Eygen, S., Dawyndt, P., Cnockaert, M.C., and Swings, J. 2002. Diversity of 746 heterotrophic bacteria isolated from microbial mats from ten Antarctic lakes. *Syst. Appl. Microbiol.*, 25: 603-610.
- Venkateswaran, K., Hattori, N., La Duc, M.T., and Kern, R. 2003. ATP as a biomarker of viable microorganisms in clean-room facilities. *J. Microbiol. Methods*, 52: 367-377.
- Yergeau, E., Newsham, K.K., Pearce, D.A., and Kowalchuk, G.A. 2007. Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. *Environ. Microbiol.*, 9: 2670-2682.

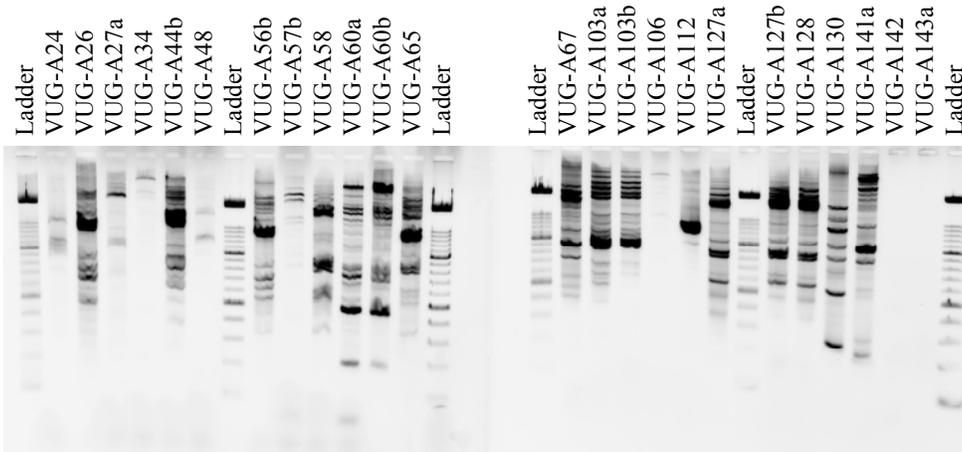
Zhang, S., Hou, S., Ma, X., Qin, D., and Chen, T. 2006. Culturable bacteria in Himalayan ice in response to atmospheric circulation. *Biogeosciences Discussions*, 3: 765-778.

Appendix B – Supplemental Data for Chapter 3

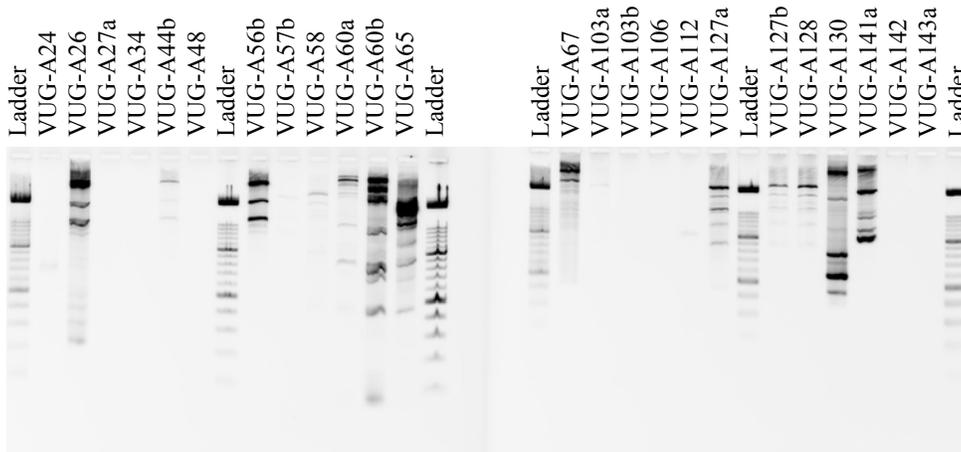
Supplemental Figure B1. ERIC- and REP-PCR results for all VUG *Hymenobacter* strains.



ERIC-PCR

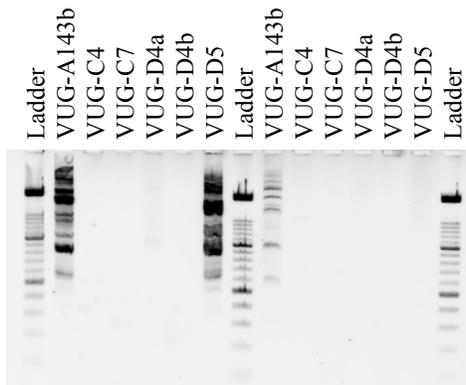


REP-PCR

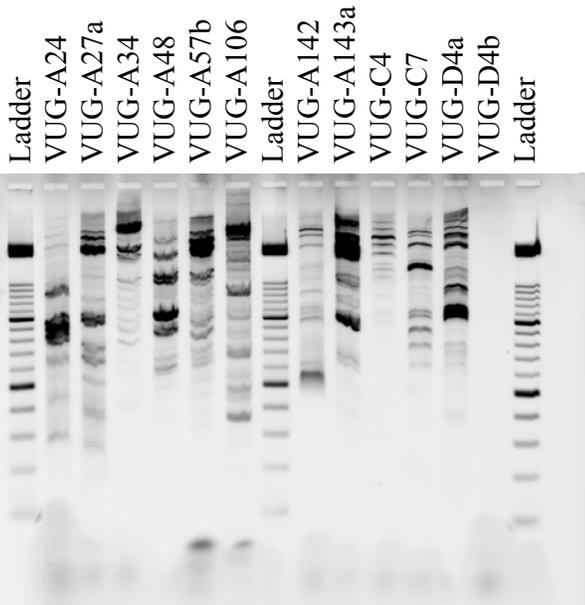


ERIC-PCR

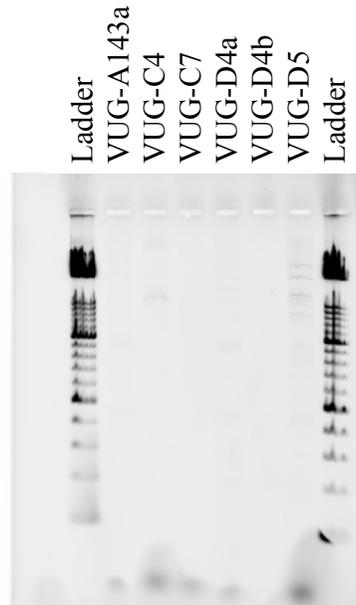
REP-PCR



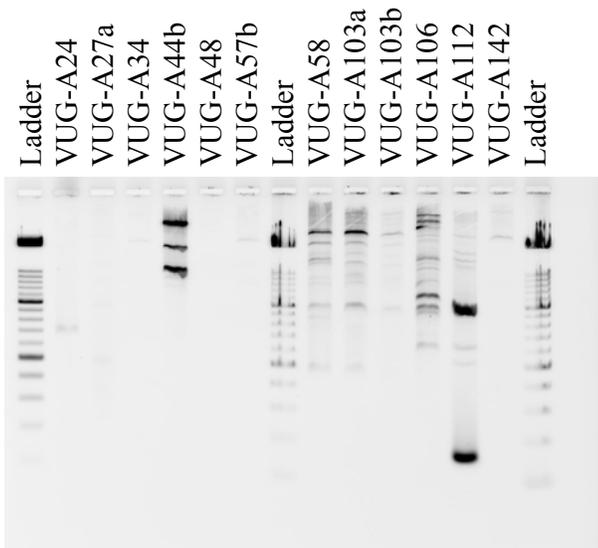
ERIC-PCR



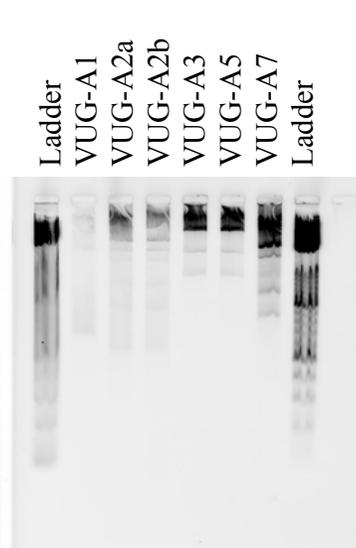
REP-PCR



REP-PCR

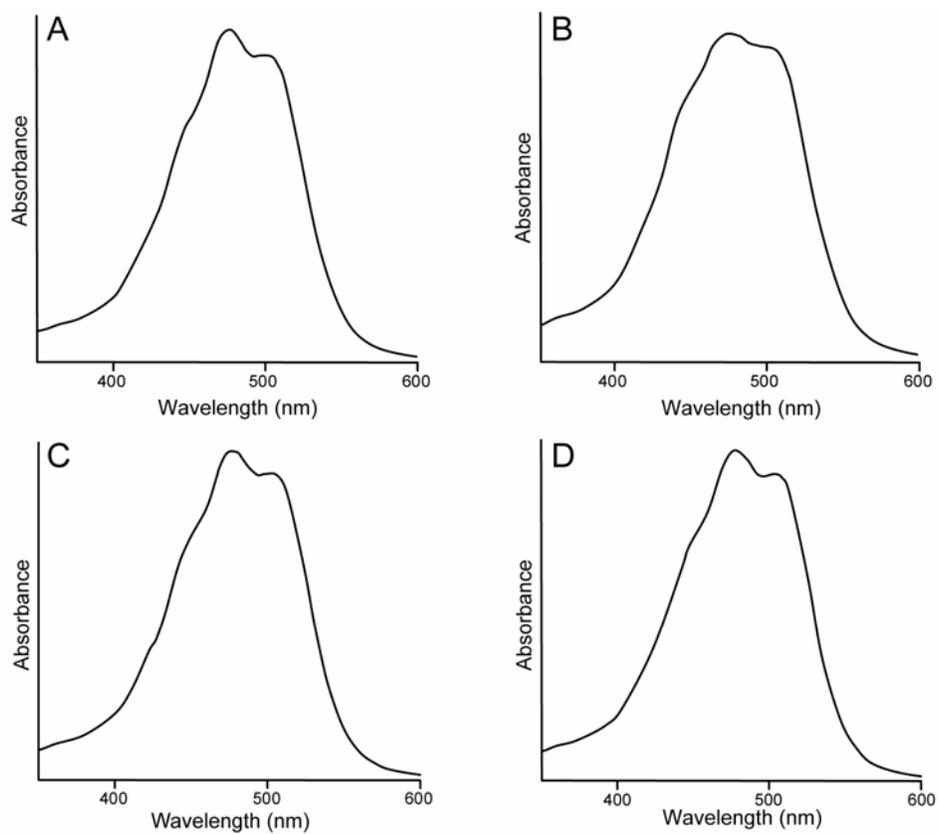


REP-PCR

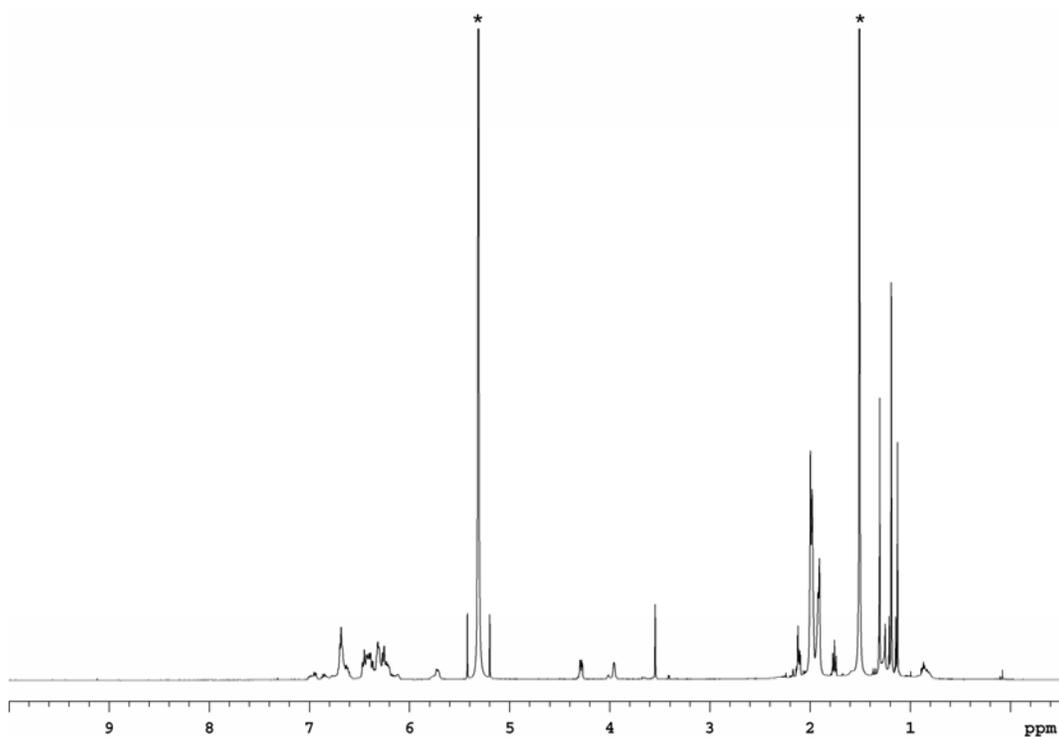


Appendix C – Supplemental Data for Chapter 5

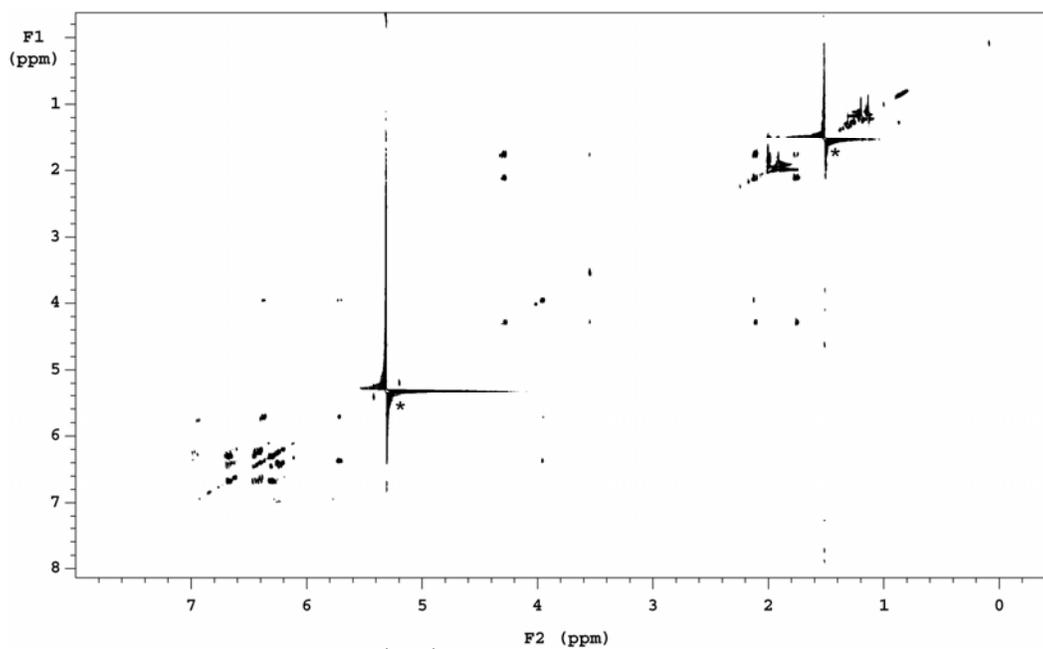
- Supplemental Figure C1: UV-Vis spectra in methanol for carotenoids 4-7
- Supplemental Figure C2: 1D-¹H NMR spectrum of carotenoid 5
- Supplemental Figure C3: 2D-¹H,¹H-TOCSY NMR spectrum of carotenoid 5
- Supplemental Figure C4: 1D-¹H NMR spectrum of carotenoid 4
- Supplemental Figure C5: 2D-¹H,¹H-TOCSY NMR spectrum of carotenoid 4
- Supplemental Figure C6: 2D-¹H,¹H-TOCSY NMR spectrum of carotenoid 6
- Supplemental Figure C7: 1D-¹H NMR spectrum of carotenoid 7
- Supplemental Figure C8: 2D-¹H,¹H-TOCSY NMR spectrum of carotenoid 7
- Supplemental Figure C9: 1D-¹H NMR spectrum of carotenoid 8 contaminated with an unidentified glycoside
- Supplemental Figure C10: 2D-¹H,¹H-TOCSY NMR spectrum of carotenoid 8 contaminated with an unidentified glycoside



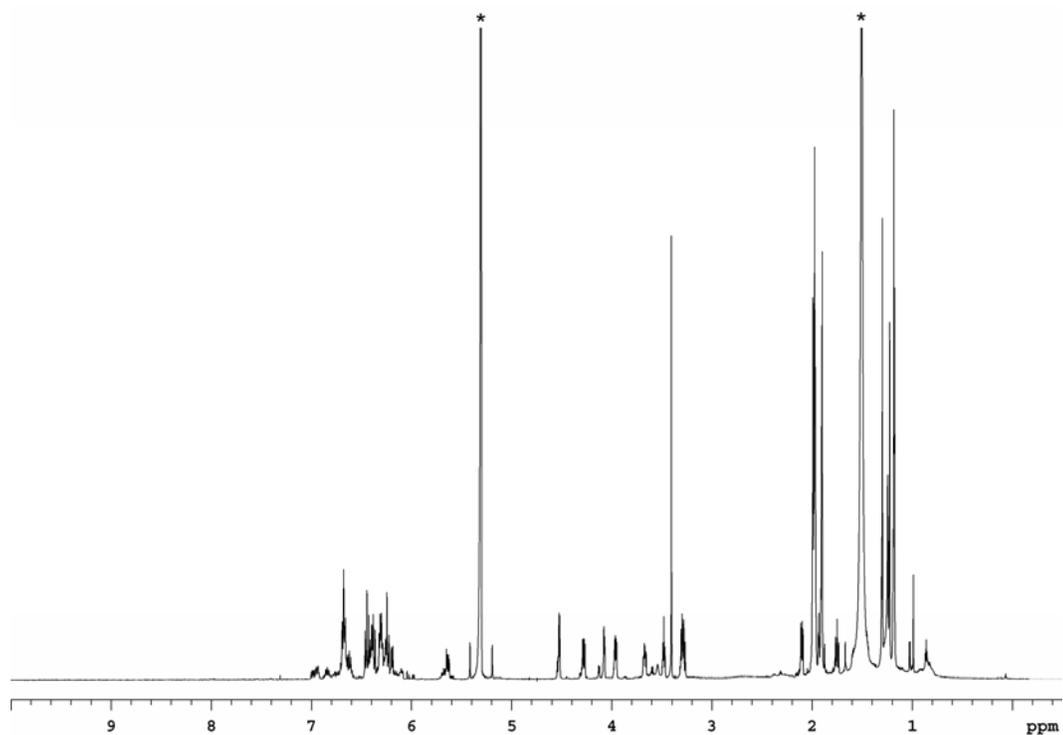
Supplemental Figure C1. UV-Vis absorption spectra (only 350-600 nm shown; the rest of the spectrum was uninformative) in methanol of carotenoids 5 (A) and 4 (B), isolated from strain VUG-A141a, and carotenoids 6 (C) and 7 (D), isolated from strain VUG-A42aa.



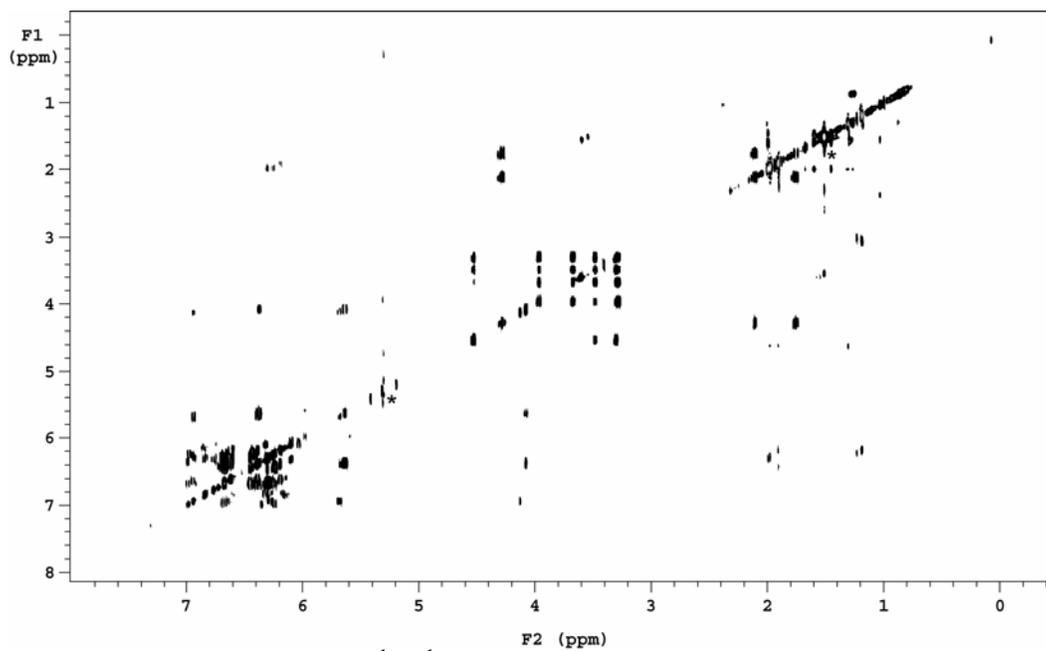
Supplemental Figure C2. 1D- ^1H NMR spectrum of carotenoid 5 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



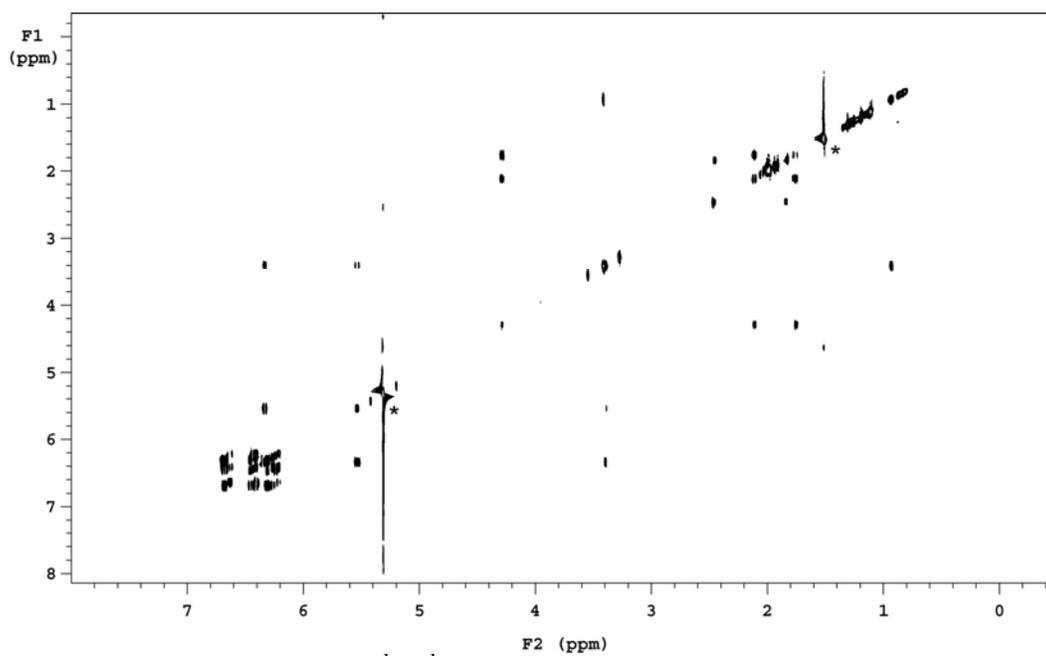
Supplemental Figure C3. 2D- ^1H , ^1H TOCSY NMR spectrum of carotenoid 5 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



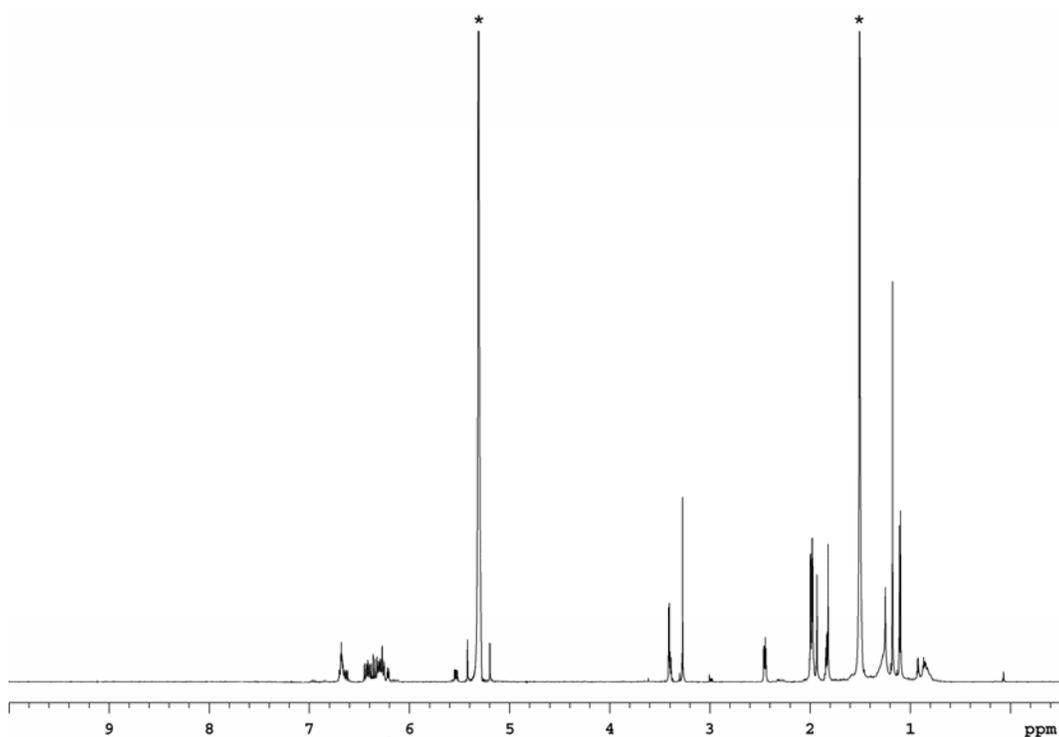
Supplemental Figure C4. 1D- ^1H NMR spectrum of carotenoid 4 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



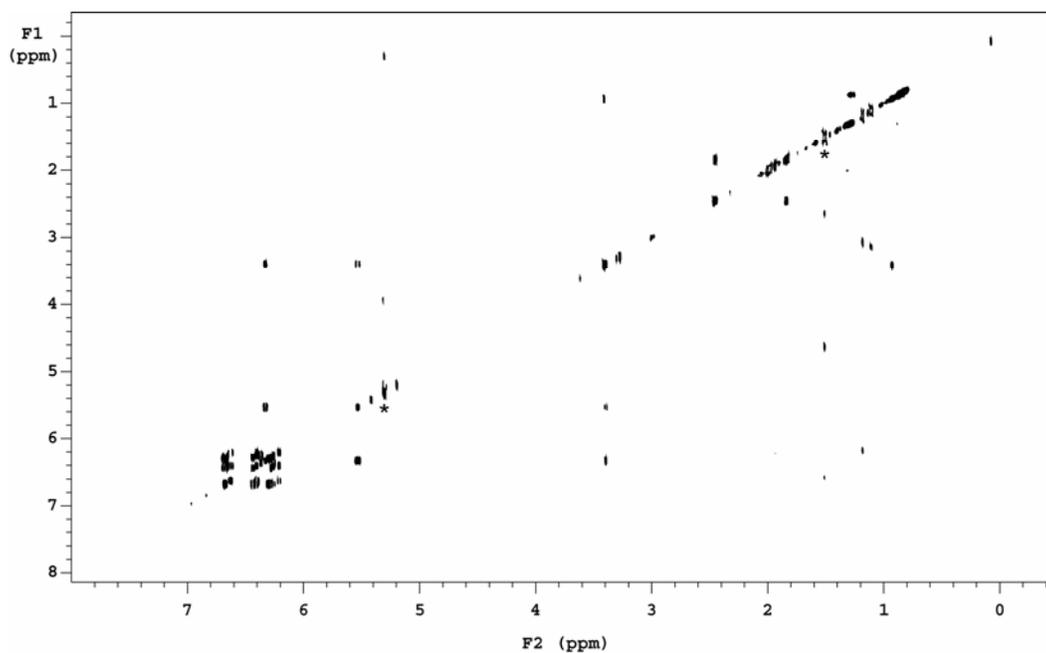
Supplemental Figure C5. 2D- ^1H , ^1H TOCSY NMR spectrum of carotenoid 4 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



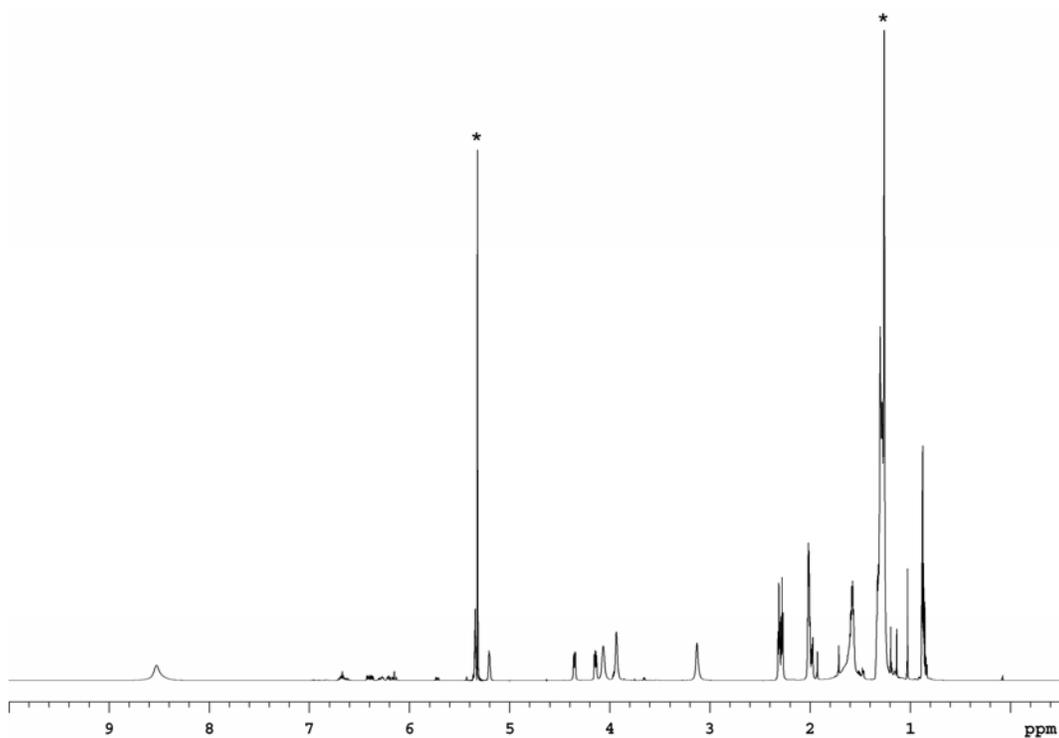
Supplemental Figure C6. 2D- ^1H , ^1H TOCSY NMR spectrum of carotenoid 6 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



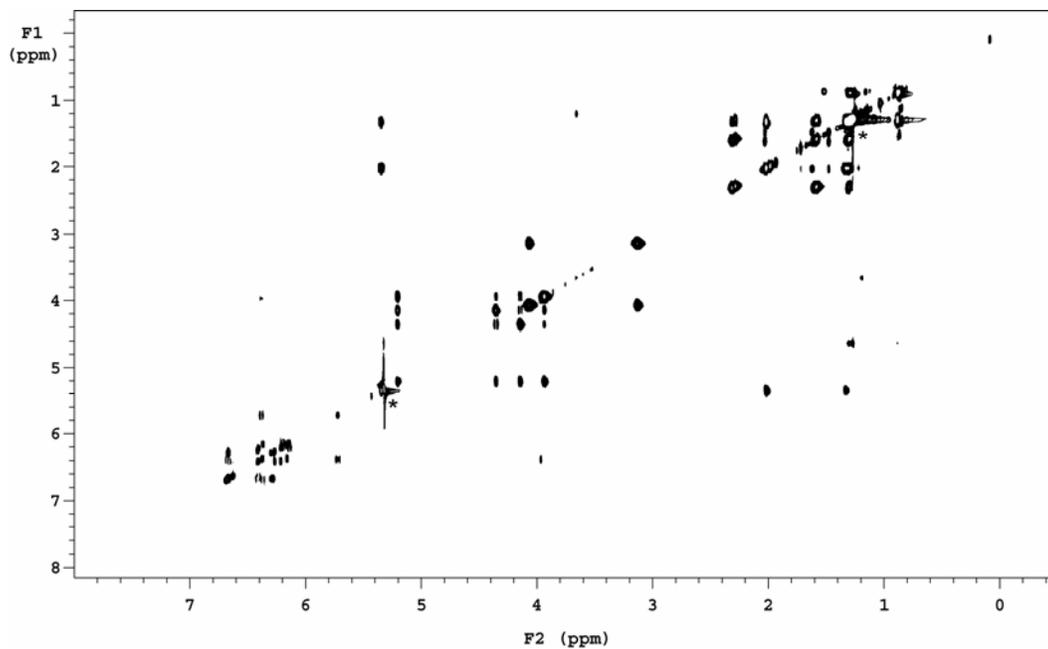
Supplemental Figure C7. 1D- ^1H NMR spectrum of carotenoid 7 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



Supplemental Figure C8. 2D- ^1H , ^1H TOCSY NMR spectrum of carotenoid 7 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



Supplemental Figure C9. 1D- ^1H NMR spectrum of carotenoid 8 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.



Supplemental Figure C10. 2D- ^1H , ^1H TOCSY NMR spectrum of carotenoid 8 dissolved in CD_2Cl_2 and referenced to the major solvent peak (defined as $\delta 5.31$). Solvent peaks are indicated with an asterisk.

Appendix D – Supplemental Data for Chapter 6

- Supplemental Table D1: Carotenoid biosynthetic proteins, their homologs and inferred and known biosynthetic pathway products.
- Supplemental Table D2: Start and end amino acids for carotenoid fussion proteins used in this study.
- Supplemental Figure D1: RAxML phylogeny of CrtY proteins.
- Supplemental Figure D2: RAxML phylogeny of CrtYcd proteins.
- Supplemental Figure D3: RAxML phylogeny of CrtL proteins.
- Supplemental Figure D4: RAxML phylogeny of CruA, CruB and CruP proteins.
- Supplemental Figure D5: RAxML phylogeny of CrtZ proteins.
- Supplemental Figure D6: RAxML phylogeny of CrtW proteins.
- Supplemental Figure D7: RAxML phylogeny of CrtC (a), CrtD (b) and CrtF (c) proteins.
- Supplemental Figure D8: RAxML phylogeny of CrtH (a) and CrtR (b) proteins.
- Supplemental Figure D9: Distributions of pairwise d_n/d_s values for phylogenetic groups described in the text expressed as a percentage of the total number of comparisons for each sequence cluster protein.
- Supplemental Figure D10: Pairwise d_n/d_s values for C40 carotenoid-producing Actinobacteria *crtYcd* (a), C50 carotenoid-producing Actinobacteria *crtYef* (b) and myxobacterial *crtB* (c), *crtC* (d), *crtD* (e) and *crtI* (f).

Supplemental Table D1. Carotenoid biosynthetic protein homologs and the (inferred) products of their corresponding biosynthetic pathways. IMG locus and GenBank accession numbers are given in the same order as their corresponding protein sequences. Carotenoids and biosynthetic proteins for which experimental evidence exists are underlined and the corresponding references indicated. Proteins leading to the production of apocarotenoids other than neurosporaxanthin are omitted. Also indicated are the presence of a detected rhodopsin homolog in an organism's genome and whether the genome analyzed was completed at the time of study.

Organism	Carotenoid biosynthesis protein homologs ^a	IMG Loci (plain text) or GenBank (<i>italics</i>) identifiers	Major carotenoid(s) produced ^{a, b}	Rhodopsin ^a	Complete genome	References
<i>α</i>-Proteobacteria						
Caulobacteriales						
<i>Brevundimonas</i> sp. SD212	<u>CrtB</u> , <u>CrtG</u> , <u>CrtI</u> , <u>CrtW</u> , <u>CrtY</u> , <u>CrtZ</u>	<i>BAD99409</i> , <i>BAD99415</i> , <i>BAD99408</i> , <i>BAD99406</i> , <i>BAD99407</i> , <i>BAD99414</i>	<u>2-Hydroxyastaxanthin</u>	N/A	N/A	(Yokoyama et al. 1996, Nishida et al. 2005)
<i>Brevundimonas vesicularis</i> DC263	<u>CrtG</u> , <u>CrtI</u> , <u>CrtW</u> , <u>CrtY</u> , <u>CrtZ</u>	<i>ABC50107</i> , <i>ABC50114</i> , <i>ABC50116</i> , <i>ABC50115</i> , <i>ABC50108</i>	<u>2,2'-Dihydroxyastaxanthin</u> , <u>2,2'-Dihydroxyadonixanthin</u>	N/A	N/A	(Tao et al. 2006a)
<i>Caulobacter</i> sp. K31	CrtB, CrtG, CrtI, CrtY, CrtZ	CaulDRAFT_3785, CaulDRAFT_4828, CaulDRAFT_3784, CaulDRAFT_3783, CaulDRAFT_3786	Caloxanthin/Nostoxanthin	N	Y	
Parvularculales						
<i>Parvularcula bermudensis</i> HTCC2503	CrtB, CrtI, CrtW, CrtY, CrtZ	PB2503_11014, PB2503_11019, PB2503_11009, PB2503_11024, PB2503_11004	Astaxanthin	N	N	
Rhizobiales - Aurantimonadaceae						
<i>Aurantimonas</i> sp. SI85-9A1	CrtB, CrtI, CrtW, CrtX, CrtY, CrtZ	SI859A1_00462, SI859A1_00461, SI859A1_00535, SI859A1_00459, SI859A1_00460, SI859A1_00536	Astaxanthin glycoside	N	N	
<i>Fulvimarina pelagi</i> HTCC2506	CrtB, CrtC, CrtD, CrtF, CrtI, CrtW, CrtX, CrtY, CrtZ	FP2506_14039, FP2506_11697, FP2506_11702, FP2506_11707, FP2506_14044, FP2506_12624, FP2506_12634, FP2506_14049, FP2506_12644	Astaxanthin glycoside, Spirilloxanthin	Y	N	
Rhizobiales - Bradyrhizobiaceae						
<i>Bradyrhizobium</i> sp. BTAi1	CrtB, CrtC, CrtD, CrtF, CrtI	BBta_6444, BBta_6442, BBta_6441, BBta_6439, BBta_6445	<u>Spirilloxanthin</u>	N	Y	(Takaichi 1999)

<i>Bradyrhizobium</i> sp. ORS278	<u>CrtB</u> (2x), <u>CrtC</u> , <u>CrtD</u> , <u>CrtE</u> , <u>CrtI</u> (2x), <u>CrtW</u> , <u>CrtY</u>	BRADO1611, <i>AAF78202</i> , <i>AAR98493</i> , <i>AAR98494</i> , BRADO1616, <i>AAR98491</i> , <i>AF218415</i> , <i>AAF78203</i> , <i>AAF78200</i>	<u>Canthaxanthin</u> , <u>Spirilloxanthin</u>	N	Y	(Takaichi 1999, Hannibal et al. 2000, Giraud et al. 2004)
<i>Rhodopseudomonas</i> <i>palustris</i> BisA53	CrtB, CrtC, CrtD, CrtF, CrtI	RPE_1316, RPE_1320, RPE_1323, RPE_1325, RPE_1315	Spirilloxanthin	N	N	
<i>Rhodopseudomonas</i> <i>palustris</i> BisB18	CrtB, CrtC, CrtD, CrtF, CrtI	RPC_1262, RPC_1288, RPC_1289, RPC_1291, RPC_1261	Spirilloxanthin	N	N	
<i>Rhodopseudomonas</i> <i>palustris</i> BisB5	CrtB, CrtC, CrtD, CrtF, CrtI	RPD_3765, RPD_3761, RPD_3760, RPD_3758, RPD_3766	Spirilloxanthin	N	N	
<i>Rhodopseudomonas</i> <i>palustris</i> CGA009	CrtB, CrtC, CrtD, CrtF, CrtI	RPA1513, RPA1517, RPA1518, RPA1520, RPA1512	Spirilloxanthin	N	N	
<i>Rhodopseudomonas</i> <i>palustris</i> HaA2	CrtB, CrtC, CrtD, CrtF, CrtI	RPB_4010, RPB_4006, RPB_4005, RPB_4003, RPB_4011	Spirilloxanthin	N	N	
<i>Rhodopseudomonas</i> <i>palustris</i> TIE-1	CrtB, CrtC, CrtD, CrtF, CrtI	<i>ZP_02303360</i> , <i>ZP_02303365</i> , <i>YP_001990709</i> , <i>ZP_02303368</i> , <i>ZP_02303359</i>	Spirilloxanthin	N	N	
Rhizobiales - Methylobacteriaceae						
<i>Methylobacterium</i> sp. 4- 46	CrtB, CrtC, CrtD, CrtF, CrtI	M446DRAFT_2531, M446DRAFT_2575, M446DRAFT_2574, M446DRAFT_2572, M446DRAFT_2532	Spirilloxanthin	Y	Y	
<i>Methylobacterium</i> <i>chloromethanicum</i> CM4	CrtB, CrtC, CrtD, CrtF, CrtI	MchlDRAFT_3516, MchlDRAFT_3862, MchlDRAFT_3773, MchlDRAFT_3771, MchlDRAFT_3517	Spirilloxanthin	N	N	
<i>Methylobacterium</i> <i>extorquens</i> AM1	<u>CrtI</u>	<i>AAQ65246</i>	?	N/A	N/A	(Van Dien et al. 2003)
<i>Methylobacterium</i> <i>extorquens</i> PA1	CrtB, CrtC, CrtD, CrtF, CrtI	MextDRAFT_0800, MextDRAFT_3418, MextDRAFT_2857, MextDRAFT_2859, MextDRAFT_0801	Spirilloxanthin	N	Y	
<i>Methylobacterium populi</i> BJ001	CrtB, CrtC, CrtD, CrtF, CrtI	<i>ZP_02198499</i> , <i>ZP_02200220</i> , <i>ZP_02200038</i> , <i>ZP_02200040</i> , <i>ZP_02198498</i>	Spirilloxanthin	N	N	
Rhizobiales - Phyllobacteriaceae						
<i>Hoeflea phototrophica</i> DFL-43	CrtA(P), CrtB, CrtC, CrtD, CrtF, CrtI	<i>ZP_02167527</i> , <i>ZP_02167509</i> , <i>ZP_02167511</i> , <i>ZP_02167512</i> , <i>ZP_02167515</i> , <i>ZP_02167508</i>	Spirilloxanthin/2,2'- Diketospirilloxanthin, Spheroidene/Spheroidenone	N	N	
Rhizobiales - Xanthobacteraceae						
<i>Xanthobacter</i> <i>autotrophicus</i> Py2	<u>CrtB</u> , <u>CrtI</u> , CrtX, <u>CrtY</u> , CrtZ	<i>AAL02001</i> , <i>AAL02000</i> , Xaut_3579, <i>AAL01999</i> , Xaut_4535	Zeaxanthin diglucoside	N	Y	(Larsen et al. 2002)

Rhodobacterales - Rhodobacteraceae

<i>α-Proteobacterium</i> HTCC2255	CrtB (2x), CrtI (2x), CrtY	OM2255_09541, OM2255_14570, OM2255_09536, OM2255_14565, OM2255_09546	<u>β-Carotene</u>	Y	N	
<i>Dinoroseobacter shibae</i> DFL 12	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	DshiDRAFT_2007, DshiDRAFT_2014, DshiDRAFT_2502, DshiDRAFT_2017, DshiDRAFT_2019, DshiDRAFT_2013	Spheroidene/ <u>Spheroidenone</u>	N	Y	(Biebl et al. 2005)
<i>Jannaschia</i> sp. CCS1	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	Jann_0145, Jann_0142, Jann_0184, Jann_0183, Jann_0181, Jann_0143	Spheroidene/Spheroidenone	N	Y	
<i>Loktanella vestfoldensis</i> SKA53	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	SKA53_12108, SKA53_12123, SKA53_12308, SKA53_12313, SKA53_12323, SKA53_12118	Spheroidene/Spheroidenone	N	N	
<i>Paracoccus</i> <i>haeundaensis</i>	<u>CrtB, CrtI, CrtW,</u> <u>CrtY, CrtZ</u>	AAY28421, AAY28420, AAY28417, AAY28419, AAY28418	<u>Astaxanthin</u>	N/A	N/A	(Lee et al. 2004, Lee and Kim 2006)
<i>Paracoccus</i> sp. N81106	<u>CrtB, CrtI, CrtW,</u> <u>CrtY, CrtZ</u>	P54975, BAA09594, BAA09591, BAA09593, BAA09592	<u>Astaxanthin</u>	N/A	N/A	(Yokoyama et al. 1994, Misawa et al. 1995)
<i>Paracoccus</i> sp. PC1	<u>CrtW, CrtZ</u>	Q44261, Q44262	<u>Astaxanthin</u>	N/A	N/A	(Yokoyama et al. 1994, Misawa et al. 1995)
<i>Paracoccus</i> <i>zeaxanthinifaciens</i> R1534	<u>CrtB, CrtI, CrtY, CrtZ</u>	AAC44849, AAC44850, AAC44851, AAC44852	<u>Zeaxanthin</u>	N/A	N/A	(Pasamontes et al. 1997, Berry et al. 2003)
<i>Rhodobacter capsulatus</i>	<u>CrtA(R), CrtB, CrtC,</u> <u>CrtD, CrtF, CrtI</u>	1613414A, 1613414C, 1613414E, 1613414F, 1613414H, 1613414B	<u>Spheroidene/Spheroidenone</u>	N	N	(Armstrong et al. 1989, Takaichi 1999)
<i>Rhodobacter</i> <i>sphaeroides</i> 2.4.1	CrtA(R), <u>CrtB, CrtC,</u> <u>CrtD, CrtF, CrtI</u>	640071482, AAB31139, AAF24292, AAF24293, AAF24295, AAF24289	<u>Spheroidene, Spheroidenone</u>	N	Y	(Lang et al. 1994, Lang et al. 1995, Takaichi 1999)
<i>Rhodobacter</i> <i>sphaeroides</i> ATCC 17025	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	Rsph17025_1022, Rsph17025_1025, Rsph17025_2041, Rsph17025_2040, Rsph17025_2038, Rsph17025_1024	Spheroidene/Spheroidenone	N	Y	
<i>Roseobacter</i> <i>denitrificans</i> OCh 114	CrtB, CrtC, CrtD, CrtF, CrtI	RD1_0119, RD1_0116, RD1_0115, RD1_0113, RD1_0120	Spirilloxanthin/ <u>2,2'-</u> <u>Diketoxanthin,</u> Spheroidene/ <u>Spheroidene</u>	N	Y	(Takaichi 1999)

<i>Roseobacter litoralis</i> Och 149	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	ZP_02142877, ZP_02142880, ZP_02142883, ZP_02142884, ZP_02142886, ZP_02142879	Spheroidene/ <u>Spheroidenone</u>	N	N	(Takaichi 1999)
<i>Roseobacter</i> sp. AzwK- 3b	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	RAZWK3B_19701, RAZWK3B_19686, RAZWK3B_19676, RAZWK3B_19671, RAZWK3B_19661, RAZWK3B_19691	Spheroidene/Spheroidenone	N	N	
<i>Roseobacter</i> sp. CCS2	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	RCCS2_09379, RCCS2_09364, RCCS2_08389, RCCS2_08384, RCCS2_08374, RCCS2_09369	Spheroidene/Spheroidenone	N	N	
<i>Roseovarius</i> sp. 217	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	ROS217_22377, ROS217_22362, ROS217_22352, ROS217_22347, ROS217_22337, ROS217_22367	Spheroidene/Spheroidenone	N	N	
<i>Roseovarius</i> sp. TM1035	CrtA(R), CrtB, CrtC, CrtD, CrtF, CrtI	RTM1035_06838, RTM1035_06853, RTM1035_12198, RTM1035_12193, RTM1035_12183, RTM1035_06848	Spheroidene/Spheroidenone	N	N	
Rhodobacterales - Rhodospirillaceae						
<i>Magnetospirillum</i> <i>magnetotacticum</i> MS-1	CrtB, CrtD, CrtI, CrtI	Magn03002575, Magn03004919, Magn03002576, Magn03006152	?	Y?	N	
<i>Rhodospirillum rubrum</i> ATCC 11170	CrtB, CrtC, CrtD, CrtF, CrtI	Rru_A0494, Rru_A2985, Rru_A2984, Rru_A2982, Rru_A0493	<u>Spirilloxanthin</u>	N	Y	(Takaichi 1999)
Rickettsiales						
Candidatus <i>Pelagibacter</i> <i>ubique</i> HTCC1002	CrtB, CrtI, CrtY	ZP_01264713, PU1002_05751, PU1002_05741	β -Carotene	Y	N	
Candidatus <i>Pelagibacter</i> <i>ubique</i> HTCC1062	CrtB, CrtI, CrtY	YP_265548, SAR11_0120, YP_265549	<u>β-Carotene^c</u>	<u>Y</u>	Y	(Béjà et al. 2000)
Sphingomonadales						
α -Proteobacterium BAL199	CrtI, CrtY	ZP_02189377, ZP_02189374	β -Carotene	Y	N	
<i>Erythrobacter litoralis</i> HTCC2594	CrtB, CrtG, CrtI, CrtW, CrtY, CrtZ	ELI_09895, ELI_12610, ELI_09885, ELI_03320, ELI_09880, ELI_03325	Erythroanthin sulphate	N	Y	
<i>Erythrobacter longus</i> OCH101	<u>CrtI</u> , <u>CrtY</u>	BAA20276, BAA20275	<u>β-Carotene, Caloxanthin, Erythroanthin sulphate, Spirilloxanthin, Zeaxanthin</u>	N/A	N/A	(Takeyama et al. 1996, Takaichi 1999)
<i>Erythrobacter</i> sp. NAP1	CrtB, CrtC, CrtD, CrtF, CrtG, CrtI, CrtW, CrtY, CrtZ	NAP1_10293, NAP1_09287, NAP1_09292, NAP1_09297, NAP1_13433, NAP1_10278, NAP1_09087, NAP1_10273, NAP1_09082	<u>Erythroanthin sulphate</u> , Spirilloxanthin		N	(Koblížek et al. 2003)
<i>Erythrobacter</i> sp. SD-21	CrtB, CrtG, CrtI, CrtY, CrtZ	ED21_21989, ED21_19052, ED21_22004, ED21_22009, ED21_23836	Caloxanthin/Nostoxanthin	N	N	

<i>Novosphingobium aromaticivorans</i> DSM 12444	CrtB, CrtG, CrtI, CrtY, CrtZ	Saro_1814, Saro_0236, Saro_1816, Saro_1817, Saro_1168	Caloxanthin/Nostoxanthin	N	Y	
<i>Sphingomonas</i> sp. SKA58	CrtB, CrtG, CrtI, CrtY, CrtZ	SKA58_15537, SKA58_00155, SKA58_15527, SKA58_15522, SKA58_17557	Caloxanthin/Nostoxanthin	N	N	
<i>Sphingopyxis alaskensis</i> RB2256	CrtB, CrtG, CrtI, CrtY, CrtZ	Sala_3132, Sala_3136, Sala_3134, Sala_3135, Sala_2128	Caloxanthin/Nostoxanthin	N	Y	
Other α-Proteobacteria						
<i>Methylomonas</i> sp. 16a	<u>CrtN</u> , <u>CrtNb</u> , <u>Ald</u>	<i>AAX46183, AAX46185, AAX46184</i>	<u>4,4'-Diapolycopene-4-oic acid</u> , <u>4,4'-Diapolycopene-4,4'-dioic acid</u>	N/A	N/A	(Tao et al. 2005)
<i>Methylophilales</i> Bacterium HTCC2181	CrtB, CrtI, CrtY	MB2181_00930, MB2181_00935, <i>ZP_01551534</i>	β -Carotene	Y	N	
<i>Thiocapsa roseopersicina</i>	<u>CrtC</u> , <u>CrtD</u> , <u>CrtF</u>	<i>AAP50935, AAP59036, AAP59038</i>	<u>Spirilloxanthin</u>	N/A	N/A	(Takaichi 1999, Kovács et al. 2003)
β-Proteobacteria						
<i>Rubrivivax gelatinosus</i> S1	<u>CrtA(P)</u> , <u>CrtB</u> , <u>CrtC</u> , <u>CrtD</u> , <u>CrtE</u> , <u>CrtI</u>	<i>AAO93123, AAB87738, AAO93124, AAC44798, AAO93114, AAO93135</i>	<u>Spirilloxanthin/2,2'-Diketoxanthin</u> , <u>Spheroidene/Spheroidenone</u>	N/A	N/A	(Ouchane et al. 1997a, Ouchane et al. 1997b, Takaichi 1999, Steiger et al. 2000, Harada et al. 2001a, Harada et al. 2001b, Pinta et al. 2003, Steiger et al. 2003, Stickforth and Sandmann 2007, Gerjets et al. 2009)
Uncultured Marine Bacterium EBO_41B09	CrtB, CrtI, CrtY	<i>ABL97760, ABL97761, ABL97759</i>	β -Carotene	Y	N/A	
γ-Proteobacteria						
<i>Congregibacter litoralis</i> KT71	CrtB, CrtC, CrtD, CrtF, CrtI	KT71_19483, KT71_19478, KT71_07854, KT71_19468, KT71_19488	<u>Spirilloxanthin</u>	N	N	(Fuchs et al. 2007)

<i>Enterobacter sakazakii</i> ATCC BAA-894	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	ESA_00342, ESA_00343, ESA_00345, ESA_00344, ESA_00341	Zeaxanthin glucoside	N/A	N/A	(Lehner et al. 2006)
<i>Enterobacteriaceae</i> Bacterium DC260	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	AAZ73131, AAZ73130, AAZ73128, AAZ73129, AAZ73132	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Sedkova et al. 2005)
<i>Enterobacteriaceae</i> Bacterium DC404	<u>CrtB</u> , <u>CrtI</u> , <u>CrtY</u> , <u>CrtZ</u>	AAZ73137, AAZ73136, AAZ73135, AAZ73138	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Sedkova et al. 2005)
<i>Enterobacteriaceae</i> Bacterium DC413	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	AAZ73150, AAZ73149, AAZ73147, AAZ73148, AAZ73151	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Sedkova et al. 2005)
<i>Enterobacteriaceae</i> Bacterium DC416	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	AAZ73143, AAZ73142, AAZ73140, AAZ73141, AAZ73144	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Sedkova et al. 2005)
<i>Halorhodospira</i> <i>halophila</i> SL1	CrtB, CrtC, CrtD, CrtF, CrtI	Hhal_1618, Hhal_1615, Hhal_1614, Hhal_1612, Hhal_1619	<u>Spirilloxanthin</u>	N	Y	(Takaichi 1999)
Marine γ - Proteobacterium HTCC2080	CrtB, CrtC, CrtD, CrtF, CrtI	MGP2080_10278, MGP2080_10273, MGP2080_10308, MGP2080_10263, MGP2080_10283	Spirilloxanthin	N	N	
Marine γ - Proteobacterium HTCC2143	CrtB, CrtI, CrtY	GP2143_03303, GP2143_03298, GP2143_03308	β -Carotene	Y	N	
Marine γ - Proteobacterium HTCC2207	CrtB, CrtI, CrtY	GB2207_10296, GB2207_10301, GB2207_10291	β -Carotene	Y	N	
<i>Marinobacter</i> sp. ELB17	CrtB, CrtI, CrtY	MELB17_06699, MELB17_06694, MELB17_06704	β -Carotene	Y	Y	
<i>Pantoea agglomerans</i> Eho10	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	AAA21264, AAA21263, AAA64979, AAA21262, AAA64983	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Hundle et al. 1991, Hundle et al. 1994)
<i>Pantoea ananatis</i> ATCC 19321	<u>CrtB</u> , <u>CrtI</u> , <u>CrtX</u> , <u>CrtY</u> , <u>CrtZ</u>	P21683, BAA14127, BAA14125, BAA14126, BAA14129	<u>Zeaxanthin glucoside</u>	N/A	N/A	(Misawa et al. 1990)
<i>Photobacterium</i> sp. SKA34	CrtB, CrtI, CrtY	SKA34_07099, SKA34_07104, SKA34_07094	β -Carotene	Y	N	
<i>Photorhabdus</i> <i>luminescens</i> subsp. <i>laumondii</i> TTO1	CrtB, CrtI, CrtY	plu4343, plu4342, plu4341	β -Carotene	N	Y	
<i>Pseudomonas stutzeri</i> A1501	CrtB, CrtI, CrtX, CrtY, CrtZ	PST_3872, PST_3873, PST_3875, PST_3874, PST_3871	Zeaxanthin glucoside	N	N	

<i>Vibrio angustum</i> S14	CrtB, CrtI, CrtY	VAS14_08785, VAS14_08780, VAS14_08790	β -Carotene	Y	N
<i>Vibrio campbellii</i> AND4	CrtB, CrtI, CrtY	ZP_02194908, ZP_02194909, ZP_02194907	β -Carotene	Y	Y
<i>Vibrio harveyi</i> ATCC BAA-1116	CrtB, CrtI, CrtY	VIBHAR_02163, VIBHAR_02162, YP_001445356	β -Carotene	Y	N
δ-Proteobacteria					
<i>Bdellovibrio bacteriovorus</i> HD100	CrtB, CrtG, CrtI, CrtY	Bd1725, Bd1729, Bd1724, Bd1730	Xanthophyll	N	Y
Myxobacteria					
<i>Myxococcus xanthus</i> DK 1622 ^d	CrtB, CrtC, CrtD, CrtI (3x) CrtW, CrtYc, CrtYd	MXAN_0896, MXAN_0898, MXAN_0897, MXAN_0895, MXAN_4052, MXAN_7517, MXAN_6049, YP_629162, YP_629161	Myxobactone, 4-Ketotorulene	N	Y
<i>Plesiocystis pacifica</i> SIR-1	CrtB, CrtC, CrtD, CrtI, CrtL, CrtU	PPSIR1_32939, PPSIR1_30569, PPSIR1_30564, PPSIR1_32944, PPSIR1_19284, PPSIR1_36107	Chlorobactene-like?	N	N
<i>Sorangium cellulosum</i>	CrtB, CrtC, CrtD, CrtI, CrtZ, CruA	YP_001611218, YP_001611221, YP_001611222, YP_001611217, YP_001611220, YP_001611219	Myxol	N	Y
<i>Stigmatella aurantiaca</i> DW4/3-1	CrtB, CrtC, CrtD, CrtI (2x), CrtYcd	STIAU_8492, STIAU_8490, STIAU_8491, STIAU_8493, STIAU_0585, ZP_01463227	Myxol-like	N	N
Unclassified Proteobacteria					
Uncultured Marine Bacterium HF10_19P19	CrtB, CrtI, CrtY	ABL60985, ABL60986, ABL60984	β -Carotene	Y	N
Uncultured Marine Bacterium HF10_25F10	CrtB, CrtI, CrtY	ABL61010, ABL61009, ABL61011	β -Carotene	Y	N
Bacteroidetes					
Unidentified Eubacterium SCB49	CrtB, CrtI, CrtYcd, CrtZ	SCB49_01402, SCB49_01397, SCB49_01412, SCB49_01407	Zeaxanthin	N	N
Flavobacteria					
<i>Cellulophaga</i> sp. MED134	CrtB, CrtI, CrtY, CrtZ	MED134_13071, MED134_13076, MED134_08681, MED134_13066	<u>Zeaxanthin</u> ^c	<u>Y</u>	N
<i>Croceibacter atlanticus</i> HTCC2559	CrtB, CrtI, CrtY, CrtZ	CA2559_00920, CA2559_00915, CA2559_00930, CA2559_00925	Zeaxanthin	N	N
Flavobacteria Bacterium BAL38	CrtB, CrtI, CrtY, CrtZ	FBBAL38_01010, FBBAL38_01005, FBBAL38_05810, FBBAL38_01015	Zeaxanthin	Y	N

(Botella et al. 1995, Iñesta et al. 2007, 2008)

(Gómez-Consarnau et al. 2007)

Flavobacteria Bacterium BBFL7	CrtA(P), CrtB, CrtD, CrtI, CrtY, CrtZ	BBFL7_00795, BBFL7_00790, BBFL7_00792, BBFL7_00789, BBFL7_00794, BBFL7_00791	Myxol	N	N	
<i>Flavobacteriales</i> Bacterium ALC-1	CrtB, CrtI, CrtYcd, CrtZ	ZP_02183100, ZP_02183098, ZP_02183102, ZP_02183101	Zeaxanthin	N	N	
<i>Flavobacteriales</i> Bacterium HTCC2170	CrtB, CrtI, CrtYcd, CrtZ	FB2170_07629, FB2170_07624, FB2170_07639, FB2170_07634	Zeaxanthin	N	N	
<i>Flavobacterium johnsoniae</i> UW101	CrtB, CrtI, CrtY, CrtZ	Fjoh_0058, Fjoh_0057, Fjoh_0926, Fjoh_0059	Zeaxanthin	N	Y	
<i>Flavobacterium psychrophilum</i> JIP02/86	CrtB, CrtI, CrtY, CrtZ	FP1450, FP1449, FP1447, FP1451	Zeaxanthin	N	Y	
<i>Gramella forsetii</i> KT0803	CrtB, CrtI, CrtY, CrtZ	orf2472, orf2471, orf2474, orf2473	Zeaxanthin	N	Y	
<i>Kordia algicida</i> OT-1	CrtB, CrtI, CrtY, CrtZ	ZP_02163393, ZP_02163392, ZP_02163354, ZP_02163394	Zeaxanthin	N	N	
<i>Leeuwenhoekiella blandensis</i> MED217	CrtB, CrtI, CrtY, CrtZ	MED217_11644, MED217_11639, MED217_17665, MED217_11649	Zeaxanthin	N	N	
Marine Bacterium P99-3	<u>CrtA(P), CrtD, CrtE, CrtI, CrtL, CrtZ</u>	BAC77674, BAC77671, AAR98496, BAC77668, BAC77673, BAC77670	<u>Myxol</u>	N/A	N/A	(Yokoyama and Miki 1995, Teramoto et al. 2003, Teramoto et al. 2004)
<i>Polaribacter irgensii</i> 23-P	CrtA(P), CrtB, CrtD, CrtI, CrtY, CrtZ	PI23P_00055, PI23P_12132, PI23P_11812, PI23P_12137, PI23P_04492, PI23P_11807	Myxol	Y	N	
<i>Psychroflexus torquis</i> ATCC 700755	CrtA(P), CrtB (2x), CrtD, CrtI, CrtY (2x), CrtYcd (2x)	P700755_03257, P700755_03232, P700755_26055, P700755_03242, P700755_03227, P700755_03252, P700755_07212, P700755_26050, P700755_32979	Myxol	Y	N	
<i>Robiginitalea biformata</i> HTCC2501	CrtB, CrtI, CrtY, CrtZ	RB2501_11842, RB2501_11847, RB2501_05565, RB2501_11837	Zeaxanthin	N	N	
<i>Tenacibaculum</i> sp. MED152	CrtA(P), CrtB, CrtD, CrtI, CrtY, CrtZ (2x)	ZP_01052136, MED152_02670, MED152_02575, MED152_02675, MED152_02660, MED152_02565, MED152_02665	<u>Myxol</u> ^c	<u>Y</u>	N	(Gómez-Consarnau et al. 2007)
Sphingobacteria <i>Algoriphagus</i> sp. KK10020C	<u>CrtI, CrtW, CrtYcd</u>	AAB88949, ABB88952, ABB88950	<u>Flexixanthin</u>	N/A	N/A	(Tao et al. 2006b)

<i>Algoriphagus</i> sp. PR1	CrtB, CrtD, CrtI, CrtW, CrtYcd, CrtZ	ALPR1_00745, ALPR1_03175, ALPR1_00750, ALPR1_00730, ALPR1_00740, ALPR1_00765	Ketomyxol-like	N	N	
<i>Cytophaga hutchinsonii</i> ATCC 33406	CrtB, CrtI, CrtYcd, CrtZ	CHU_2036, CHU_2033, CHU_2039, CHU_2038	Zeaxanthin	N	Y	
<i>Microscilla marina</i> ATCC 23134	CrtB, CrtD, CrtI, CrtYcd, CrtZ	M23134_07222, M23134_07225, M23134_07221, M23134_07228, M23134_07227	Myxol-like	N	N	
<i>Pedobacter</i> sp. BAL39	CrtB, CrtI, CrtY, CrtZ	PBAL39_25495, PBAL39_25490, PBAL39_23642, PBAL39_25510	Zeaxanthin	N	N	
<i>Salinibacter ruber</i> DSM 13855	CrtB, CrtI (2x), CrtO, CrtYcd	SRU_1430, SRU_0743, SRU_2060, SRU_1502, YP_445624	<u>Salinixanthin</u>	<u>Y</u>	Y	(Lutnaes et al. 2002, Balashov et al. 2005)
Firmicutes						
Bacillales						
<i>Bacillus</i> sp. NRRL B- 14911	CrtM, CrtN (2x), CrtNb, CrtO(at), CrtQ(gt)	B14911_18170, B14911_18150, B14911_20973, B14911_18165, ZP_01169745, ZP_01169743	Staphyloxanthin	N	N	
<i>Bacillus pumilus</i> SAFR- 032	CrtM, CrtN, CrtNb	YP_001486256, BPUM_1011, YP_001487908	4,4'-Diapolycopene oxide/4,4'- Diaponeurosporene oxide	N	Y	
<i>Bacillus selenitireducens</i> MLS10	CrtM, CrtN, CrtNb	ZP_02170848, ZP_02170846, ZP_02171655	4,4'-Diapolycopene oxide/4,4'- Diaponeurosporene oxide	N	N	
<i>Oceanobacillus</i> <i>ihyensii</i> HTE831	CrtM, CrtN, CrtNb, CrtO(at), CrtQ(gt)	OB2460, OB2461, OB2459, NP_693378, NP_693379	Staphyloxanthin	N	Y	
<i>Exiguobacterium</i> <i>sibiricum</i> 255-15	CrtM, CrtN, CrtNb	ZP_00538724, ExigDRAFT_1694, ExigDRAFT_2090	4,4'-Diapolycopene oxide/4,4'- Diaponeurosporene oxide	Y	Y	
<i>Staphylococcus aureus</i> <i>aureus</i> Newman ^c	<u>CrtM</u> , <u>CrtN</u> , <u>CrtNb</u> , <u>CrtO(at)</u> , <u>CrtQ(gt)</u>	CAA52097, CAA52098, Q2FV57, NWMN_2465, Q53590	<u>Staphyloxanthin</u>	N	Y	(Wieland et al. 1994, Pelz et al. 2005, Tao et al. 2005)
<i>Staphylococcus</i> <i>haemolyticus</i> JCSC1435	CrtM, CrtN, CrtNb, CrtO(at), CrtQ(gt)	SH0490, SH0491, SH0488, YP_252402, SH0489	Staphyloxanthin	N	Y	
Lactobacillales						
<i>Carnobacterium</i> sp. AT7	CrtM, CrtN, CrtNb, CrtO(at), CrtQ(gt)	ZP_02184019, ZP_02184020, ZP_02184018, ZP_02184015, ZP_02184016	Staphyloxanthin	N	N	

<i>Lactobacillus plantarum</i> WCFS1	CrtM, CrtN	lp_3263, lp_3262	4,4'-Diapolycopene/4,4'- Diaponeurosporene	N	Y	
<i>Leuconostoc citreum</i> KM20	CrtM, CrtN	YP_001728065, YP_001728064	4,4'-Diapolycopene/4,4'- Diaponeurosporene	N	Y	
<i>Leuconostoc mesenteroides</i> subsp. <i>mesenteroides</i> ATCC 8293	CrtM, CrtN	LEUM_1047, LEUM_1046	4,4'-Diapolycopene/4,4'- Diaponeurosporene	N	Y	
Clostridiales-Heliobacteriaceae						
<i>Heliobacterium modesticaldum</i> Icel	CrtN	YP_001679882	<u>4,4'-Diaponeurosporene</u>	N	?	(Takaichi et al. 1997)
Mollicutes-Acholeplasmatales						
<i>Acholeplasma laidlawii</i> PG-8A	CrtM, CrtN	YP_001621382, YP_001621383	4,4'-Diapolycopene/4,4'- Diaponeurosporene	N	Y	
Actinobacteria						
Marine Actinobacterium PHSC20C1	CrtB, CrtEb, CrtI, CrtYe, CrtYf	A20C1_09079, A20C1_09059, A20C1_09074, ZP_01129023, ZP_01129022	Decaprenoxanthin/C.P.450	N	N	
<i>Rubrobacter xylanophilus</i> DSM 9941	CrtB, CrtI, CrtY (2x)	Rxyl_0844, Rxyl_0845, Rxyl_2038, YP_645780	β -Carotene	Y	Y	
Corynebacterineae						
<i>Corynebacterium diphtheriae</i> NCTC 13129	CrtB, CrtI	DIP1870, DIP1871	?	N	Y	
<i>Corynebacterium efficiens</i> YS-314	CrtB, CrtEb, CrtI, CrtYe, CrtYf	CE0641, CE0637, CE0640, CE0638, CE0639	Decaprenoxanthin	N	Y	
<i>Corynebacterium glutamicum</i> ATCC 13032 ^f	<u>CrtB</u> (2x), <u>CrtEb</u> , <u>CrtI</u> (2x, one broken), <u>CrtYe</u> , <u>CrtYf</u>	cg0721, cg2672, cg0717, cg0720, cg2668+cg2670 (fused), cg0718, cg0719	<u>Decaprenoxanthin</u>	N	Y	(Krubasik et al. 2001)
<i>Corynebacterium jeikeium</i> K411	CrtB, CrtI	jk0515, jk0735	?	N	Y	
<i>Dietia</i> sp. CQ4	<u>CrtB</u> , <u>CrtEbYe</u> (LitAB), <u>CrtI</u> , <u>CrtL</u> , <u>CrtYf</u> (LitC)	ABD24399, ABD24402, ABD24400, ABD24404, ABD24401	<u>Canthaxanthin</u> , C.P.450	N/A	N/A	(Tao et al. 2007)

<i>Gordonia</i> sp. TM414	<u>CrtB</u> , <u>CrtI</u>	BAC75675, BAC75676	?	N/A	N/A	(Matsui and Maruhashi 2004)
<i>Mycobacterium aurum</i> A+	<u>CrtB</u> , <u>CrtI</u> , <u>CrtU</u> , <u>CrtYc</u> , <u>CrtYd</u>	CAB94795, CAB94794, CAB94798, CAB94797, CAB94796	<u>Isorenieratine</u>	N/A	N/A	(Viveiros et al. 2000)
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10 ^h	CrtB, CrtI, CrtU, CrtYc, CrtYd	MAP3071, MAP3070, MAP3075, MAP3072, MAP3073	Isorenieratine	N	Y	
<i>Mycobacterium gilvum</i> PYR-GCK	CrtB, CrtI, CrtL, CrtU	Mflv_1846, Mflv_1847, Mflv_0950, Mflv_1844	Isorenieratine	N	Y	
<i>Mycobacterium smegmatis</i> MC2 155	CrtB, CrtI, CrtU	MSMEG_2346, MSMEG_2347, MSMEG_2344	Isorenieratine?	N	Y	
<i>Mycobacterium</i> sp. MCS ^g	CrtB, CrtI, CrtU, CrtYc, CrtYd	Mmcs_5076, Mmcs_5075, Mmcs_5080, Mmcs_5077, Mmcs_5078	Isorenieratine	N	Y	
<i>Mycobacterium ulcerans</i> Agy99	CrtB, CrtYc, CrtYd	MUL_0375, MUL_0376, MUL_0377	?	N	Y	
<i>Mycobacterium vanbaalenii</i> PYR-1	CrtB, CrtI, CrtL, CrtU	Mvan_1579, Mvan_1578, Mvan_5914, Mvan_1580	Isorenieratine	N	Y	
<i>Nocardia farcinica</i> IFM 10152	CrtB (2x), CrtI (2x), CrtL (2x), CrtO	nfa17350, nfa43980, nfa17370, nfa43990, nfa7290, nfa34900, nfa17530	Canthaxanthin	N	Y	
<i>Rhodococcus erythropolis</i> AN12	<u>CrtI</u> , <u>CrtL</u> , <u>CrtO</u>	AAW23161, AAR98749, AAW23159	<u>4-Keto-γ-carotene</u>	N/A	N/A	(Tao and Cheng 2004, Tao et al. 2004)
<i>Rhodococcus</i> sp. RHA1	CrtB, CrtI, CrtL, CrtO	RHA1_ro01109, RHA1_ro01107, RHA1_ro07203, RHA1_ro01101	4-Keto- γ -carotene	N	Y	
Frankineae						
<i>Frankia alni</i> ACN14a	CrtB, CrtI, CrtYc, CrtYd	FRAAL2154, FRAAL2160, YP_715252, YP_715253	β -Carotene	N	Y	
<i>Frankia</i> sp. CcI3	CrtB, CrtI, CrtYc, CrtYd	Franci3_1383, Franci3_1387, YP_482155, YP_482154	β -Carotene	N	Y	
<i>Frankia</i> sp. EAN1pec	CrtB, CrtI, CrtYc, CrtYd	Franean1DRAFT_1529, Franean1DRAFT_1533, YP_001506187, YP_001506188	β -Carotene	N	Y	
<i>Kineococcus radiotolerans</i> SRS30216	CrtB, CrtI, CrtY	Krad_3229, Krad_3228, Krad_0091	β -Carotene	Y	Y	

Micrococcineae						
<i>Brevibacterium linens</i> ATCC 9175	<u>CrtB</u> , <u>CrtI</u> , <u>CrtU</u> , <u>CrtYe</u> , <u>CrtYd</u> , ORF10	AAF65581, AAF65582, AAF65586, AAF65588, AAF65587, AAF65589	<u>3,3'-Dihydroxyisorenieratine</u>	N/A	N/A	(Kohl et al. 1983, Krubasik and Sandmann 2000, Cheng 2006)
<i>Brevibacterium linens</i> BL2	CrtB, CrtI, CrtU, CrtYe, CrtYd, ORF10	BlinB01002637, BlinB01002636, BlinB01002631, ZP_00378971, ZP_00378970, ZP_00378969	3,3'-Dihydroxyisorenieratine	N	N	
<i>Clavibacter</i> <i>michiganensis</i> subsp. <i>michiganensis</i> NCPPB 382 ⁱ	CrtB, CrtEb, CrtI, CrtYe, CrtYf	CMM_2887, CMM_2884, CMM_2887, YP_001223630, YP_001223631	Decaprenoxanthin/C.P.450	N	Y	
<i>Leifsonia xyli</i> subsp. <i>xyli</i> CTCB07	CrtB, CrtEb, CrtI, CrtYe, CrtYf	Lxx15620, Lxx15580, Lxx15610, YP_062469, YP_062470	Decaprenoxanthin/C.P.450	N	Y	
<i>Arthrobacter aurescens</i> TC1	CrtB, CrtEb, CrtI	AAur_0319, AAur_0315, AAur_0318	Linear C50 carotenoids	N	Y	
Micromonosporineae						
<i>Salinispora arenicola</i> CNS205	CrtB, CrtI, CrtU, CrtY	SareDRAFT_4686, SareDRAFT_4684, SareDRAFT_0271, SareDRAFT_2473	Isorenieratine	N	Y	
<i>Salinispora tropica</i> CNB-440	CrtB, CrtI, CrtU, CrtY	Strop_4441, Strop_4439, Strop_0241, Strop_2408	Isorenieratine	N	Y	
Pseudonocardineae						
<i>Saccharopolyspora</i> <i>erythraea</i> NRRL 2338	CrtB (2x), CrtI, CrtL, CrtU	SACE_3269, SACE_3539, SACE_1713, SACE_2184, SACE_3271	Isorenieratine	N	Y	
Streptomycineae						
<i>Streptomyces avermitilis</i> MA-4680	CrtB, CrtI, CrtU, CrtY	SAV1024, SAV1023, SAV1019, SAV1021	Isorenieratine	N	Y	
<i>Streptomyces coelicolor</i> A3(2)	CrtB, CrtI, CrtU, CrtY	SCO0187, NP_639818, SCO0186, SCO0189, SCO0191	Isorenieratine	N	Y	
<i>Streptomyces griseus</i> NCBI 3933	<u>CrtB</u> , <u>CrtI</u> , <u>CrtU</u> , <u>CrtY</u>	P54977, CAA64850, CAA64853, CAA64855	<u>Isorenieratine</u>	N/A	N/A	(Krügel et al. 1999)
<i>Streptomyces rochei</i> 7434AN4 plasmid pSLA2-L	CrtB, CrtI, CrtU, CrtY	pSLA2-L_p105, pSLA2-L_p106, pSLA2-L_p110, pSLA2-L_p108	Isorenieratine	N/A	N/A	

Streptosporangineae

<i>Thermobifida fusca</i> YX	CrtB, CrtI, CrtU, CrtY	Tfu_3076, Tfu_3075, Tfu_3090, Tfu_3088	Isorenieratine	N	Y
------------------------------	------------------------	--	----------------	---	---

Cyanobacteria**Chroococcales**

<i>Crocospaera watsonii</i> WH 8501	CrtB, CrtD, CrtH, CrtP, CrtQ, CrtR, CruA, CruE, CruF, CruG, CruH, CruP	CwatDRAFT_0948, CwatDRAFT_6111, CwatDRAFT_4423, CwatDRAFT_0947, CwatDRAFT_5404, CwatDRAFT_5424, CwatDRAFT_6211, CwatDRAFT_3579, CwatDRAFT_6133, CwatDRAFT_6134, CwatDRAFT_4939, CwatDRAFT_1341	β -Carotene, Myxol-like, Synechoxanthin, Zeaxanthin	N	N
<i>Cyanothece</i> sp. CCY 0110	CrtB, CrtD, CrtG, CrtH, CrtO, CrtP, CrtQ, CrtR, CruA, CrtE, CrtF, CrtG, CrtH, CruP	CY0110_11242, CY0110_22577, CY0110_09580, CY0110_24336, CY0110_15365, Cy0110_11237, CY0110_26552, CY0110_08481, CY0110_10722, CY0110_29874, CY0110_00310, CY0110_00315, CY0110_21165, CY0110_12337	β -Carotene, Caloxanthin/Nostoxanthin, Echinenone/Canthaxanthin, Myxol-like, Synechoxanthin, Zeaxanthin	Y	N
<i>Microcystis aeruginosa</i> NIES-843	CrtB, CrtH, CrtO, CrtP, CrtQ, CrtR, CrtX, CruA, CruF, CruG, CruP	YP_001656306, YP_001656250, YP_001660920, YP_001656307, YP_001660324, YP_001655751, YP_001655539, YP_001655787, YP_001661102, YP_001661103, YP_001660458	β -Carotene, Echinenone/Canthaxanthin, Myxol-like, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. BL107	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	BL107_05144, BL107_11616, BL107_15440, BL107_05149, BL107_04944, BL107_08054, BL107_14110	β -Carotene, Echinenone/Canthaxanthin, Zeaxanthin	N	N
<i>Synechococcus</i> sp. CC9311	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	sync_2607, sync_1140, sync_0974, sync_2608, sync_2568, sync_0336, sync_1804	β -Carotene, Echinenone/Canthaxanthin, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. CC9605	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR	Syncc9605_2394, Syncc9605_1681, Syncc9605_1941, Syncc9605_2395, Syncc9605_2356, Syncc9605_0286	β -Carotene, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. CC9902	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	Syncc9902_0299, Syncc9902_1423, Syncc9902_0724, Syncc9902_0298, Syncc9902_0335, Syncc9902_2058, Syncc9902_0972	β -Carotene, Echinenone/Canthaxanthin, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. JA-2- 3Ba(2-13)	CrtB, CrtH, CrtP, CrtQ, CrtR, CruA, CruF, CruG, CruP	CYB_1695, CYB_1298, CYB_1694, CYB_1060, CYB_0102, CYB_0376, CYB_0539, CYB_0176, CYB_2530	β -Carotene, Myxol-like, Zeaxanthin	N	Y

<i>Synechococcus</i> sp. JA-3-3Ab	CrtB, CrtH, CrtP, CrtQ, CrtR, CruA, CruF, CruG, CruP	CYA_0317, CYA_1857, CYA_0316, CYA_0668, CYA_1931, CYA_0185, CYA_1532, CYA_0955, CYA_1571	β -Carotene, Myxol-like, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. PCC 7002	CrtB, CrtH, CrtP, CrtQ, CrtR, CrtW, CruA, CruE, CruF, CruG, CruH, CruP	YP_001735179, YP_001735135, YP_001735178, YP_001733792, YP_001734175, YP_001736033, YP_001733316, YP_001734502, YP_001735274, YP_001735273, YP_001735480, YP_001735389	<u>β-Carotene, Caloxanthin, Synechoxanthin, Zeaxanthin</u>	N	Y
<i>Synechococcus</i> sp. RCC307	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	SynRCC307_0239, SynRCC307_1542, SynRCC307_0743, SynRCC307_0238, SynRCC307_0275, SynRCC307_2209, SynRCC307_1993	β -Carotene, Echinenone/Canthxanthin, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. RS9916	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR	RS9916_35877, RS9916_31147, RS9916_31637, RS9916_35882, RS9916_35677, RS9916_39311	β -Carotene, Zeaxanthin	N	N
<i>Synechococcus</i> sp. RS9917	CrtB, CrtG, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	RS9917_07105, RS9917_09626, RS9917_00762, RS9917_01237, RS9917_07100, RS9917_07310, RS9917_03663, RS9917_00687	β -Carotene, Caloxanthin/Nostoxanthin, Echinenone/Canthxanthin, Zeaxanthin	N	N
<i>Synechococcus</i> sp. WH 5701	CrtB, CrtG, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	WH5701_00830, WH5701_02055, WH5701_09029, WH5701_08084, WH5701_00835, WH5701_00665, WH5701_01215, WH5701_04005	β -Carotene, Caloxanthin/Nostoxanthin, Zeaxanthin	N	N
<i>Synechococcus</i> sp. WH 7803	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	SynWH7803_2269, SynWH7803_1380, SynWH7803_1588, SynWH7803_2273, SynWH7803_2224, SynWH7803_0337, SynWH7803_0928	β -Carotene, Echinenone/Canthxanthin, Zeaxanthin	N	Y
<i>Synechococcus</i> sp. WH 7805	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	WH7805_11093, WH7805_03627, WH7805_04751, WH7805_11088, WH7805_11278, WH7805_07481, WH7805_01197	β -Carotene, Echinenone/Canthxanthin, Zeaxanthin	N	N
<i>Synechococcus</i> sp. WH 8102	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CrtW	SYNW2256, SYNW0901, SYNW0728, SYNW2257, SYNW2213, SYNW0291, SYNW1368	β -Carotene, Echinenone/Canthxanthin, Zeaxanthin	N	Y

(Maresca et al. 2007, Takaichi and Mochimaru 2007, Maresca et al. 2008a)

<i>Synechococcus elongatus</i> PCC 7942 ^j	<u>CrtB</u> , <u>CrtG</u> , CrtH, <u>CrtL</u> , <u>CrtP</u> , CrtQ, CrtR, CruP	P37269, Synpcc7942_0680, Synpcc7942_1246, CAA52677, CAA39004, Synpcc7942_1512, Synpcc7942_2439, Synpcc7942_0652	<u>β-Carotene, Zeaxanthin, Caloxanthin/Nostoxanthin</u>	N	Y	(Chamovitz et al. 1991, Chamovitz et al. 1992, Cunningham Jr. et al. 1994, Takaichi and Mochimaru 2007, Maresca et al. 2008a) (Martínez-Férez and Vioque 1992, Martínez-Férez et al. 1994, Fernández-González et al. 1997, Breitenbach et al. 1998, Masamoto et al. 1998, Breitenbach et al. 2001, Masamoto et al. 2001, Maresca et al. 2008a)
<i>Synechocystis</i> sp. PCC 6803	CrtB, <u>CrtD</u> , CrtG, <u>CrtH</u> , <u>CrtO</u> , <u>CrtP</u> , <u>CrtQ</u> , <u>CrtR</u> , CrtX, CruA, <u>CruE</u> , CruF, <u>CrtG</u> , <u>CruH</u> , CruP	slr1255, BAA16840, slr0224, BAA10798, BAA10561, CAA44452, P74306, BAA17468, NP_439972, sl10659, sl10253, sl10814, sl11004, cbaB, sl0147	<u>β-Carotene, Caloxanthin/Nostoxanthin, Echinenone, Myxol, Synechoxanthin, Zeaxanthin</u>	N	Y	(Takaichi and Mochimaru 2007, Iwai et al. 2008)
<i>Thermosynechococcus elongatus</i> BP-1	CrtB, CrtD, CrtG, CrtP, CrtQ, CrtR, CruA, CruF, <u>CruG</u>	tl11560, tl10232, tlr1917, tl11561, tl10337, tlr1900, tlr1139, tlr0414, tlr2019	<u>β-Carotene, Caloxanthin/Nostoxanthin, Hydroxymyxol 2'-fucoside, Myxol 2'-fucoside, Zeaxanthin/β-Cryptoxanthin</u>	N	Y	(Steiger et al. 2005, Tsuchiya et al. 2005)
Gloeobacteria						
<i>Gloeobacter violaceus</i> PCC 7421	<u>CrtB</u> , CrtD, <u>CrtL</u> , CrtO, CrtW, CruA, CruE, CruF, CruG, CruH, CruP	BAC89685, gl12874, BAC88808, gvip032, gvip239, gl12484, gl11923, glr1356, glr1357, gl11922, gl13598	<u>β-Carotene, Oscillol, Synechoxanthin</u>	Y	Y	(Takaichi and Mochimaru 2007, Iwai et al. 2008)

Nostocales

<i>Anabaena variabilis</i> ATCC 29413	CrtB, CrtD, CrtH, CrtO, CrtP, CrtQ, CrtR, CrtW (2x), CruA, CruE, CruF, CruH, CruP	Ava_4794, Ava_2342, Ava_3112, Ava_1581, Ava_4795, Ava_0200, Ava_1693, Ava_2048, Ava_3888, Ava_3214, Ava_0036, Ava_1513, Ava_1827, Ava_4521	<u>β-Carotene, Canthaxanthin, Echinenone, 4-Hydroxymyxol, Myxol, Synechoxanthin</u>	N	Y	(Takaichi et al. 2006)
<i>Nodularia spumigena</i> CCY9414	CrtB, CrtD, CrtH, CrtO, CrtP, CrtQ, CrtR, CrtW, CruA, CruE, CruF, CruG, CruH	N9414_14563, N9414_14318, N9414_19217, N9414_21450, N9414_14558, N9414_04980, N9414_01572, N9414_07726, N9414_16776, N9414_21696, N9414_16701, N9414_16696, N9414_03388	β -Carotene, Canthaxanthin/Echinenone, Ketomyxol-like, Synechoxanthin, Zeaxanthin	N	N	
<i>Nostoc</i> sp. PCC 7120	CrtB, CrtD, CrtH, <u>CrtO</u> , CrtP, CrtQ, <u>CrtQ-1</u> , CrtR, <u>CrtW</u> (2x), CruA, CruE, CruF, CruG, CruH, CruP	alr1833, all5123, alr2064, all3744, alr1832, all2382, <i>BAA05091</i> (plasmid), alr4009, alr3189, <i>BAB78246</i> (plasmid), alr0920, alr2785, all0144, all0143, all3866, alr3524	<u>β-Carotene, Echinenone, Ketomyxol glycoside, Myxol glycoside, Synechoxanthin</u>	<u>Y</u>	Y	(Linden et al. 1993, Jung et al. 2003, Mochimaru et al. 2005, Takaichi et al. 2005)
<i>Nostoc punctiforme</i> PCC 73102	CrtB, CrtD, CrtH, CrtO (2x), CrtP, CrtQ, CrtR, <u>CrtW</u> (2x), CruA, CruF, CruG, CruP	Npun02003603, Npun02007228, Npun02004596, Npun02000238, Npun02000982, Npun02003602, Npun02000603, Npun02006805, <i>ZP_00111258</i> , <i>ZP_00345866</i> , Npun02000486, Npun02002145, Npun02002146, Npun02004577	<u>β-Carotene, Echinenone, Ketomyxol glycoside, Myxol glycoside</u>	N	N	(Steiger and Sandmann 2004, Takaichi et al. 2005)
Oscillatoriales						
<i>Lyngbya</i> sp. PCC 8106	CrtB, CrtD, CrtH, CrtI, CrtO, CrtP, CrtQ, CrtR, CruA, CruE, CruF, CruG, CruH, CruP	L8106_05750, L8106_04976, L8106_26227, L8106_14465, L8106_26402, L8106_05755, L8106_07831, L8106_30215, L8106_12075, L8106_26267, L8106_14400, L8106_14395, L8106_26272, L8106_18117	β -Carotene, Canthaxanthin/Echinenone, Myxol-like, Synechoxanthin, Zeaxanthin	N	N	
<i>Trichodesmium</i> <i>erythraeum</i> IMS101	CrtB, CrtD, CrtH, CrtP, CrtQ, CrtR, CrtX, CruA, CruP	Tery_4010, Tery_4343, Tery_2192, Tery_4011, Tery_3954, Tery_2925, <i>YP_721820</i> , Tery_0494, Tery_1762	β -Carotene, Zeaxanthin	N	Y	

Prochlorales

<i>Prochlorococcus marinus</i> AS9601	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	A9601_01601, A9601_12211, A9601_11691, A9601_06891, A9601_01611, A9601_01331, A9601_02571	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9211	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	P9211_08067, P9211_05912, P9211_03117, P9211_05097, P9211_08062, P9211_08202, P9211_07547	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9215	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	P9215_01601, P9215_12511, P9215_11991, P9215_07161, P9215_01611, P9215_01331, P9215_02581	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9301	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	P9301_01621, P9301_12221, P9301_11701, P9301_06601, P9301_01631, P9301_01321, P9301_02581	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9312	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	PMT9312_0145, PMT9312_1126, PMT9312_1075, PMT9312_0633, PMT9312_0146, PMT9312_0118, PMT9312_0238	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9313 ^k	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	PMT2003, PMT1051, PMT1773, PMT1123, PMT2004, PMT1968, PMT1816	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> MIT 9515	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	P9515_01711, P9515_12061, P9515_11541, P9515_06991, P9515_01721, P9515_01291, P9515_02681	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> NATL2A ¹	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	PMN2A_1509, PMN2A_0636, PMN2A_0688, PMN2A_0073, PMN2A_1510, PMN2A_1484, PMN2A_1603	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> subsp. <i>marinus</i> CCMP1375	CrtB, CrtH, CrtLb, CrtLe, CrtP, CrtQ, CrtR	Pro0166, Pro0584, Pro1136, Pro0790, Pro0167, Pro0136, Pro0266	α -Carotene, β -Carotene, Zeaxanthin	N	Y	
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> CCMP1986	CrtB, CrtH, <u>CrtLb</u> , <u>CrtLe</u> , CrtP, CrtQ, CrtR	PMM0143, PMM1115, <i>CAE19093</i> , <i>CAE19092</i> , PMM0144, PMM0115, PMM0236	α -Carotene, β -Carotene, <u>Zeaxanthin</u>	N	Y	(Stickforth et al. 2003)

Unclassified							
<i>Acaryochloris marina</i> MBIC11017	CrtB, CrtH, CrtL, CrtP, CrtQ, CrtR, CruA (2x), CruP	YP_001519114, YP_001515816, YP_001520326, YP_001519115, YP_001517998, YP_001517943, YP_001519700, YP_001516943, YP_001516710	<u>α-Carotene, Zeaxanthin</u>	N	Y	(Miyashita et al. 1997)	
Chlorobi							
<i>Chlorobium chlorochromatii</i> CaD3	CrtB, CrtC, CrtH, CrtP, CrtQ, CruA, CruC, CruD	Cag_1175, Cag_0393, Cag_1888, Cag_1188, Cag_1590, Cag_0265, Cag_0212, Cag_0595	Chlorobactene, 1'- Hydroxychlorobacetene/ glycoside/glycoside laurate	N	Y		
<i>Chlorobium ferrooxidans</i> DSM 13031	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruC, CruD	CferDRAFT_0237, CferDRAFT_1001, CferDRAFT_0390, CferDRAFT_0138, CferDRAFT_0317, CferDRAFT_2005, CferDRAFT_0992, CferDRAFT_0710, CferDRAFT_0466	<u>Chlorobactene, 1'- Hydroxychlorobacetene/</u> glycoside/glycoside laurate	N	N	(Takaichi 1999)	
<i>Chlorobium limicola</i> DSM 245	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruC, CruD	ClimDRAFT_1659, ClimDRAFT_1225, ClimDRAFT_1699, ClimDRAFT_2229, ClimDRAFT_2229, ClimDRAFT_1832, ClimDRAFT_2111, ClimDRAFT_0833, ClimDRAFT_0278	Chlorobactene, 1'- Hydroxychlorobacetene/ glycoside/glycoside laurate	N	N		
<i>Chlorobium phaeobacteroides</i> BS1	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruB, CruC, CruD	Cphamn1DRAFT_2436, Cphamn1DRAFT_2230, Cphamn1DRAFT_2570, Cphamn1DRAFT_2471, Cphamn1DRAFT_2632, Cphamn1DRAFT_2517, Cphamn1DRAFT_3315, Cphamn1DRAFT_2874, Cphamn1DRAFT_3003, Cphamn1DRAFT_2103	Chlorobactene, 1'- Hydroxychlorobacetene/ glycoside/glycoside laurate, Isorenieratene	N	N		
<i>Chlorobium phaeobacteroides</i> DSM 266	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, <u>CruA, CruB, CruC,</u> CruD	Cpha266_1738, Cpha266_0428, Cpha266_0874, Cpha266_1173, Cpha266_0830, Cpha266_0659, Cpha266_0192, Cpha266_0474, Cpha266_2356, Cpha266_1013	<u>Chlorobactene, 1'- Hydroxychlorobacetene/</u> glycoside/glycoside laurate, <u>Isorenieratene</u>	N	Y	(Takaichi 1999, Maresca et al. 2008b)	
<i>Chlorobium tepidum</i> TLS	<u>CrtB, CrtC, CrtH,</u> <u>CrtP, CrtQ, CrtU,</u> <u>CruA, CruC, CruD</u>	AAM72615, AAM71547, AAM71888, AAM72043, AAM72642, AAM71569, AAM71699, AAM73205, AAM72202	<u>Chlorobactene, 1'- Hydroxychlorobacetene/</u> <u>glycoside/glycoside laurate</u>	N	Y	(Takaichi 1999, Frigaard et al. 2004, Maresca and Bryant 2006, Maresca et al. 2007)	
<i>Pelodictyon luteolum</i> DSM 273	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruC, CruD	Plut_1356, Plut_1720, Plut_0626, Plut_1283, Plut_1415, Plut_0435, Plut_1700, Plut_0242, Plut_1188	<u>Chlorobactene, 1'- Hydroxychlorobacetene/</u> glycoside/glycoside laurate, <u>Isorenieratene</u>	N	N	(Takaichi 1999)	

<i>Pelodictyon phaeoclathratiforme</i> BU-1	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruB, CruC, CruD	PphaDRAFT_2098, PphaDRAFT_0849, PphaDRAFT_0104, PphaDRAFT_2540, PphaDRAFT_2291, PphaDRAFT_1677, PphaDRAFT_0832, PphaDRAFT_2724, PphaDRAFT_2155, PphaDRAFT_2526	Chlorobactene, 1'-Hydroxychlorobacetene/glycoside/glycoside laurate, <u>Isorenieratine</u>	N	N	(Takaichi 1999)
<i>Prosthecochloris aestuarii</i> DSM 271	CrtB, CrtC, CrtD, CrtH, CrtP, CrtU, CruA, CruC, CruD	PaesDRAFT_0157, PaesDRAFT_0561, PaesDRAFT_1494, PaesDRAFT_1466, PaesDRAFT_1265, PaesDRAFT_0420, PaesDRAFT_1729, PaesDRAFT_2266, PaesDRAFT_0285	Chlorobactene, 1'-Hydroxychlorobacetene/glycoside/glycoside laurate, <u>Isorenieratine</u>	N	N	(Takaichi 1999)
<i>Prosthecochloris vibrioformis</i> DSM 265	CrtB, CrtC, CrtH, CrtP, CrtQ, CrtU, CruA, CruC, CruD	Cvib_0699, Cvib_1502, Cvib_1153, Cvib_1032, Cvib_1233, Cvib_0486, Cvib_1484, Cvib_0308, Cvib_0769	Chlorobactene, 1'-Hydroxychlorobacetene/glycoside/glycoside laurate	N	Y	(Takaichi 1999)
Chloroflexi						
<i>Chloroflexus aggregans</i> DSM 9485	CrtI, CrtO, CrtY	CaggDRAFT_0197, CaggDRAFT_2957, CaggDRAFT_2028	<u>β-Carotene, γ-Carotene, 1-OH-γ-Carotene/glucoside, 4-Keto-1-OH-γ-Carotene</u>	N	N	(Takaichi 1999)
<i>Chloroflexus aurantiacus</i> J-10-fl	CrtB, CrtI, CrtO, CrtY	CaurDRAFT_2153, CaurDRAFT_2173, CaurDRAFT_270, CaurDRAFT_0959	<u>β-Carotene, γ-Carotene, 1-OH-γ-Carotene/glucoside/FA</u>	N	Y	(Takaichi 1999)
<i>Herpetosiphon aurantiacus</i> ATCC 23779	CrtB, CrtI, CrtO, CruA	HaurDRAFT_5057, HaurDRAFT_1159, HaurDRAFT_3603, HaurDRAFT_1188	<u>β-Carotene, γ-Carotene, 1-OH-γ-Carotene/glucoside/FA</u>	N	Y	(Kleining and Reichenbach 1977)
<i>Roseiflexus castenholzii</i> DSM 13941	CrtB, CrtI (2x), CrtL, CrtO	Reas_1752, Reas_1585, Reas_3488, Reas_2711, Reas_1486	<u>Methoxy-keto-myxocoxanthin, Keto-myxocoxanthin glucoside fatty acid ester</u>	N	Y	(Takaichi et al. 2001)
<i>Roseiflexus</i> sp. RS-1	CrtB, CrtI (2x), CrtL, CrtO	RoseRS_2117, RoseRS_0943, RoseRS_2155, RoseRS_2643, RoseRS_3475	Methoxy-keto-myxocoxanthin, Keto-myxocoxanthin glucoside fatty acid ester	Y	Y	
Deinococcus/Thermus						
<i>Deinococcus geothermalis</i> DSM 11300	CrtB, CrtD, CrtI, CrtL, CrtO	Dgeo_0523, Dgeo_2306, Dgeo_0524, Dgeo_0857, Dgeo_2310	Deinoxanthin	N	Y	

<i>Deinococcus radiodurans</i> R1	CrtB, <u>CrtD</u> , <u>CrtI</u> , <u>CrtL</u> , <u>CrtO</u>	DR0862, DR2250, AAF10439, AAF10377, AAF09686	<u>Deinoxanthin</u>	N	Y	(Tao and Cheng 2004, Tao et al. 2004, Xu et al. 2007, Tian et al. 2008)
<i>Thermus thermophilus</i> HB27	<u>CrtB (plasmid)</u> , CrtI, CrtYcd, <u>P450</u>	P37270, TT_P0066, YP_00643, CYP175A1	Zeaxanthin	N	Y	(Tabata et al. 1994, Blasco et al. 2004)
<i>Thermus thermophilus</i> HB8	<u>CrtB (plasmid)</u> , CrtI, CrtYcd, P450	TTHB101, TTHB109, YP_145343, TTHB103	Zeaxanthin	N	Y	(Tabata et al. 1994)
Planctomycetes						
<i>Gemmata obscuriglobus</i> UQM 2246	CrtN, CrtNb	ZP_02732075, ZP_02732075	4,4'-Diapolycopene oxide/4,4'- Diaponeurosporene oxide	N	?	
<i>Rhodopirellula baltica</i> SH 1	CrtN, CrtNb	NP_869339, NP_870237	4,4'-Diapolycopene oxide/4,4'- Diaponeurosporene oxide	N		
Uncultured Marine Bacterium HF10_49E08	CrtBI, CrtY	ABL97829, ABL97830	β -Carotene	Y	N/A	
Crenarchaeota						
<i>Metallosphaera sedula</i> DSM 5348	CrtB, CrtI, CrtYcd, CrtZ	YP_001191163, Msed_1073, Msed_1076, YP_001191162	Zeaxanthin	N	N	
<i>Picrophilus torridus</i> DSM 9790	CrtB, CrtI, CrtYcd, CrtZ	YP_024313, PTO1532, PTO1534, YP_024309	Zeaxanthin	N	Y	
<i>Sulfolobus acidocaldarius</i> DSM 639	CrtB, CrtI, CrtYcd, CrtZ	YP_256333, Saci_1732, Saci_1735, YP_256332	Zeaxanthin	N	Y	
<i>Sulfolobus solfataricus</i> P2	CrtB, CrtI, <u>CrtYcd</u> , CrtZ	NP_344224, SSO2907, SSO2904, SSO2906	Zeaxanthin	N	Y	(Hemmi et al. 2003)
Euryarchaeota						
<i>Methanoculleus marisnigri</i> JR1	CrtB, CrtEb, CrtI	YP_001046034, YP_001046035, Memar_0116	Linear C50 carotenoids	N	Y	
<i>Methanothermobacter thermautotrophicus</i> Delta H	CrtB, CrtEb, CrtI	NP_276914, NP_276915, MTH1807	Linear C50 carotenoids	N	Y	
Uncultured Marine Bacterium HF10_29C11	CrtB, CrtI, CrtY	ABL97779, ABL97778, ABL97780	β -Carotene	N/A	N/A	

Haloarchaea						
<i>Haloarcularia marismortui</i> ATCC 43049	CrtB, CrtEb, CrtI (2x), CrtYcd	rrnAC2069, rrnAC0320, rrnAC0321, rrnAC1902, YP_136628	β -Carotene ^c , Linear C50 carotenoids	Y	Y	
<i>Halobacterium</i> sp. NRC-1	CrtB (2x), CrtEb, CrtI (2x)	VNG1458G, VNG1680G, VNG1682C, VNG1684G, VNG1755G	Linear C50 carotenoids	Y	Y	
<i>Haloquadratum walsbyi</i> DSM 16790	CrtB, CrtEb, CrtI (2x), CrtYcd	HQ2860A, HQ2862A, HQ1794A, HQ2863A, YP_656805	β -Carotene, Linear C50 carotenoids	Y	Y	
<i>Halorubrum lacusprofundi</i> ATCC 49239	CrtB, CrtEb, CrtI, CrtYcd	HlacDRAFT_0809, HlacDRAFT_1351, HlacDRAFT_1352, ZP_02017097	β -Carotene, Linear C50 carotenoids	Y	N	
<i>Natronomonas pharaonis</i> DSM 2160	CrtB, CrtEb, CrtI (2x), CrtYcd	NP4770A, NP4766A, NP0204A, NP4764A, YP_325986	β -Carotene, Linear C50 carotenoids	Y	Y	
Fungi						
<i>Aspergillus niger</i>	CrtBYcd, CrtI, CAO-2, YLO-1	XP_001391172, XP_001391204, XP_001401639, XP_001389346	β -Carotene, Neurosporaxanthin	Y	Y	
<i>Aspergillus oryzae</i> RIB40	CrtBYcd, CrtI, CAO-2, YLO-1	XP_001824519, XP_001824518, XP_001821249, XP_001816713	β -Carotene, Neurosporaxanthin	Y	Y	
<i>Cercospora nicotianae</i> ATCC 18366	<u>CrtI</u>	AAB86988	<u>β-Carotene</u>	N/A	N	(Daub and Payne 1989, Ehrenshaft and Daub 1994) (Avalos et al. 1985,
<i>Gibberella fujikuroi</i> IMI58289	<u>CrtBYcd (CarRA), CrtI (CarB), CAO-2 (CrtT)</u>	CAD19988, CAD19989, CAL90971	β -Carotene, Neurosporaxanthin	N/A	N	Linnemannstöns et al. 2002, Prado-Cabrero et al. 2007)
<i>Gibberella zeae</i> PH-1	CrtBYcd, CrtI, CAO-2, YLO-1	XP_383242, XP_383241, XP_382801, XP_390136	β -Carotene, Neurosporaxanthin	Y	Y	
<i>Mucor circinelloides</i>	<u>CrtBM (CarRP)</u>	Q9UUQ6	<u>β-Carotene</u>	N/A	N	(Fraser et al. 1996, Velayos et al. 2000)

<i>Neurospora crassa</i>	<u>CrtBYcd (AL-2), CrtI (AL-1), CAO-2, YLO-1</u>	CAE76609, P21334, XP_001727958, XP_957628	<u>β-Carotene, Neurosporaxanthin</u>	N/A	N	(Schmidhauser et al. 1990, Schmidhauser et al. 1994, Saelices et al. 2007, Estrada et al. 2008)
<i>Phaeosphaeria nodorum</i> SN15	CrtBYcd, CrtI, CAO-2, YLO-1	XP_001791029, XP_001791032, XP_001792684, XP_001796173	β-Carotene, Neurosporaxanthin	Y	N	
<i>Phycomyces blakesleeanus</i>	<u>CrtBYcd (CarRA), CrtI (CarB)</u>	CAB86388, P54982	<u>β-Carotene</u>	N/A	N	(Shlomain et al. 1991, Ruiz-Hidalgo et al. 1997, Arrach et al. 2001)
<i>Podospora anserina</i>	CrtBYcd, CrtI, CAO-2, YLO-1	XP_00190692, XP_001906937, XP_00193611, XP_001903919	β-Carotene, Neurosporaxanthin	Y	Y	
<i>Ustilago maydis</i> 521	CrtBYcd, CrtI	XP_762434, XP_760357	β-Carotene	Y	Y	
<i>Xanthophyllomyces dendrorhous</i>	<u>CrtBYcd (CrtYB), CrtI</u>	AAV33923, AAA19428	<u>Astaxanthin</u>	N/A	N	(Andrews and Starr 1976, Verdoes et al. 1999a, Verdoes et al. 1999b)
Photosynthetic Eukaryotes						
<i>Chlamydomonas reinhardtii</i>	PSY, CRTISO, LYCB, LYCE, PDS, ZDS, BKT, CHYB, CruP (2x), ZEP	XP_001701192, XP_001698231, AAX54906, XP_001696529, XP_001690859, XP_001700786, XP_001698699, XP_001698698, XP_001692181, XP_001696289, AAO34404	β-Carotene, Epoxides, Lutein	Y?	Y	
<i>Cyanidioschyzon merolae</i>	PSY, CRTISO, LYCB, PDS, ZDS, CrtR, CruP	CMM166C, CMN268C, CMK050C, CMK151C, CMT061C, CMV041C, CMC032C	<u>Zeaxanthin</u>	N/A	Y	(Cunningham Jr. et al. 2007)
<i>Dunaliellia salina</i>	<u>LYCB</u>	ACA34345	<u>β-Carotene</u>	N/A	Y	(Mil'ko 1963, Ramos et al. 2008)
<i>Galdieria sulphuraria</i>	PSY, CRTISO, LYCB, PDS, CrtR	Contig06203.g3.t1, Gs12840.1, Contig02802.g56.t1, Gs46720.1, Gs39970.1	Zeaxanthin	N/A	N	

<i>Haematococcus pluvialis</i>	<u>PSY</u> , <u>BKT</u> , <u>CHYB</u>	<i>AAK15621, AAT35555, AAD54243</i>	<u>Astaxanthin</u>	N/A	N/A	(Kajiwara et al. 1995, Lotan and Hirschberg 1995, Linden 1999, Steinbrenner and Linden 2001)
<i>Ostreococcus lucimarinus</i> CCE9901	PSY, LYCB, LYCE, PDS (2x), CHYB, CruP, ZEP, VDE	<i>XP_001418049, XP_001422490, XP_001422489, XP_001421697, XP_001420014, XP_001419973, XP_001415446, ABO99857, ABO99997</i>	β -Carotene, Epoxides, Lutein	N	Y	
<i>Phaeodactylum tricorutum</i>	PSY, LYCB, PDS, ZDS, CruP, ZEP, VDE	<i>estExt_fgensch1_pg.C_chr_50182, estExt_Phatr1_ua_kg.C_chr_10036, fgensch1_pg.C_chr_24000086 (short), estExt_gwp_gw1.C_chr_80030 (short), fgensch1_pg.C_chr_4000121, fgensch1_pg.C_chr_4000464, fgensch1_pg.C_chr_4000463, estExt_thaps1_ua_kg.C_chr_50183, thaps1_ua_kg.chr_2000231, fgensch1_pg.C_chr_6000679 (short), fgensch1_pg.C_chr_1000517 (short), estExt_thaps1_ua_kg.C_chr_60282, thaps1_ua_kg.chr_8000103</i>	β -Carotene, Epoxides, Lutein	N/A	Y	
<i>Thalassiosira pseudonana</i>	PSY, LYCB, PDS, ZDS, ZEP, VDE	<i>estExt_fgensch1_pg.C_chr_50182, estExt_Phatr1_ua_kg.C_chr_10036, fgensch1_pg.C_chr_24000086 (short), estExt_gwp_gw1.C_chr_80030 (short), fgensch1_pg.C_chr_4000121, fgensch1_pg.C_chr_4000464, fgensch1_pg.C_chr_4000463, estExt_thaps1_ua_kg.C_chr_50183, thaps1_ua_kg.chr_2000231, fgensch1_pg.C_chr_6000679 (short), fgensch1_pg.C_chr_1000517 (short), estExt_thaps1_ua_kg.C_chr_60282, thaps1_ua_kg.chr_8000103</i>	β -Carotene, Epoxides, Lutein	N/A	Y	

^aUnderlined data indicates evidence present in the literature, with the relative references cited in the right-most column.

^bOnly pathway endproducts are indicated, unless a specific literature reference is cited, due to the difficulty in identifying accumulatory intermediates solely from sequence homology.

^c β -Carotene production is considered to be demonstrated in these organisms by the demonstration of a functional rhodopsin without the addition of exogenous carotenoid cofactors.

^d*Myxococcus xanthus* DK 1622 is considered representative of *M. xanthus* DK 1050, for which experimental evidence but no genome sequence is available, due to nearly 100% sequence homology between strains.

^e*Staphylococcus aureus* Newman is considered representative of *S. aureus* str. COL, JH1, JH9, MRSA252 (no CrtM), MSSA476, MW2, Mu3, Mu50, N315 (no CrtNb), NCTC 8325 and USA300 due to nearly 100% sequence homology between strains.

^f*Corynebacterium glutamicum* ATCC 13032 (Bielefeld) is considered representative of *C. glutamicum* str. ATCC 13032 (Kitasato), R. and MJ233C, for which experimental evidence but no second CrtI homolog exists, due to nearly 100% sequence homology between strains.

^g*Mycobacterium* sp. MCS is considered representative of *Mycobacterium* spp. KMS and JLS due to nearly 100% sequence homology between strains.

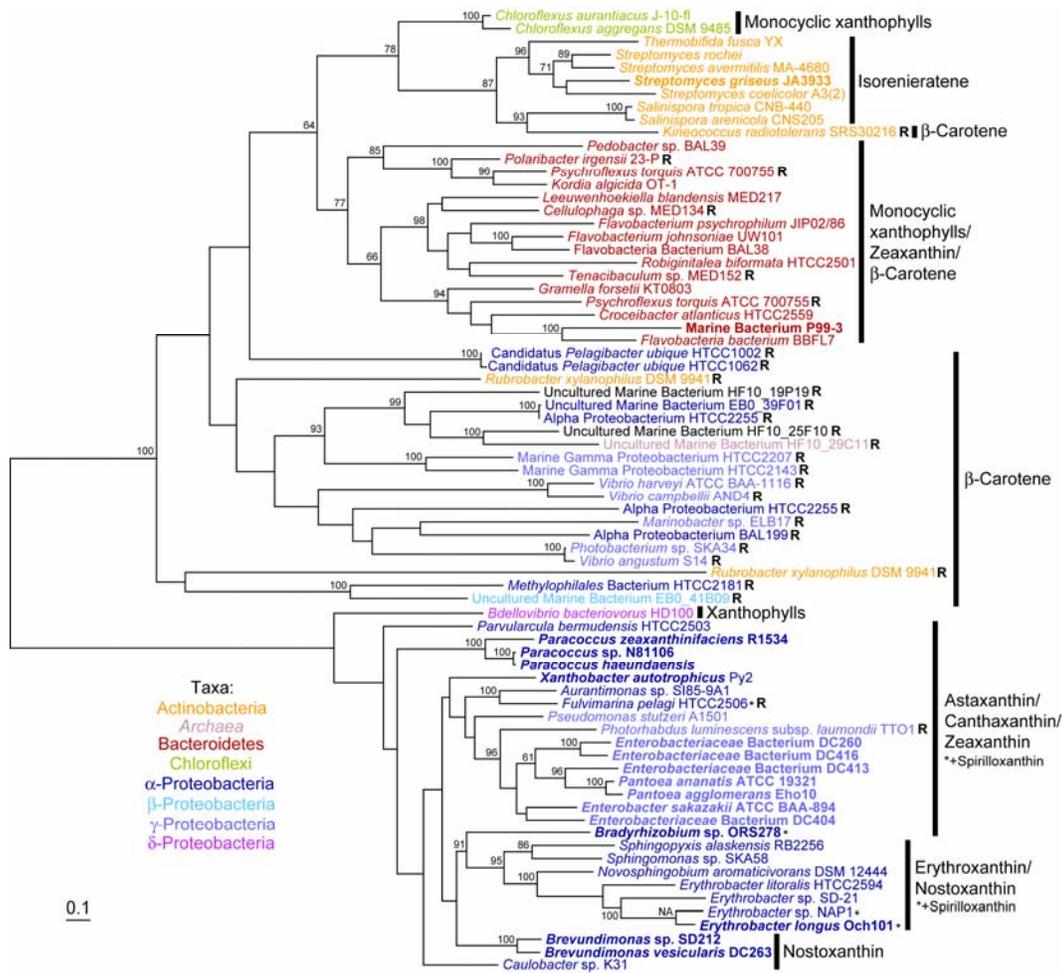
^h*Mycobacterium avium* subsp. Paratuberculosis K-10 is considered representative of *M. avium* 104 due to nearly 100% sequence homology between strains.

ⁱ*Clavibacter michiganensis michiganensis* NCPPB 382 is considered representative of *C. michiganensis* subsp. *sepedonicus* due to nearly 100% sequence homology between strains.

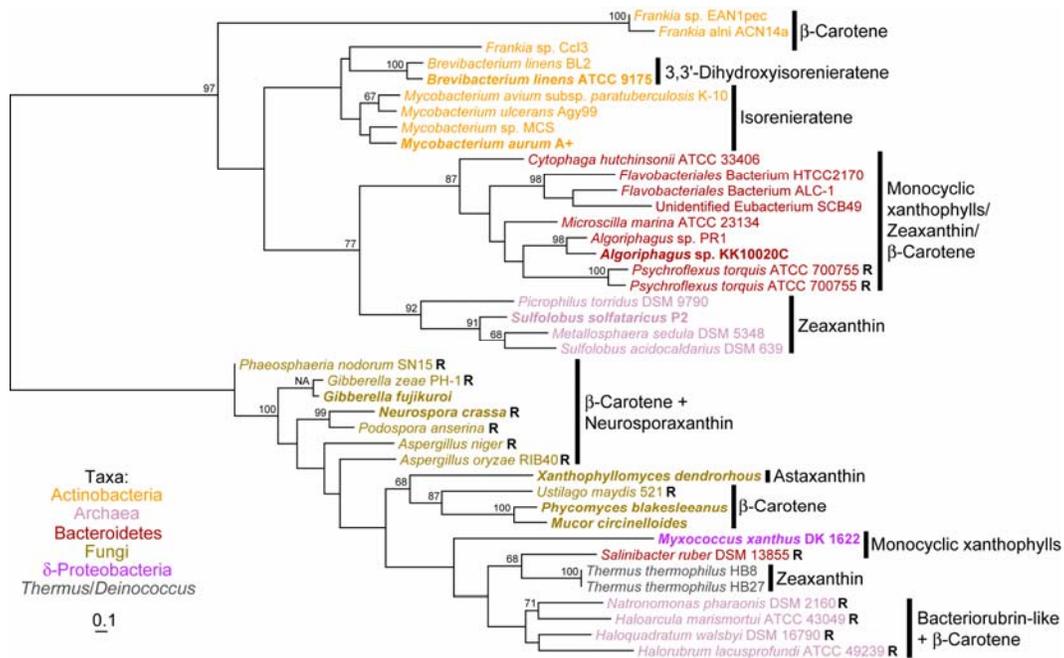
^j*Synechococcus elongatus* PCC 7942 is considered representative of *S. elongatus* PCC 6301 due to nearly 100% sequence homology between strains.

^l*Prochlorococcus marinus* MIT 9313 is considered representative of *P. marinus* MIT 9303 due to nearly 100% sequence homology between strains.

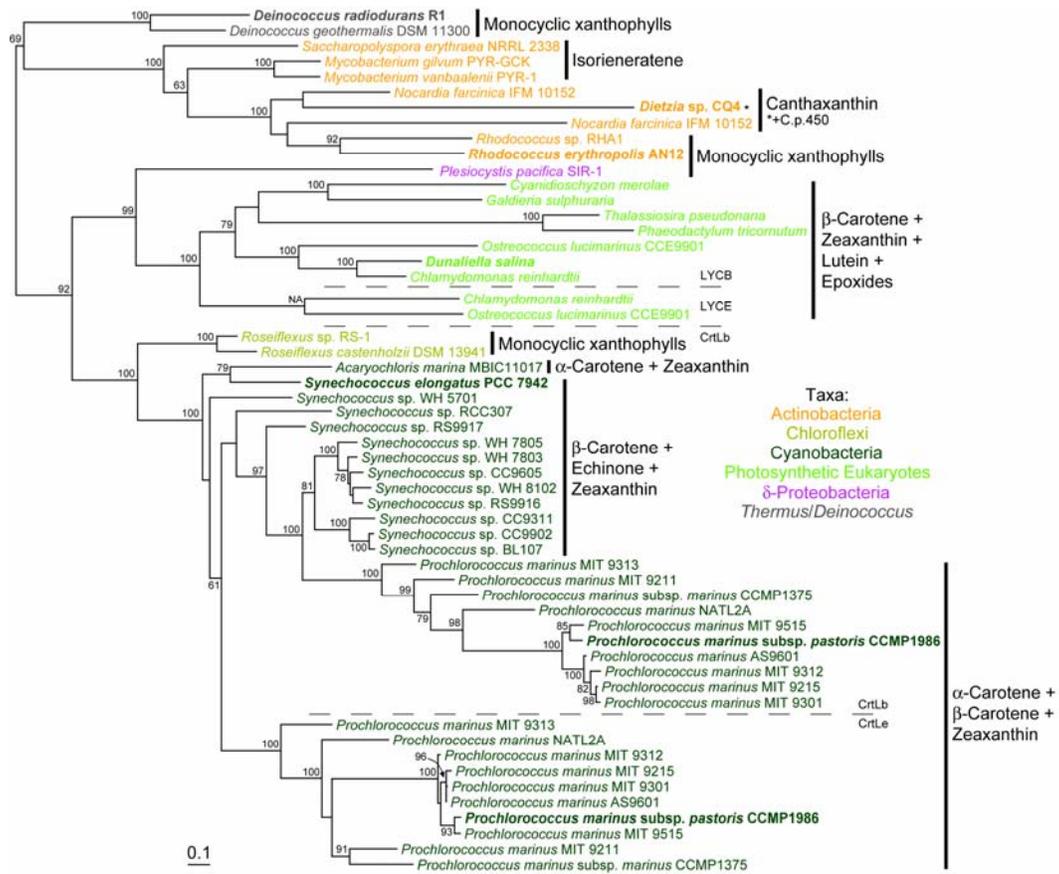
^k*Prochlorococcus marinus* NATL2A is considered representative of *P. marinus* NATL1A due to nearly 100% sequence homology between strains.



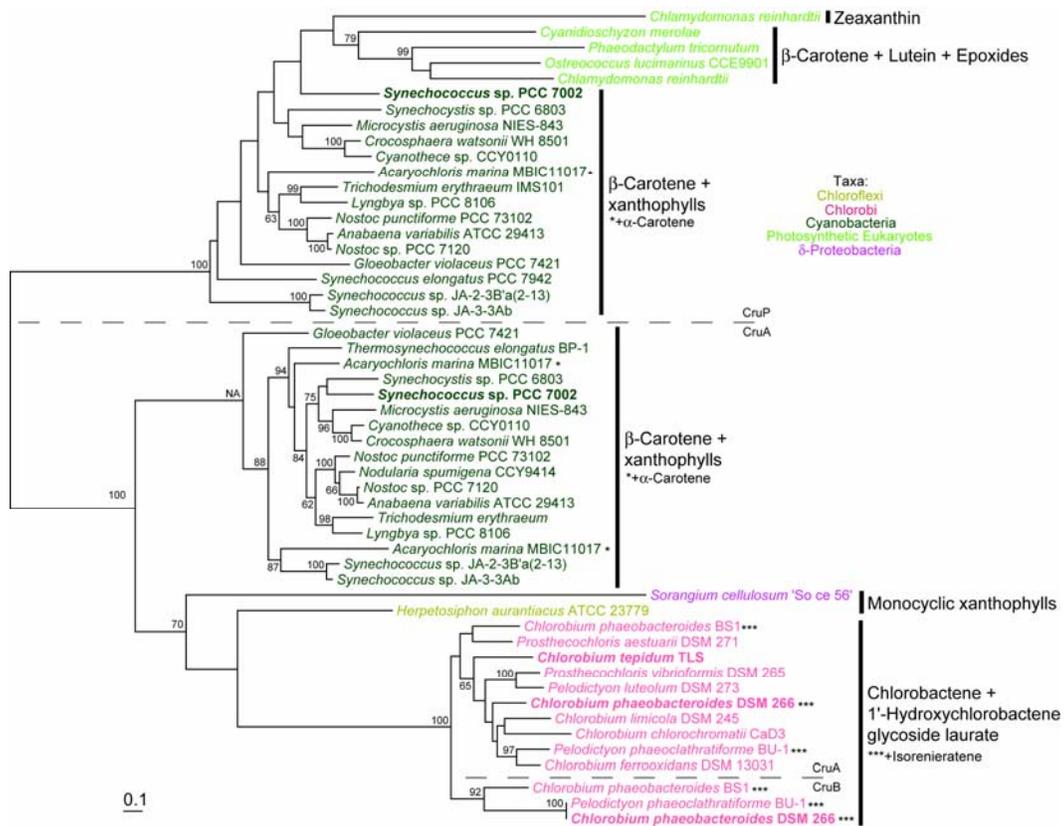
Supplemental Figure D1. Phylogenetic tree of CrY protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Genomes containing a rhodopsin homolog are indicated by an "R" and sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given. Because of its long branch length the CrY sequence for uncultured marine bacterium HF10_49E08, although homologous to other CrY sequences, was excluded.



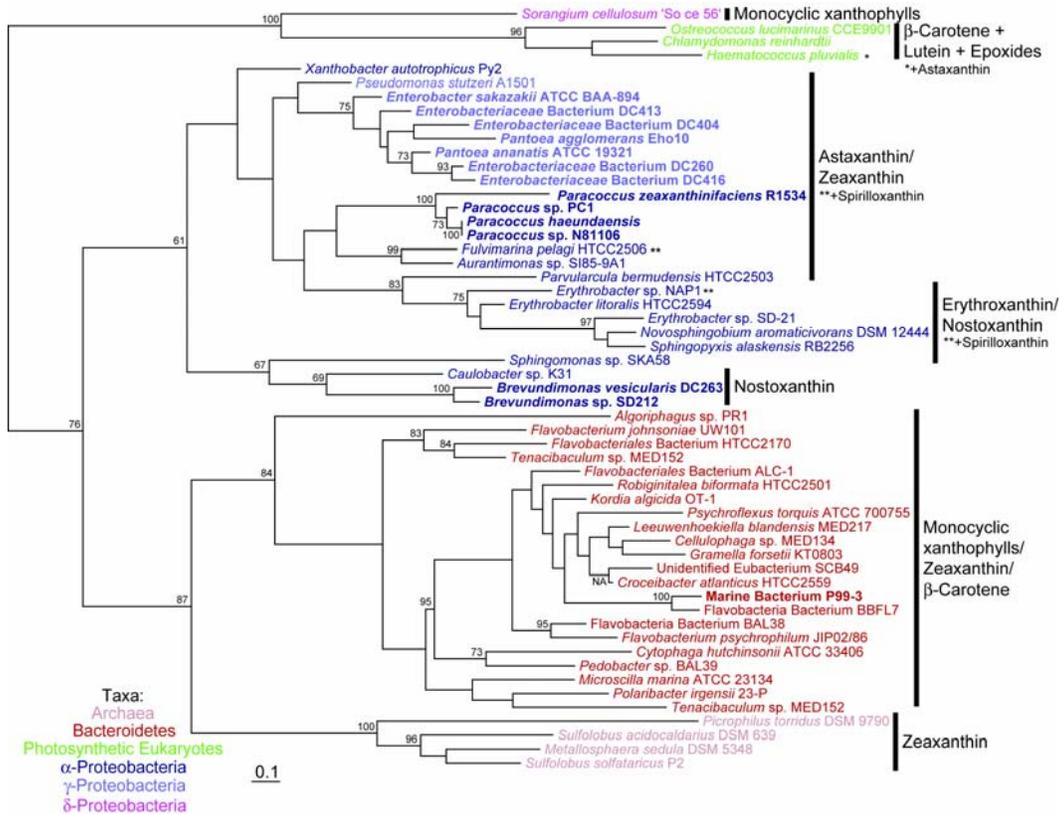
Supplemental Figure D2. Phylogenetic tree of CrtYcd protein sequences constructed using RAxML. Sequences present as separate subunits were artificially fused prior to alignments. Fungal CrtBM fusion proteins sequences were artificially cleaved according to Supplemental Table D2. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Genomes containing a rhodopsin homolog are indicated by an “R” and sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



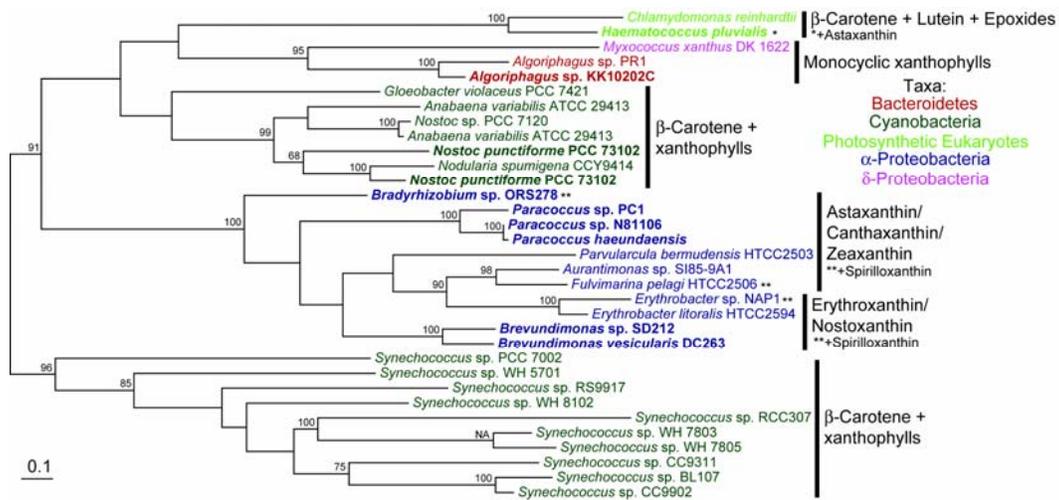
Supplemental Figure D3. Phylogenetic tree of CrtL protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



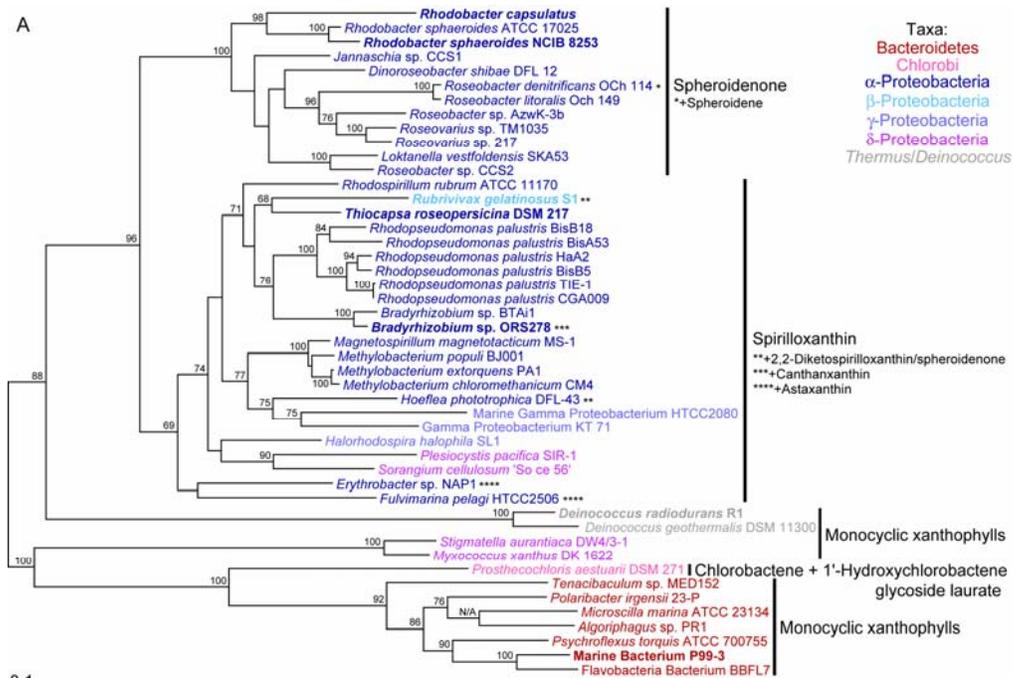
Supplemental Figure D4. Phylogenetic tree of CruA, CruB and CruP protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



Supplemental Figure D5. Phylogenetic tree of CrtZ protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.

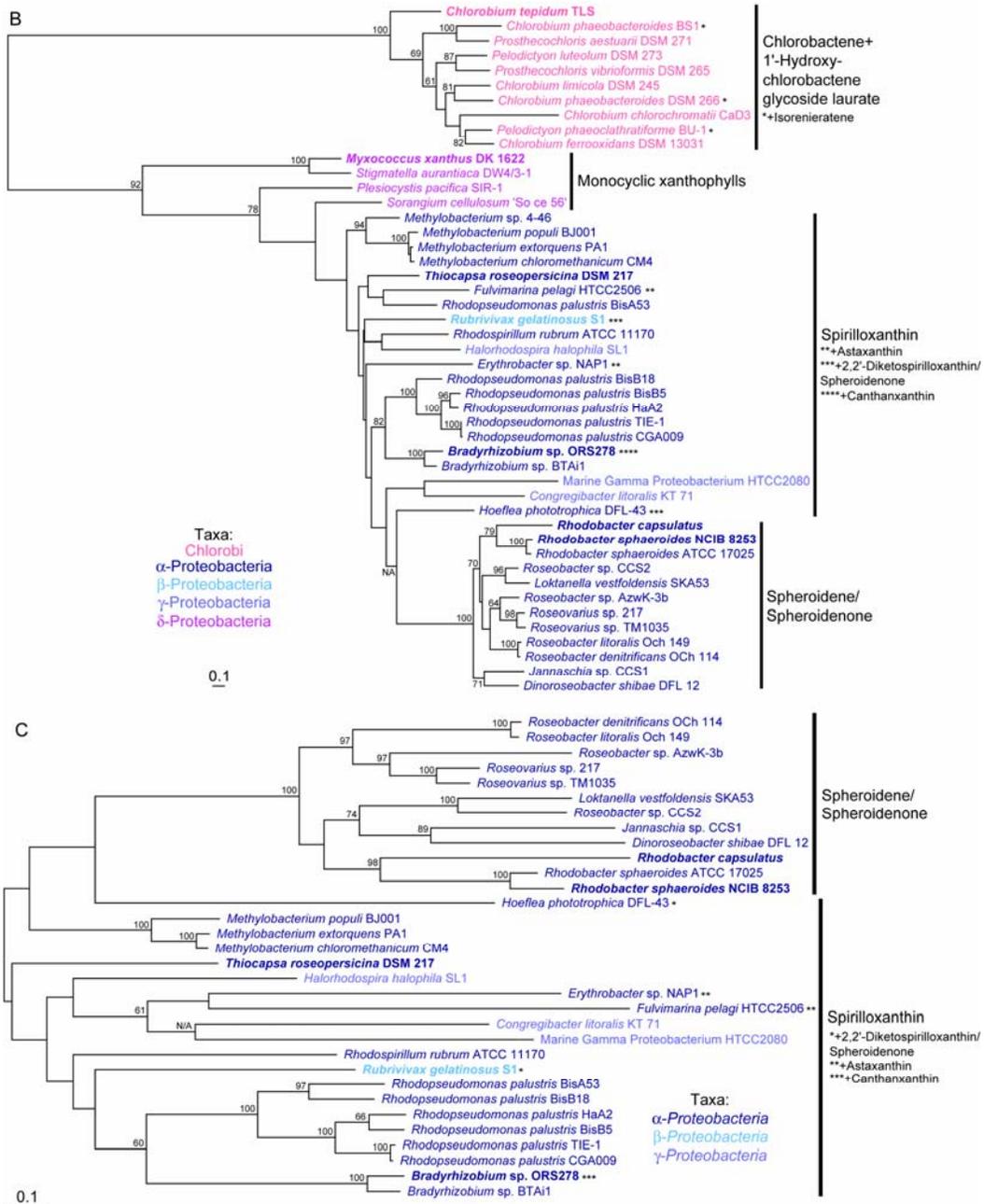


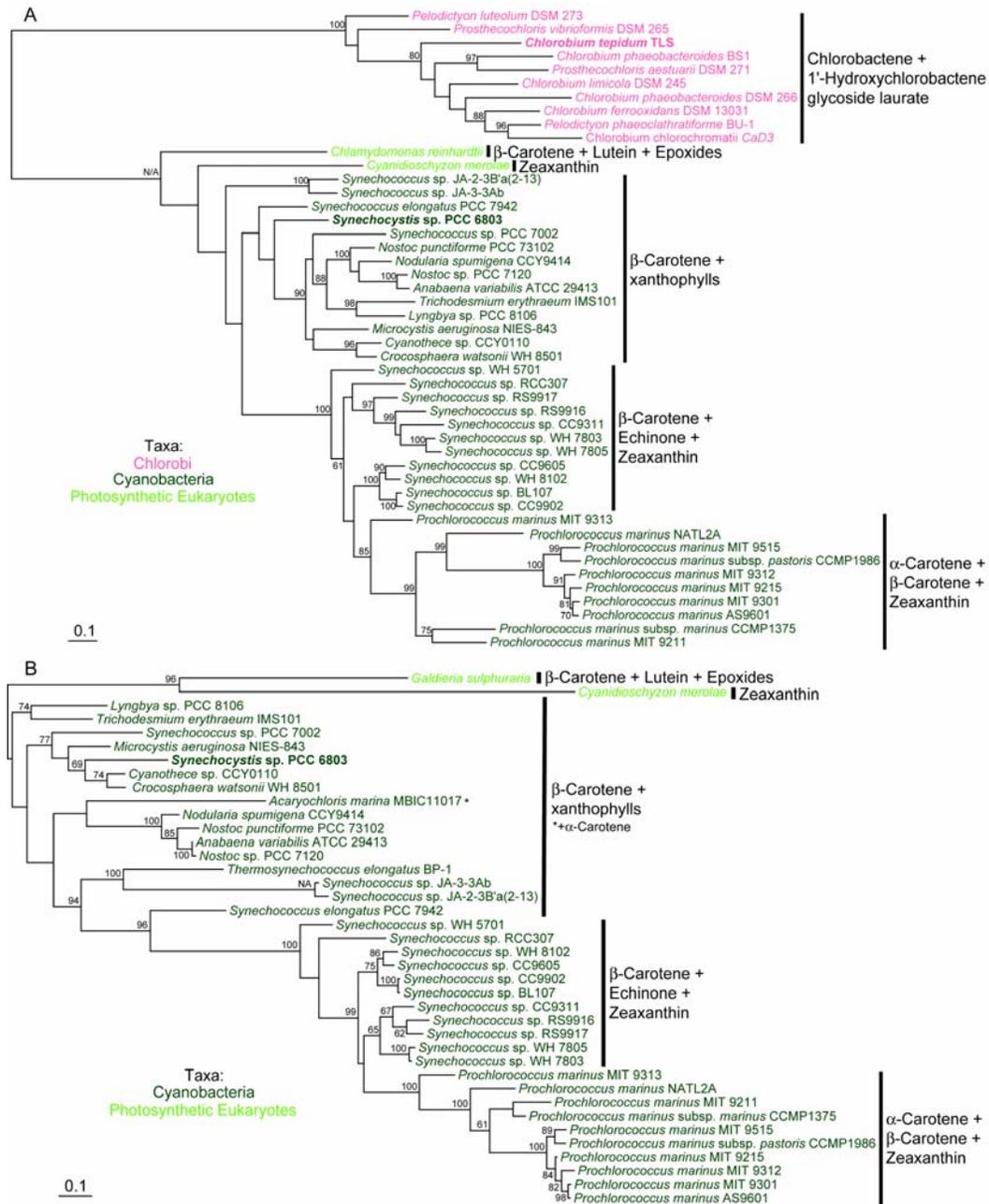
Supplemental Figure D6. Phylogenetic tree of CrtW protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



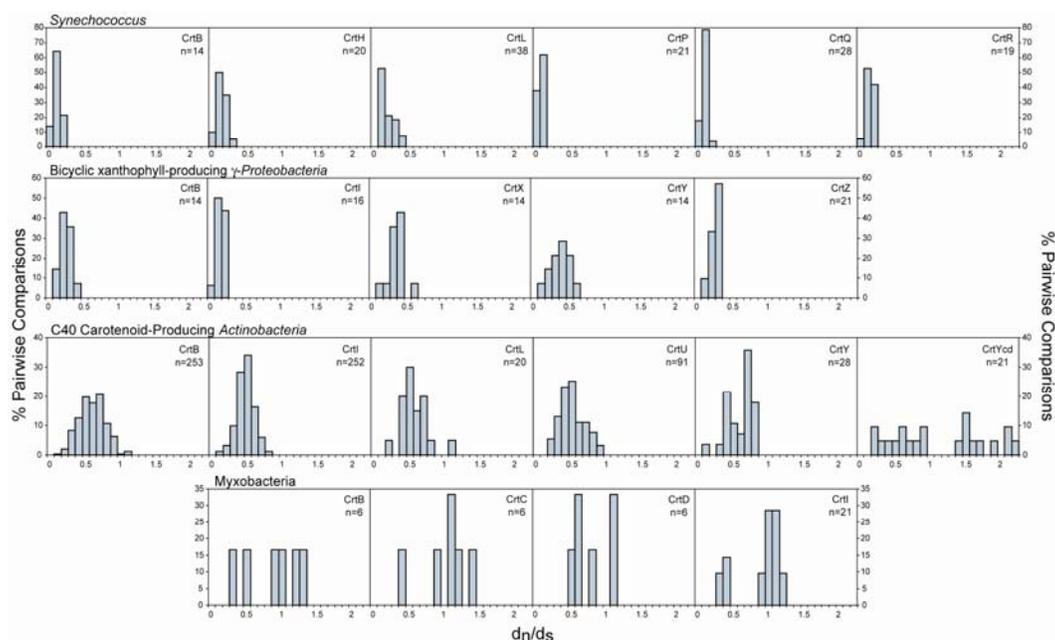
Supplemental Figure D7. Phylogenetic tree of CrtD (A), CrtC (B) and CrtF (C) protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given. In the CrtD tree all cyanobacterial sequences were omitted as they are more closely related to CrtH sequences than the other CrtD sequences.

Supplemental Figure D7 continued.

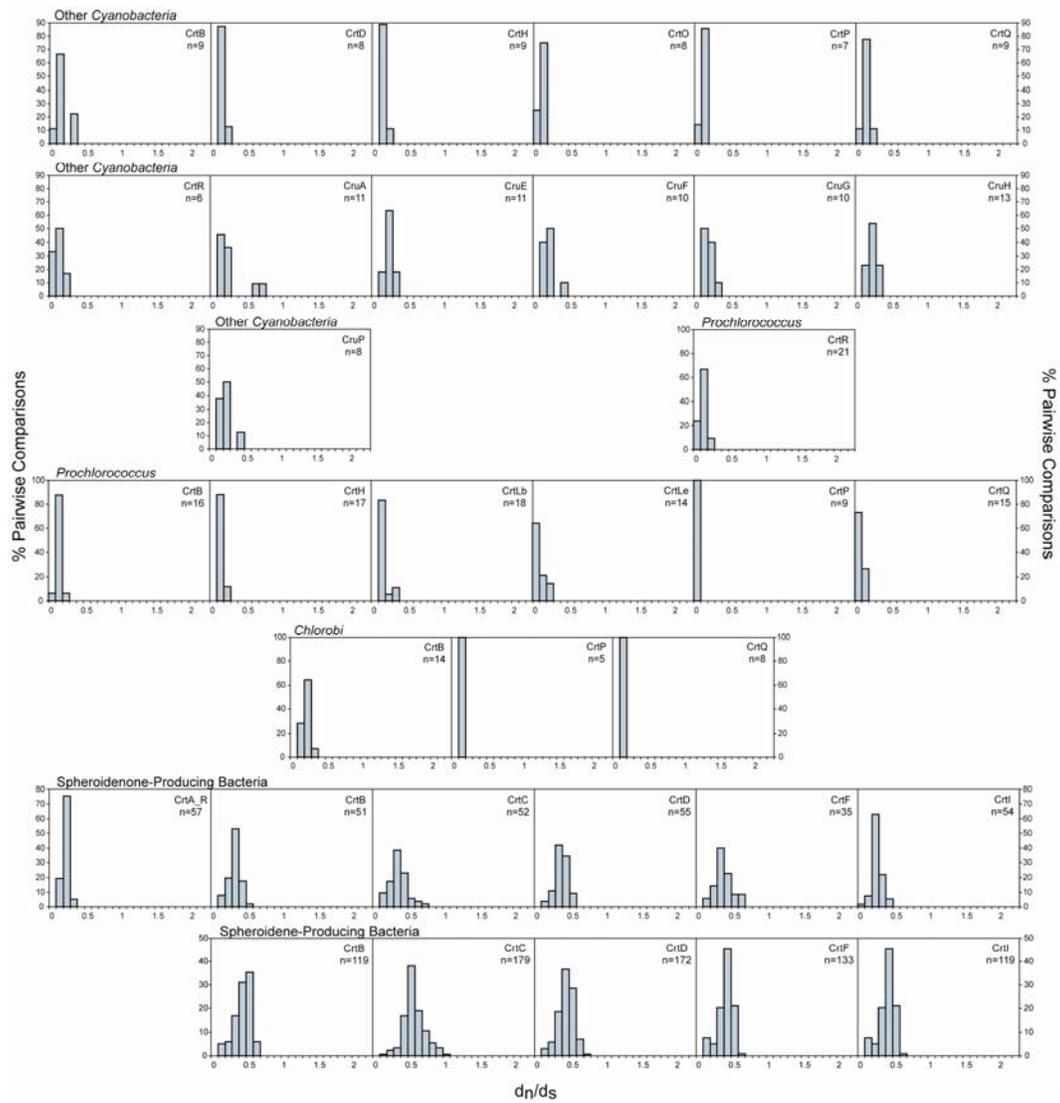




Supplemental Figure D8. Phylogenetic tree of CrtH (A) and CrtR (B) protein sequences constructed using RAxML. Bootstrap values $\geq 60\%$ are indicated as a percentage of the automatically determined number of replicates determined using the CIPRES web portal. Sequences with genetically or biochemically demonstrated functions are bolded. Carotenoids typical of each lineage are indicated to the right of each clade, with exceptions indicated by asterisks. The scale bar represents 10% sequence divergence. The tree shown is rooted to its midpoint to maximise the clarity of intraclade relationships. NA indicates the ML basal node for which no bootstrap value was given.



Supplemental Figure D9. Distributions of pairwise d_n/d_s values, rounded to one decimal place, for phylogenetic groups described in the text, expressed as a percentage of the total number of comparisons (n) for each sequence cluster protein. Only values with $d_n > 0.01$ and $d_s < 1.5$ were included; note that this underestimates the values at the lower end of the distributions shown, especially for Cyanobacteria and Chlorobi. Results for *Synechococcus*, bicyclic xanthophyll-producing γ -Proteobacteria, C40 carotenoid-producing Actinobacteria and *Myxobacteria* are shown in Figure 6.7.



Supplemental Figure D9 continued.

Supplemental Figure D10. Pairwise d_n/d_s values for: (a) C40 carotenoid-producing Actinobacteria *crtYcd*; (b) C50 carotenoid-producing Actinobacteria *crtYef* and myxobacterial *crtB* (c), *crtC* (d), *crtD* (e) and *crtI* (f). Matrices are one-sided, with cells of the opposite side filled with a dash. Bolded values are those highlighted in the text. In some cases a pairwise comparison of two sequences otherwise determined to have a high d_n/d_s values yielded an unexpectedly low d_n/d_s value; these ratios are italicized. NC indicates comparisons for which MEGA 4.0 could not calculate d_s value.

A. C40 Carotenoid-producing Actinobacteria *crtYcd*

	<i>Frankia</i> sp. Ccl3	<i>Frankia alni</i> ACN14a	<i>Frankia</i> sp. EAN1pec	<i>Mycobacterium aurum</i> A+	<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	<i>Mycobacterium</i> sp. MCS
<i>Frankia alni</i> ACN14a	1.523022	-	-	-	-	-
<i>Frankia</i> sp. EAN1pec	0.196231	1.866216	-	-	-	-
<i>Mycobacterium aurum</i> A+	1.668994	<i>0.855869</i>	2.099379	-	-	-
<i>Mycobacterium avium</i> subsp. <i>paratuberculosis</i> K-10	0.643956	1.522868	0.816431	2.123245	-	-
<i>Mycobacterium</i> sp. MCS	0.747638	2.198642	0.861472	1.596491	0.414443	-
<i>Mycobacterium ulcerans</i> Agy99	0.501406	1.392157	0.559184	1.485261	0.23934	0.308671

B. C50 Carotenoid-producing Actinobacteria *crtYef*

	<i>Corynebacterium efficiens</i> YS-314	<i>Corynebacterium glutamicum</i> ATCC 13032	<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPFB 382	Marine Actinobacterium PHSC20C1	<i>Leifsonia xyli</i> subsp. <i>xyli</i> str. CTCB07
<i>Corynebacterium glutamicum</i> ATCC 13032	NC	-	-	-	-
<i>Clavibacter michiganensis</i> subsp. <i>michiganensis</i> NCPFB 382	0.810289	NC	-	-	-
Marine Actinobacterium PHSC20C1	$d_s > 1.5$	NC	0.818693	-	-
<i>Leifsonia xyli</i> subsp. <i>xyli</i> str. CTCB07	0.657615	NC	0.44763	0.948099	-
<i>Dietzia</i> sp. CQ4	1.955801	NC	1.998487	$d_s > 1.5$	1.770889

C. Myxobacteria *crtB*

	<i>Sorangium cellulosum</i> So ce 56'	<i>Myxococcus xanthus</i> DK 1622	<i>Stigmatella aurantiaca</i> DW4/3-1
<i>Myxococcus xanthus</i> DK 1622	1.317324	-	-
<i>Stigmatella aurantiaca</i> DW4/3-1	1.012788	0.301676	-
<i>Plesiocystis pacifica</i> SIR-1	0.471429	1.171018	0.946352

D. Myxobacteria *crtC*

	<i>Sorangium cellulosum</i> 'So ce 56'	<i>Myxococcus xanthus</i> DK 1622	<i>Stigmatella aurantiaca</i> DW4/3-1
<i>Myxococcus xanthus</i> DK 1622	1.132099	-	-
<i>Stigmatella aurantiaca</i> DW4/3-1	1.411848	0.392027	-
<i>Plesiocystis pacifica</i> SIR-1	0.855153	1.087117	1.167082

E. Myxobacteria *crtD*

	<i>Sorangium cellulosum</i> 'So ce 56'	<i>Myxococcus xanthus</i> DK 1622	<i>Plesiocystis pacifica</i> SIR-1
<i>Myxococcus xanthus</i> DK 1622	1.080882	-	-
<i>Plesiocystis pacifica</i> SIR-1	0.59375	1.056098	-
<i>Stigmatella aurantiaca</i> DW4/3-1	1.009467	0.323124	1.046838

F. Myxobacteria *crtI*

	<i>Sorangium cellulosum</i> 'So ce 56'	<i>Myxococcus xanthus</i> DK 1622 (<i>crtI</i> -like)	<i>Plesiocystis pacifica</i> SIR-1	<i>Myxococcus xanthus</i> DK 1622 (<i>crtI</i>)	<i>Myxococcus xanthus</i> DK 1622 (<i>crtIb</i>)	<i>Stigmatella aurantiaca</i> DW4/3-1 (<i>crtIb</i>)
<i>Myxococcus xanthus</i> DK 1622 (<i>crtI</i> -like)	0.432967	-	-	-	-	-
<i>Plesiocystis pacifica</i> SIR-1	0.392694	0.412391	-	-	-	-
<i>Myxococcus xanthus</i> DK 1622 (<i>crtI</i>)	1.041018	0.985788	1.055707	-	-	-
<i>Myxococcus xanthus</i> DK 1622 (<i>crtIb</i>)	1.166667	0.911435	1.092288	1.025526	-	-
<i>Stigmatella aurantiaca</i> DW4/3-1 (<i>crtIb</i>)	1.184828	1.00618	1.013746	0.969697	0.296296	-
<i>Stigmatella aurantiaca</i> DW4/3-1 (<i>crtI</i>)	1.139104	0.909305	1.073816	0.26087	1.117293	1.098431

Literature cited:

- Andrews, A.G., and Starr, M.P. 1976. (3R,3R')-Astaxanthin from the yeast *Phaffia rhodozyma*. *Phytochemistry*, 15: 1009-1011.
- Armstrong, G.A., Alberti, M., Leach, F., and Hearst, J.E. 1989. Nucleotide sequence, organization, and nature of the protein products of the carotenoid biosynthesis gene cluster of *Rhodobacter capsulatus*. *Mol. Gen. Genet.*, 216: 254-268.
- Arrach, N., Fernández-Martín, R., Cardá-Olmedo, E., and Avalos, J. 2001. A single gene for lycopene cyclase, phytoene synthase, and regulation of carotene biosynthesis in *Phycomyces*. *Proc. Nat. Acad. Sci. USA*, 98: 1687-1692.
- Avalos, J., Casadesús, J., and Cerdá-Olmedo, E. 1985. *Gibberella fujikuroi* mutants obtained with UV radiation and *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine. *Appl. Environ. Microbiol.*, 49: 187-191.
- Balashov, S.P., Imasheva, E.S., Boichenko, V.A., Antón, J., Wang, J.M., and Lanyi, J.K. 2005. Xanthorhodopsin: a proton pump with a light-harvesting carotenoid antenna. *Science*, 309: 2061-2064.
- Béjà, O., Aravind, L., Koonin, E.V., Suzuki, M.T., Hadd, A., Nguyen, L.P., Jovanovich, S.B., Gates, C.M., Feldman, R.A., Spudich, J.L., Spudich, E.N., and DeLong, E.F. 2000. Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, 289: 1902-1906.
- Berry, A., Janssens, D., Hümbelin, M., Jore, J.P.M., Hoste, B., Cleenwerck, I., Vancanneyt, M., Bretzel, W., Mayer, A.F., Lopez-Ulibarri, R., Shanmugam, B., Swings, J., and Pasamontes, L. 2003. *Paracoccus zeaxanthinifaciens* sp. nov., a zeaxanthin-producing bacterium. *Int. J. Syst. Evol. Microbiol.*, 53: 231-238.
- Biebl, H., Allgaier, M., Tindall, B.J., Koblizek, M., Lünsdorf, H., Pukall, R., and Wagner-Döbler, I. 2005. *Dinoroseobacter shibae* gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. *Int. J. Syst. Evol. Microbiol.*, 55: 1089-1096.
- Blasco, F., Kauffmann, L., and Schmid, R.D. 2004. CYP175A1 from *Thermus thermophilus* HB27, the first β -carotene hydroxylase of the P450 superfamily. *Appl. Microbiol. Biotechnol.*, 64: 671-674.
- Botella, J.A., Murillo, F.J., and Ruiz-vázquez, R. 1995. A cluster of structural and regulatory genes for light-induced carotenogenesis in *Myxococcus xanthus*. *Eur. J. Biochem.*, 233: 238-248.

- Breitenbach, J., Fernández-González, B., Vioque, A., and Sandmann, G. 1998. A higher-plant type ζ -carotene desaturase in the cyanobacterium *Synechocystis* PCC6803. *Plant Mol. Biol.*, 36: 725-732.
- Breitenbach, J., Vioque, A., and Sandmann, G. 2001. Gene *sll0033* from *Synechocystis* 6803 encodes a carotene isomerase involved in the biosynthesis of all-*E* lycopene. *Z. Naturforsch. C*, 56: 915-917.
- Chamovitz, D., Pecker, I., and Hirschberg, J. 1991. The molecular basis of resistance to the herbicide norflurazon. *Plant Mol. Biol.*, 16: 967-974.
- Chamovitz, D., Misawa, N., Sandmann, G., and Hirschberg, J. 1992. Molecular cloning and expression in *Escherichia coli* of a cyanobacterial gene coding for phytoene synthase, a carotenoid biosynthesis enzyme. *FEBS Lett.*, 296: 305-310.
- Cheng, Q. 2006. Structural diversity and functional novelty of new carotenoid biosynthesis genes. *J. Ind. Microbiol. Biotechnol.*, 33: 552-559.
- Cunningham Jr., F.X., Sun, Z., Chamovitz, D., Hirschberg, J., and Gantt, E. 1994. Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp. strain PCC7942. *Plant Cell*, 6: 1107-1121.
- Cunningham Jr., F.X., Lee, H., and Gantt, E. 2007. Carotenoid biosynthesis in the primitive red alga *Cyanidioschyzon merolae*. *Eukaryotic Cell*, 6: 533-545.
- Daub, M.E., and Payne, G.A. 1989. The role of carotenoids in resistance of fungi to cercosporin. *Phytopathology*, 79: 180-185.
- Ehrenshaft, M., and Daub, M.E. 1994. Isolation, sequence, and characterization of the *Cercospora nicotianae* phytoene dehydrogenase gene. *Appl. Environ. Microbiol.*, 60: 2766-2771.
- Estrada, A.F., Youssar, L., Scherzinger, D., Al-Babili, S., and Avalos, J. 2008. The *ylo-1* gene encodes an aldehyde dehydrogenase responsible for the last reaction in the *Neurospora* carotenoid pathway. *Mol. Microbiol.*, 69: 1207-1220.
- Fernández-González, B., Sandmann, G., and Vioque, A. 1997. A new type of asymmetrically acting β -carotene ketolase is required for the synthesis of echinenone in the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.*, 272: 9728-9733.

- Fraser, P.D., Ruiz-Hidalgo, M.J., Lopez-Matas, M.A., Alvarez, M.I., Eslava, A.P., and Bramley, P.M. 1996. Carotenoid biosynthesis in wild type and mutant strains of *Mucor circinelloides*. *Biochim. Biophys. Acta*, 1289: 203-208.
- Frigaard, N.-U., Maresca, J.A., Yunker, C.E., Jones, A.D., and Bryant, D.A. 2004. Genetic manipulation of carotenoid biosynthesis in the green sulfur bacterium *Chlorobium tepidum*. *J. Bacteriol.*, 186: 5210-5220.
- Fuchs, B.M., Spring, S., Teeling, H., Quast, C., Wulf, J., Schattenhofer, M., Yan, S., Ferreira, S., Johnson, J., Glöckner, F.O., and Amann, R. 2007. Characterization of a marine gammaproteobacterium capable of aerobic anoxygenic photosynthesis. *Proc. Nat. Acad. Sci. USA*, 104: 2891-2896.
- Gerjets, T., Steiger, S., and Sandmann, G. 2009. Catalytic properties of the expressed acyclic carotenoid 2-ketolases from *Rhodobacter capsulatus* and *Rubrivivax gelatinosus*. *Biochim. Biophys. Acta*, 1791: 125-131.
- Giraud, E., Hannibal, L., Fardoux, J., Jaubert, M., Jourand, P., Dreyfus, B., Sturgis, J.N., and Verméglio, A. 2004. Two distinct *crt* gene clusters for two different functional classes of carotenoid in *Bradyrhizobium*. *J. Biol. Chem.*, 279: 15076-15083.
- Gómez-Consarnau, L., González, J.M., Coll-Lladó, M., Gourdon, P., Pascher, T., Neutze, R., Pedrós-Alió, C., and Pinhassi, J. 2007. Light stimulates growth of proteorhodopsin-containing marine flavibacteria. *Nature*, 445: 210-213.
- Hannibal, L., Lorquin, J., D'Ortoli, N.A., Garcia, N., Chaintreuil, C., Masson-Boivin, C., Dreyfus, B., and Giraud, E. 2000. Isolation and characterization of canthaxanthin biosynthesis genes from the photosynthetic bacterium *Bradyrhizobium* sp. strain ORS278. *J. Bacteriol.*, 182: 3850-3853.
- Harada, J., Nagashima, K.V.P., Takaichi, S., Misawa, N., Matsuura, K., and Shimada, K. 2001a. Phytoene desaturase, *CrtI*, of the purple photosynthetic bacterium, *Rubrivivax gelatinosus*, produces both neurosporene and lycopene. *Plant Cell Physiol.*, 42: 1112-1118.
- Harada, J., Takaichi, S., Nagashima, K.V.P., Matsuura, K., and Shimada, K. 2001b. Functional analysis of spheroidene mono-oxygenase, *CrtA*, of the purple photosynthetic bacterium, *Rubrivivax gelatinosus*. In 12th International Congress on Photosynthesis(ed.). Brisbane, Australia. CSIRO, pp. S2-025.
- Hemmi, H., Ikejiri, S., Nakayama, T., and Nishino, T. 2003. Fusion-type lycopene β -cyclase from a thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochem. Biophys. Res. Commun.*, 305: 586-591.

- Hundle, B., Alberti, M., Nievelstein, V., Beyer, P., Klening, H., Armstrong, G.A., Burke, D.H., and Hearst, J.E. 1994. Functional assignment of *Erwinia herbicola* Eho10 carotenoid genes expressed in *Escherichia coli*. *Mol. Gen. Genet.*, 245: 406-416.
- Hundle, B.S., Beyer, P., Kleinig, H., Englert, G., and Hearst, J.E. 1991. Carotenoids of *Erwinia herbicola* and an *Escherichia coli* HB101 strain carrying the *Erwinia herbicola* carotenoid gene cluster. *Photochem. Photobiol.*, 54: 89-93.
- Iniesta, A.A., Cervantes, M., and Murillo, F.J. 2007. Cooperation of two carotene desaturases in the production of lycopene in *Myxococcus xanthus*. *FEBS J.*, 274: 4306-4314.
- Iniesta, A.A., Cervantes, M., and Murillo, F.J. 2008. Conversion of the lycopene monocyclus of *Myxococcus xanthus* into a bicyclase. *Appl. Microbiol. Biotechnol.*, 79: 793-802.
- Iwai, M., Maoka, T., Ikeuchi, M., and Takaichi, S. 2008. 2,2'- β -Hydroxylase (CrtG) is involved in carotenogenesis of both nostoxanthin and 2-hydroxymyxol 2'-fucoside in *Thermosynechococcus elongatus* strain BP-1. *Plant Cell Physiol.*, 49: 1678-1687.
- Jung, K.-H., Trivedi, V.D., and Spudich, J.L. 2003. Demonstration of a sensory rhodopsin in eubacteria. *Mol. Microbiol.*, 47: 1513-1522.
- Kajiwarra, S., Kakizono, T., Saito, T., Kondo, K., Ohtani, T., Nishido, N., Nagai, S., and Misawa, N. 1995. Isolation and functional identification of a novel cDNA for astaxanthin biosynthesis from *Haematococcus pluvialis*, and astaxanthin synthesis in *Escherichia coli*. *Plant Mol. Biol.*, 29: 343-352.
- Kleining, H., and Reichenbach, H. 1977. Carotenoid glucosides and menaquinones from the gliding bacterium *Herptosiphon giganteus* Hp. a2. *Arch. Microbiol.*, 112: 307-310.
- Kobližek, M., Bèjà, O., Bidigare, R.R., Christensen, S., Benetiz-Nelson, B., Vetriani, C., Kolber, M.K., Falkowski, P.G., and Kolber, Z.S. 2003. Isolation and characterization of *Erythrobacter* sp. strains from the upper ocean. *Arch. Microbiol.*, 180: 327-338.
- Kohl, W., Achenbach, H., and Reichenbach, H. 1983. The pigments of *Brevibacterium linens*: aromatic carotenoids. *Phytochemistry*, 22: 207-210.

- Kovács, Á.T., Rákhely, G., and Kovács, K.L. 2003. Genes involved in the biosynthesis of photosynthetic pigments in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*. *Appl. Environ. Microbiol.*, 69: 3093-3102.
- Krubasik, P., and Sandmann, G. 2000. 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.
- Krubasik, P., Kobayashi, M., and Sandmann, G. 2001. Expression and functional analysis of a gene cluster involved in the synthesis of decaprenoxanthin reveals the mechanisms for C₅₀ carotenoid formation. *Eur. J. Biochem.*, 268: 3702-3708.
- Krügel, H., Krubasik, P., Weber, K., Saluz, H.P., and Sandmann, G. 1999. 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. *Biochim. Biophys. Acta*, 1439: 57-64.
- Lang, H.P., Cogdell, R.J., Gardiner, A.T., and Hunter, C.N. 1994. Early steps in carotenoid biosynthesis: sequences and transcriptional analysis of the *crtI* and *crtB* genes of *Rhodobacter sphaeroides* and overexpression and reactivation of *crtI* in *Escherichia coli* and *R. sphaeroides*. *J. Bacteriol.*, 176: 3859-3869.
- Lang, H.P., Cogdell, R.J., Takaichi, S., and Hunter, C.N. 1995. Complete DNA sequence, specific Tn5 insertion map, and gene assignment of the carotenoid biosynthesis pathway of *Rhodobacter sphaeroides*. *J. Bacteriol.*, 177: 2064-2073.
- Larsen, R.A., Wilson, M.M., Guss, A.M., and Metcalf, W.W. 2002. Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. *Arch. Microbiol.*, 178: 193-201.
- Lee, J.H., Kim, Y.S., Choi, T.-J., Lee, W.J., and Kim, Y.T. 2004. *Paracoccus haeundaensis* sp. nov., a Gram-negative, halophilic, astaxanthin-producing bacterium. *Int. J. Syst. Evol. Microbiol.*, 54: 1699-1702.
- Lee, J.T., and Kim, Y.T. 2006. Cloning and characterization of the astaxanthin biosynthesis gene cluster from the marine bacterium *Paracoccus haeundaensis*. *Gene*, 370: 86-95.
- Lehner, A., Grimm, M., Rattei, T., Ruepp, A., Frishman, D., Manzardo, G.G.G., and Stephan, R. 2006. Cloning and characterization of *Enterobacter*

- sakazakii* pigment genes and *in situ* spectroscopic analysis of the pigment. FEMS Microbiol. Lett., 265: 244-248.
- Linden, H., Vioque, A., and Sandmann, G. 1993. Isolation of a carotenoid biosynthesis gene coding for ζ -carotene desaturase from *Anabaena* PCC 7120 by heterologous complementation. FEMS Microbiol. Lett., 106: 99-104.
- Linden, H. 1999. Carotenoid hydroxylase from *Haematococcus pluvialis*: cDNA sequence, regulation and functional complementation. Biochim. Biophys. Acta, 1446: 203-212.
- Linnemannstöns, P., Prado, M.M., Fernández-Martín, R., Tudzynski, B., and Avalos, J. 2002. A carotenoid biosynthesis gene cluster in *Fusarium fujikuroi*: the genes *carB* and *carRA*. Mol. Genet. Genomics, 267: 593-602.
- Lotan, T., and Hirschberg, J. 1995. Cloning and expression in *Escherichia coli* of the gene encoding β -C-4-oxygenase, that converts β -carotene to the ketocarotenoid canthaxanthin in *Haematococcus pluvialis*. FEBS Lett., 364: 125-128.
- Lutnaes, B.F., Oren, A., and Liaaen-Jensen, S. 2002. New C₄₀-carotenoid acyl glycoside as principal carotenoid in *Salinibacter ruber*, an extremely halophilic eubacterium. J. Nat. Prod., 65: 1340-1343.
- Maresca, J.A., and Bryant, D.A. 2006. Two genes encoding new carotenoid-modifying enzymes in the green sulfur bacterium *Chlorobium tepidum*. J. Bacteriol., 188: 6217-6223.
- Maresca, J.A., Graham, J.E., Wu, M., Eisen, J.S., and Bryant, D.A. 2007. Identification of a fourth family of lycopene cyclases in photosynthetic bacteria. Proc. Nat. Acad. Sci. USA, 104: 11784-11789.
- Maresca, J.A., Graham, J.E., and Bryant, D.A. 2008a. The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. Photosynth. Res., 97: 121-140.
- Maresca, J.A., Romberger, S.P., and Bryant, D.A. 2008b. Isorenieratene biosynthesis in green sulfur bacteria requires the cooperative actions of two carotenoid cyclases. J. Bacteriol., 190: 6384-6391.
- Martínez-Férez, I., Fernández-González, B., Sandmann, G., and Vioque, A. 1994. Cloning and expression in *Escherichia coli* of the gene for phytoene synthase from the cyanobacterium *Synechocystis* sp. PCC6803. Biochim. Biophys. Acta, 1218: 145-152.

- Martínez-Férez, I.M., and Vioque, A. 1992. Nucleotide sequence of the phytoene desaturase gene from *Synechocystis* sp. PCC 6803 and characterization of a new mutation which confers resistance to the herbicide norflurazon. *Plant Mol. Biol.*, 18: 981-983.
- Masamoto, K., Misawa, N., Kaneko, T., Kikuno, R., and Toh, H. 1998. β -Carotene hydroxylase gene from the cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiol.*, 39: 560-564.
- Masamoto, K., Wada, H., Kaneko, T., and Takaichi, S. 2001. Identification of a gene required for *cis*-to-*trans* carotene isomerization in carotenogenesis of the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.*, 42: 1398-1402.
- Matsui, T., and Maruhashi, K. 2004. Isolation of carotenoid-deficient mutant from alkylated dibenzothiophene desulfurizing nocardioform bacteria, *Gordonia* sp. TM414. *Curr. Microbiol.*, 48: 130-134.
- Mil'ko, E.S. 1963. Effect of various environmental factors on pigment production in the alga *Dunaliella salina*. *Mikrobiologiya*, 32: 299-307.
- Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., and Harashima, K. 1990. Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *J. Bacteriol.*, 172: 6704-6712.
- Misawa, N., Satomi, Y., Kondo, K., Yokoyama, A., Kajiwara, S., Saito, T., Ohtani, T., and Miki, W. 1995. Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene cluster and astaxanthin biosynthetic pathway proposed at the gene level. *J. Bacteriol.*, 177: 6575-6584.
- Miyashita, H., Adachi, K., Kurano, N., Ikemoto, H., Chihara, M., and Miyachi, S. 1997. Pigment composition of a novel oxygenic photosynthetic prokaryote containing chlorophyll *d* as the major chlorophyll. *Plant Cell Physiol.*, 38: 274-281.
- Mochimaru, M., Masukawa, H., and Takaichi, S. 2005. The cyanobacterium *Anabaena* sp PCC 7120 has two distinct beta-carotene ketolases: CrtO for echinenone and CrtW for ketomyxol synthesis. *FEBS Lett.*, 579: 6111-6114.
- Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., Komemushi, S., Miki, W., and Misawa, N. 2005. Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2,2'- β -

- hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanthophylls. *Appl. Environ. Microbiol.*, 71: 4286-4296.
- Ouchane, S., Picaud, M., Vernotte, C., and Astier, C. 1997a. Photooxidative stress stimulates illegitimate recombination and mutability in carotenoid-less mutants of *Rubrivivax gelatinosus*. *The EMBO Journal*, 16: 4777-4787.
- Ouchane, S., Picaud, M., Vernotte, C., Reiss-Husson, F., and Astier, C. 1997b. Pleiotropic effects of *puf* interposon mutagenesis on carotenoid biosynthesis in *Rubrivivax gelatinosus*. *J. Biol. Chem.*, 272: 1670-1676.
- Pasamontes, L., Hug, D., Tessier, M., Hohmann, H.-P., Schierle, J., and van Loon, A.P.G.M. 1997. Isolation and characterization of the carotenoid biosynthesis genes of *Flavobacterium* sp. strain R1534. *Gene*, 185: 35-41.
- Pelz, A., Wieland, K.-P., Putzbach, K., Hentschel, P., Albert, K., and Götz, F. 2005. Structure and biosynthesis of staphyloxanthin from *Staphylococcus aureus*. *J. Biol. Chem.*, 280: 32493-32498.
- Pinta, V., Ouchane, S., Picaud, M., Takaichi, S., Astier, C., and Reiss-Husson, F. 2003. Characterization of unusual hydroxy- and ketocarotenoids in *Rubrivivax gelatinosus*: involvement of enzyme CrtF or CrtA. *Arch. Microbiol.*, 179: 354-362.
- Prado-Cabrero, A., Estrada, A.F., Al-Babili, S., and Avalos, J. 2007. Identification and biochemical characterization of a novel carotenoid oxygenase: elucidation of the cleavage step in the *Fusarium* carotenoid pathway. *Mol. Microbiol.*, 64: 448-460.
- Ramos, A., Coesel, S., Marques, A., Rodrigues, M., Baumgartner, A., Noronha, J., Rauter, A., Brenig, B., and Varela, J. 2008. Isolation and characterization of a stress-inducible *Dunaliella salina* *Lcy-β* gene encoding a functional lycopene β-cyclase. *Appl. Microbiol. Biotechnol.*, 79: 819-828.
- Ruiz-Hidalgo, M.J., Benito, E.P., Sandmann, G., and Eslava, A.P. 1997. The phytoene dehydrogenase gene of *Phycomyces*: regulation of its expression by blue light and vitamin A. *Mol. Gen. Genet.*, 253: 734-744.
- Saelices, L., Youssar, L., Holdermann, I., Al-Babili, S., and Avalos, J. 2007. Identification of the gene responsible for torulene cleavage in the *Neurospora* carotenoid pathway. *Mol. Genet. Genomics*, 278: 527-537.

- Schmidhauser, T.J., Lauter, F.R., Russo, V.E.A., and Yanofsky, C. 1990. Cloning, sequence, and photoregulation of *al-1*, a carotenoid biosynthetic gene of *Neurospora crassa*. *Mol. Cell. Biol.*, 10: 5064-5070.
- Schmidhauser, T.J., Lauter, F.-R., Schumacher, M., Zhou, W., Russo, V.E.A., and Yanofsky, C. 1994. Characterization of *al-2*, the phytoene synthase gene of *Neurospora crassa*. *J. Biol. Chem.*, 269: 12060-12066.
- Sedkova, N., Tao, L., Rouvière, P.E., and Cheng, Q. 2005. Diversity of carotenoid synthesis gene clusters from environmental *Enterobacteriaceae* strains. *Appl. Environ. Microbiol.*, 71: 8141-8146.
- Shlomain, P., Ben-Amotz, A., and Margalith, P. 1991. Production of carotene stereoisomers by *Phycomyces blakesleeanus*. *Appl. Microbiol. Biotechnol.*, 34: 458-462.
- Steiger, S., Astier, C., and Sandmann, G. 2000. Substrate specificity of the expressed carotenoid 3,4-desaturase from *Rubrivivax gelatinosus* reveals the detailed reaction sequence to spheroidene and spirilloxanthin. *Biochem. J.*, 349: 635-640.
- Steiger, S., Mazet, A., and Sandmann, G. 2003. Heterologous expression, purification, and enzymatic characterization of the acyclic carotenoid 1,2-hydratase from *Rubrivivax gelatinosus*. *Arch. Biochem. Biophys.*, 414: 51-58.
- Steiger, S., and Sandmann, G. 2004. Cloning of two carotenoid ketolase genes from *Nostoc punctiforme* for the heterologous production of canthaxanthin and astaxanthin. *Biotechnol. Lett.*, 26: 813-817.
- Steiger, S., Jackisch, Y., and Sandmann, G. 2005. Carotenoid biosynthesis in *Gloeobacter violaceus* PCC4721 involves a single crtI-type phytoene desaturase instead of typical cyanobacterial enzymes. *Arch. Microbiol.*, 184: 207-214.
- Steinbrenner, J., and Linden, H. 2001. Regulation of two carotenoid biosynthesis genes coding for phytoene synthase and carotenoid hydroxylase during stress-induced astaxanthin formation in the green alga *Haematococcus pluvialis*. *Plant Physiol.*, 125: 810-817.
- Stickforth, P., Steiger, S., Hess, W.R., and Sandmann, G. 2003. A novel type of lycopene ϵ -cyclase in the marine cyanobacterium *Prochlorococcus marinus* MED4. *Arch. Microbiol.*, 179: 409-415.

- Stickforth, P., and Sandmann, G. 2007. Kinetic variations determine the product pattern of phytoene desaturase from *Rubrivivax gelatinosus*. Arch. Biochem. Biophys., 461: 235-241.
- Tabata, K., Ishida, S., Nakahara, T., and Hoshino, T. 1994. A carotenogenic gene cluster exists on a large plasmid in *Thermus thermophilus*. FEBS Lett., 341: 251-255.
- Takaichi, S., Inoue, K., Akaike, M., Kobayashi, M., Oh-oka, H., and Madigan, M.T. 1997. The major carotenoid in all known species of heliobacteria is the C30 carotenoid 4,4'-diaponeurosporene, not neurosporene. Arch. Microbiol., 168: 277-281.
- Takaichi, S. 1999. Carotenoids and carotenogenesis in anoxygenic photosynthetic bacteria. In The photochemistry of carotenoids. H.A. Frank, A.J. Young, G. Britton, and R.J. Cogdell. (ed.) Kluwer Academic Publishers, New York, NY. pp. 39-69.
- Takaichi, S., Moaka, T., Yamada, M., Matsuura, K., Haikawa, Y., and Hanada, S. 2001. Absence of carotenes and presence of a tertiary methoxy group in a carotenoid from a thermophilic filamentous photosynthetic bacterium *Roseiflexus castenholzii*. Plant Cell Physiol., 42: 1355-1362.
- Takaichi, S., Mochimaru, M., Moaka, T., and Katoh, H. 2005. Myxol and 4-ketomyxol 2'-fucosides, not rhamnosides, from *Anabaena* sp. PCC 7120 and *Nostoc punctiforme* PCC 73102, and proposal for the biosynthetic pathway of carotenoids. Plant Cell Physiol., 46: 497-504.
- Takaichi, S., Mochimaru, M., and Moaka, T. 2006. Presence of free myxol and 4-hydroxymyxol and Absence of myxol Glycosides in *Anabaena variabilis* ATCC 29413, and proposal of biosynthetic pathway of carotenoids. Plant Cell Physiol., 47: 211-216.
- Takaichi, S., and Mochimaru, M. 2007. Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. Cell. Mol. Life Sci., 64: 2607-2619.
- Takeyama, H., Sunarjo, J., Yamada, A., Matsumura, H., Kusakabe, E., and Matsunaga, T. 1996. β -Carotene production in a novel hydrogen-producing marine photosynthetic bacterium *Rhodovulum sulfidophilum* expressing the *Erythrobacter longus* OCh101 *crtI* and *crtY* genes. J. Mar. Biotechnol., 4: 224-229.
- Tao, L., and Cheng, Q. 2004. Novel β -carotene ketolases from non-photosynthetic bacteria for canthaxanthin synthesis. Mol. Genet. Genomics, 272: 530-537.

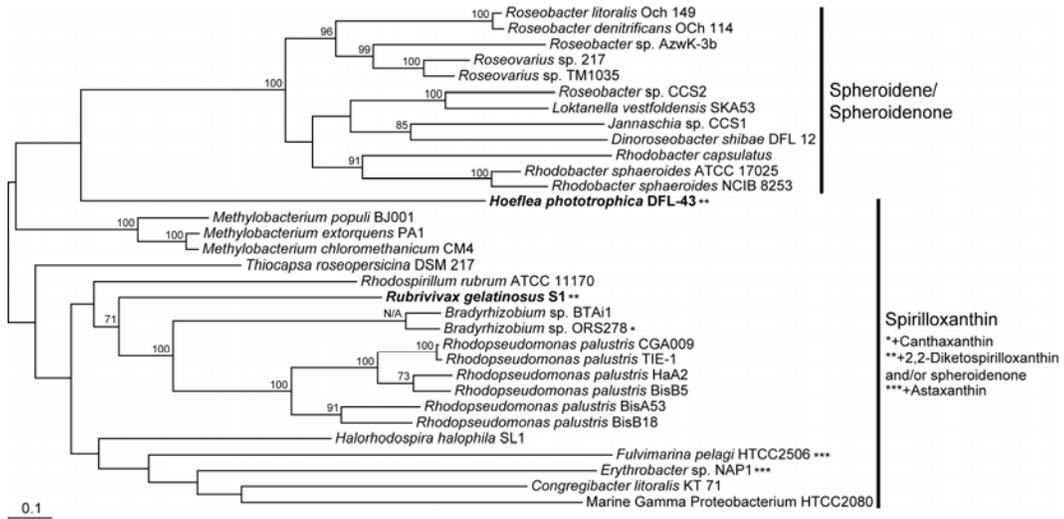
- Tao, L., Picataggio, S., Rouvière, P.E., and Cheng, Q. 2004. Asymmetrically acting lycopene β -cyclases (CrtLm) from non-photosynthetic bacteria. *Mol. Genet. Genomics*, 271: 180-188.
- Tao, L., Schenzle, A., Odom, J.M., and Cheng, Q. 2005. Novel carotenoid oxidase involved in biosynthesis of 4,4'-diapolycopene dialdehyde. *Appl. Environ. Microbiol.*, 71: 3294-3301.
- Tao, L., Rouvière, P.E., and Cheng, Q. 2006a. A carotenoid synthesis gene cluster from a non-marine *Brevundimonas* that synthesizes hydroxylated astaxanthin. *Gene*, 379: 101-108.
- Tao, L., Yao, H., Kasai, H., Misawa, N., and Cheng, Q. 2006b. A carotenoid synthesis gene cluster from *Algoriphagus* sp. KK10202C with a novel fusion-type lycopene β -cyclase gene. *Mol. Genet. Genomics*, 276: 79-86.
- Tao, L., Yao, H., and Cheng, Q. 2007. Genes from a *Dietzia* sp. for synthesis of C₄₀ and C₅₀ β -cyclic carotenoids. *Gene*, 386: 90-97.
- Teramoto, M., Takaichi, S., Inomata, Y., Ikenaga, H., and Misawa, N. 2003. Structural and functional analysis of a lycopene β -monocyclase gene isolated from a unique marine bacterium that produces myxol. *FEBS Lett.*, 545: 120-126.
- Teramoto, M., Rähler, N., Misawa, N., and Sandmann, G. 2004. 1-Hydroxy monocyclic carotenoid 3,4-dehydrogenase from a marine bacterium that produces myxol. *FEBS Lett.*, 570: 184-188.
- Tian, B., Sun, Z., Xu, Z., Shen, S., Wang, H., and Hua, Y. 2008. Carotenoid 3',4'-desaturase is involved in carotenoid biosynthesis in the radioresistant bacterium *Deinococcus radiodurans*. *Microbiology*, 154: 3697-3706.
- Tsuchiya, T., Takaichi, S., Misawa, N., Maoka, T., Miyashita, H., and Mimuro, M. 2005. The cyanobacterium *Gloeobacter violaceus* PCC 7421 uses bacterial-type phytoene desaturase in carotenoid biosynthesis. *FEBS Lett.*, 579: 2125-2129.
- Van Dien, S.J., Marx, C.J., O'Brien, B.N., and Lidstrom, M.E. 2003. Genetic characterization of the carotenoid biosynthetic pathway in *Methylobacterium extorquens* AM1 and isolation of a colorless mutant. *Appl. Environ. Microbiol.*, 69: 7563-7566.
- Velayos, A., Eslava, A.P., and Iturriaga, E.A. 2000. A bifunctional enzyme with lycopene cyclase and phytoene synthase activities is encoded by the *carRP* gene of *Mucor circinelloides*. *Eur. J. Biochem.*, 267: 5509-5519.

- Verdoes, J.C., Krubasik, P., Sandmann, G., and van Ooyen, A.J.J. 1999a. Isolation and functional characterisation of a novel type of carotenoid biosynthetic gene from *Xanthophyllomyces dendrorhous*. *Mol. Gen. Genet.*, 262: 453-461.
- Verdoes, J.C., Misawa, N., and van Ooyen, A.J.J. 1999b. Cloning and characterization of the astaxanthin biosynthetic gene encoding phytoene desaturase of *Xanthophyllomyces dendrorhous*. *Biotechnol. Bioeng.*, 63: 750-755.
- Viveiros, M., Krubasik, P., Sandmann, G., and Houssaini-Iraqi, M. 2000. Structural and functional analysis of the gene cluster encoding carotenoid biosynthesis in *Mycobacterium aurum* A+. *FEMS Microbiol. Lett.*, 187: 95-101.
- Wieland, B., Feil, C., Gloria-Maercker, E., Thumm, G., Lechner, M., Bravo, J.-M., Poralla, K., and Götz, F. 1994. Genetic and biochemical analyses of the biosynthesis of the yellow carotenoid 4,4'-diaponeurosporene of *Staphylococcus aureus*. *J. Bacteriol.*, 176: 7719-7726.
- Xu, Z., Tian, B., Sun, Z., Lin, J., and Hua, Y. 2007. Identification and functional analysis of a phytoene desaturase gene from the extremely radioresistant bacterium *Deinococcus radiodurans*. *Microbiology*, 153: 1642-1652.
- Yokoyama, A., Izumida, H., and Miki, W. 1994. Production of astaxanthin and 4-ketozeaxanthin by the marine bacterium *Agrobacterium aurantiacum*. *Biosci. Biotechnol. Biochem.*, 58: 1842-1844.
- Yokoyama, A., and Miki, W. 1995. Isolation of myxol from a marine bacterium *Flavobacterium* sp. associated with a marine sponge. *Fish. Sci.*, 61: 684-686.
- Yokoyama, A., Miki, W., Izumida, H., and Shizuri, Y. 1996. New trihydroxy-keto-carotenoids isolated from an astaxanthin-producing marine bacterium. *Biosci. Biotechnol. Biochem.*, 60: 200-203.

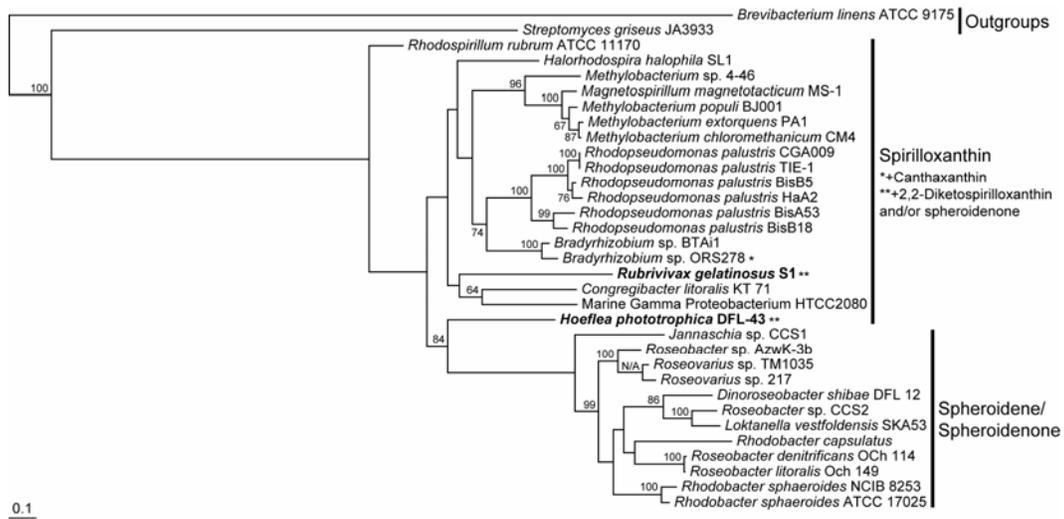
Appendix E – Supplemental Data for Chapter 7

- Supplemental Figure E1: RAxML for purple bacterial CrtB (a), CrtC (b), CrtF (c) and CrtI (d) protein sequences
- Supplemental Figure E2: Plots of d_n and d_s for purple bacterial CrtA, CrtB, CrtC, CrtD, CrtF and CrtI using the (a) Nei-Gojobori with the Jukes-Cantor correction, (b) Li-Wu-Lou, (c) Pamilo-Bianchi-Li and (d) Kumar methods

(c)



(d)



Supplemental Figure E2. Plots of pair-wise non-synonymous and synonymous substitution rates for primarily spheroidenone-producing bacteria (circles) and primarily spirilloxanthin-producing bacteria (squares). Values from comparisons involving *Hoeflea phototrophica* DFL-43 and *Rubrivivax gelatinosus* S1 are indicated in orange and blue respectively. Also shown are trend lines for d_n/d_s ratios of 1 and 0.2. Plots were constructed using the (a) Nei-Gojobori with the Jukes-Cantor correction, (b) Li-Wu-Lou, (c) Pamilo-Bianchi-Li and (d) Kumar methods. For plots generated using *crtA*, *crtI* and the Nei-Gojobori with the Jukes-Cantor correction method see Figure 3. See also the summary and comparative statistics in Table 1 and Supplementary Table 1.

(a)

