

ECOLOGICAL GENOMICS OF FILAMENTOUS ANOXYGENIC
PHOTOTROPHIC BACTERIA INHABITING GEOTHERMAL
SPRINGS IN YELLOWSTONE NATIONAL PARK

by

Christian Gerald Klatt

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

in

Ecology and Environmental Sciences

MONTANA STATE UNIVERSITY
Bozeman, Montana

May, 2012

© Copyright

by

Christian Gerald Klatt

2012

All Rights Reserved

APPROVAL

of a dissertation submitted by

Christian Gerald Klatt

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to The Graduate School.

Dr. David M. Ward

Approved for the Department of Land Resources and Environmental Sciences

Dr. Tracy M. Sterling

Approved for The Graduate School

Dr. Carl A. Fox

STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted “the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part.”

Christian Gerald Klatt

May, 2012

DEDICATION

I dedicate this work to my best friend and life partner, Carrie Taylor. Her patience and encouragement have kept me on track to reaching my goals, and she has always gently reminded me to contemplate my true path in life.

The Voice of the Ancient Bard

Youth of delight! come hither
And see the opening morn,
Image of Truth new-born.
Doubt is fled, and clouds of reason,
Dark disputes and artful teasing.
Folly is an endless maze;
Tangled roots perplex her ways;
How many have fallen there!
They stumble all night over bones of the dead;
And feel —they know not what but care;
And wish to lead others, when they should be led.

~William Blake,
Songs of Experience 1794

ACKNOWLEDGEMENTS

First and foremost, I thank Dr. David Ward for providing the opportunity to work on these projects, and I'm grateful for his mentorship and instruction over the years. Additionally, this work could not have been done without data and insight provided by Dr. Don Bryant, and I'll fondly remember our shared excitement in the initial discovery of a gene predicted to encode a subunit of a peculiar type-I photosystem reaction center early in our analysis of the metagenomic data from Octopus and Mushroom Springs. I gratefully acknowledge support from an Integrative Graduate Education and Research Traineeship award in Geobiological Systems (NSF Grant #DGE 0654336) in the second and third years of my program. I am also grateful for the support of past and present members of the Ward Lab, including Mary Bateson, Eric Becraft, Melanie Melendrez, and Jason Wood. Jason has taught me to embrace the machine, and I am indebted to him for providing robust foundations of code upon which I have subsequently built rickety data structures. He made much of the bioinformatic analyses in Chapter 3 possible. I'm thankful for the mentorship and collaboration with Dr. Bill Inskeep in both the IGERT program and the metagenomics project presented in Chapter 4. Also, I thank Jay (Zhenfeng) Liu from the Bryant Lab for sharing techniques and alignments in the analysis of the metatranscriptomic data in Chapter 5. I have indicated other funding support for projects with the corresponding chapters in this thesis.

TABLE OF CONTENTS

1. INTRODUCTION	1
2. COMPARATIVE GENOMICS PROVIDES EVIDENCE FOR THE 3-HYDROXYPROPIONATE AUTOTROPHIC PATHWAY IN FILAMENTOUS ANOXYGENIC PHOTOTROPHIC BACTERIA AND IN HOT SPRING MICROBIAL MATS	13
Contribution of Authors and Co-Authors.....	13
Manuscript Information Page.....	14
Summary.....	15
Introduction	15
Results and Discussion	19
Genome Annotation Evidence of 3-OHP Pathway in FAP Isolates	19
Similarity in Genes and Gene Order in <i>Chloroflexus</i> and <i>Roseiflexus</i>	27
Absence of Alternative Autotrophic Pathways	28
Environmental Genomic Analysis	29
Conclusions	31
Experimental Procedures.....	32
Metagenome Library Construction and Assembly	32
BLAST Comparisons	33
Hidden Markov Model Analysis.....	34
Phylogenetic Analysis.....	34
Acknowledgements	34
3. COMMUNITY ECOLOGY OF HOT SPRING CYANO- BACTERIAL MATS: PREDOMINANT POPULATIONS AND THEIR FUNCTIONAL POTENTIAL.....	36
Contribution of Authors and Co-Authors.....	36
Manuscript Information Page.....	38
Abstract	39
Introduction	39
Methods.....	42
Collection, Preliminary Sequence Analysis, and Metagenomic Sequencing ...	42
Metagenome Assembly and Annotation	43
Clustering and Characterization of Assemblies	44
BLASTN Recruitment and Synteny with Reference Genomes.....	44
Results.....	45
Major Populations and their Functional Potential	46
Patterns of Metagenomic Diversity	61

TABLE OF CONTENTS – CONTINUED

Evidence of Homologous Recombination	63
Discussion	65
Linkage Between Community Composition and Potential Community Function	66
Description of Functional Guilds	68
Diversity Within Scaffold Clusters	69
Insights Into Genome Evolution	70
Conclusion.....	71
Acknowledgements	72
 4. COMMUNITY STRUCTURE AND FUNCTION OF HIGH- TEMPERATURE PHOTOTROPHIC MICROBIAL MATS INHABITING DIVERSE GEOTHERMAL ENVIRON- MENTS	 74
Contribution of Authors and Co-Authors.....	74
Manuscript Information Page.....	76
Abstract	77
Introduction	78
Results	81
Geochemical and Physical Context	81
Analysis of Metagenome Sequences.....	84
Phylogenetic Analysis of Metagenome Assemblies.....	86
Chloroflexi Diversity and Distribution.....	93
Geochemical Influences on Community Composition.....	96
Functional Analysis of Predominant Sequence Assemblies	96
Discussion	101
Conclusion.....	104
Materials and Methods.....	104
Sample Collection and Geochemical Analyses.....	104
DNA Extraction and Preparation	105
Pre-Assembly Metagenomic Sequence Analyses	106
Sequence Assembly and Annotation.....	106
Ribosomal RNA Sequence Analyses.....	107
Statistical Analyses	108
Sequence Availability.....	108
 5. TEMPORAL PATTERNING OF IN SITU GENE EXPRESSION IN UNCULTIVATED PHOTO-	

TABLE OF CONTENTS – CONTINUED

TROPHIC CHLOROFLEXI INHABITING AN ALKALINE SILICEOUS GEOTHERMAL SPRING.....	109
Contribution of Authors and Co-Authors.....	109
Manuscript Information Page.....	110
Abstract.....	111
Introduction	112
Materials and Methods.....	114
Metagenomic Analyses.....	114
Collection and Preparation of Microbial Mat Samples.....	115
Nucleic Acid Extraction and Analysis.....	116
cDNA Synthesis	117
Alignment and Statistical Analyses of cDNA Sequences	117
Clustering and Visualization of Gene Expression Patterns.....	119
Results and Discussion	119
Metagenomes of FAP Populations	119
Metatranscriptomes of FAP Populations	126
Photosynthesis	128
Bacteriochlorophyll Biosynthesis.....	129
Electron Transport Complexes	131
Mixotrophy and the TCA/3-OHP Cycles.....	135
Alternative Reactions Involving CO ₂	138
Glycolysis/Gluconeogenesis	140
Heterotrophic Carbon Assimilation and Storage	141
Nitrogen and Hydrogen Metabolism	144
Conclusions	145
6. CONCLUSIONS AND RELATION TO OTHER COLLABORATIVE WORK.....	147
APPENDICES	154
APPENDIX A: Chapter 2 Appendix	155
APPENDIX B: Chapter 3 Appendix	158
APPENDIX C: Chapter 4 Appendix	226
APPENDIX D: Chapter 5 Appendix	230
REFERENCES CITED.....	233

LIST OF TABLES

Table		Page
2.1	Isolate Organisms Investigated in this Study.	19
2.2	Percent Amino Acid Identity and Similarity of ORFs Coding for Experimentally Characterized (bold) and Uncharacterized Enzymes of the 3-OHP pathway in <i>C. aurantiacus</i> to Orthologs of Chloroflexi Isolate Genomes.	21
3.1	Assembly Statistics of Scaffold Clusters $\geq 20\ 000$ bp in Length.	47
3.2	Comparison of Metagenomic Analyses Based on Genome Recruitment and Assembly	50
3.3	Phylogenetic Marker Genes and Gunctional Genes in Assembly Clusters.	51
3.4	Relationship Between Predominant Phylogenetic Groups, Functional Potential and Functional Guilds.	53
4.1	Sample Location, Aqueous Geochemical Parameters and Physical Context of Six, High-temperature Phototrophic Microbial Communities in Yellowstone National Park (YNP)	82
4.2	Properties of Metagenomic Scaffold Clusters as Demarcated with Oligonucleotide Composition.	89
4.3	Phylogenetic Distribution of Phototrophic, Autotrophic, and Sulfur Cycling Genes in Metagenomes	99
5.1	Genome and Metagenome Scaffolds Used in the Analysis of Metatranscriptomes.	121
5.2	Expected <i>Chloroflexus</i> spp. Genes Absent in the <i>Chloroflexus</i> Metagenome Scaffolds.	123
5.3	Expression Categories of Genes Involved in Electron Transport.	133

LIST OF FIGURES

Figure		Page
1.1	Phylogenies of Extant Phototrophic Bacteria.	5
1.2	Diel Model of FAP Physiology.....	9
2.1	The 3-Hydroxypropionate Pathway as Proposed for <i>Chloroflexus aurantiacus</i>	17
2.2	Locations of Genes on Isolate Genome and Metagenome Contigs.....	23
2.3	Partial Alignment and Phylogeny of Prokaryotic Carboxyltransferases...	24
2.4	Per Cent Amino Acid Identity of Metagenome Sequences Encoding 3-OHP Pathway Genes to Homologues in the <i>C. aurantiacus</i> and <i>Roseiflexus</i> sp. RS-1 Genomes	31
3.1	Network Map of Core Scaffold Clusters Observed in Celera Assemblies..	48
3.2	Histograms of Recruited Metagenomic Sequences.	49
3.3	PufL and PufM Phylogeny and Genomic Context.	56
3.4	Position of Metagenomic Sequence Alignments on <i>Synechococcus</i> sp. A Genome	62
3.5	Synteny Conservation Between the <i>Synechococcus</i> sp. Strain A Genome and Metagenomic Sequences and other Genomes.....	64
4.1	Site Photographs of the Microbial Mats Selected for Metagenome Sequencing in the Current Study	83
4.2	Percent G+C Content of Individual Metagenome Sequences	85
4.3	Oligonucleotide Frequency Principal Components Ordination of Assemblies from BLVA_5 and BLVA_20	88
4.4	Scaffold Oligonucleotide Frequency Similarity Network	90
4.5	Comparison of the Distribution of Phylogenetic Marker Genes from Metagenomes and from 16S rRNA Clones.....	92
4.6	Comparison of Chloroflexi Phylogenetic Marker Genes from Metagenomes and Chloroflexi 16S rRNA Clones	94
4.7	Unrooted Neighbor-joining Phylogenetic Trees of Chloroflexi 16S rRNA Sequences from PCR Clone Libraries.....	95
4.8	Ordination of Geochemical and Community Distance Matrices.....	97

LIST OF FIGURES – CONTINUED

Figure		Page
5.1	Major Transcription Categories.....	120
5.2	Total Transcript Abundance Levels of <i>Roseiflexus</i> and <i>Chloroflexus</i> Transcripts.	127
5.3	Expression of Phototrophy Genes.....	129
5.4	The Integrated TCA and 3-OHP Pathways for Mixotrophic Metabolism.	137
5.5	A Diel Model of Central Carbon Metabolism in <i>Roseiflexus</i> spp.....	139
6.1	Daytime Guild Interactions Derived from Flux Models.	153

ABSTRACT

The filamentous anoxygenic phototrophic bacteria (FAPs) are dominant members of many phototrophic microbial mat communities in geothermal springs. In non-sulfidic springs, FAPs are known to primarily utilize photoheterotrophic metabolism, where they incorporate organic carbon sources such as glycolate or acetate, which are byproducts of cyanobacterial metabolism. Cultures of *Chloroflexus aurantiacus* have also been shown to be capable of photoautotrophic metabolism via the 3-hydroxypropionate pathway in culture. FAPs in non-sulfidic springs have been shown to take up bicarbonate, and this behavior is stimulated by light, H₂, and H₂S. However, previously investigated mat communities contain FAPs that are more closely related to *Roseiflexus* spp. which have not demonstrated autotrophic growth in culture. This work aimed to *i*) determine whether *Roseiflexus* spp. isolates and uncultured FAPs contain genes necessary for autotrophy, *ii*) compare the community structures of FAPs in different environments, and *iii*) observe patterns in gene transcription over an entire diel period, which may indicate how these organisms physiologically acclimate to changing environmental conditions.

Comparisons among multiple genomes revealed that *Roseiflexus* spp. contain genes necessary for the 3-hydroxypropionate pathway. A metagenomic investigation of the dominant constituents of the communities in Octopus Spring and Mushroom Spring resulted in the discovery of novel phototrophic organisms. Functional attributes were assigned to eight dominant ecological guilds, including three previously unknown phototrophic bacteria belonging to Kingdoms Acidobacteria, Chlorobi, and Chloroflexi. Metagenomic sequencing of six communities from diverse geochemical environments revealed the presence of FAPs and other phototrophic bacteria, however there was evidence that some FAPs were unique to particular springs. Examination of transcripts produced by FAPs inhabiting Mushroom Spring indicated that genes related to phototrophy are most highly expressed at night, which presumably allows for phototrophic metabolism in the morning. Additionally, FAPs are predicted to utilize carbon and energy storage compounds such as polyglucose, wax esters, and polyhydroxyalkanoates. Based upon the transcription profiles of relevant genes, a model of their carbon and energy metabolism is proposed. Taken together, these genomic, metagenomic, and metatranscriptomic studies have advanced the understanding of FAP diversity and both the community and physiological ecology in geothermal springs.

CHAPTER 1

INTRODUCTION

Photoautotrophy, defined as the utilization of light for energy coupled with the biological incorporation of inorganic carbon, is the primary material and energetic input for the vast majority of ecosystems on Earth. Notable exceptions to these photoautotrophic systems are geochemical or thermal ecosystems where chemolithotrophic metabolisms are the primary sources of carbon and energy, however, photoautotrophic organisms have also adapted to thermal environments. A defining characteristic of these 'extremophilic' phototrophic microbial communities is the absence of grazing consumers; above temperatures of approximately 42 to 50 °C, environmental conditions often exceed the physiological adaptations of eukaryotic organisms such that they are typically excluded from these environments (Wickstrom and Castenholz, 1973). This exclusion of grazers results in the formation of thick mats (on the order of millimeters to centimeters) of densely packed cells ($\sim 10^{10}$ cells cm^{-3}) (Bauld and Brock, 1973; Brock, 1978; Ward *et al.*, 1989b, 1992). These phototrophic microbial mats are generally less diverse than mesophilic communities, which makes them tractable for studies aimed at establishing links between the diversity within microbial communities and the functions catalyzed by community members that drive the cycling of material and energy.

Thermophilic phototroph communities can be found in hot springs all over the world, including Iceland (Castenholz, 1969b, 1976; Jørgensen and Nelson, 1988; Skirnisdottir *et al.*, 2000), Japan (Nakagawa and Fukui, 2002; Hanada, 2003), New Zealand (Castenholz, 1976), and North America, especially in Oregon (Wickstrom and Castenholz, 1985; Richardson and Castenholz, 1987) and in numerous hot springs

of Yellowstone National Park in Wyoming (Brock, 1978; Ward *et al.*, 1989b). With the exception of eukaryotic algae of the Order Cyanidiales that can inhabit acidic springs at temperatures up to $\sim 55^{\circ}\text{C}$ (Ferris *et al.*, 2005; Toplin *et al.*, 2008), the constituents of these mats are typically strictly prokaryotic. The upper temperature limit for the distribution of thermophilic cyanobacteria is typically 72°C (Brock and Brock, 1968) and is lower in the presence of sulfide (Castenholz, 1977, 1978). Anoxygenic phototrophic bacteria can inhabit alkaline to neutral springs ($\text{pH} \sim 4.5 - 9$) with temperatures ranging from $\sim 45 - 72^{\circ}\text{C}$, while sulfide concentrations influence the upper temperatures at which these organisms are found (Castenholz, 1977; Castenholz and Pierson, 1995). Yellowstone mats have been studied intensively using both cultivation-based (Bauld and Brock, 1974; Pierson and Castenholz, 1974a; Madigan *et al.*, 1974; Madigan and Brock, 1975; Pierson *et al.*, 1985; Giovannoni *et al.*, 1987) and molecular-based methods (Ward *et al.*, 1990; Ward, 1998; Nübel *et al.*, 2002; Boomer *et al.*, 2002; Miller *et al.*, 2009).

Two alkaline siliceous hot springs in the Lower Geyser Basin of Yellowstone, Mushroom Spring and Octopus Spring, have been particularly well studied with molecular methods. These studies have revealed that the most abundant community members inhabiting the effluent channels of these springs are a mix of oxygenic and anoxygenic phototrophic bacteria. The former are unicellular cyanobacteria most closely related to the cultured isolates *Synechococcus* spp. strains A and B', which co-inhabit these mats together with anoxygenic phototrophs of the bacterial Kingdom Chloroflexi. The filamentous anoxygenic phototrophs (FAPs) from this latter group were once thought to be close relatives of the isolate *Chloroflexus aurantiacus*, given that this organism was the first FAP to be cultivated from springs such as these (Pierson and Castenholz, 1974a). The application of molecular techniques to describe the community structure of these and other low-sulfide alkaline-siliceous springs revealed that,

while *Chloroflexus* spp. were present, a distinct group of FAPs belonging to the sister genus *Roseiflexus* was also present (Weller *et al.*, 1992) and members of this group were found to be the more dominant FAPs at temperatures below 65 °C in Octopus Spring, Mushroom Spring, and Fairy Geyser mats (Nübel *et al.*, 2002; Boomer *et al.*, 2002).

Understanding the contemporary community structures and functions of these mats is important for interpreting how ancient phototrophic microbial mats, which were lithified to form stromatolite fossils (Doemel and Brock, 1974; Des Marais, 1991), may have formed and persisted. The FAPs in Kingdom Chloroflexi are significant with respect to their potential contribution to mat building processes in ancient microbial mats, which underscores the need to understand their role in modern mats such that geochemical signatures in stromatolites may be interpreted correctly. Mats that were preserved in the Precambrian geologic record were prominent before ~1 GYA, and their decline is attributed to the evolution of grazing eukaryotic organisms (Walter and Heys, 1985). Prior to the evolution of oxygen-evolving photosynthesis by ancestral cyanobacteria, it is thought that these mats were predominantly composed of anoxygenic phototrophs (Olson, 2006); however, there is also evidence for ancient mats composed of both oxygenic and anoxygenic phototrophs (Awramik, 1992).

Of all known organisms capable of chlorophyll-based phototrophic metabolism (as opposed to phototrophic metabolisms based upon rhodopsin-mediated proton translocation), Chloroflexi occupy the most basal lineage (i.e. closest to the last universal common ancestor of the three domains of life) based on comparative analysis of 16S rRNA sequences (Figure 1.1; Oyaizu *et al.* 1987; Woese 1987). Similar to anoxygenic phototrophs belonging to various lineages of α -, β -, and γ - Proteobacteria (the so-called purple non-sulfur and purple sulfur bacteria), FAPs utilize a type-2, or quinone-based phototrophic reaction center (RC) homologous to photosystem (PS) II

in cyanobacteria and plants. These reaction centers share a common evolutionary origin with the type-1 FeS-based RCs homologous to PS I in oxygenic phototrophs, and the RCs of anoxygenic phototrophs such as phototrophic Chlorobi and gram-positive Heliobacteria (Figure 1.1B; Bruce *et al.* 1982; Yamada *et al.* 2005). This phylogenetic position has implied that Chloroflexi are descendants of the most ancestral lineage of bacteria capable of phototrophy (Castenholz and Pierson, 1995), however it is possible that phototrophy could have later been acquired by the Chloroflexi via horizontal gene transfer.

Phylogenetic analyses of loci encoding heat-shock proteins (Hsp70 and Hsp70) suggest that other phototrophic groups were possibly more ancestral, however these analyses still support Chloroflexi as being members of the most ancestral lineage of the type-2 RC-containing phototrophic bacteria (Gupta *et al.* 1999; but see a contrasting view from the phylogeny of chlorophyll biosynthesis genes that suggest that the proteobacterial phototrophs were the most ancestral in Xiong *et al.* 1998, 2000). Genome-wide phylogenetic analysis has shown that horizontal gene exchange has indeed occurred among the different phototrophic lineages, leading to inconsistent inferences depending upon the loci chosen (Figure 1.1C, Raymond *et al.* 2002); irregardless, these same studies have revealed that the phylogenies inferred from a plurality of orthologous genes among phototrophic organisms are consistent with those of the early studies of 16S rRNA (Raymond *et al.*, 2003). These results strongly suggest that ancestral Chloroflexi were integral community members of ancient phototrophic microbial mats, with or without oxygen-evolving cyanobacteria.

Chloroflexus and *Roseiflexus* spp. are ecologically and physiologically similar in their capacity for photoheterotrophic and aerobic respiratory metabolisms. In studies of pure cultures of *Chloroflexus aurantiacus*, it was found that cells grew most rapidly with light and minimal media supplemented with short chain organic acids,

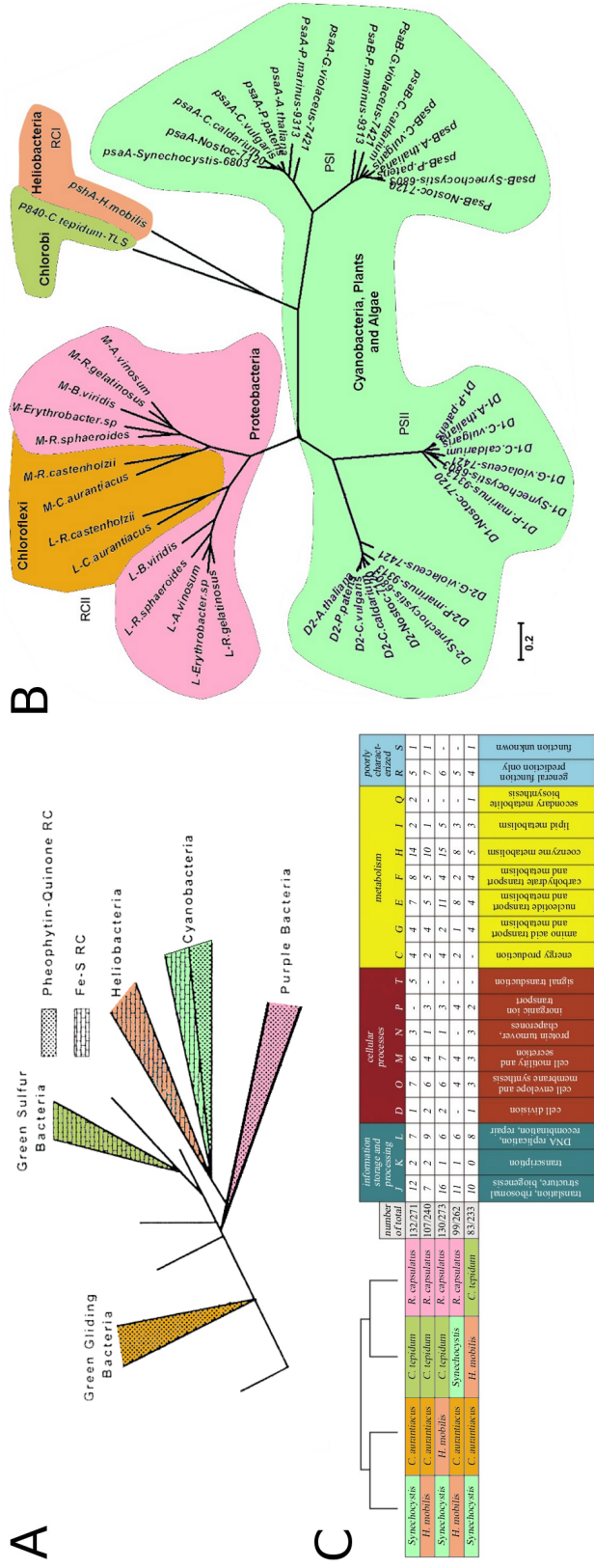


Figure 1.1: Phylogenies of Extant Phototrophic Bacteria. A) A least-squares distance-based phylogenetic tree based on 16S rRNA sequences of phototrophic bacteria with the corresponding reaction center types indicated as pheophytin-quinone RC (type-2) and Fe-S RC (type-1). Figure adapted from Blankenship (1992). B) An unrooted neighbor-joining phylogeny based on photosynthetic reaction center protein sequences, with phylogenetic groups colored as in A. Figure adapted from Sadekar *et al.* (2006). C) Whole genome analyses of orthologs found in four phototrophic bacteria; the numbers of orthologs in the table on the right are broken up into gene categories (Clusters of Orthologous Groups) that support the example unrooted trees on the left. Table from Raymond *et al.* (2003).

hexose sugars, and amino acids (Madigan *et al.*, 1974). The same was found for *Roseiflexus castenholzii*, however undefined media containing yeast extract supported the most rapid growth, followed by citrate, lactate, glucose, and casamino acids (Hanada *et al.*, 2002; van der Meer *et al.*, 2010). These results supported the inference that populations of FAPs in their natural environments primarily exhibit a photoheterotrophic metabolism during the day when light is available. Subsequent experiments determined that cells with filamentous morphology photoassimilate organic acids (most notably acetate) when mat organisms were incubated with radiolabeled compounds (Anderson *et al.*, 1987). Cyanobacteria were determined to be the primary source of these low-molecular weight organic acids, which they excrete as a byproduct of polyglucose fermentation (Nold and Ward, 1996). In addition to fermentation products, cyanobacteria were also found to excrete the compound glycolate, which is produced as a byproduct of photorespiration (i.e., the oxygenase activity of the ribulose biphosphate carboxylase/oxygenase, or RuBisCO, enzyme) (Bateson and Ward, 1988) due to the high oxygen concentrations in these mats during peak daylight hours (Revsbech and Ward, 1984). Filamentous cells were found to photoassimilate glycolate as well (Bateson and Ward, 1988), supporting the hypothesis that FAPs utilize a range of organic carbon substrates that are cross-fed from cyanobacteria to support photoheterotrophic metabolism during the day. The physiological ecology of FAPs during the night was less clear. Some of the first culture studies of *Chloroflexus aurantiacus* revealed that aerobic respiratory growth occurred in the dark (Pierson and Castenholz, 1974a; Madigan *et al.*, 1974), and these observations influenced early inferences that FAPs aerobically respire at night in situ; it was even suggested that FAPs use their gliding motility to migrate to the surface of the mat at night in response to the need to overcome diffusion limitations of O₂ (Brock, 1978).

One important difference in growth experiments on organisms of these genera was the unique ability for *Chloroflexus* spp. cultures to grow photoautotrophically, with HCO_3^- as the sole carbon source and either H_2S (Madigan and Brock, 1977; Giovannoni *et al.*, 1987) or H_2 (Holo and Sirevåg, 1986) as electron donors; no such photoautotrophic growth has yet been demonstrated for *Roseiflexus* spp. cultures (Hanada *et al.*, 2002; van der Meer *et al.*, 2010). Cell extracts of autotrophically grown *Chloroflexus aurantiacus* did not have ribulose biphosphate carboxylase or ATP citrate lyase activity, such that these organisms were hypothesized to use a pathway for reduction of CO_2 other than the reductive pentose phosphate pathway (i.e., Calvin-Benson-Bassham cycle) or the reductive tricarboxylic acid pathway, respectively (Holo and Sirevåg, 1986). Subsequent investigations elucidated the 3-hydroxypropionate (3-OHP) pathway (Strauss and Fuchs, 1993; Herter *et al.*, 2002b), which utilizes the novel enzymes malonyl-CoA reductase (Hügler *et al.*, 2002) and propionyl-CoA synthase (Alber and Fuchs, 2002) and shares biochemical reactions with fatty acid biosynthesis (acetyl-CoA and propionyl-CoA carboxylases) and the tricarboxylic acid cycle (succinate dehydrogenase and fumarate hydratase; Zarzycki *et al.* 2009).

Given the autotrophic potential of some FAPs, there was interest as to whether these organisms were contributing to primary production in these mats through use of the 3-OHP pathway. Field studies in Yellowstone springs that were focused on the natural abundance of stable isotopes in lipid biomarkers diagnostic of FAPs indicated that the 3-OHP pathway could be occurring in these mats (van der Meer *et al.*, 2000). In low-sulfide systems where cyanobacteria are present, the primary input of inorganic carbon into biomass is assumed to be through the cyanobacterial reductive pentose phosphate pathway, which imparts an isotopic signature that is 20-25 ‰ lighter in $\delta^{13}\text{C}$ (i.e., relatively depleted in ^{13}C due to the kinetic isotope effect characteristic

of the reaction catalyzed by RuBisCO compared to the isotopic composition of the source pool of inorganic carbon; Madigan *et al.* 1989; Sakata *et al.* 1997). Assuming that FAPs primarily use organic carbon derived from cyanobacterial photosynthesis, it was thought that the isotopic composition of their lipids would be similar to those of cyanobacteria. Contrastingly, the $\delta^{13}\text{C}$ of carbon fixed from the 3-OHP pathway was known to impart less isotopic discrimination than the reductive pentose phosphate pathway from studies of autotrophically grown *Chloroflexus aurantiacus* cultures (Holo and Sirevåg, 1986; van der Meer *et al.*, 2001), and thus heavier isotopic signatures in Chloroflexi-specific lipids would potentially indicate FAP autotrophy. The $\delta^{13}\text{C}$ s of cyanobacteria-specific lipids such as n-C₁₇ alkanes were found to exhibit $\delta^{13}\text{C}$ values of -34-36 ‰, whereas FAP-specific C_{31:3} alkenes and wax esters exhibited $\delta^{13}\text{C}$ values ranging from -9 to -24 ‰ (van der Meer *et al.*, 2000, 2003), suggesting the possibility that FAPs conduct photoautotrophy in situ. This isotopic difference was subsequently corroborated in a study in which a Percoll density gradient centrifugation was used to separate FAPs and cyanobacteria based upon differences in the density of their cells; this effectively separated the mat into a green fraction that was ~60-fold enriched in cyanobacterial cells, and a brown fraction that was ~2-fold enriched in FAPs. The isotopic composition of the cyanobacteria-dominated fraction exhibited a lighter isotopic composition (relatively depleted in ¹³C) compared to the FAP-dominated fraction, especially with respect to the specific lipid biomarkers mentioned above (van der Meer *et al.*, 2007). Finally, evidence of FAP autotrophy was most definitively demonstrated by showing incorporation of isotopically labeled $\text{H}^{13}\text{CO}_3^-$ into FAP biomarkers, especially when incubated with H₂ or H₂S as a source of electrons (van der Meer *et al.*, 2005); these labeling studies also suggested that FAPs have higher rates of bicarbonate incorporation in the morning compared to the afternoon (Figure 1.2; van der Meer *et al.* 2007).

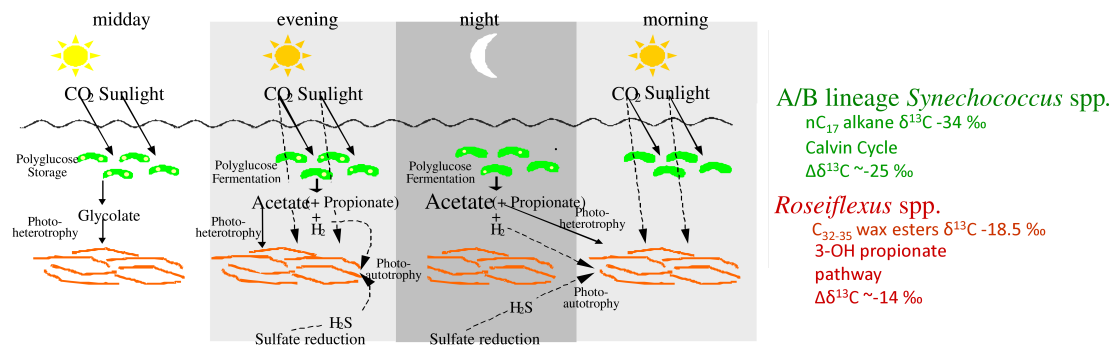


Figure 1.2: Diel Model of FAP Physiology. During the day, *Synechococcus* spp. are responsible for the majority of inorganic carbon input by way of the Calvin-Benson-Bassham cycle, which imparts a relatively lighter isotopic composition to cyanobacterial-specific lipid biomarkers. FAPs couple the uptake of glycolate with photic energy input during the day, while switching to *Synechococcus* spp. fermentation products such as acetate and propionate during the night. FAPs are predicted to be photoautotrophic during the evening and morning when electron donors such as H₂ and H₂S are most readily available, and this autotrophy via the 3-OHP pathway imparts heavier isotopic signatures to wax esters specific to FAPs. Adapted from van der Meer *et al.* (2005).

Despite the fact that *Roseiflexus* spp. have never been successfully grown autotrophically in culture, it was of particular interest whether they had the potential for CO₂/HCO₃⁻ fixation, given their dominance at lower temperatures in Octopus Spring and Mushroom Spring. The above-mentioned lipid biomarkers did not differentiate between *Chloroflexus* or *Roseiflexus* spp., such that it was still an open question as to whether *Roseiflexus* spp. were photoautotrophic in situ. The genomic sequencing of the *Roseiflexus* sp. RS-1 isolate combined with the random shotgun metagenomic sequencing of DNA extracted from the mat communities of Mushroom Spring and Octopus Spring (van der Meer *et al.*, 2010) enabled me to determine whether this isolate and its relatives in the mat community, like their *Chloroflexus* spp. relatives, were capable of utilizing the 3-OHP pathway for autotrophy. The results of this initial investigation are presented in Chapter 2.

While metagenomic sequencing was utilized to obtain evidence concerning the autotrophic potential for native *Chloroflexus* and *Roseiflexus* spp., this method simultaneously produced genomic data that allowed me to analyze the functional potential of the most dominant members of the Octopus Spring and Mushroom Spring communities. In Chapter 3, the context of these FAP-containing microbial communities (as revealed by metagenomic sequencing) is presented, in which particular phylogenetic groups that were previously detected by ribosomal RNA-based molecular approaches could now be categorized into various functional groups. Moreover, these studies revealed the presence of two novel phototrophic bacteria in these mats that were not previously known to science. Overall, these findings led to inferences as to how *Roseiflexus* and *Chloroflexus* spp. each partition the environment into unique ecological niches with respect to their sympatric community members in these alkaline-siliceous hot springs.

The alkaline-siliceous springs have been extensively characterized, but FAPs are found in a diversity of environments within different geochemical and community contexts. Phototrophic Chloroflexi and other anoxygenic phototrophs are able to withstand higher levels of sulfide than cyanobacteria at temperatures above 50 °C, and anoxygenic phototrophic mats devoid of oxygenic phototrophs can be found above this temperature where sulfide concentration ranges from 30 to 130 μM (Castenholz, 1977; Giovannoni *et al.*, 1987; Ward *et al.*, 1992). These mats are geochemically distinct in that they are not subject to the diel fluctuations in oxygen concentrations that are experienced in mats with cyanobacteria present. The sulfidic carbonate springs at Mammoth Terraces in the northern part of Yellowstone Park support anoxygenic mats such as these, and previously there had been very little characterization of these communities with molecular-sequencing techniques (Ward *et al.*, 1997). Characterization of nearby chemolithotroph-dominated communities has revealed subdominant

populations of phototrophs e.g. (Fouke *et al.*, 2003), however, these communities are not visibly similar to phototroph-dominated mats. In addition to alkaline silicious and sulfidic carbonate springs, FAPs occupy a variety of other geothermal habitats in Yellowstone including iron-rich anoxic springs such as Chocolate Pots in the Gibbon River drainage, intermittently warm splash mats at Fairy Geyser, and larger thermal stream environments such as those found at White Creek (the latter two systems are located in the Lower Geyser Basin). A broader metagenomic survey of five different phototrophic Chloroflexi habitats is presented in Chapter 4, such that the same approach allowing links to be made between community structure and function could be applied to a more diverse set of geothermal environments.

The functional versatility of FAPs may in part explain their ubiquity among the phototrophic mat sites that are described in Chapter 4, however it remained unclear how FAPs temporally regulate their metabolism to cope with changing environmental conditions in a particular location. While genomes and metagenomes supported the hypothesis that Chloroflexi in these mats are capable of photoautotrophy, photoheterotrophy, and aerobic chemoorganotrophy, the inferences of these metabolisms remained at the state of testable hypotheses that needed to be supported by additional lines of evidence. Metatranscriptomics, or the sequencing of cDNA synthesized from whole-community extractions of RNA (both ribosomal RNA (rRNA) and messenger RNA (mRNA)) was applied to determine if the genes that were predicted to be involved in common physiological functions were co-ordinately transcribed. An initial pilot experiment that was conducted on 60 °C Mushroom Spring mat samples collected at key times of day (i.e., evening, predawn, low-light morning and high-light morning periods) confirmed that the two novel community members belonging to kingdoms Chlorobi and Chloroflexi whose phototrophic potential was detected by metagenomics indeed expressed genes involved in the assembly of phototrophic RCs

and the production of bacteriochlorophylls (Liu *et al.*, 2011b). It was surprising to discover that three key genes involved in the 3-OHP pathway in both *Chloroflexus* and *Roseiflexus* spp. were most highly transcribed in high light when the mat was highly oxic (Bryant *et al.*, 2012). This was significant considering the results of previous studies, which indicated that bicarbonate incorporation into FAP-specific lipids occurred most rapidly during the morning and evening low-light transition periods, when reductant in the form of H_2 was more readily available (van der Meer *et al.*, 2003). A second metatranscriptomic study was implemented to more closely examine the temporal transcription patterns of the Mushroom Spring community over an entire diel cycle in which higher temporal resolution was achieved by sampling on an hourly basis. Chapter 5 presents the results of this metatranscriptomic study for the *Roseiflexus* and *Chloroflexus* spp., which provided the basis for a model of the physiological strategies that these FAPs implement to obtain carbon and energetic resources in response to fluctuations in their availability.

In summary, the work represented in this dissertation aimed to contribute to an understanding of the diversity, and ecological physiology, and community ecology of phototrophic Chloroflexi populations in their native habitats. Chapter 6 highlights the major conclusions that were enabled using these genomic, metagenomic, and metatranscriptomic approaches. Additional projects relevant to the aim of this thesis are also summarized in this chapter. Finally, remaining questions and future directions for research are discussed.

CHAPTER 2

COMPARATIVE GENOMICS PROVIDES EVIDENCE FOR THE 3-HYDROXYPROPIONATE AUTOTROPHIC PATHWAY IN FILAMENTOUS ANOXYGENIC PHOTOTROPHIC BACTERIA AND IN HOT SPRING MICROBIAL MATS

Contribution of Authors and Co-Authors

Manuscript in Chapter 2

Author: Christian G. Klatt

Contributions: Designed the study, conducted the experiments, collected and analyzed output data, and wrote the manuscript. Sequencing was performed by The Institute for Genomic Research (TIGR, now the J. Craig Venter Institute)

Co-author: Donald A. Bryant

Contributions: Obtained funding, assisted with experimental design, discussed the results and edited the manuscript at all stages.

Co-author: David M. Ward

Contributions: Obtained funding, assisted with experimental design, discussed the results and edited the manuscript at all stages.

Manuscript Information Page

Christian G. Klatt, Donald A. Bryant, and David M. Ward

Journal Name: Environmental Microbiology

Status of Manuscript:

___ Prepared for submission to a peer-reviewed journal

___ Officially submitted to a peer-reviewed journal

___ Accepted by a peer-reviewed journal

X Published in a peer-reviewed journal

Published by the Society for Applied Microbiology in 2007, Issue 9 pages 2067-2078.

Summary

Stable carbon isotope signatures of diagnostic lipid biomarkers have suggested that *Roseiflexus* spp., the dominant filamentous anoxygenic phototrophic bacteria inhabiting microbial mats of alkaline siliceous hot springs, may be capable of fixing bicarbonate via the 3-hydroxypropionate pathway, which has been characterized in their distant relative, *Chloroflexus aurantiacus*. The genomes of three filamentous anoxygenic phototrophic Chloroflexi isolates (*Roseiflexus* sp. RS-1, *Roseiflexus castenholzii* and *Chloroflexus aggregans*), but not that of a non-photosynthetic Chloroflexi isolate (*Herpetosiphon aurantiacus*), were found to contain open reading frames that show a high degree of sequence similarity to genes encoding enzymes in the *C. aurantiacus* pathway. Metagenomic DNA sequences from the microbial mats of alkaline siliceous hot springs also contain homologues of these genes that are highly similar to genes in both *Roseiflexus* spp. and *Chloroflexus* spp. Thus, *Roseiflexus* spp. appear to have the genetic capacity for carbon dioxide reduction via the 3-hydroxypropionate pathway. This may contribute to heavier carbon isotopic signatures of the cell components of native *Roseiflexus* populations in mats compared with the signatures of cyanobacterial cell components, as a similar isotopic signature would be expected if *Roseiflexus* spp. were participating in photoheterotrophic uptake of cyanobacterial photosynthate produced by the reductive pentose-phosphate cycle.

Introduction

The microbial mats that develop in the effluent channels of alkaline siliceous hot springs of Yellowstone National Park are model systems for the study of microbial community ecology, and they are valuable modern analogues to ancient stromato-

lite formations (Ward *et al.*, 1998, 2006; van der Meer *et al.*, 2000). Based on our molecular and microscopic studies of Octopus and Mushroom Springs, these mat communities are dominated by two groups of phototrophs at 60 and 65 °C: unicellular cyanobacteria (*Synechococcus* spp.) and filamentous anoxygenic phototrophs (FAPs) related to *Chloroflexus* and *Roseiflexus* spp. (Nübel *et al.*, 2002). Based on growth in culture (Madigan *et al.*, 1974; Pierson and Castenholz, 1974b) and in situ experiments showing light stimulated uptake of radiolabelled organic substrates (Sandbeck and Ward, 1981; Anderson *et al.*, 1987; Bateson and Ward, 1988), it was previously suggested that FAPs in these mats predominantly use photoheterotrophic metabolism to assimilate low-molecular weight organic compounds cross-fed from the cyanobacteria (Ward *et al.*, 1987). However, stable carbon isotope signatures in lipid biomarkers diagnostic of *Chloroflexus aurantiacus* and *Roseiflexus* spp. (van der Meer *et al.*, 2001, 2002) were found to be isotopically heavier than those typically observed for cyanobacteria (van der Meer *et al.*, 2000, 2003). This was surprising for a situation involving cross-feeding of metabolites between organisms, in which case similar isotopic signatures would be expected in cell components of both organisms. The heavier isotopic signature of the biomarkers of FAPs in the mat was taken as possible evidence for autotrophic metabolism by a mechanism similar to the autotrophic pathway in *C. aurantiacus* (van der Meer *et al.*, 2000, 2003).

Chloroflexus aurantiacus strain OK-70-fl has been grown photoautotrophically in culture (Madigan and Brock, 1977; Sirevåg and Castenholz, 1979), under which conditions it fixes bicarbonate via the proposed 3-hydroxypropionate (3-OHP) pathway, as outlined in Figure 2.1 (Strauss and Fuchs, 1993; Alber and Fuchs, 2002; Herter *et al.*, 2002a; Hügler *et al.*, 2002; Friedmann *et al.*, 2006b,a).

The 3-OHP pathway discriminates less against heavier isotopes of carbon (incorporated as bicarbonate) than does the Calvin cycle. This leads to the synthesis

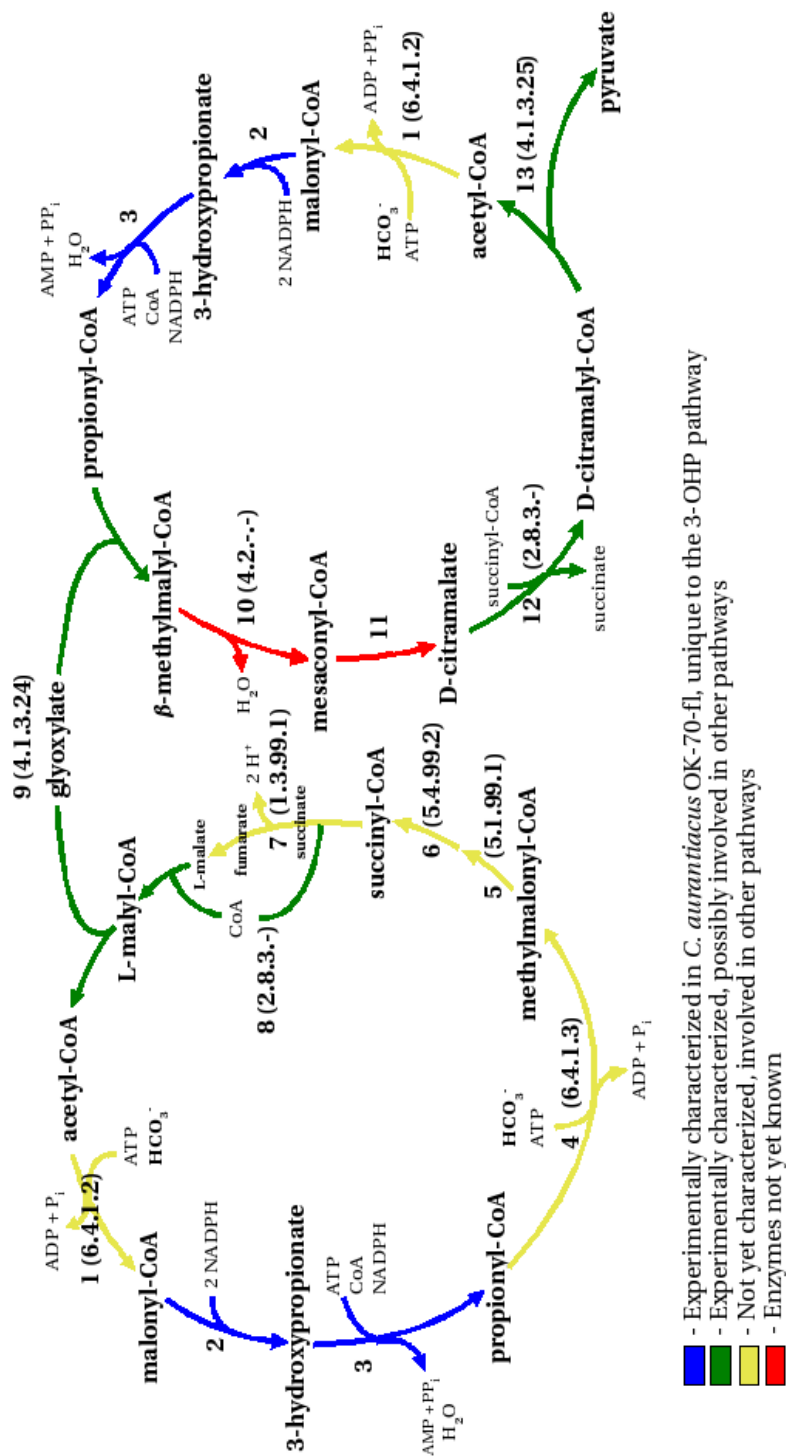


Figure 2.1: The 3-Hydroxypropionate Pathway as Proposed for *Chloroflexus aurantiacus*. Enzymatic steps are coloured in reference to the level of their characterization, and known enzyme classification (E.C.) numbers are indicated. Enzymes: 1, acetyl-CoA carboxylase; 2, malonyl-CoA reductase; 3, propionyl-CoA synthase; 4, propionyl-CoA carboxylase; 5, methylmalonyl-CoA epimerase; 6, methylmalonyl-CoA mutase; 7, succinate dehydrogenase and fumarate hydratase; 8, succinyl-CoA : L-malate-CoA transferase; 9, L-malyl-CoA/β-methylmalyl-CoA lyase; 10, proposed β-methylmalyl-CoA dehydratase; 11, postulated mesaconyl-CoA-transforming enzymes; 12, succinyl-CoA : D-citramalate CoA transferase; 13, D-citramalyl-CoA lyase (adapted from Friedmann *et al.* (2006b,a)).

of organic compounds that are relatively enriched in ^{13}C ($\Delta\delta^{13}\text{C} \sim 14\text{‰}$) compared with those produced by the Calvin cycle ($\Delta\delta^{13}\text{C} \sim 20$ to 25‰) (Holo and Sirevåg, 1986; Madigan *et al.*, 1989; van der Meer *et al.*, 2001). The heavy isotopic signatures of the lipid biomarkers of FAPs in these mats suggested that autotrophy by FAPs using the 3-OHP pathway may be an important mechanism for the input of isotopically heavy carbon in these communities. Incorporation of $^{13}\text{CO}_2$ into FAP lipid biomarkers, and stimulation of this activity by H_2 and sulfide, also supported the possibility of anoxygenic photoautotrophy and suggested that these organisms may be using this metabolism during low-light periods (van der Meer *et al.*, 2005). The interpretation that FAPs are photoautotrophic in situ is complicated by the observations that (i) *Roseiflexus* spp. are more abundant than *Chloroflexus* spp. in these mats (Nübel *et al.*, 2002) and (ii) isolates of *Roseiflexus* spp. have not been shown to grow photoautotrophically (Hanada *et al.*, 2002; Madigan *et al.*, 2005). Additionally, other Chloroflexi have not been shown to be autotrophic in culture (e.g. the phototrophic *Chloroflexus aggregans* and the non-phototrophic *Herpetosiphon aurantiacus*; (Holt and Lewin, 1968; Hanada *et al.*, 1995). Some phototrophic Chloroflexi use other carbon fixation pathways, such as *Oscillochloris trichoides*, which uses the reductive pentose phosphate pathway for autotrophy (Ivanovsky *et al.*, 1999; Berg *et al.*, 2005) and *Chlorothrix halophila*, in which activities that distinguish the 3-OHP pathway could not be demonstrated (Klappenbach and Pierson, 2004). Forthcoming genomic data indicate the presence of ribulose 1,5-bisphosphate carboxylase/oxygenase and phosphoribulokinase in *Chlorothrix halophila*, suggesting this organism also uses the Calvin cycle for autotrophy (D. Bryant, unpublished). Several Chloroflexi genomes have recently been sequenced as part of a Joint Genome Institute/Department of Energy project to survey the properties of FAPs. The draft genomes of three FAPs, *C. aggregans*, *Roseiflexus* sp. RS-1, *Roseiflexus castenholzii*, and one non-

Table 2.1: Isolate Organisms Investigated in this Study.

Organism	Isolation source	Reference
<i>Chloroflexus aurantiacus</i> J-10-fl	Sokokura, Hakone area, Japan	Pierson and Castenholz (1974a)
<i>Chloroflexus aggregans</i> MD-66	Okukinu Meotobuchi hot spring, Tochigi Pct, Japan	Hanada <i>et al.</i> (1995)
<i>Roseiflexus</i> sp. RS-1	Octopus Spring, WY, USA	Madigan <i>et al.</i> (2005)
<i>Herpetosiphon aurantiacus</i> DSM 785	Birch Lake, MN, USA	Holt and Lewin (1968)

photosynthetic Chloroflexi isolate, *H. aurantiacus* (Table 2.1), were compared with the existing genome sequence of *C. aurantiacus* J-10-fl to determine whether these organisms have homologues of genes shown to be involved in 3-OHP autotrophy in *C. aurantiacus* (Alber and Fuchs, 2002; Herter *et al.*, 2002a; Hügler *et al.*, 2002; Friedmann *et al.*, 2006b,a).

Putative homologues were then used to screen a metagenomic sequence database for Octopus and Mushroom Springs (obtained as part of an NSF Frontiers in Integrative Biological Research project; <http://landresources.montana.edu/FIBR/>; <http://www.tigr.org/tdb/ENVMGX/YNPHS/index.html>; Bhaya *et al.* 2007) to determine the in situ genetic capacity for the 3-OHP pathway. Once identified, metagenomic homologues of these genes were compared with the sequences of the cultured isolates, particularly *Roseiflexus* sp. strain RS-1, which is a genetically relevant isolate compared with Octopus Spring *Roseiflexus* populations (Madigan *et al.*, 2005).

Results and Discussion

Genome Annotation Evidence of 3-OHP Pathway in FAP Isolates

Figure 2.1 shows the bicyclic reactions that have been postulated to comprise the 3-OHP pathway for CO₂ fixation in *C. aurantiacus* and indicates the level to which the steps in the pathway have been experimentally characterized. Homologues of all these genes were found in the genomes of the three phototrophs we examined but not

in the genome of *H. aurantiacus*. This inference is based on amino acid identities and similarities derived from BLASTP analyses (Table 2.2) and from matching to profile hidden Markov models (HMMs) in the PFAM and TIGRFAM databases.

Steps 1 and 4 (acyl-CoA carboxylases). Evidence for genes encoding the acyl carboxylase enzymes proposed for steps 1 (acetyl-CoA carboxylase) and 4 (propionyl-CoA carboxylase) is presented together and we refer to homologues of these genes as acetyl-CoA/propionyl-CoA carboxylases, because the substrate specificity for these enzymes is unknown. All five analysed genomes contained open reading frames (ORFs) that correspond to the functional domains of bacterial acetyl-CoA carboxylases (for a review, see Cronan and Waldrop 2002), which is not surprising given that these genes are also involved in fatty acid metabolism (Strauss and Fuchs, 1993) and are not diagnostic of the 3-OHP pathway. The functional domains of acetyl-CoA carboxylases include the biotin carboxylase (BC) subunit AccC, the biotin carboxyl carrier protein (BCCP) subunit AccB, and the α and β subunits of the carboxyltransferase components (CT $_{\alpha}$ and CT $_{\beta}$) AccA and AccD, respectively (Li and Cronan, 1992b,a; Best and Knauf, 1993; Marini *et al.*, 1995; Kimura *et al.*, 2000; Kiatpapan *et al.*, 2001). Additional evidence for the putative *accC* ORFs includes a conserved N-terminal sequence A₈NRGEIA₁₄ and a glycine-rich region with the sequence GGGG(K/R)G, consistent with other BC subunits of acyl-CoA carboxylases (Chuakrut *et al.*, 2003). Open reading frames annotated as *accB* shared the biotin binding site motif EAMKM, and the lysine residues predicted to be biotin binding sites have glycine and proline residues flanking them as seen in other BCCP sequences (Samols *et al.*, 1988; Chuakrut *et al.*, 2003). *Roseiflexus* sp. RS-1 has two copies of the *accA* and *accD* as determined by HMMs, and they are 73 and 62% identical to each other at the amino acid level respectively. The *accA* and *accD* most closely related to the sequences in *C. aurantiacus* are reported in Table 2.2. The colocalization of

Table 2.2: Percent Amino Acid Identity and Similarity of ORFs Coding for Experimentally Characterized (bold) and Uncharacterized Enzymes of the 3-OHP pathway in *C. aurantiacus* to Orthologs of Chloroflexi Isolate Genomes.

<i>C. aurantiacus</i> gene	Gene name	Step in pathway	% amino acid identity/similarity			
			<i>C. aggregans</i>	<i>Roseiflexus</i> sp. RS-1	<i>R. castenholzii</i>	<i>H. aurantiacus</i>
Acetyl/propionyl-CoA carboxylase, carboxyltransferase alpha subunit	<i>accA</i>	1/4	92/97	68/80	69/83	58/76
Acetyl/propionyl-CoA carboxylase, carboxyltransferase beta subunit	<i>accD</i>	1/4	90/96	67/81	68/82	62/77
Acetyl/propionyl-CoA carboxylase, biotin carboxyl carrier protein subunit	<i>accB</i>	1/4	88/93	60/70	55/69	54/69
Acetyl/propionyl-CoA carboxylase, biotin carboxylase subunit	<i>accC</i>	1/4	94/97	75/88	74/88	70/83
Acetyl/propionyl-CoA carboxylase, carboxyltransferase subunit	CT3	1/4	98/99	86/94	85/92	82/89
Malonyl-CoA reductase	-	2	88/93	58/70	58/70	ND
Propionyl-CoA synthase	-	3	90/95	71/82	71/81	ND
Methylmalonyl-CoA epimerase	-	5	93/98	65/81	65/80	55/71
Methylmalonyl-CoA mutase, C-terminus	-	6	96/97	84/94	82/91	85/90
Methylmalonyl-CoA mutase, N-terminus	-	6	96/98	91/96	91/95	ND
Methylmalonyl-CoA mutase, N-terminus	-	6	79/88	65/78	67/79	ND
Methylmalonyl-CoA mutase, N-terminus	-	6	95/98	78/85	77/84	72/84
Succinate dehydrogenase/fumarate reductase, b-cytochrome subunit	-	7	93/98	47/61	46/61	48/64
Succinate dehydrogenase/fumarate reductase, FeS subunit	-	7	95/97	70/81	70/81	70/84
Succinate dehydrogenase/fumarate reductase FeS subunit	-	7	97/99	81/88	81/88	76/86
Fumarate hydratase	-	7	96/97	86/91	85/91	74/84
Succinyl-CoA:L-malyl-CoA transferase	<i>smtA</i>	8	94/97	91/96	91/96	ND
Succinyl-CoA:L-malyl-CoA transferase	<i>smtB</i>	8	95/97	88/95	88/96	ND
L-malyl-CoA/β-methylmalyl-CoA lyase	<i>mclA</i>	9	95/98	94/98	93/97	32/50
Succinyl-CoA:D-citramalate CoA transferase	<i>sct</i>	12	93/98	78/88	94/96	ND
D-citramalyl-CoA lyase	<i>ccl</i>	13	82/89	65/79	83/90	ND

Enzymes catalysing steps 1, 4, 6, 7 and 8 are putatively encoded by multiple genes, and acetyl-CoA/propionyl-CoA carboxylase subunits have multiple paralogs. ND, not detected.

an *accC* gene downstream of *accDA* in the *Roseiflexus* sp. RS-1 and *R. castenholzii* genomes provided evidence that these particular genes are likely to be subunits of the same carboxylase. Additionally, these genes are adjacent to genes whose products are predicted to encode enzymes that catalyse steps 2 and 3, suggesting that this carboxylase is involved the 3-OHP pathway (Figure 2.2).

Bacterial propionyl-CoA carboxylases and bifunctional acyl-CoA carboxylases have been characterized in the actinomycetes (Hunaiti and Kolattukudy, 1982; Rodríguez and Gramajo, 1999; Rodríguez *et al.*, 2001; Diacovich *et al.*, 2002, 2004; Gago *et al.*, 2006; Lin *et al.*, 2006; Daniel *et al.*, 2007), and they contain a minimum of two different subunits: the BC and BCCP domains are encoded within the α -subunit (AccA/PccA), and the CT domain lies within the β -subunit (AccB/PccB). Bifunctional acyl-CoA carboxylases have also been described in a proposed alternative 3-OHP pathway in the archaeal family *Sulfolobaceae* in the Crenarchaeota (Menendez *et al.*, 1999; Chuakrut *et al.*, 2003; Hügler *et al.*, 2003b; Alber *et al.*, 2006; Hallam *et al.*, 2006). Diacovich and colleagues (2004) used the crystal structure of the *Streptomyces coelicolor* PccB and site-directed mutagenesis to determine which residues impart substrate specificity for acetyl-CoA and propionyl-CoA. Their findings suggest that bulky hydrophobic residues at position 422 of PccB in *S. coelicolor* (position 473 in Figure 2.3A) allow for both acetyl and propionyl-CoA to enter the binding pocket of the active site, whereas an aspartate residue at this position has less affinity for acetyl-CoA. This insight was coupled with a phylogenetic analysis of all FAP ORFs that are predicted to encode a carboxyltransferase domain (Figure 2.3) to predict carboxyltransferase substrate specificity. From these data, FAP carboxyltransferase ORFs labelled CT3 are predicted to have higher substrate affinity for propionyl-CoA based on the aspartate residue at position 473 (Figure 2.3A) and the fact that these

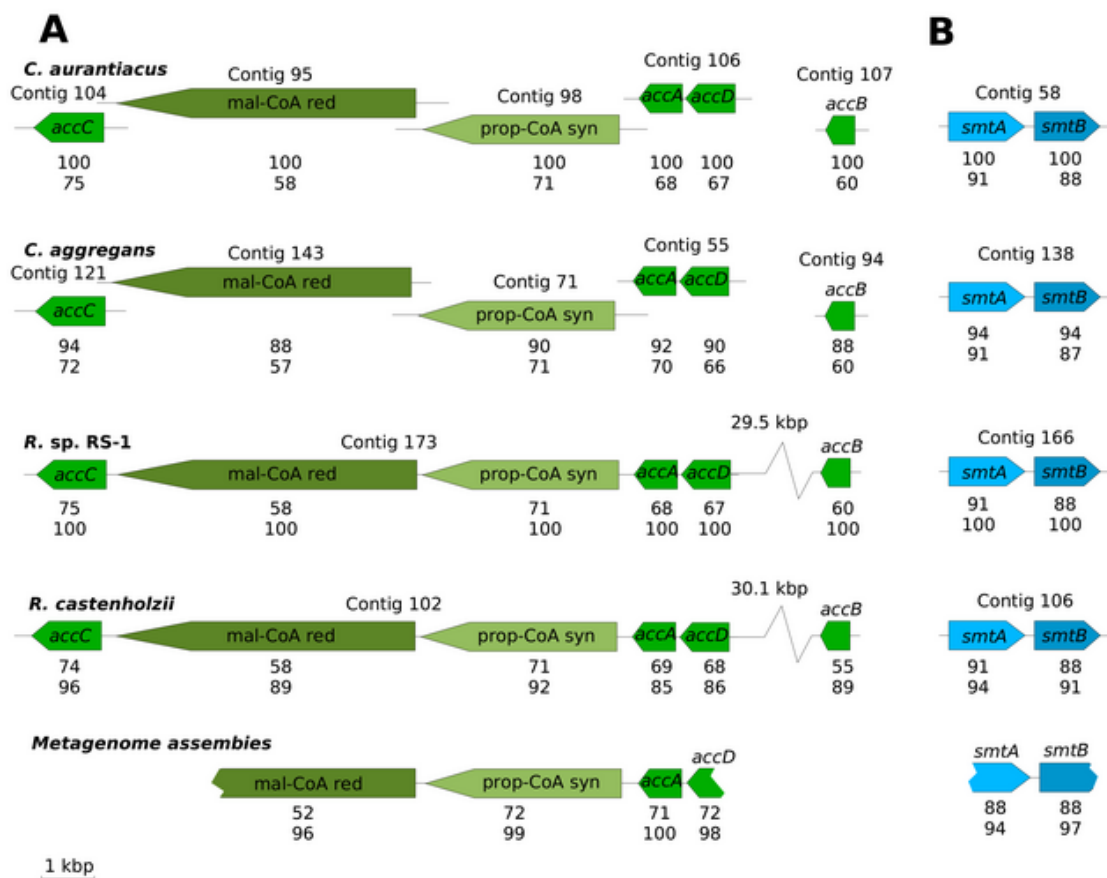


Figure 2.2: Locations of Genes on Isolate Genome and Metagenome Contigs. A. Acetyl/propionyl-CoA carboxylase (*acc*), malonyl-CoA reductase (*mal-CoA red*), and propionyl-CoA synthase (*prop-CoA syn*). B. Succinyl-CoA : L-malate CoA transferase (*smtAB*). Zigzag cut-offs represent the ends of fragments of the gene included in the contig. Amino acid identities to *C. aurantiacus* (top) and *Roseiflexus* sp. RS-1 (bottom) are indicated under each gene.

sequences cluster with known propionyl-CoA specific carboxyltransferases (Figure 2.3B). Filamentous anoxygenic phototroph sequences CT2 and CT4 are predicted to be involved in both acetyl-CoA and propionyl-CoA carboxylase activity, as evidenced by the hydrophobic residue at position 473 and their clustering with bifunctional acyl-CoA carboxylases (Figure 2.3B). Open reading frames labelled CT1 are phylo-

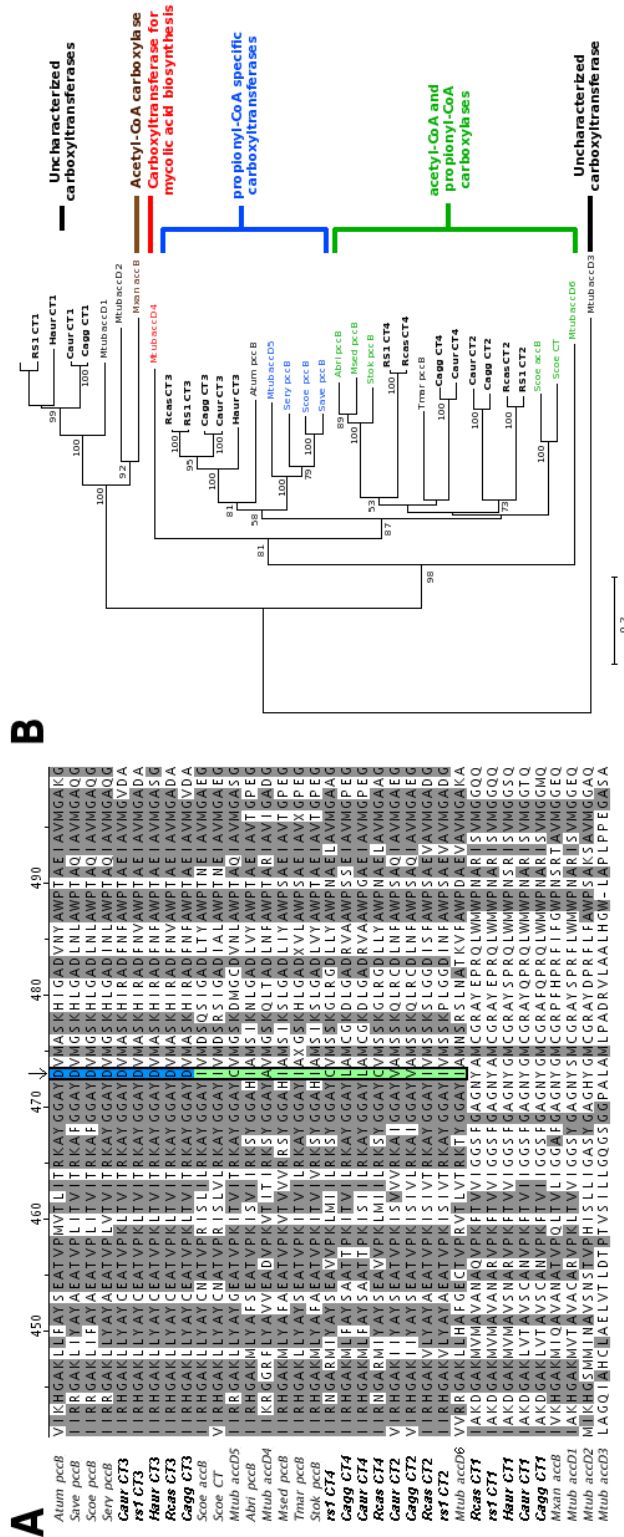


Figure 2.3: A. Partial alignment of Chloroflexi (bold) and experimentally characterized prokaryotic carboxyltransferases. The marked residue at position 473 imparts substrate specificity in *Streptomyces coelicolor*. Shaded green residues indicate $\geq 50\%$ amino acid consensus. Blue aspartate residues show predicted preferential specificity for propionyl-CoA, while green hydrophobic residues indicate predicted specificity for both acetyl-CoA and propionyl-CoA. B. Phylogenetic analysis of prokaryotic carboxyltransferases. This unrooted neighbour-joining tree shows bootstrap values over 50% (out of 1000 replicates). Horizontal branch lengths are proportional to inferred evolutionary distances, with the scale bar indicating the number of substitutions per site. Names in bold refer to Chloroflexi sequences, while coloured names indicate proteins that have been experimentally characterized with respect to substrate specificity for acetyl-CoA or propionyl-CoA. Organisms include the following: Abri, *Acidithiobacillus brierleyi*; Atum, *Agrobacterium tumefaciens*; Cagg, *C. aggregans*; Caur, *C. aurantiacus*; Haur, *H. aurantiacus*; Msed, *Metallosphaera sedula*; Mtu, *Mycobacterium tuberculosis*; Mxan, *Mycrococcus xanthus*; Rcas, *R. castenholzii*; rsl, *Roseiflexus* sp. RS-1; Sery, *Saccharopolyspora erythraea*; Save, *Streptomyces avermitilis*; Scoe, *S. coelicolor*; Stok, *Sulfolobus tokodaii*; Tmar, *Thermus maritimus*.

genetically distant from experimentally characterized acyl-CoA carboxylases, and the function of these carboxyltransferases remains unexplored. It should be noted that each analysed genome has multiple copies of putative BC, BCCP and CT subunits as determined from HMMs, and these subunits could combine to form isoenzymes with varying substrate specificities.

Step 2 (malonyl-CoA reductase). Malonyl-CoA reductase catalyses the NADPH-dependent two-step reduction of malonyl-CoA to 3-hydroxypropionate via a malonate semialdehyde intermediate (Hügler *et al.*, 2002). Open reading frames identified as homologues to this gene had statistically significant hits to PFAM models indicating domains conserved in NAD-dependent epimerases and short-chain aldehyde/alcohol dehydrogenases, consistent with earlier investigations of the function of this enzyme in *C. aurantiacus* OK-70-fl (Hügler *et al.*, 2002).

Step 3 (propionyl-CoA synthase). The trifunctional enzyme propionyl-CoA synthase activates 3-hydroxypropionate to 3-hydroxypropionyl-CoA, which is then converted to acrylyl-CoA and reduced to propionyl-CoA (Alber and Fuchs, 2002). According to profile HMMs from the PFAM database, this enzyme shares the conserved domain structure of other enoyl-CoA hydratases and includes an AMP binding site, which is consistent with the findings of Alber and Fuchs (2002). Additionally, a NAD(P)H binding motif of (GXGX₂AX₃A) was found in the sequences from all four phototrophic genomes, with *C. aggregans* sequence exhibiting two such motifs. *Herpetosiphon aurantiacus* does not contain any ORFs that have sequence similarity as statistically significant as our expectation value cut-off to malonyl-CoA reductase or propionyl-CoA synthase (Table 2.2).

Steps 5 to 7 (methylmalonyl-CoA epimerase, methylmalonyl-CoA mutase, succinate dehydrogenase and fumarate hydratase). The enzymes in the pathway that convert methylmalonyl-CoA to succinyl-CoA (steps 5 and 6) are also used to oxidize

fatty acid chains with an odd number of carbons, while those catalysing the conversion of succinyl-CoA to L-malate (step 7) are also components of the TCA cycle. Evidence of their putative function comes in the form of highly specific (equivalog-level) profile HMMs from the TIGRFAM database. Homologues to genes encoding enzymes catalysing these three enzymatic steps were also found in *H. aurantiacus*.

Step 8 (succinyl-CoA : L-malate-CoA transferase). Two subunits make up the enzyme succinyl-CoA : L-malateCoA transferase (SmtA and SmtB), which is a Type III CoA transferase (Friedmann *et al.*, 2006a). This family level function was predicted by a PFAM HMM for both SmtA and SmtB in each Chloroflexi genome except that of *H. aurantiacus*.

Step 9 (L-malyl-CoA/ β -methylmalyl-CoA lyase). The proposed 3-OHP pathway is bicyclic in that the glyoxylate produced in the first cycle acts as the intermediate that is used in a second cycle to produce pyruvate. L-malyl-CoA/ β -methylmalyl-CoA lyase has been demonstrated to have the dual function of cleaving L-malyl-CoA to acetyl-CoA and glyoxylate (thus completing the first cycle), and then condensing glyoxylate with propionyl-CoA to produce β -methylmalyl-CoA (which begins the second cycle) (Herter *et al.*, 2002a). Sequences showing similarity to the L-malyl-CoA/ β -methylmalyl-CoA lyase were also predicted to have aldolase/citrate lyase activity, consistent with the results of Herter and colleagues (2002). A homologue of L-malyl-CoA/ β -methylmalyl-CoA lyase (step 9) in *C. aurantiacus* was also found in *H. aurantiacus*. However, these predicted proteins were not very similar in sequence, and therefore these gene products may not share the same function (Table 2.2).

Steps 10 and 11. The enzymes proposed to convert β -methylmalyl-CoA via the intermediate mesaconyl-CoA to D-citramalate (steps 10 and 11) have not yet been identified and characterized.

Steps 12 and 13 (succinyl-CoA/D-citramalate CoA transferase and D-citramalyl-CoA lyase). A second Type III CoA transferase has been shown to catalyse the reaction in step 12 in which D-citramalate is converted to D-citramalyl-CoA (Friedmann *et al.*, 2006b). Homologues to this sequence in *C. aurantiacus* are also predicted to have Type III CoA transferase activity as determined from HMMs. A D-citramalyl-CoA lyase gene is adjacent to this gene in *C. aurantiacus*, and its function in catalysing step 13 in the pathway was predicted and confirmed by Friedmann and colleagues (2006b). The PFAM model for this sequence is not sufficiently specific to identify the CoA lyase function in the genomes of the four phototrophs.

Similarity in Genes and Gene Order in *Chloroflexus* and *Roseiflexus*

All of the genes encoding enzymes of the 3-OHP pathway in *C. aurantiacus* have greater amino acid identities and similarities to their homologues in *C. aggregans* than to the homologues in the two *Roseiflexus* spp. (Table 2.2). This is consistent with the greater phylogenetic relatedness between the two *Chloroflexus* species than between *Chloroflexus* and *Roseiflexus* species. The 16S rRNA sequences of *C. aurantiacus* is 92% identical to that of *C. aggregans*, but it is only 83% identical to those of both *Roseiflexus* sp. RS-1 and *R. castenholzii*. The order and direction of ORFs predicted to encode enzymes of the 3-OHP pathway provided additional evidence that these genes are used in this pathway. For instance, Figure 2.2 shows contigs in the draft *Roseiflexus* spp. genomes in which ORFs encoding subunits of an acetyl-CoA/ propionyl-CoA carboxylase (step 1 or 4) are adjacent to those encoding the unique 3-OHP pathway genes malonyl-CoA reductase (step 2) and propionyl-CoA synthase (step 3). In *Chloroflexus* spp., only *accA* and *accD* are adjacent on the same contig, while the *accC*, malonyl-CoA reductase and propionyl-CoA synthase

genes are each found on different contigs and are surrounded by neighbouring genes that do not encode enzymes in the pathway. The observed synteny between isolates of each species, but the absence of synteny between isolates of the two different genera, is also consistent with the greater phylogenetic distance separating *Roseiflexus* strains and *Chloroflexus* strains.

Absence of Alternative Autotrophic Pathways

To determine whether these organisms have the potential to use carbon fixation pathways other than the 3-OHP pathway, TBLASTN was used to query the genome sequences for evidence of other carboxylase genes. As defined by the criteria described in the Experimental procedures, none of the five genomes that were analysed in this study appear to contain homologues to (i) ribulose-1,5-bisphosphate carboxylase/oxygenase (Calvin-Benson-Bassham cycle) from *O. trichoides* (GenBank Accession AAZ52657), (ii) carbon monoxide dehydrogenase (Accession P31896) and acetyl-CoA synthase (Accession P27988) from experimentally characterized protein sequences (both in the Wood-Ljungdahl or reductive acetyl-CoA pathway) or (iii) ATP-dependent citrate lyase (Accessions AAM72322 and AAM72321), and 2-oxoglutarate : ferredoxin oxidoreductase from *Chlorobium tepidum* (Accessions AAM71411 and AAM71410) (both in the reductive tricarboxylic acid cycle) (data not shown). Despite the lack of evidence of other autotrophic pathways in these genomes, all *Chloroflexus* and *Roseiflexus* genomes contain an ORF that is homologous to pyruvate : flavodoxin/ferredoxin oxidoreductase, an enzyme that can be used to either decarboxylate pyruvate, or synthesize pyruvate by carboxylation of acetyl-CoA in an anapleurotic pathway (Raymond, 2005). The latter reaction was proposed to operate in an autotrophic reductive cycle of dicarboxylic acids in *C. aurantiacus* strain B-3 (Ugolkova and Ivanovsky, 2000). These genomic comparisons have allowed us

to identify the potential of three phototrophic Chloroflexi to perform the 3-OHP pathway for autotrophy despite the present limitation of not being able to grow these isolates autotrophically in culture. The inability to grow these strains autotrophically could result from the failure to identify a suitable electron donor for autotrophic growth, or it could reflect the possibility that these Chloroflexi strains can only grow mixotrophically by oxidizing organic compounds while at the same time fixing some CO₂ via the 3-OHP pathway. An analagous strategy is used by the aerobic anoxygenic phototroph *Roseobacter denitrificans*, which lacks a definitive autotrophic pathway, yet still demonstrates light-stimulated uptake of CO₂ (Swingley *et al.*, 2007). *Roseiflexus castenholzii* was tested for photoautotrophic growth using Na₂S₂O₃ and Na₂S as electron donors (Hanada *et al.*, 2002) but it is possible that *Roseiflexus* spp. are capable of using H₂ as an electron donor as evidenced by a putative membrane bound Group 1 [Ni-Fe] uptake hydrogenase enzyme in both *R. castenholzii* and *Roseiflexus* sp. RS-1 genomes (Accession numbers: *R. castenholzii*, ZP_01531052 and ZP_01531053; *Roseiflexus* sp. RS-1, ZP_01357085 and ZP_01357084) (Vignais *et al.*, 2001). Other than the above-mentioned difference in gene organization, this study revealed the potential for a 3-OHP pathway in the three FAPs studied, and this pathway is similar to the proposed pathway in *C. aurantiacus*. A similar comparative approach of genomic and metagenomic data has been applied to find evidence of the 3-OHP pathway in organisms of the Crenarchaeota (Hallam *et al.*, 2006), which use an alternative malonyl-CoA reductase enzyme (Alber *et al.*, 2006) and use a modified 3-OHP pathway (Hügler *et al.*, 2003a).

Environmental Genomic Analysis

The 3-OHP pathway homologues identified in both *C. aurantiacus* and *Roseiflexus* sp. RS-1 genomes showed high amino acid sequence identity to the translations of

environmental DNA sequences obtained by shotgun cloning and clone-end sequencing from Octopus and Mushroom Spring mat samples collected from sites with average temperatures of 60 °C and 65 °C (Figure 2.4). It is clear that homologues of genes involved in the 3-OHP pathway from both *Chloroflexus* and *Roseiflexus* spp. are present in the mat. For each gene, multiple homologous reads with different sequences were observed. Reads encoding homologues more closely related to *Roseiflexus* sp. RS-1 (126 reads $\geq$ 90% amino acid identity to *Roseiflexus* sp. RS-1) outnumber those to *C. aurantiacus* (61 reads $\geq$ 90% amino acid identity for *C. aurantiacus*). Reads encoding homologues that are more closely related to *C. aurantiacus* genes are more abundant in the high temperature (65 °C) clone libraries, consistent with previous data showing greater relative abundance of *Roseiflexus* spp. at 60 °C, and a greater abundance of *Chloroflexus* spp. at the higher temperature (Nübel *et al.*, 2002). The lower sequence identity of metagenomic homologues (Figure 2.4) to *C. aurantiacus* strain J-10-fl protein sequences may be due to the phylogenetic distance separating this Japanese isolate and populations inhabiting Yellowstone hot springs (Nübel *et al.*, 2002). Metagenome read sequences that are less than 80% identical to either isolate are too phylogenetically distant to infer their function. The colocalization of 3-OHP pathway genes on a contig assembled from the metagenome provided additional evidence of autotrophic capability in uncultured *Roseiflexus* spp. (Figure 2.2). This contig contains four 3-OHP pathway genes, including those encoding two acyl-CoA carboxylase subunits and the diagnostic enzymes malonyl-CoA reductase and propionyl-CoA synthase, and these are arranged in the same order as found in *Roseiflexus* isolate genomes. A BLASTX comparison of translated metagenomic sequences to homologous amino acid sequences in the isolate genomes indicate that the genes on this contig are more closely related to genes of *Roseiflexus* sp. RS-1 than to genes of *C. aurantiacus* (Figure 2.2). The *smtAB* (step 8) homologues (Friedmann

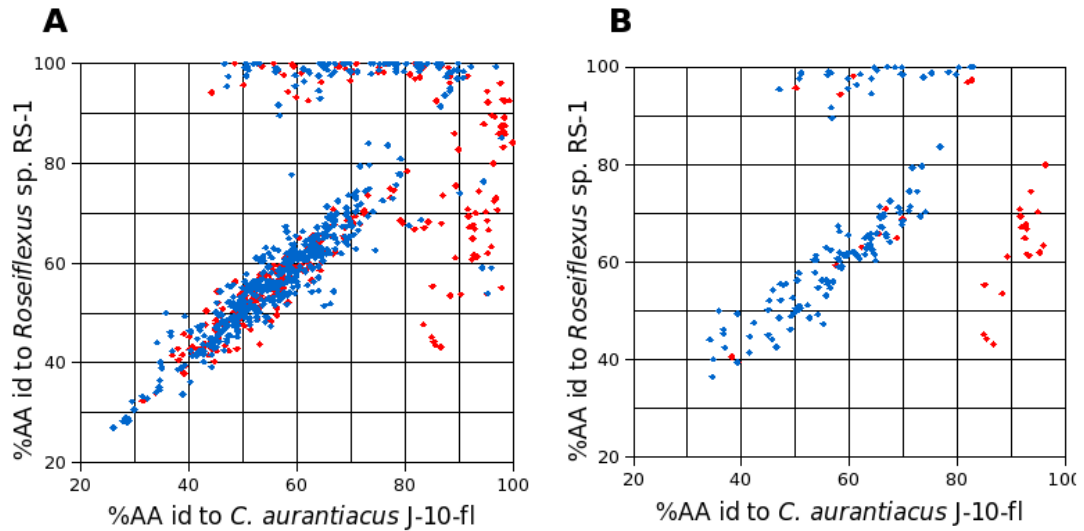


Figure 2.4: Per Cent Amino Acid Identity of Metagenome Sequences Encoding 3-OHP Pathway Genes to Homologues in the *C. aurantiacus* and *Roseiflexus* sp. RS-1 Genomes. Blue and red symbols indicate metagenome reads from low temperature (average 60°C), and high temperature (65°C) sites respectively. A. All homologues putatively involved in the 3-OHP pathway in *C. aurantiacus* and *Roseiflexus* sp. RS-1 (811 reads in a reciprocal TBLASTN/BLASTX search). B. The subset of 172 reads hitting malonyl-CoA reductase or propionyl-CoA synthase, which catalyse the two unique steps of the pathway.

et al., 2006a) were similarly found to be adjacent on a 1.8-kb contig assembled from the metagenome and showed 94% and 88% amino acid identity to *smtA* homologues and 97% and 88% amino acid identity to the *smtB* homologues of *Roseiflexus* sp. RS-1 and *C. aurantiacus*, respectively (Figure 2.2).

Conclusions

The results reported here support the hypothesis that the dominant *Roseiflexus* populations in the microbial mats of alkaline siliceous hot springs have the capacity to fix inorganic carbon via the 3-OHP pathway. These results provide a basis for

inferences made previously from other evidence that autotrophy via this pathway is one mechanism that can lead to heavier ^{13}C signatures in *Roseiflexus* spp. biomarkers compared with *Synechococcus* spp. biomarkers. The crucial next steps will be to verify these in silico predictions by demonstrating the expression of genes encoding 3-OHP pathway enzymes in the mat. We will use the gene sequences we have reported here to study the contributions of *Roseiflexus* and *Chloroflexus* populations to the overall inorganic carbon fixation in these mats over a diel time-course. These studies will use enzyme activity measurements combined with quantitative reverse transcription polymerase chain reaction analysis of mRNA transcripts, which is an approach that has been successfully used to measure in situ expression of *Synechococcus* genes in these same mats (Steunou *et al.*, 2006).

Experimental Procedures

Metagenome Library Construction and Assembly

Core samples (0.5 cm²) of variable depth were taken in the afternoons of 2 October 2003 and 5 November 2004 from 60 °C and 65 °C cyanobacterial mats in Octopus and Mushroom Springs in Yellowstone National Park, Wyoming, USA. These were sectioned in the field into ~1 mm thick depth intervals using a razor blade and quickly frozen on liquid nitrogen. Two lysis protocols were used: (i) a mechanical bead-beating lysis and (ii) an enzymatic lysis using lysozyme and proteinase K (details can be found in Appendix A). The mechanical bead-beating procedure was insufficient to lyse the cells completely, resulting in an over-representation of DNA from organisms that are easily lysed. Thus, clone libraries constructed from mechanically lysed cells are assumed to be atypical of the mat environment in terms of the relative abundance of organisms sampled. Two sets of metagenomic clone libraries were constructed from

the DNA extracted from the enzymatically and mechanically lysed cells. The first set was from the top 1 mm of the cores sampled, resulting in the $\sim 167\,500$ kb of DNA sequence reported in D. Bhaya and colleagues (2007). The second data set came from deeper layers of Mushroom Spring cores, and added an additional 18 700 kb of sequence. Appendix A details the metagenomic library sources and layers analysed in this study. The Celera assembler was used on the subset of metagenome reads from Octopus Spring resulting in 5757 contigs with an average size of 2.4 kb. These sequences are available on a website: <http://www.tigr.org/tdb/ENVMGX/YNPHS/index.html>

BLAST Comparisons

The isolate genome sequences were screened for homologues to genes involved in experimentally characterized steps of the 3-OHP pathway in *C. aurantiacus* OK-70-fl (GenBank Accession numbers AAS20429, AAL47820, ABF14399 and ABF14400) using TBLASTN <http://blast.wustl.edu> against the genome contigs. Open reading frames exhibiting alignments at least 100 amino acids long and having expectation values more significant than 1×10^{-15} were then reciprocally used in a BLASTP search against the NCBI nr database. The same method was used to query the genomes for the presence of carboxylases involved in alternative autotrophic pathways. For steps in the 3-OHP pathway that have not yet been characterized, ORFs were selected for putative gene products that corresponding to predicted functions in the pathway via profile hidden Markov models (see below). All identified ORFs were queried against the *C. aurantiacus* J-10-fl genome in a reciprocal BLASTP analysis with the parameters `hitdist = 40`, `wordmask = seg`, and `postsw` set to obtain the values listed in Table 2.2. Homologous sequences among metagenomic reads were found using protein sequences from the genomes as a query in a TBLASTN search against the nucleotide

metagenome database. Reads produced alignments of less than 100 amino acids long and with expectation values greater than 1×10^{-15} and reads that resulted from the biased mechanical lysis protocol were not analysed. A reciprocal BLASTX search of nucleotide metagenome reads against the translated genome peptide databases was used to verify that each read aligned to the original query sequence as the best scoring match.

Hidden Markov Model Analysis

Support for the annotations of the ORFs predicted to encode proteins of the 3-OHP pathway came from the program HMMER 2.3.2 (Eddy, 1998) (<http://selab.janelia.org/>), which determines statistically significant matches to profile hidden Markov models (HMMs). This program is used to screen ORFs for conserved domains, which provide functional evidence for gene annotation. Only models that scored above the trusted cut-offs in the curated TIGRFAM and PFAM profile HMM databases were used.

Phylogenetic Analysis

Experimentally characterized carboxyltransferase sequences were obtained from GenBank. An alignment was constructed using Clustalx and was manually edited in MEGA3.1 (Kumar *et al.*, 2004). The neighbour-joining tree was constructed using MEGA3.1 with 1000 bootstrap replicates.

Acknowledgements

This work was funded by the NASA Exobiology Program (NAG5-8824), the Montana State University Thermal Biology Institute (NASA NAG5-8807) and an NSF

Frontiers in Integrative Biological Research award (EF-0328698) to D.M.W. This work was also funded by NSF Grant MCB-0523100 to D.A.B. We thank J.F. Heidelberg (University of Southern California) for creating the metagenomic sequence assemblies analysed in this work, S. Hanada (National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan), M.T. Madigan (Southern Illinois University) and B.K. Pierson (University of Puget Sound) for providing *Chloroflexus aggregans*, *Chlorothrix halophila*, *Roseiflexus* sp. RS-1, and *R. castenholzii* cultures used to obtain genomes, M.M. Bateson (Montana State University) for providing DNA extracts, M.A. McClure (Montana State University) for assistance with the phylogenetic work, and J.W. Peters (Montana State University) for helpful suggestions. We thank R.E. Blankenship (Washington University), B.K. Pierson, and P. Richardson at the Department of Energy Joint Genome Institute (http://genome.jgi-psf.org/mic_home.html) for producing and giving us permission to use the genome sequence of *Chloroflexus aurantiacus* J-10-fl.

CHAPTER 3

COMMUNITY ECOLOGY OF HOT SPRING CYANOBACTERIAL MATS: PREDOMINANT POPULATIONS AND THEIR FUNCTIONAL POTENTIAL

Contribution of Authors and Co-Authors

Manuscript in Chapter 3

Author: Christian G. Klatt

Contributions: Designed the study, conducted the experiments, collected and analyzed output data and wrote the manuscript.

Co-author: Jason M. Wood

Contributions: Wrote computer programs for data processing, conducted the experiments and edited the manuscript.

Co-author: Douglas B. Rusch

Contributions: Processed and provided sequencing data, discussed the results and implications and edited the manuscript.

Co-author: Mary M. Bateson

Contributions: Assisted with experimental design and analyses, assisted in conducting field experiments, discussed the results and implications and edited the manuscript.

Co-author: Natsuko Hamamura

Contributions: Conducted the denaturing gradient gel electrophoresis experiment and edited the manuscript.

Co-author: John F. Heidelberg

Contributions: Obtained funding and edited the manuscript.

Co-author: Arthur R. Grossman

Contributions: Obtained funding and edited the manuscript.

Co-author: Devaki Bhaya

Contributions: Obtained funding and edited the manuscript.

Co-author: Frederick M. Cohan

Contributions: Obtained funding and edited the manuscript.

Co-author: Michael Kühl

Contributions: Obtained funding and edited the manuscript.

Co-author: Donald A. Bryant

Contributions: Obtained funding, discussed the results, and edited the manuscript at all stages.

Co-author: David M. Ward

Contributions: Obtained funding, assisted with experimental design, assisted in conducting field experiments, discussed the results and edited the manuscript at all stages.

Manuscript Information Page

Christian G. Klatt, Jason M. Wood, Douglas B. Rusch, Mary M. Bateson, Natsuko Hamamura, John F. Heidelberg, Arthur R. Grossman, Devaki Bhaya, Frederick M. Cohan, Michael Köhl, Donald A. Bryant, and David M. Ward.

Journal Name: The ISME Journal

Status of Manuscript:

---- Prepared for submission to a peer-reviewed journal

---- Officially submitted to a peer-reviewed journal

---- Accepted by a peer-reviewed journal

X Published in a peer-reviewed journal

Published by the International Society for Microbial Ecology in 2011, Issue 5 pages 1262-1278.

Abstract

Phototrophic microbial mat communities from 60 °C and 65 °C regions in the effluent channels of Mushroom and Octopus Springs (Yellowstone National Park, Wyoming USA) were investigated with shotgun metagenomic sequencing. Analyses of assembled metagenomic sequences resolved six dominant chlorophototrophic populations and permitted the discovery and characterization of undescribed but predominant community members and their physiological potential. Linkage of phylogenetic marker genes and functional genes revealed novel chlorophototrophic bacteria belonging to uncharacterized lineages within the order Chlorobiales and within the Kingdom Chloroflexi. The latter is the first chlorophototrophic member of Kingdom Chloroflexi that lies outside the monophyletic group of chlorophototrophs of the Order Chloroflexales. Direct comparison of unassembled metagenomic sequences to genomes of representative isolates revealed extensive genetic diversity, genomic rearrangements, and novel physiological potential in native populations compared to genomic references. *Synechococcus* spp. metagenomic sequences exhibited a high degree of synteny with the reference genomes of *Synechococcus* spp. strains A and B', but synteny declined with decreasing sequence relatedness to these references. There was evidence of horizontal gene transfer among native populations, but the frequency of these events was inversely proportional to phylogenetic relatedness.

Introduction

The cyanobacterial mats of alkaline siliceous hot springs in Yellowstone National Park (Supplementary Figures 1A and 1B; all Supplementary information, figures, and tables can be found in Appendix B) have been studied for several decades as models

for understanding the composition, structure and function of microbial communities (Brock, 1978; Ward *et al.*, 1987, 1992, 2002, 2012b). Simple and stable microbial communities containing dense populations of unicellular cyanobacteria (*Synechococcus* spp.) form in effluent channels of these springs between temperatures of 71-75°C (the upper temperature limit of the phototrophic mats) and ~50°C.

Analysis of 16S ribosomal RNA (rRNA) gene sequences demonstrated the poor relationship of initially cultivated isolates and predominant native populations. For instance, the predominant *Synechococcus* spp. of these mats (A/B lineage) had $\leq 92\%$ nucleotide identity at the 16S rRNA locus to the cultivated representatives available at that time (Ward *et al.*, 1990). Similarly, based on cultivation and pigment analyses (Bauld and Brock, 1973; Pierson and Castenholz, 1974a), it was once thought that *Chloroflexus* spp., which in culture use bacteriochlorophylls (BChl) *c* and *a* to support photoheterotrophy (Pierson and Castenholz, 1974a) or photoautotrophy (Holo and Sirevåg, 1986; Strauss and Fuchs, 1993), were the dominant anoxygenic phototrophic bacteria in these mats. However, 16S rRNA studies revealed the importance of *Roseiflexus* spp. (Nübel *et al.*, 2002), organisms which contain BChl *a* but lack BChl *c* and grow axenically as photoheterotrophs (Hanada *et al.*, 2002), although they possess genes encoding the enzymes of the 3-hydroxypropionate autotrophic pathway (Klatt *et al.*, 2007). In these cases the inference of chlorophototrophic physiologies (i.e., Chls are obligately required for phototrophy, in contrast to retinal-mediated proton translocation) could be made because oxygenic chlorophototrophs and anoxygenic chlorophototrophic Chloroflexales comprise monophyletic groups defined by 16S rRNA phylogeny. These predictions were confirmed with more recent cultivation and genomic analyses of *Synechococcus* spp. and *Roseiflexus* spp. isolates closely related to native mat populations (Allewalt *et al.*, 2006; Bhaya *et al.*, 2007; van der Meer *et al.*, 2010).

The inference of functional potential from 16S rRNA phylogeny is more problematic when sequences do not belong to groups that are monophyletic with respect to function. For instance, based on the observation that some 16S rRNA sequences retrieved from the mats fell just outside the monophyletic clade of known Chlorobiales, Ferris and Ward (1997) suggested the possible presence of bacteria closely related to green sulfur bacteria. Targeted analyses of photosynthetic reaction center genes provided evidence in support of this hypothesized functional group (Bryant *et al.*, 2007), but there was no way to associate the functional genes directly with the phylogenetic marker gene. Despite the successful retrieval of Chlorobiales from other thermal environments (Wahlund *et al.*, 1991; Madigan *et al.*, 2005), this organism has to date evaded cultivation. Interestingly, the search for photosynthetic reaction center genes in the mats led to the discovery of the first known chlorophototrophic member of Kingdom Acidobacteria, *Candidatus* Chloracidobacterium thermophilum (Bryant *et al.* 2007; usage of kingdom for major Domain sublineages sensu Ward *et al.* 2012a). The inference of potential chlorophototrophy was based on the discovery of a metagenomic clone containing an insert of mat DNA with both phylogenetic marker and functional genes. Because cultivated acidobacteria were not previously known to be phototrophic, inferences concerning the potential for phototrophy could not have been made before this discovery. Studies of an enrichment culture of *Ca. C. thermophilum* (Bryant *et al.*, 2007) and its genome (Bryant *et al.*, 2012) confirmed the inferences made from genetic data.

In this study, we used assembly of metagenomic sequences, combined with oligonucleotide frequency distributions and cluster analysis of scaffolds, to identify phylogenetically distinctive populations inhabiting Octopus Spring and Mushroom Spring mats. Oligonucleotide frequency patterns contain phylogenetic information (Pride *et al.*, 2003; Teeling *et al.*, 2004) and have been used as a tool to determine phylo-

genetic signatures in metagenomic data from microbial communities (Woyke *et al.*, 2006; Wilmes *et al.*, 2008; Dick *et al.*, 2009; Inskeep *et al.*, 2010). Annotation of open reading frames (ORFs) was used to identify phylogenetically and functionally informative genes in the scaffolds. We used the sequenced genomes of selected organisms, many of which have been cultivated from these or similar hot spring environments, and some of which are close relatives of predominant native populations in these mats, to recruit metagenomic sequences (Supplementary Table 1). This combined approach enabled us to (i) discover new major populations of uncultivated community members; (ii) explore differences in the functional potential of native populations as compared with closely related isolates; and (iii) observe differences in genomic content and synteny among closely related populations. This study also created a foundation for a companion study using metatranscriptomics to describe in situ gene expression in the chlorophototrophic taxa (Liu *et al.*, 2011b), the results of which strongly support our functional inferences and expand upon in situ gene expression studies of these mats (Steunou *et al.*, 2006, 2008; Jensen *et al.*, 2011).

Methods

Here we present the experimental approaches; Supplementary Information Section 3 contains the technical details of the methods used.

Collection, Preliminary Sequence Analysis, and Metagenomic Sequencing

Microbial mats were collected from Mushroom Spring (44.5386°N, 110.7979°W) on 2 October 2003 and from Octopus Spring (44.5340°N, 110.7978°W) on 5 November 2004 (Bryant *et al.*, 2007) at sites with average temperatures of $\sim 60^{\circ}\text{C}$ and

$\sim 65^{\circ}\text{C}$. *Synechococcus* spp. genotypes B' and A are the dominant cyanobacterial 16S rRNA sequences at these temperatures, respectively. Samples were collected and sectioned vertically into approximately 1 mm-thick layers, which were frozen then stored at -80°C until further analysis. After enzymatic lysis of cells in the top green layer, DNA was extracted and sequences were characterized by PCR amplification of cyanobacterial 16S rRNA genes and subsequent analysis with denaturing gradient gel electrophoresis to verify the presence of *Synechococcus* A and B'-like genotypes (Supplementary Figure 3). Extracted DNA was sheared into $\sim 1\text{-}3$ kbp and $\sim 10\text{-}12$ kbp fragments, which were used to prepare four metagenomic libraries which correspond to low and high temperature samples from Octopus Spring low or Mushroom Spring, respectively. End sequences of cloned inserts were produced by Sanger sequencing at the J. Craig Venter Institute (JCVI, Rockville, MD).

Metagenome Assembly and Annotation

Metagenome assembly and annotation. The metagenomic sequences were assembled into scaffolds using the Celera assembler (Miller *et al.*, 2008) with the error rate set to 8% for the purpose of assembling non-identical close relatives, and the `utgGenomeSize` set to 2 000 000. Phylogenetic and functional marker genes in assemblies were identified using the programs AMPHORA (Wu and Eisen, 2008), the JCVI annotation pipeline (Tanenbaum *et al.*, 2010), or BLAST Altschul *et al.* (1990) using known reference sequences as queries. All annotations are inferences based upon multiple lines of evidence produced using the tools listed above, but their functions are considered hypotheses for future biochemical characterization.

Clustering and Characterization of Assemblies

Oligonucleotide patterns were determined to obtain phylogenetic signals (Teeling *et al.*, 2004) by counting the frequencies of all possible tri-, tetra-, penta-, and hexa-nucleotide combinations for each scaffold 20 000 bp. Frequency counts were normalized by the length of the respective scaffold and subjected to k-means clustering (Kanungo *et al.*, 2002) with the a priori value of k equal to 8 (see Supplementary Information Section 3 for rationale). Scaffolds that clustered together with $\geq 90\%$ of 100 bootstrap trials were mapped using Cytoscape (Shannon *et al.*, 2003). Many scaffolds formed associations with core clusters at less stringent thresholds, but, except where noted, these were not included in the cluster analysis described here.

BLASTN Recruitment and Synteny with Reference Genomes

Metagenomic sequences were used as queries in a custom BLASTN search to a selected database of twenty genomes from organisms isolated from thermal springs, known to be functionally and/or phylogenetically related to indigenous mat populations and/or processes, or representative of phylogenetic groups not otherwise included (Supplementary Table 1). The percent nucleotide identity (% NT ID) of metagenomic sequences relative to the reference genome that recruited them was used to identify those that could be confidently associated with the reference organism, taking into account the % NT ID between the genomes of strains of named species and genera (approximately $>70\%$ NT ID among species of named genera; see Supplementary Information Section 3 and Supplementary Figure 7). The end sequences of a clone were considered jointly recruited if the sequences were recruited by the same genome or were considered disjointly recruited if their end sequences were recruited by different reference genomes. The end sequences of jointly recruited

clones were considered syntenous when the sequences had the same orientation and were separated by a distance on the reference genome that was similar to the size of the DNA fragments used to construct the metagenomic library. Jointly recruited sequences that did not meet both of these criteria were considered non-syntenous. The details of this process are described in Supplementary Information Section 3.

Results

Sanger sequencing of samples from all sites and temperatures yielded 167 Mbp of metagenomic sequence data. Assembly resulted in 5 769 scaffolds, totaling 33 Mbp, which were produced from 67 Mbp (40%) of the total sequence dataset. Cluster analysis of oligonucleotide frequencies was used to characterize 394 scaffolds that were $\geq 20\,000$ bp in length, totaling 20.2 Mbp (Table 3.1). Prior to assembly, recruitment by reference genomes above the specified % NT ID cutoffs indicated in Table 3.2 accounted for 102 Mbp of the total sequence dataset (61%). Scaffold clusters accounted for an additional 13 Mbp (7.8%) of the total unassembled metagenomic sequences that were not recruited to reference genomes above % NT ID cutoffs. Thus, we could confidently assign 69% of the total metagenomic sequences to known taxa or novel phylogenetic clusters by combining these approaches; 31% of the metagenomic sequences are currently of unknown origin. Consistent with the failure to detect 18S rRNA sequences at these temperatures (Liu *et al.*, 2011b), no eukaryotic sequences were observed. Aside from a relative underrepresentation of sequences from *Ca. C. thermophilum* (Supplementary Figure 4), pyrosequencing of SSU rDNA amplicons from environmental DNA showed taxonomic profiles that were similar to those for cDNA sequences produced from rRNAs for the metatranscriptome studies (Liu *et al.*,

2011b). Sequences likely originating from archaea were present, but these organisms are not in high abundance in the upper photic layer of these mats.

Major Populations and their Functional Potential

Clustering on the basis of oligonucleotide frequency revealed eight scaffold clusters (Figure 3.1 and Table 3.1). Phylogenetic affiliations of these clusters were inferred from (i) direct co-clustering with reference genomes (Figure 3.1); (ii) clusters being comprised of sequences recruited by a reference genome at high % NT ID (Figure 3.2, Table 3.2 and Supplementary Table 7); and (iii) the presence of phylogenetically informative marker genes within the clusters (Figure 3.1 and Table 3.3). The metabolic potentials of organisms associated with these clusters were inferred from functional genes they contained (Table 3.4).

(i) Oxygenic Chlorophototrophs: Cluster 1 contained scaffolds that were strongly associated with the *Synechococcus* spp. strains A and B' genomes and included cyanobacterial phylogenetic marker genes and functional genes that were indicative of oxygenic photosynthesis, the Calvin-Benson-Bassham cycle, and genes involved in nitrogen and phosphorus acquisition that were previously described (Bhaya *et al.*, 2007; Steunou *et al.*, 2006, 2008). Most (86%) of these metagenomic sequences were jointly recruited and were more closely related to either the *Synechococcus* sp. strain A or B' genome (Supplementary Figure 8). The cyanobacterial scaffolds in these bins accounted for 19.7% of the total assembled sequence data (Table 3.2), which was the largest amount assigned to any particular group of organisms. Differences between these cyanobacterial scaffolds and the *Synechococcus* spp. isolate genomes were found and give evidence for functional diversity. Scaffolds from native

Table 3.1: Assembly Statistics of Scaffold Clusters $\geq 20\,000$ bp in Length.

Cluster	Phylogenetic Affiliation	Number of Scaffolds	Mbp Sequence in Cluster	Mean Scaffold Length (Kbp)	Mean depth of coverage (read depth)
1	<i>Synechococcus</i> spp.	68	3.54	52.1	4.7
2	<i>Roseiflexus</i> spp.	78	4.26	54.7	3.7
3	<i>Chloroflexus</i> spp.	59	1.91	32.4	1.7
4	<i>Candidatus</i> C. thermophilum-like organisms	10	3.2	319	3.7
5	Chlorobiales-like organisms	32	2.82	88	5.7
6	Anaerolineae-like organisms	46	2.31	50.3	3.2
7	Unknown Cluster 1	27	1.42	52.4	2.2
8	Unknown Cluster 2	39	1.62	41.6	4.4

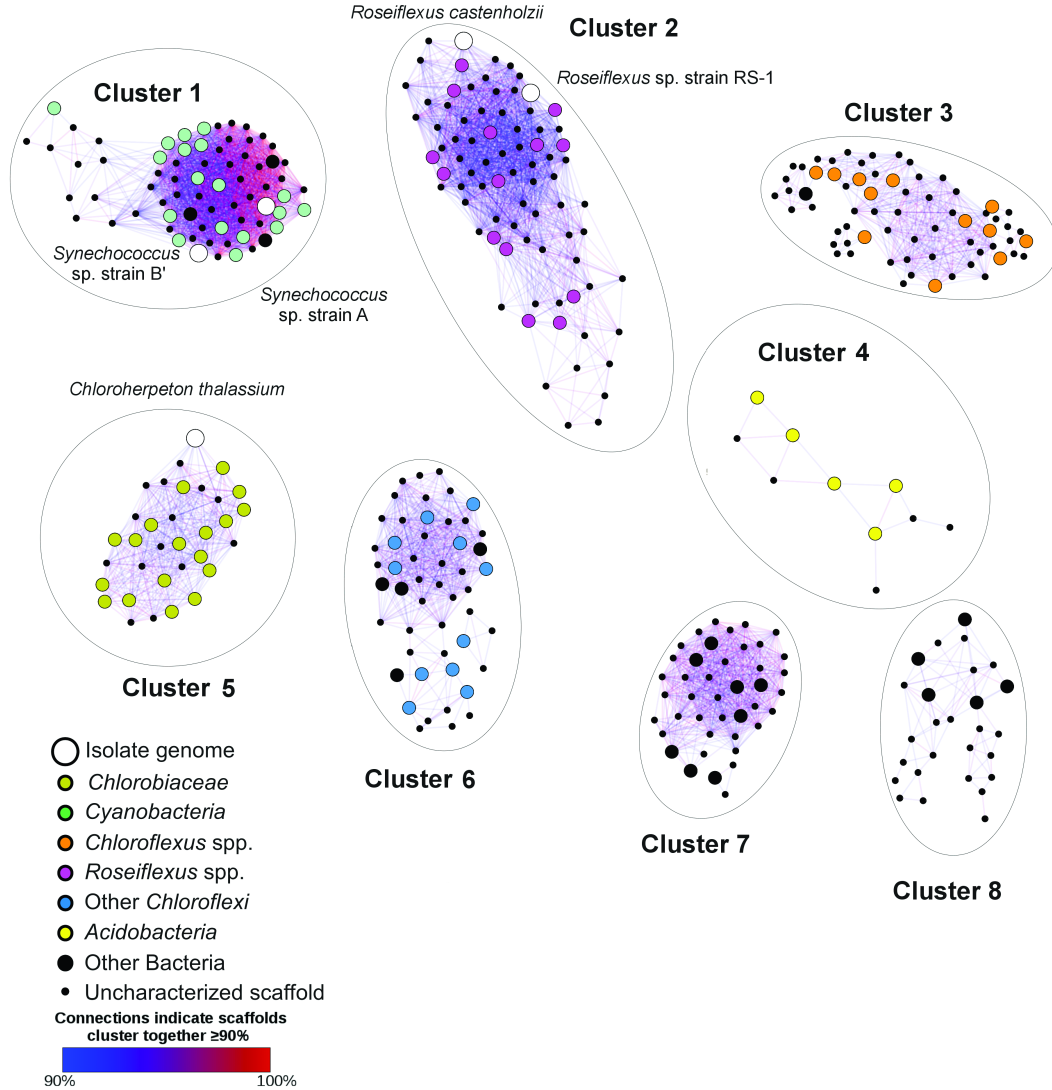


Figure 3.1: Network Map of Core Scaffold Clusters Observed in Celera Assemblies. Scaffolds with similar oligonucleotide frequency profiles that group together in the same cluster are connected by lines colored to indicate the percentage of times they cluster together (in $\geq 90\%$ of 100 trials). Isolate genomes included in this analysis are indicated by large white circles, whereas metagenomic scaffolds that contain characterized phylogenetic marker genes are marked as medium-sized circles colored according to taxonomic grouping. The area of each ellipse is proportional to the amount of metagenomic sequence data contained within each respective scaffold cluster.

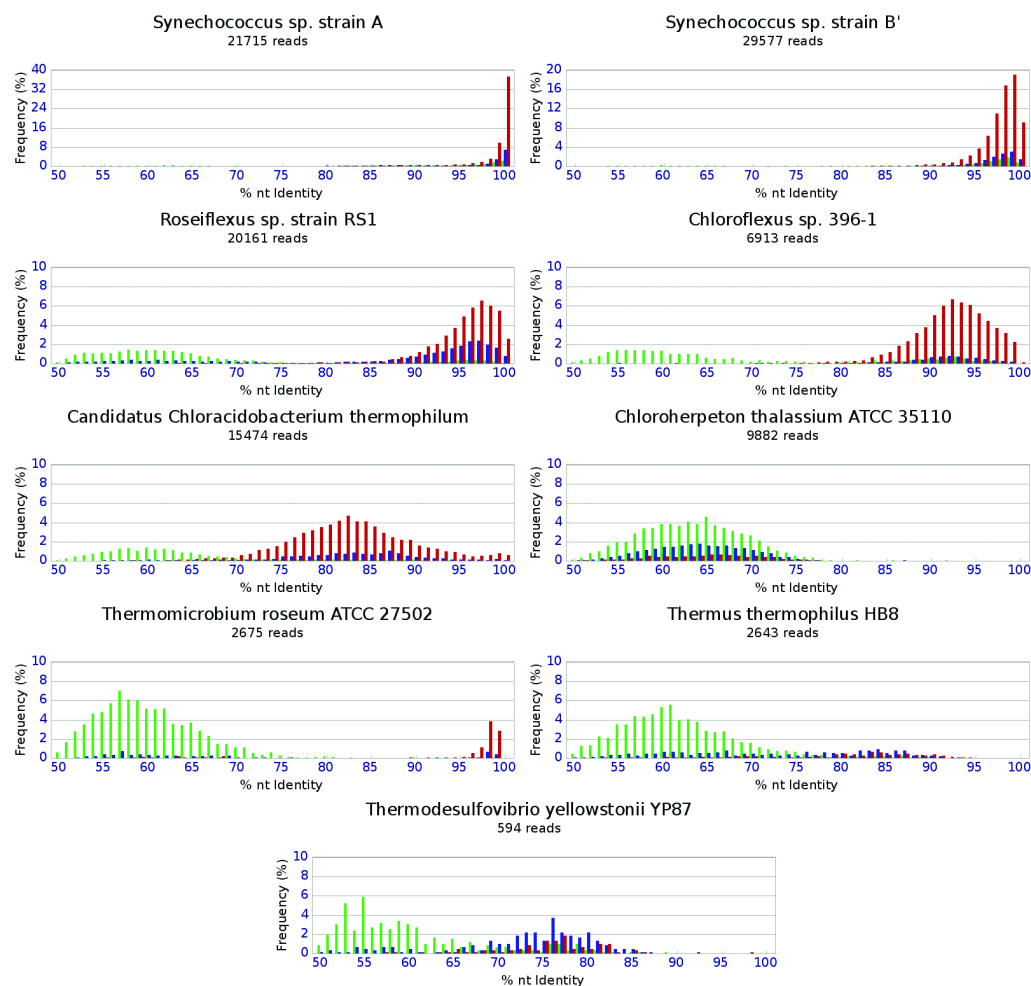


Figure 3.2: Histograms of Recruited Metagenomic Sequences. Histograms of disjointly-recruited (green), jointly recruited syntenous (red) and jointly recruited non-syntenous (blue) metagenomic sequences that can be associated confidently with a reference genome presented as a function of their % NT ID relative to reference genomes that recruited them in BLASTN analysis.

Table 3.2: Comparison of Metagenomic Analyses Based on Genome Recruitment and Assembly¹

Reference genome	Reference Genome Size Mb	Mean $\pm$ s.d. % NT ID of recruited sequences	% NT ID range used in analyses ²	% of individual metagenomic sequences recruited ³				% of total assembled sequences	
				MS low	MS high	OS low	OS high		total
<i>Synechococcus</i> sp. strain A	2.93	94.1 $\pm$ 10.8	92-100	7.03	22.1	6.36	21.1	11.0	9.78
<i>Synechococcus</i> A ⁴	-	-	83-92	0.63	2.92	1.23	1.14	1.57	1.15
<i>Synechococcus</i> sp. strain B'	3.04	95.6 $\pm$ 7.4	90-100	22.1	1.15	24.9	1.10	17.7	8.75
<i>Roseiflexus</i> sp. strain RS1	5.80	84.6 $\pm$ 16.1	80-100	16.0	9.58	7.52	14.9	9.17	9.42
<i>Chloroflexus</i> sp. strain 396-1	5.2	90.4 $\pm$ 6.6	65-100	0.77	16.0	0.99	1.91	4.63	4.62
<i>Cand. Chloracidobacterium thermophilum</i>	3.7	78.5 $\pm$ 11.0	70-100	9.66	2.15	10.35	6.69	8.10	9.23
<i>Chloroherpeton thalassium</i>	3.29	63.4 $\pm$ 5.3	50-100	6.46	2.11	11.64	3.48	8.41	7.65
<i>Thermomicrobium roseum</i>	2.93	64.0 $\pm$ 12.7	75-100	0.21	0.66	0.03	0.15	0.21	0.05
<i>Thermus thermophilus</i>	2.11	73.6 $\pm$ 11.6	75-100	0.08	0.80	0.22	0.27	0.35	0.13
<i>Thermodesulfobivrio yellowstonii</i>	2.00	73.4 $\pm$ 8.0	75-100	0.46	0.15	0.04	0.12	0.11	0.03

¹ All results met the criterion of having an e-value more significant than 10^{-10} for WU-BLASTN parameters M=3, N=-2, and a database size of 68 Mbp.

² Relative to the reference genome.

³ The abbreviations for the four metagenomes are as follows: MS = Mushroom Spring; OS = Octopus Spring; low = 60°C average, high = 65°C average.

⁴ Recruited by the *Synechococcus* sp. A genome with 83 to 92% NT ID.

Table 3.3: Phylogenetic Marker Genes and Functional Genes in Assembly Clusters.

Cluster: Phylogeny	Phylogenetic genes ¹	Functional genes ²	Pathways/functions
1: <i>Synechococcus</i> spp.	16S rRNA <i>pyrG</i> <i>recA</i> <i>rplB</i> , <i>rplC</i> , <i>rplD</i> , <i>rplE</i> , <i>rplF</i> , <i>rplK</i> , <i>rplL</i> , <i>rplN</i> , <i>rplM</i> , <i>rplP</i> , <i>rplT</i> <i>rpoB</i> <i>rpsB</i> , <i>rpsE</i> , <i>rpsI</i> , <i>rpsJ</i> , <i>rpsK</i> , <i>rpsM</i> , <i>rpsS</i> <i>snpB</i> <i>tsf</i>	<i>chlB</i> , <i>chlG</i> , <i>chlI</i> , <i>psaA</i> , <i>psaL</i> , <i>psbB</i> , <i>psbC</i> <i>cpcABCDEFGHI</i> , <i>apcABC</i> , <i>nblA</i> <i>rbtX</i> , <i>cbtS</i> , <i>cbtL</i> , FBP aldolase, PRK <i>ctaC</i> , <i>ctaD</i> , <i>ctaE</i> <i>nifH</i> <i>narB</i> <i>amtB</i> <i>pstABCS</i> <i>phnCEGHILJ</i>	oxygenic chlorophototrophy Calvin-Benson-Bassham Cycle oxygen respiration nitrogen fixation nitrate metabolism ammonium transport phosphate transport phosphonate metabolism
2: <i>Roseiflexus</i> spp.	16S rRNA <i>nusA</i> <i>rplA</i> , <i>rplB</i> , <i>rplC</i> , <i>rplD</i> , <i>rplE</i> , <i>rplN</i> , <i>rplP</i> , <i>rplT</i> <i>rpsB</i> , <i>rpsC</i> , <i>rpsI</i> , <i>rpsS</i> <i>tsf</i>	<i>bchB</i> , <i>bchC</i> , <i>bchD</i> , <i>bchE</i> , <i>bchF</i> , <i>bchG</i> , <i>bchH-1</i> , <i>bchH-2</i> , <i>bchI</i> , <i>bchJ</i> , <i>bchL</i> , <i>bchM</i> , <i>bchN</i> , <i>bchP</i> , <i>bchT</i> -like, <i>bchX</i> , <i>bchY</i> , <i>bchZ</i> <i>pufB</i> , <i>pufC</i> , <i>pufLM</i> <i>cyoB</i> , <i>coxA</i> , <i>coxB</i> <i>mch</i> , <i>mcl</i> , <i>mcr</i> , <i>mct</i> , <i>meh</i> , <i>pcs</i> , <i>smtA</i> , <i>smtB</i> <i>hydAB</i> <i>amtB</i>	Bchl <i>a</i> biosynthesis anoxygenic chlorophototrophy oxygen respiration 3-hydroxypropionate pathway hydrogenase ammonium transport
3: <i>Chloroflexus</i> spp.	<i>infC</i> <i>nusA</i> <i>pgk</i> <i>recA</i> <i>rplT</i>	<i>bchB</i> , <i>bchG</i> , <i>bchH-1</i> , <i>bchL</i> , <i>bchN</i> , <i>bchP</i> , <i>bchS</i> , <i>bchY</i> , slr1923-homolog <i>mcr</i> , <i>pcs</i> <i>pufC</i> <i>cyoB</i> , <i>coxC</i> <i>sqr</i> <i>hydAB</i> <i>pstABCS</i>	Bchl <i>a</i> and <i>c</i> biosynthesis 3-hydroxypropionate pathway anoxygenic chlorophototrophy oxygen respiration reduced S oxidation hydrogenase phosphate transport
4: <i>Candidatus C. thermophilum</i> -like organisms	16S rRNA <i>dnaG</i> <i>frr</i> <i>nusA</i> <i>pgk</i> <i>pyrG</i> <i>recA</i> <i>rplE</i> , <i>rplM</i> , <i>rplT</i> <i>rpsB</i> , <i>rpsI</i> <i>snpB</i>	<i>bchD</i> , <i>bchE</i> , <i>bchF</i> , <i>bchL</i> , <i>bchU</i> , <i>acsF</i> , <i>bchI</i> , <i>bchM</i> , <i>bchX</i> , <i>bchB</i> , <i>bchN</i> , <i>bchK</i> , <i>bchC</i> , <i>bchZ</i> , <i>bchY</i> , <i>bchG</i> , <i>bchP</i> <i>csmA</i> <i>amtB</i>	Bchl <i>a</i> and <i>c</i> biosynthesis chlorosome biosynthesis ammonium transport

continued on next page

continued from previous page

Cluster: Phylogeny	Phylogenetic genes ¹	Functional genes ²	Pathways/functions
5: Chlorobiales-like organisms	16S rRNA <i>frr</i> <i>rplB</i> , <i>rplC</i> , <i>rplD</i> , <i>rplF</i> , <i>rplK</i> , <i>rplL</i> , <i>rplN</i> , <i>rplP</i> , <i>rplS</i> , <i>rplT</i> <i>rpsE</i> , <i>rpsI</i> , <i>rpsJ</i> , <i>rpsK</i> , <i>rpsM</i> <i>smpB</i> <i>tsf</i>	<i>bchB</i> , <i>bchC</i> , <i>bchD</i> , <i>bchF</i> , <i>bchG</i> -homolog, <i>bchH</i> , <i>bchH</i> -homolog, <i>bchI</i> , <i>bchK</i> , <i>bchL</i> , <i>bchP</i> , <i>bchR</i> , <i>bchX</i> , <i>bchY</i> , <i>bchZ</i> <i>fmoA</i> <i>pscA</i> , <i>pscB</i> , <i>csmC</i> <i>norE</i> -like cyt. c oxidase <i>amtB</i>	Bchl <i>a</i> and <i>c</i> biosynthesis Anoxygenic chlorophototrophy oxygen respiration ammonium transport
6: Anaerolineae-like organisms	16S rRNA <i>frr</i> <i>infC</i> <i>pgk</i> <i>pyrG</i> <i>recA</i> <i>rplA</i> , <i>rplK</i> , <i>rplL</i> , <i>rplM</i> , <i>rplT</i> <i>rpoB</i> <i>rpsI</i> , <i>rpsM</i> <i>smpB</i> <i>tsf</i>	<i>bchF</i> , <i>bchG</i> , <i>bchI</i> , <i>bchP</i> , <i>bchS</i> -like, <i>bchX</i> , <i>bchY</i> , <i>bchZ</i> <i>pufL</i> , <i>pufM</i> , <i>pufC</i> <i>coxA</i> , <i>coxB</i> <i>sqr</i>	Bchl <i>a</i> biosynthesis Anoxygenic chlorophototrophy oxygen respiration reduced S oxidation
7: Unknown Cluster 1	<i>dnaG</i> <i>rplE</i> , <i>rplN</i> , <i>rplP</i> , <i>rpsC</i> , <i>rpsK</i> , <i>rpsM</i> , <i>rpsS</i>	<i>glcD</i> , <i>glcE</i> <i>acs</i> <i>coxA</i> , <i>coxB</i> , <i>coxC</i>	glycolate oxidation acetate metabolism oxygen respiration
8: Unknown Cluster 2	<i>dnaG</i> <i>infC</i> <i>pgk</i> <i>pyrG</i> <i>recA</i> <i>rplB</i> , <i>rplC</i> , <i>rplK</i> , <i>rplL</i> , <i>rplM</i> , <i>rplN</i> , <i>rplP</i> <i>rpsC</i> , <i>rpsE</i> , <i>rpsJ</i> , <i>rpsM</i>	<i>coxA</i> , <i>coxB</i> , <i>coxC</i>	oxygen respiration

¹Phylogenetic marker genes identified with AMPHORA and/or phylogenetic analysis.

²Functional genes identified with BLAST and annotated using hidden markov models, genomic context, and/or phylogenetic analysis (see Supplementary Information Section 3).

Table 3.4: Relationship Between Predominant Phylogenetic Groups, Functional Potential and Functional Guilds.

Phylogenetic group	Chlorophylls	Carbon metabolism	Relative Temperature Distribution	Possible Electron Donor Utilization	Possible Electron Acceptor Utilization	Functional guild ¹
A-like and B'-like <i>Synechococcus</i> spp.	Chl <i>a</i>	Autotrophy	60 and 65°C ²	H ₂ O	O ₂	600-700 nm oxygenic phototrophs
<i>Roseiflexus</i> -like FAPs	BChl <i>a</i>	Mixotrophy ³	60 and 65°C	H ₂	O ₂	850-950 nm mixotrophs
<i>Chloroflexus</i> -like FAPs	Major: BChl <i>c</i> , Minor: BChl <i>a</i>	Mixotrophy	60 <65°C	H ₂ , HS ⁻ , S ₂ O ₃ ²⁻ , S ⁰	O ₂	700-750 nm mixotrophs
<i>C. thermophilum</i> -like spp.	Major: BChl <i>c</i> , Minor: BChl <i>a</i> , Chl <i>a</i>	Heterotrophy	60 >65°C	ND	O ₂	700-750 nm heterotrophs
Chlorobiales-like spp.	Major: BChl <i>c</i> , Minor: BChl <i>a</i> , Chl <i>a</i>	Mixotrophy	60 >65°C	ND	O ₂	700-750 nm mixotrophs
Anaerolineae-like spp.	BChl <i>a</i> ⁴	Unknown	60 and 65°C	HS ⁻ , S ₂ O ₃ ²⁻ , S ⁰	O ₂	BChl <i>a</i> /N-IR mixotrophs
Cluster 7 spp.	ND	heterotrophy	60 and 65°C	Glycolate, acetate	O ₂	Aerobic chemoorganoheterotrophs
Cluster 8 spp.	ND	heterotrophy	60 <65°C	????	O ₂	Aerobic chemoorganoheterotrophs

¹Ranges in which the absorption maxima of the light-harvesting systems of these guilds are maximal in the red to near-infrared region of the electromagnetic spectrum.

²B'-like sequences were much more predominant at 60°C than at 65°C; A-like sequences were observed at 60°C and were predominant at 65°C.

³Mixotrophy is referring to both heterotrophic and autotrophic growth, perhaps simultaneously (Bryant et al., 2011).

⁴Insufficient evidence currently exists to determine whether this organism can synthesize other chlorophylls and to know its principal absorption range in the near-IR.

Synechococcus sp. strain A-like populations contained genes encoding *feoAB* (involved in Fe^{2+} transport) and genes homologous to the characterized bacterial enzymes urea carboxylase (*ureA*) and allophanate hydrolase (*atzF*; involved in the degradation of urea into ammonia and CO_2), both of which are not found in the *Synechococcus* sp. strain A genome (Supplementary Table 9) (Kanamori *et al.*, 2004; Cheng *et al.*, 2005).

(ii) Filamentous Anoxygenic Chlorophototrophs: Cluster 2 scaffolds had similar oligonucleotide frequencies to both the *Roseiflexus* sp. strain RS1 and *R. castenholzii* genomes, and they were predominantly comprised of sequences recruited by the *Roseiflexus* sp. strain RS1 genome (98%, with a mean of 95% NT ID; Supplementary Table 7). Many conserved phylogenetic marker genes, with sequences almost identical to homologs in the *Roseiflexus* sp. RS1 genome, were found on Cluster 2 scaffolds (Table 3.4). Most of the Cluster 2 sequences were jointly recruited by the *Roseiflexus* sp. strain RS1 genome with more than 80% NT ID (Figure 3.2), which was above the mean from a comparison of *Roseiflexus* sp. strain RS1 and *R. castenholzii* homologs (Supplementary Information Section 3). This observation implies that a large proportion of scaffolds are represented by sequences from a diverse assemblage of *Roseiflexus* spp. and is consistent with the diversity of sequences directly recruited by the *Roseiflexus* sp. strain RS1 genome by BLASTN independently of metagenomic assembly (Figure 3.2). One scaffold in Cluster 2 contained a diagnostic fused *pufLM* gene that encodes both of the type-2 photosystem reaction center polypeptides (*pufL* and *pufM* are characteristically fused in *Roseiflexus* spp.; Youvan *et al.* 1984; Yamada *et al.* 2005 (Figure 3.3). There were *recA* sequences highly similar to the *Roseiflexus* sp. strain RS1 *recA* in the metagenome (Supplementary Figure 10), but these were not encoded

on the large scaffolds included in the cluster analysis. Suggesting that these organisms have the capability to fix inorganic carbon, Cluster 2 also contained eight ORFs homologous to *Roseiflexus* spp. genes encoding key enzymes in the 3-hydroxypropionate pathway (Klatt *et al.*, 2007). Like *Roseiflexus* sp. strain RS1, *Roseiflexus* spp. native to the mat may have the potential to use H₂ as an electron donor because Cluster 2 scaffolds contain homologs of bidirectional [NiFe]-hydrogenases (*hydAB*) (Table 3.4, van der Meer *et al.* 2010. One ORF homologous to a *nifH* gene in the *Roseiflexus* sp. strain RS1 genome was also observed.

Oligonucleotide compositions of Cluster 3 scaffolds were not similar to any sequenced isolate genomes above the 90% bootstrap cutoff; however, the phylogenetic and functional marker genes they contained indicated that these scaffolds were contributed by *Chloroflexus* spp. Most (82%) of the metagenomic sequences comprising these scaffolds were recruited at a high degree of similarity (Table 3.3) by the genome of *Chloroflexus* sp. strain 396-1, which is currently the most representative cultivated organism compared to the native *Chloroflexus* spp. in these mats (van der Meer *et al.*, 2010). Most (85%) of the metagenome sequences recruited by the *Chloroflexus* sp. strain 396-1 genome were jointly recruited sequences that had a mean % NT ID of 91.3 $\pm$ 5.3 % (Figure 3.2). One Cluster 3 scaffold contained a *pufC* homolog adjacent to *bchP* and *bchG*, consistent with the *Chloroflexus* sp. 396-1 genome (93 % NT ID, 100 % AA ID) (Figure 3.3). Overlapping metagenome sequences were missing upstream of the *pufC* open reading frame, so it could not be confirmed whether the native *Chloroflexus* spp. have the *pufBAC* operon structure observed in other *Chloroflexus* spp. (Watanabe *et al.*, 1995). However, the co-localized *bchG* and *bchP* genes and high % NT ID to *Chloroflexus* sp. 396-1 are consistent with this inference derived from oligonucleotide clustering (Figure 3.3). Homologs of genes involved in both BChl *c* and *a* biosynthesis were present in Cluster 3, indicating that the native *Chloroflexus*

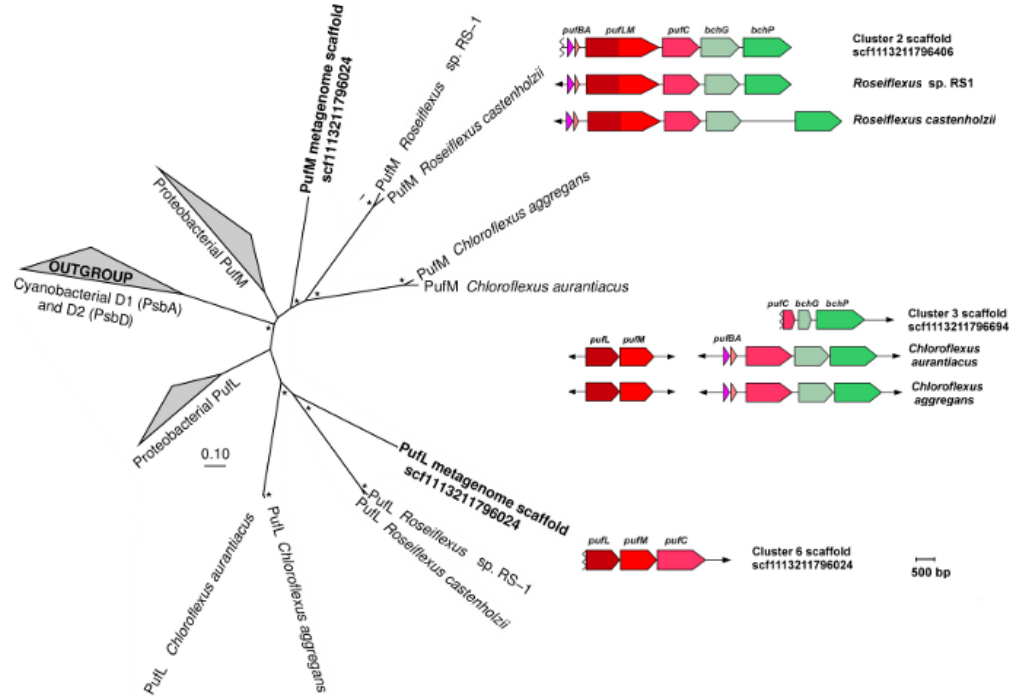


Figure 3.3: PufL and PufM Phylogeny and Genomic Context. The neighbor-joining phylogenetic tree of PufL and PufM sequences from a novel Chloroflexi metagenomic scaffold from Cluster 6 and from sequenced genomes is marked with asterisks at nodes which reflect bootstrap support (1000 replications). A more detailed tree is shown as Supplementary Figure 12. The genomic context of genes encoding the type-2 reaction center and light harvesting polypeptides in metagenomic scaffolds and chromosomes of *Chloroflexus* and *Roseiflexus* isolates is also displayed. Jagged lines indicate positions on scaffolds that are interrupted by a lack of overlapping sequence data between contigs.

spp. are physiologically similar to known isolates with respect to light-harvesting strategies (Bryant and Frigaard, 2006; Frigaard and Bryant, 2006; Bryant *et al.*, 2012) (Table 3.4). Sequences encoding two key enzymes in the 3-hydroxypropionate pathway, and most closely related to homologs in the *Chloroflexus* sp. strain 396-1 genome, were present on Cluster 3 scaffolds. This suggests that *Chloroflexus* spp. in the mats may be capable of carbon fixation by the 3-hydroxypropionate pathway. Cluster 3 contained a homolog of sulfide-quinone oxidoreductases (*sqr*) in *Chloroflexus* spp., which suggested that these organisms might oxidize sulfide to polysulfides (Bryant *et al.*, 2012).

(iii) Candidatus Chloracidobacterium spp.: Cluster 4 contained five scaffolds containing phylogenetic marker genes with best matches to Acidobacteria (including a *recA* sequence labeled RecA Cabt in Supplementary Figure 10). These scaffolds had distinct oligonucleotide frequency patterns as compared to the *Ca. C. thermophilum* genome, of which a detailed analysis will be published separately (Garcia Costas *et al.*, 2012), despite the fact that 97% of the sequences from these scaffolds were recruited by this genome with a mean of 82.5% NT ID (Figure 3.2 and Supplementary Table 7). Genes involved in BChl and chlorosome biosynthesis were observed on these scaffolds, and a gene encoding a type-1 photosynthetic reaction center gene (*pscA*) was observed when the clustering stringency was lowered to 80%. Although the number of Cluster 4 scaffolds was small, these scaffolds were the largest produced by the Celera assembler (the average size was >300 000 bp, and the largest was 1.6 Mbp; see Table 3.1). The *Ca. C. thermophilum* genome recruited 8.1% of all unassembled metagenome sequences, 90.8% of which were jointly recruited (Figure 3.2). The % NT ID distribution of these sequences suggested that, while there are

native mat organisms nearly identical to the *Ca. C. thermophilum* isolate at some loci (Figure 3.2), most Cluster 4 sequences are derived from organisms more distantly related to *Ca. C. thermophilum* than are two species of the genera we investigated (Supplementary Information Section 3). The high proportion of syntenous, jointly recruited metagenome sequences from the genome recruitment analysis was evidence for conservation of synteny within this population, which probably contributed in part to the longer than average assemblies.

(iv) Chlorobiales-like Organisms: Cluster 5 scaffolds had oligonucleotide frequency signatures similar to that of the *Chloroherpeton thalassium* genome (Figure 3.1) and contained phylogenetic marker and functional genes (Table 3.4) that are typical of members of the Chlorobiales. The genome of *C. thalassium* recruited 8.4% of the metagenomic sequences across all temperature-spring combinations, most of which were from low-temperature samples and were disjointly recruited (Table 3.2 and Figure 3.2). Although they were not found on scaffolds >20 kbp, many *recA* sequences were recruited that, like the *C. thalassium recA* sequence, form an out-group to the clade that contains the well-characterized chlorophototrophs in the order Chlorobiales (Supplementary Figure 10). The 63.4% mean NT ID to *C. thalassium* homologs was approximately equal to the % NT ID of homologs belonging to different genera within a kingdom-level lineage (Figure 3.2, Supplementary Information Section 3). Hence, phylogenetic information alone did not provide high confidence that these sequences were derived from members of the Chlorobiales. Functional genes found on the scaffolds of this cluster clarified the potential physiological properties of this population. In particular, one scaffold contained a gene encoding a homolog of the Fenna-Matthews-Olson protein, which is a BChl *a*-binding antenna protein

involved in anoxygenic photosynthesis and only known to occur in members of the Chlorobiales and chlorophototrophic *Acidobacteria* (Bryant *et al.*, 2007, 2012). Other Cluster 5 scaffolds contained homologs of the reaction center subunit gene *pscA* (OS GSB PscA, Bryant *et al.* 2007), *pscB*, *pscD* as well as *csmC*, a gene encoding a chlorosome envelope protein that has no homologs in other chlorosome-containing chlorophototrophs and thus is currently diagnostic for Chlorobiales (Bryant *et al.*, 2012).

(v) Novel Anaerolineae-like Chlorophototroph: Cluster 6 scaffolds were not similar in oligonucleotide composition to any isolate genome but contained phylogenetic marker genes associated with bacteria from Kingdom Chloroflexi (Figure 3.1). The RDP Bayesian Classifier assigned a full-length 16S rRNA sequence in this cluster to the taxonomic Class Anaerolineae with 95% confidence, and this observation was supported by phylogenetic analysis (see Supplementary Figure 11). Furthermore, genes encoding ribosomal proteins and *recA* genes (Table 3.4) supported this kingdom-level phylogenetic assignment. In particular, a *recA* gene associated with assembly Cluster 6 (RecA 6, Supplementary Figure 10) is phylogenetically earlier diverging than the monophyletic clade containing known chlorophototrophic Chloroflexales (e. g., *Roseiflexus* and *Chloroflexus* spp.). Several genes involved in anoxygenic chlorophototrophy were encoded on the same scaffold as the 16S rRNA gene in Cluster 6. This cluster also contained *bchXYZ* genes encoding the subunits of the light-independent chlorophyllide reductase, an enzyme required for the biosynthesis of BChl *a* (Chew and Bryant, 2007), as well as other BChl biosynthesis genes (*bchD*, *bchF*, *bchH* and *bchI*) common to BChl *a* and BChl *c* biosynthetic pathways. A separate scaffold in this cluster contained non-fused *pufL* and *pufM* sequences homologous to Chloroflexi

sequences but in a unique genomic context (Figure 3.3). Phylogenetic analysis of the PufL and PufM sequences showed that, in comparison to those of known filamentous anoxygenic chlorophototrophs (FAPs) in the Chloroflexales, these sequences occupy novel and/or basal positions in a phylogenetic tree (Figure 3.3, Supplementary Figure 12). When compared to their closest homologs in *Chloroflexus* and *Roseiflexus* spp. genomes, these PufL and PufM sequences had amino acid identities of 48 to 62%, respectively. Assembly-independent BLASTN analysis revealed that the metagenome sequences comprising Cluster 6 scaffolds had lower (60-66)% NT ID to the Chloroflexi genomes. Approximately 33% of the sequences comprising the Cluster 6 scaffolds were not recruited by any reference genome above established cutoffs, and thus were null bin sequences (see Supplementary Table 7).

(vi) Novel Putatively Chemoorganotrophic Populations: Scaffolds in Clusters 7 and 8 did not have oligonucleotide frequencies similar to any tested isolate genomes, and contained functional and phylogenetic marker genes (including RecA 7 in Supplementary Figure 10) with very distant relationships to sequences in currently available public databases. Most metagenomic sequences contained in these scaffolds were not recruited by a reference genome above the specified cutoff and were assigned to the null bin, but some sequences were recruited at low % NT ID by multiple genomes (Supplementary Table 7). Clusters 7 and 8 did not contain any genes homologous to those specific for chlorophototrophy. Both clusters contained genes encoding *caa3*-type cytochrome c oxidases, which suggested the potential for aerobic oxidative phosphorylation exists in the organisms contributing these sequences. Cluster 7 additionally included scaffolds encoding glycolate oxidase (*glcD*) and acetyl-CoA synthetase (*acs*) genes (Table 3.3). Thus, the organisms contributing these sequences may have the

potential for aerobic chemoorganotrophy with glycolate and/or acetate as an electron donor. No assembly clusters corresponded to organisms related to *Thermomicrobium roseum*, *Thermus thermophilus* or *Thermodesulfovibrio yellowstonii*, but the genomes of these isolates recruited sequences above 75% NT ID (Table 3.2; Figure 3.2). All other reference genomes recruited a low number of sequences with low % NT ID values (Supplementary Figure 13). Approximately 20% of metagenomic sequences could not be associated with any reference genome above an e-value cutoff of 10^{-10} with the specified parameters and were assigned to the null bin.

Patterns of Metagenomic Diversity

(i) Multiple Populations in Recruitment Bins: Recruitment analysis of the metagenomic clones from the 65°C Mushroom Spring sample revealed at least two populations, one with >96% NT ID and one with 83-92% NT ID relative to the *Synechococcus* sp. strain A genome (Figure 3.4). The more divergent sequences were likely contributed by A'-like *Synechococcus* spp., as they exhibited >98% NT ID with homologs in a metagenome produced by pyrosequencing from a 68°C sample from Mushroom Spring, known to be dominated by these genotypes (Ferris *et al.* 2003; Supplementary Information Section 6 and Supplementary Figures 3 and 14). These accounted for only 1.57% of the A-like sequences in all metagenomes (Table 3.2).

(ii) Synteny Versus Relatedness: There was a positive relationship between the degree of genetic relatedness and the conservation of synteny in both metagenomic sequences and genomic reference sequences as compared to *Synechococcus* sp. strain A (Figure 3.5). Metagenomic sequences originating from A-like organisms (i.e., $\geq 92\%$ NT ID with the *Synechococcus* sp. strain A genome) displayed greater synteny with

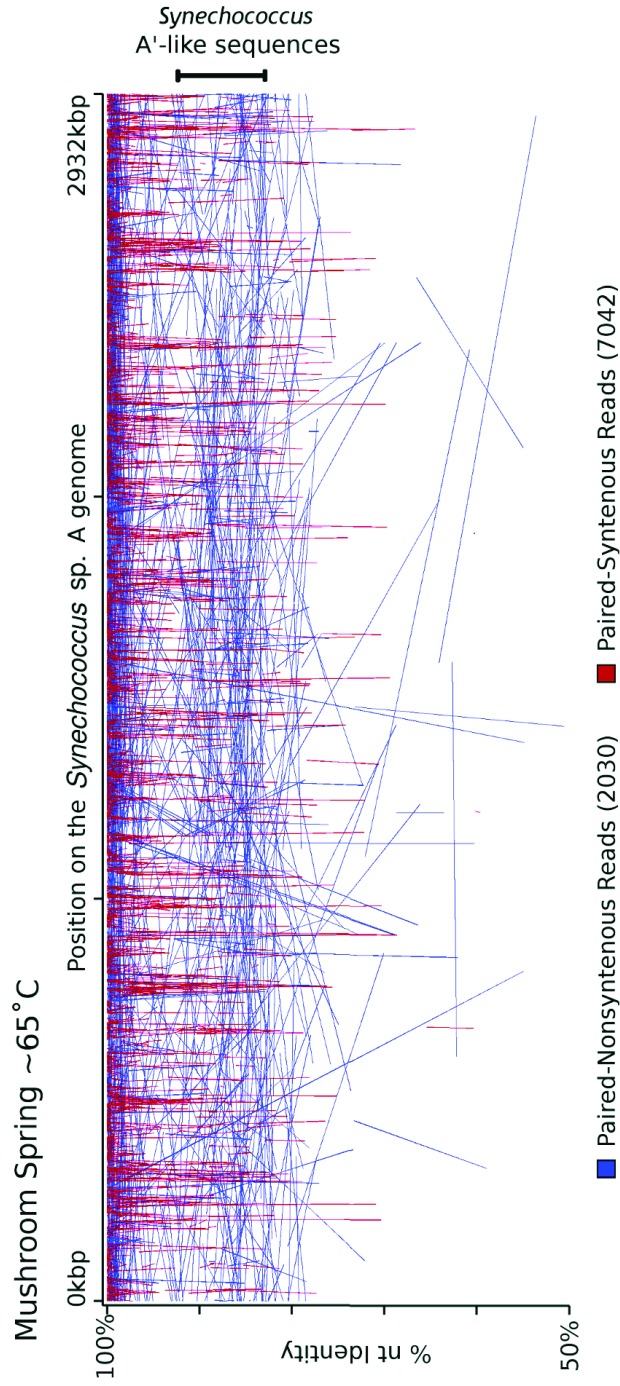


Figure 3.4: Position of Metagenomic Sequence Alignments on *Synechococcus* sp. A Genome. Position of alignments and the corresponding % NT ID to the *Synechococcus* sp. A genome of syntenous (red) and nonsynonymous (blue) sequences jointly recruited by the *Synechococcus* sp. A genome from the Mushroom Sp. ~ 65°C metagenome. Each end sequence is connected by a line to its clone mate. Sequences suspected to originate from *Synechococcus* sp. A'-like populations ranging from 83 to 92 % NT ID are indicated on the right side of the graph.

respect to the *Synechococcus* sp. strain A genome than did sequences associated with A'-like organisms (i.e., 83-92% NT ID with the *Synechococcus* sp. strain A genome), which in turn displayed higher synteny than did B'-like sequences (i.e., comparing sequences that had $\geq 90\%$ NT ID to the *Synechococcus* sp. strain B' genome with homologs in the *Synechococcus* sp. strain A genome). To assess synteny with more distantly related isolate genomes, we compared paired end sequences of simulated metagenomic fragments (comprised of sequence fragments from representative cyanobacterial isolate genomes fractionated to reflect the range of sizes and abundances of our Sanger metagenome clone inserts) with the *Synechococcus* sp. strain A genome (Supplementary Information Section 3). Synteny between the *Synechococcus* sp. strain A and B' genomes was nearly identical to that observed empirically, but synteny between the *Synechococcus* sp. strain A genome and the more distantly related genomes was almost undetectable (Figure 3.5).

Evidence of Homologous Recombination

Metagenomic clones, whose disjointly recruited ends can each be confidently associated with different reference genomes, provided evidence for possible past gene exchange between A-like *Synechococcus* spp. and members of the *Synechococcus* A' and B' lineages, as well as between these cyanobacteria and FAPs or *Ca. C. thermophilum*. The relative percentage of clones, whose end sequences could be confidently associated with *Synechococcus* sp. strain A on one end and with other populations on the other end, decreased from 26% of all A'-like sequences (i. e., 83 to 92% NT ID to *Synechococcus* sp. strain A; no isolate genome is available from this organism type) to 4.5% of all *Synechococcus* sp. strain B'-like sequences (i. e., $>90\%$ NT ID to *Synechococcus* sp. strain B'), to 1.1% of sequences associated with a more distantly related cyanobacterial reference genome (i. e., *Thermosynechococcus*

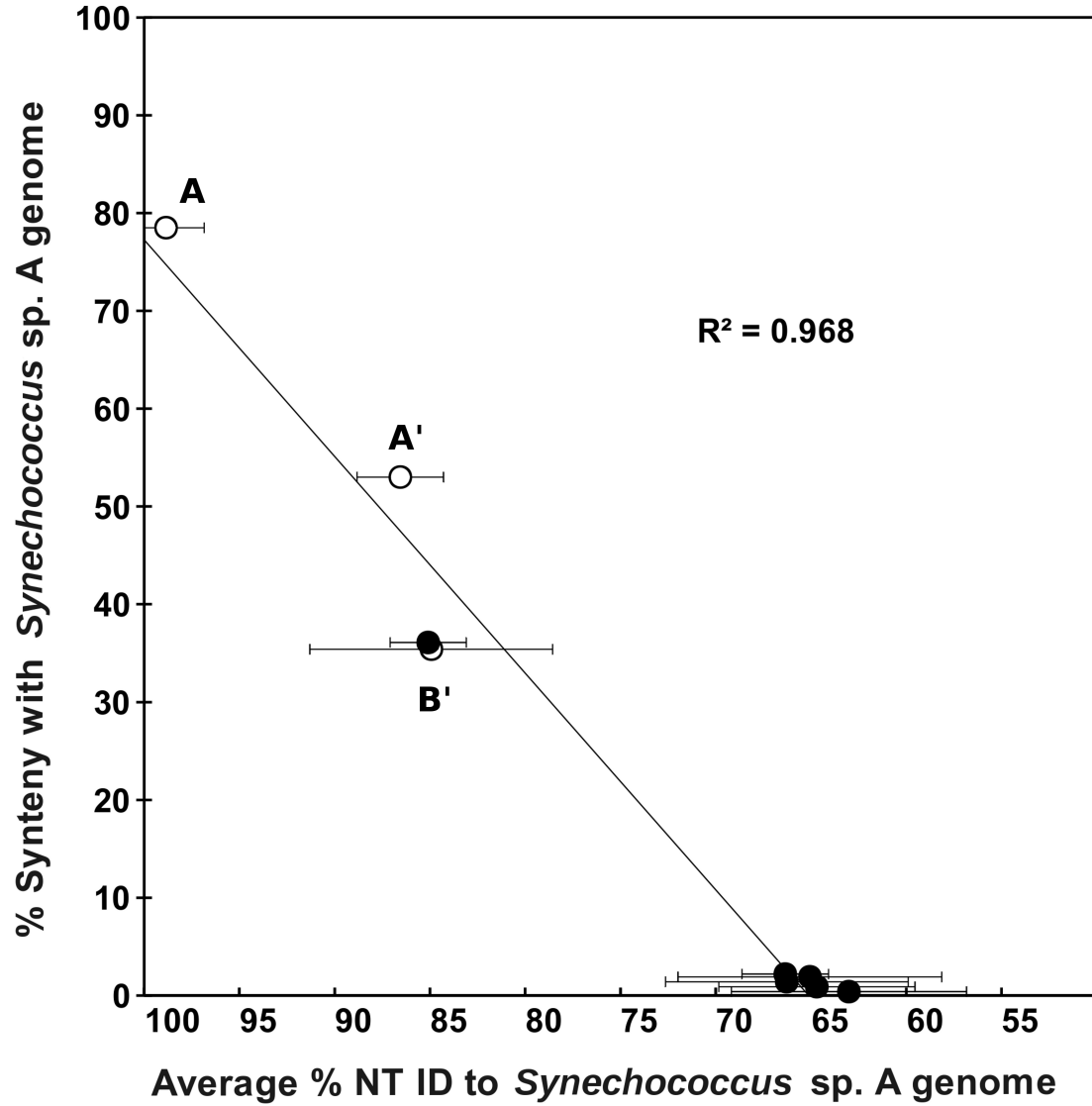


Figure 3.5: Synteny Conservation Between the *Synechococcus* sp. Strain A Genome and Metagenomic Sequences and other Genomes. Open circles represent alignments of metagenomic sequences relative to the *Synechococcus* sp. strain A genome. Metagenome sequences were categorized as *Synechococcus* A, A', or B' based on % NT ID ranges to the *Synechococcus* spp. strain A and B' recruitment bins. Closed circles represent alignments of genome sequences from cultivated cyanobacteria (*Thermosynechococcus elongatus*, *Gloeobacter violaceus*, *Synechococcus* sp. strain WH8102, *Nostoc* sp. strain PCC7120) and the outgroup organism *Roseiflexus* sp. RS-1 relative to the *Synechococcus* sp. strain A genome. These genome fragments were generated in silico to represent the same proportion of insert sizes observed in the distribution of metagenome sequences that were recruited by the *Synechococcus* sp. A genome.

elongatus BP-1), and to 0.2% of sequences associated with yet more distantly related genomes (i. e., *Roseiflexus* sp. strain RS1, *Chloroflexus* sp. strain 396-1, or *Ca. C. thermophilum*). Many of these disjointly recruited metagenome sequences encoded CRISPR-associated proteins putatively involved in adaptive responses to phage predation. Some recombination events among cyanobacteria and more distantly related organisms may thus be indicative of phage-host interactions (Supplementary Table 9; Heidelberg *et al.* 2009). Other disjointly recruited cyanobacterial sequences encoded transposases on the linked paired-end sequences that were recruited to bacterial genomes other than from cyanobacteria. Such mobile genetic elements may even be transferred across distant lineages (Supplementary Table 9). These putative homologous recombination events were more frequently observed between closely related populations, e.g., between *Synechococcus* sp. strain A and A' populations (Figures 3.4 and 3.5).

Discussion

This 167-Mbp metagenome study of the green mat layer of Octopus and Mushroom Springs resulted in depth-of-coverage estimates between $\sim 1.7X$ and $\sim 5.7X$ for the eight dominant populations demarcated by scaffold clustering (Table 3.1). The complexity of this metagenome was relatively limited compared to the metagenome of a non-thermal, hypersaline phototrophic, microbial mat from Guerrero Negro in Baja California Sur, Mexico (~ 105 Mbp total metagenomic sequence; Kunin *et al.* 2008, which did not produce assemblies greater than $\sim 8\,400$ bp in length. Metagenomic studies of less complex microbial communities have benefited from the assembly of metagenomic sequence data to identify and characterize the function of novel community members for which reference genomes of closely related organisms are not

available (e. g., Tyson *et al.* 2004; Simmons *et al.* 2008; Dick *et al.* 2009; Inskeep *et al.* 2010; Denef *et al.* 2010). The structure of the Octopus and Mushroom Spring communities enabled us to use similar strategies to link community composition and potential function in these mats by resolving the phylogenetic and genomic context of individual functional genes, which led to the assignment of metabolic characteristics for microorganisms previously known only by the presence of 16S rRNA sequences.

Linkage Between Community Composition and Potential Community Function

The observation of assembly clusters with genes that indicated metabolic properties consistent with *Synechococcus* spp., *Roseiflexus* spp., *Chloroflexus* spp. and *Ca. C. thermophilum* was expected. However, the ability to associate functional potential with phylogeny also enabled us to link genes indicative of anoxygenic chlorophototrophy with a Chlorobiales-like population, and thus to confirm suspicions based on 16S rRNA sequence data that were not definitive and on a *pscA* sequence that previously could not be linked to phylogenetic markers. The ability to link functional and phylogenetic markers through assembly also enabled the discovery of three new predominant populations of organisms in this mat, which is remarkable because this system has been studied by numerous microbiologists over many decades.

One newly discovered population (Cluster 6), which has the functional potential for anoxygenic chlorophototrophy, is most closely related to cultured chemoorganotrophic bacteria isolated from thermal environments belonging to the classes Anaerolineae and Caldilineae within Kingdom Chloroflexi (Sekiguchi *et al.*, 2003; Hugenholtz and Stackebrandt, 2004; Yamada *et al.*, 2006, 2007). We detected 16S rRNA sequences of these populations (Supplementary Figure 4 and Liu *et al.* 2011b) but were unable to infer from them a phototrophic phenotype, since these lineages of Kingdom Chloroflexi

had not previously been known to contain phototrophic organisms. The novel population forms an outgroup to the currently known FAPs within Order Chloroflexales and sequences of nonphototrophic Chloroflexi (Supplementary Figure 11). Before this discovery, chlorophototrophy in Chloroflexi was thought to be restricted to the Chloroflexales, which seemed to have evolved from a chemoorganotrophic common ancestor of this group and the non-phototrophic organisms in Order Herpetosiphonales. The discovery of chlorophototrophy in another deeply rooted branch of Kingdom Chloroflexi suggests that it is plausible that chlorophototrophy was an ancestral trait in Kingdom Chloroflexi that was subsequently lost in some descendant lineages. Possible ancestral traits in Kingdom Chloroflexi can be inferred from properties shared between the newly discovered Anaerolineae-like chlorophototroph and members of Chloroflexales. All contain genes needed for BChl *a* synthesis and type-2 photosynthetic reaction centers similar to those of Proteobacteria, but some members (e. g., *Chloroflexus* spp.) also have chlorosomes, a trait shared with Chlorobiales and one member of the Acidobacteria (Bryant *et al.*, 2012). It is not yet known whether the newly discovered chlorophototroph has the capability of producing BChl *c* and chlorosomes.

Genes indicating chlorophototrophic metabolism were not found on metagenomic scaffolds of two other newly discovered populations corresponding to Clusters 7 and 8, yet these scaffolds provide an estimated depth-of-coverage that is greater than that of *Chloroflexus* spp. represented by Cluster 3 in which nine phototrophy genes were observed. Genes for oxidation of reduced inorganic compounds were not observed, but these organisms apparently possess genes that encode enzymes involved in aerobic respiratory metabolism. One of these populations has the genes necessary for oxidation of glycolate and acetate, which are known to be produced and excreted by mat cyanobacteria and can be metabolized by other community members (Bateson and Ward, 1988; Nold and Ward, 1996; van der Meer *et al.*, 2005).

Pyrosequencing of cDNA from reverse-transcribed rRNA (Liu *et al.*, 2011b) showed that most rRNAs (~88%) dominating the upper green layer of the mat are derived from the same eight phylogenetic groups identified in the metagenome. The linkage of these rRNA sequences to shotgun metagenomic data have allowed us to assign functional roles for the predominant populations in the upper green layer of the Octopus and Mushroom Springs.

Description of Functional Guilds

Our analysis of the attributes of eight distinct assembly clusters (Table 3.4) provided evidence for the functions of major taxa, which we assigned to functional guilds according to their partitioning of environmental resources and conditions (Table 3.4). Cyanobacteria conduct oxygenic photosynthesis using the visible light spectrum, but other chlorophotrophic groups have the potential to harvest near infrared light. For instance, *Roseiflexus* spp. have the genes to produce BChl *a* harvesting 850-900 nm light. Three phylogenetic groups share the potential to produce BChl *c*. The Chlorobiales-like population contained genes essential for producing chlorosomes, which are also known to occur in *Chloroflexus* spp. and *Ca. C. thermophilum* isolates (Pierson and Castenholz, 1974a; Bryant *et al.*, 2007). These observations suggest that these three populations harvest primarily 700-750 nm light. Further niche partitioning undoubtedly explains the co-existence of different types of phototrophs using similar light wavelengths. One possibility is that different members of a functional guild differ in terms of carbon metabolism. For instance, among phototrophs using 700-750 nm light, native *Chloroflexus* spp. have the genetic potential for carbon fixation via the 3-hydroxypropionate pathway (Klatt *et al.*, 2007; Bryant *et al.*, 2012), but most *Cfx. aurantiacus* strains achieve higher growth rates in culture with photoheterotrophic metabolism (Pierson and Castenholz, 1974a; Madigan *et al.*, 1974) and

may conduct mixotrophic rather than autotrophic carbon metabolism in situ (Bryant *et al.*, 2012). However, *Ca. C. thermophilum* and the Chlorobiales population do not appear capable of autotrophic metabolism and are more likely heterotrophic. Another possible explanation for niche differentiation among these phototrophs is temperature adaptation. *Chloroflexus* spp. sequences were relatively more abundant in the 65°C metagenome, whereas *Ca. C. thermophilum* and Chlorobiales-like organisms were relatively more abundant in the 60°C metagenome (Table 3.2 and Supplementary Figure 6). At this time, *Ca. C. thermophilum* spp. and Chlorobiales-like organisms cannot be placed into separate functional guilds on the basis of differences in light harvesting, carbon metabolism or temperature preference. Differences in electron donor utilization could also be involved in niche partitioning, but deeper metagenomic sequencing, coupled with genetic and physiological studies, will be required to test this hypothesis. Differences in the timing of gene expression provide additional clues to explain the co-existence of populations that cannot be separated based on putative physiological differences inferred from gene content (Liu *et al.*, 2011b).

Diversity Within Scaffold Clusters

The taxonomic resolution of the phylogenetic groups defined by scaffold clustering in this study is approximately at the level of named genera. However, population genetics studies of uncultivated *Synechococcus* spp. from Octopus and Mushroom Spring have indicated the presence of numerous, genetically distinct ecotypes within the A-like and B'-like lineages that occupy discrete positions along environmental gradients (e. g., light and temperature (Melendrez *et al.*, 2011; Becraft *et al.*, 2011) and exhibit complex metabolic regulation over the diel cycle (Liu *et al.*, 2011b). Consistent with these findings, the genetic and functional differences in metagenomic *Synechococcus* spp. populations in comparison to the two cyanobacterial isolate genomes revealed

ecological heterogeneity within closely related phylogenetic groups. The discovery of ferrous iron transporter homologs in *Synechococcus* sp. A-like populations (this study), and in B'-like populations (Bhaya *et al.*, 2007), as well as the presence of these genes in the *Roseiflexus* sp. strain RS1 genome (van der Meer *et al.*, 2010), suggests that the ability to utilize Fe^{2+} might be a common adaptation among mat community members. The presence of genes for an alternative pathway for urea metabolism in the metagenomic A-like *Synechococcus* provides additional evidence that urea may be an important nitrogen-containing nutrient in these mats (Bhaya *et al.*, 2007).

Overall, there were few examples of functional genes present in native populations but absent in the genomes of sequenced isolates; however, it is clear that ecological diversification also occurs through mechanisms other than differences in gene content. For example, adaptations to temperature (Miller and Castenholz, 2000; Allewalt *et al.*, 2006) may be based on adaptive nucleotide substitutions (Miller, 2003; Ward *et al.*, 2012b). The metagenomic diversity with respect to the *Roseiflexus* sp. RS1 genome likely encompasses multiple ecologically distinct *Roseiflexus* spp., such as those exhibiting different distributions along the flow path in these mats (Ferris and Ward, 1997; Nübel *et al.*, 2002; Ward *et al.*, 2006).

Insights Into Genome Evolution

Comparisons of metagenomic sequences and genomes of representative mat isolates also yielded insights into genome diversity among closely related populations. Cyanobacterial genomes are less syntenous with each other at a given degree of sequence divergence compared to other taxonomic groups (Rocha, 2006; Frangeul *et al.*, 2008). The time of divergence between *Synechococcus* spp. strains A and B' has been long enough to exhibit nearly a complete lack of synteny (Bhaya *et al.*, 2007), yet it

is apparent that *Synechococcus* spp. more closely related to either *Synechococcus* sp. strain A or B' are more syntenous to their respective closest relative. Both synteny and the number of disjointly recruited metagenomic clones, which might document past recombination events, decrease as the genetic relatedness between two organisms decreases. The latter trend is consistent with empirical findings in *Bacillus* and *Streptococcus* spp., which demonstrated that recombination rates declined as the genetic distances between organisms increased (Roberts and Cohan, 1993; Majewski *et al.*, 2000). Our results suggested that homologous recombination between populations as divergent as *Synechococcus* spp. strains A and B' has generally been uncommon ($\sim 5\%$ of the total number of sequences recruited by either *Synechococcus* sp. strain A or B'). Comparative genomic studies have shown that, while gene transfer among cyanobacteria is evident (Zhaxybayeva *et al.*, 2006), these events have been infrequent and do not obscure inferences about phylogenetic relationships in this kingdom (Kettler *et al.*, 2007; Swingley *et al.*, 2008; Zhaxybayeva *et al.*, 2009; Melendrez *et al.*, 2011).

Conclusion

This metagenomic study revealed that the chlorophototrophic communities inhabiting the effluent channels of Octopus and Mushroom Springs were more phylogenetically and physiologically diverse than was known on the basis of light microscopy, traditional cultivation methods and previous 16S rRNA surveys. The combination of depth of coverage and limited diversity enabled metagenomic assemblies leading to (i) the confirmation of a novel chlorophototrophic member of Chlorobiales in these mats and (ii) the discovery of several novel populations, including a chlorophototroph in a novel lineage of Chloroflexi and two types of putatively chemoorganotrophic com-

munity members more representative of native populations than currently cultivated chemoorganotrophic isolates. This effectively doubled the number of predominant populations known to inhabit the mat. Deeper coverage metagenomes are in production that will further enhance our understanding of the physiological potential of the dominant members of this microbial mat community. The availability of genomes of isolates closely related to native populations enabled (i) discovery of functions not represented by the isolates and (ii) evidence that breakdown of synteny and the exchange of genetic information are functions of how much populations have diverged. Finally, the results of these analyses provide the foundation for interpreting the metatranscriptome of Mushroom Spring mat over a portion of the diel cycle in an accompanying study (Liu *et al.*, 2011b).

Acknowledgements

This research was supported by the National Science Foundation Frontiers in Integrative Biology Research Program (EF-0328698) and IGERT Program in Geobiological Systems (DGE 0654336), the National Aeronautics and Space Administration Exobiology Program (NAG5-8824, -8807 and NX09AM87G) and the U.S. Department of Energy (DOE), Office of Biological and Environmental Research (BER), as part of BERs Genomic Science Program 395 (GSP) [This contribution originates from the GSP Foundational Scientific Focus Area (FSFA) at the Pacific Northwest National Laboratory (PNNL); contract #112443]. We appreciate the support and assistance of National Park Service personnel at Yellowstone National Park. We thank Marcus B. Jones at the J. Craig Venter Institute for his help using the PCT Barocycler for cell lysis. D. A. B. additionally and gratefully acknowledges support from the National Science Foundation (MCB-0523100), Dept. of Energy (DE-FG02-94ER20137), and

the Joint Genome Institute for support in obtaining genomic sequences mentioned herein.

CHAPTER 4

COMMUNITY STRUCTURE AND FUNCTION OF HIGH-TEMPERATURE
PHOTOTROPHIC MICROBIAL MATS INHABITING DIVERSE
GEOTHERMAL ENVIRONMENTSContribution of Authors and Co-Authors

Manuscript in Chapter 4

Author: Christian G. Klatt

Contributions: Designed the study, conducted the experiments, collected and analyzed data and wrote the manuscript.

Co-author: William P. Inskeep

Contributions: Obtained funding, assisted with experimental design, assisted in conducting field experiments and edited the manuscript at all stages.

Co-author: Zackary Jay

Contributions: Assisted in conducting field experiments, collected and analyzed data, and edited the manuscript.

Co-author: Douglas B. Rusch

Contributions: Obtained funding, collected and analyzed data, and edited the manuscript.

Co-author: Susannah G. Tringe

Contributions: Obtained funding, collected and analyzed data, and edited the manuscript.

Co-author: Mary N. Parenteau

Contributions: Assisted in conducting field experiments, collected and analyzed data,

and edited the manuscript.

Co-author: David M. Ward

Contributions: Obtained funding, assisted in conducting field experiments and edited the manuscript at all stages.

Co-author: Sarah M. Boomer

Contributions: Obtained funding, assisted in conducting field experiments and edited the manuscript.

Co-author: Donald A. Bryant

Contributions: Obtained funding and edited the manuscript.

Co-author: Scott R. Miller

Contributions: Obtained funding, assisted in conducting field experiments and edited the manuscript.

Manuscript Information Page

Christian G. Klatt, William P. Inskeep, Zackary Jay, Douglas B. Rusch, Susannah G. Tringe, Mary N. Parenteau, David M. Ward, Sarah M. Boomer, Donald A. Bryant, and Scott R. Miller

Journal Name: Geobiology

Status of Manuscript:

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

Published by Blackwell Publishing.

Abstract

Six phototrophic microbial mat communities inhabiting geothermal springs in Yellowstone National Park were studied with metagenomic sequencing, which provided new insights into the structure and functional gene content of these microbial communities within a range of different geochemical contexts. These communities were sampled from the sulfidic Bath Lake Vista Annex near Mammoth Springs (BLVA_5 and BLVA_20), a high-iron anoxic spring source at Chocolate Pots (CP_7), and three neutral-alkaline springs in the Lower and Middle Geyser Basins (White Creek, WC_6; Mushroom Spring, MS_15; Fairy Geyser splash mat FG_16). Ribosomal rRNA clone libraries were constructed in parallel with random shotgun metagenomic Sanger sequencing from these six communities, which averaged ~ 53 Mbp of metagenomic sequence data per community. Assembled scaffolds that were subjected to oligonucleotide frequency-based clustering revealed the dominant community members represented by these metagenomes. Novel chlorophototrophic bacteria of Order Chlorobiales were observed at CP_7, and cyanobacterial populations of *Synechococcus* and *Mastigocladus* spp. were observed in CP_7 and WC_6. Sequences originating from organisms in Kingdom Chloroflexi were found in all six phototrophic mats, and genes predicted to function in bacteriochlorophyll biosynthesis and the 3-hydroxypropionate autotrophic pathway showed low sequence similarity to those from any characterized chlorophototrophs. Metagenomic sequencing and assembly of these microbial communities has provided links between phylogenetically and functionally informative genes, such that comparisons could be made of the functional attributes of major populations present among these springs. The geochemical limitations placed upon community structure are predicted to impact which functional groups are dominant

in a given community, which correspondingly limit the possible interactions among community members and may in turn impact rates of biogeochemical cycling.

Introduction

Although the cultivation and subsequent genome sequencing of relevant microorganisms from the environment provides reference information for the physiological capabilities of individual community members, many naturally occurring microorganisms have eluded isolation, due in part to a poor understanding of the chemical, physical and biotic factors defining their realized niches (Rappé and Giovannoni, 2003). Moreover, much of the sequence diversity revealed by amplification of specific gene targets (e.g. 16S rRNA) is susceptible to biases inherent in primer-design and PCR protocols. The random shotgun sequencing of DNA extracted from entire microbial communities avoids the biases inherent in PCR-based sequencing while simultaneously sampling both phylogenetically and functionally informative genes. This linkage between phylogeny and function enables the discovery of novel organisms and allows for predictions to be made regarding their functional attributes. For example, three phylogenetically distinct chlorophototrophs were discovered in prior metagenome analysis of phototrophic mats in YNP (Chapter 3). Two of these organisms belong to the Kingdoms Chlorobi and Chloroflexi, but lie outside their respective monophyletic clades of known phototrophic organisms within these lineages and phototrophic functions could not have been inferred from rRNA analysis (Chapter 3). This is especially true for the third novel phototroph recently discovered from metagenomics; *Candidatus* Chloracidobacterium thermophilum represents the only known occurrence of chlorophototrophy in the entire Kingdom Acidobacteria (Bryant *et al.*, 2012). Consequently, metagenome sequencing and subsequent bioin-

formatic analyses provides an opportunity to integrate geochemical and physiological processes in conceptual and computational models of microbial interaction and function (Taffs *et al.*, 2009), as well as to postulate detailed biochemical linkages among individual community members.

High-temperature phototrophic microbial mats have served as models for studying microbial community structure and function including investigations of microbial community composition (Miller *et al.*, 2009), the ecophysiology of novel isolates (Pierson and Castenholz, 1974a; Miller and Castenholz, 2000; Pierson and Parenteau, 2000; Allewalt *et al.*, 2006; Bryant *et al.*, 2007; Parenteau and Cady, 2010; van der Meer *et al.*, 2010), comparative genomics, metagenomics and metatranscriptomics (Bhaya *et al.*, 2007; Klatt *et al.*, 2007, 2011; Liu *et al.*, 2011b), community network modeling (Taffs *et al.*, 2009), natural phage-host interactions (Heidelberg *et al.*, 2009), and theoretical mechanisms of evolution (Ward *et al.*, 2008). The high temperature and relative geochemical stability of geothermal phototrophic mats provide the opportunity for understanding environmental factors controlling community composition (Brock, 1978; Ward *et al.*, 1989b; Ward and Castenholz, 2000; Ward *et al.*, 2012b). Prior investigations have revealed that temperature, pH and sulfide are among the most important environmental variables dictating differences in phototrophic mat community structure (Castenholz, 1976, 1977; Ward *et al.*, 1992; Castenholz and Pierson, 1995; Madigan *et al.*, 2005; Cox *et al.*, 2011). The presence of sulfide is an important factor controlling phototroph distribution and was used in the current study to separate communities dominated by anoxygenic phototrophs from those dominated by oxygenic phototrophs (i.e., cyanobacteria). Oxygenic and/or anoxygenic photoautotrophs are generally the predominant primary producers in geothermal mats ranging from ~ 50 - 72°C and acidic to alkaline pH (5 - 9), and support a diverse array of heterotrophic, fermentative, sulfate-respiring, and/or methanogenic organisms whose physiological

attributes are critical for understanding community function (Nold and Ward, 1996; Jackson *et al.*, 1973; Ward *et al.*, 1998; Brock and Freeze, 1969; Zeikus and Wolfe, 1972; Zeikus *et al.*, 1979, 1983; Henry *et al.*, 1994; Taffs *et al.*, 2009). Cyanobacteria are limited in their habitat range in that they are not generally found in acidic or sulfidic environments (Castenholz, 1976, 1977). However, filamentous anoxygenic phototrophs (FAPs) of the Kingdom Chloroflexi exhibit a wider habitat range than other phototrophic bacteria, and closely related Chloroflexi (>97% identity of 16S rRNA gene) with different phenotypes have been cultured from geothermal environments. For example, FAPs isolated from a high-sulfide (>100 μM) spring devoid of cyanobacteria (*Chloroflexus* sp. GCF strains) were found to prefer photoautotrophic growth using sulfide as an electron donor (Giovannoni *et al.*, 1987). In contrast, most other cultured *Chloroflexus* spp. from low-sulfide environments prefer to grow photoheterotrophically in culture (Pierson and Castenholz, 1974a; Madigan *et al.*, 1974) utilizing organic compounds produced by co-inhabiting cyanobacteria. Consequently, more detailed functional information is necessary to understand the role of different Chloroflexi populations observed in situ.

The overall goal of this study was to investigate the underlying environmental factors and potential physiological adaptations important in defining the microbial community structure and function of different types of phototrophic mats in high-temperature systems common in YNP. The specific objectives were to *i*) utilize metagenome sequencing and bioinformatic analyses to determine the community composition of high-temperature phototrophic mats in YNP, *ii*) identify key metabolic attributes of the major phototrophic organisms present in these communities, and *iii*) evaluate the predominant environmental and/or geochemical attributes that contribute to niche differentiation in thermophilic phototrophic mats. The phototrophic

communities sampled in the current study were chosen in part to capture several of the predominant mat types distributed across the YNP geothermal ecosystem.

Results

Geochemical and Physical Context

The predominant differences among the six phototrophic microbial mat communities in this study include geochemical and environmental characteristics such as pH, dissolved sulfide, temperature, and the specific mat-layer sampled (Table 4.1). For example, temperature varies across these six sites (e.g., 40 – 60 °C), and four of the geothermal sites contain no measurable dissolved sulfide (DS), while two samples from Bath Lake Vista Annex (BLVA_5 and BLVA_20, exhibiting different microbial communities as discussed below) are from sub-oxic sulfidic environments (DS $\sim$ 117 μ M). Although the dissolved O₂ content near the source (and sample location) of Chocolate Pots Spring (CP_7) was below detection (<1 μ M), this spring does not contain measurable DS (Table 4.1), but contains high concentrations of ferrous Fe ($\sim$ 76 μ M) that result in the precipitation of Fe(III)-oxides upon discharge and reaction with O₂ (Figure 4.1). The phototrophic mat obtained from White Creek (WC_6) exists within an oxygenated alkaline-siliceous geothermal drainage channel that lacks detectable DS. The site was included in the study to target cyanobacteria related to *Mastidocladus*-like populations that have been the focus of prior work at this location (Miller *et al.*, 2006, 2007, 2009). Samples from Mushroom Spring (MS_15) and Fairy Geyser (FG_16) were obtained from within laminated phototrophic mats after removal of the top layer (see Methods). Dissection of these mats was performed to focus purposely on filamentous anoxygenic phototrophs (FAPs) known to increase in abundance at depths within the mat and below surface layers that are dominated

Table 4.1: Sample Location, Aqueous Geochemical Parameters and Physical Context of Six, High-temperature Phototrophic Microbial Communities in Yellowstone National Park (YNP).

Location	°C			μM							mM				Coordinates	Date of Collection
	T	pH	DS	DO	As	Fe	Mn	Mg	NH ₄ ⁺	NO ₃ ⁻	Na ⁺	Ca ⁺	Cl ⁻	SO ₄ ²⁻		
Bath Lake Vista Annex-Green (BLVA.5)	57	6.2	117	<3	24	0.7	0.2	2132	40	4.8	3.9	8.8	4.4	5.6	44.96505 -110.71173	September 28, 2007
Bath Lake Vista Annex-Purple (BLVA.20)	54	6.2	117	<3	23	0.7	b.d.	2625	40	5.6	5.5	9.8	5.7	7.3	44.96505 -110.71173	May 14, 2008
White Creek (WC.6)	52	8.2	<3	188	5	1.7	4.7	13.8	1.9	2.1	3.6	0.4	1.8	0.23	44.53150 -110.79767	August 24, 2007
Chocolate Pots (CP.7)	52	6.2	<3	<3	9	75.5	24	58.7	4.2	2.2	4.1	0.5	0.89	0.23	44.71008 -110.74134	October 25, 2007
Mushroom Spring (MS.15)	60	8.2	<3	141	26	<1	0.1	0.8	4.4	6.6	12.6	0.02	7.3	0.18	44.53869 -110.79797	December 15, 2007
Fairy Geyser (FG.16)	~40	9.1	<3	31	13	<1	b.d.	3.5	1.3	6.7	9.4	0.02	5.2	0.18	44.54217 -110.86133	October 25, 2007
Correlation ² (r ²)	0.887 *			0.987 ***	0.7194 **		0.837 **	0.9728 *	0.9874 **	0.928 *				0.964 *		

¹ DS=dissolved sulfides; DO= dissolved oxygen; b.d.=below detection level.

² Correlation significance values: * = (p<0.05), ** = (p<0.01), *** = (p<0.001).

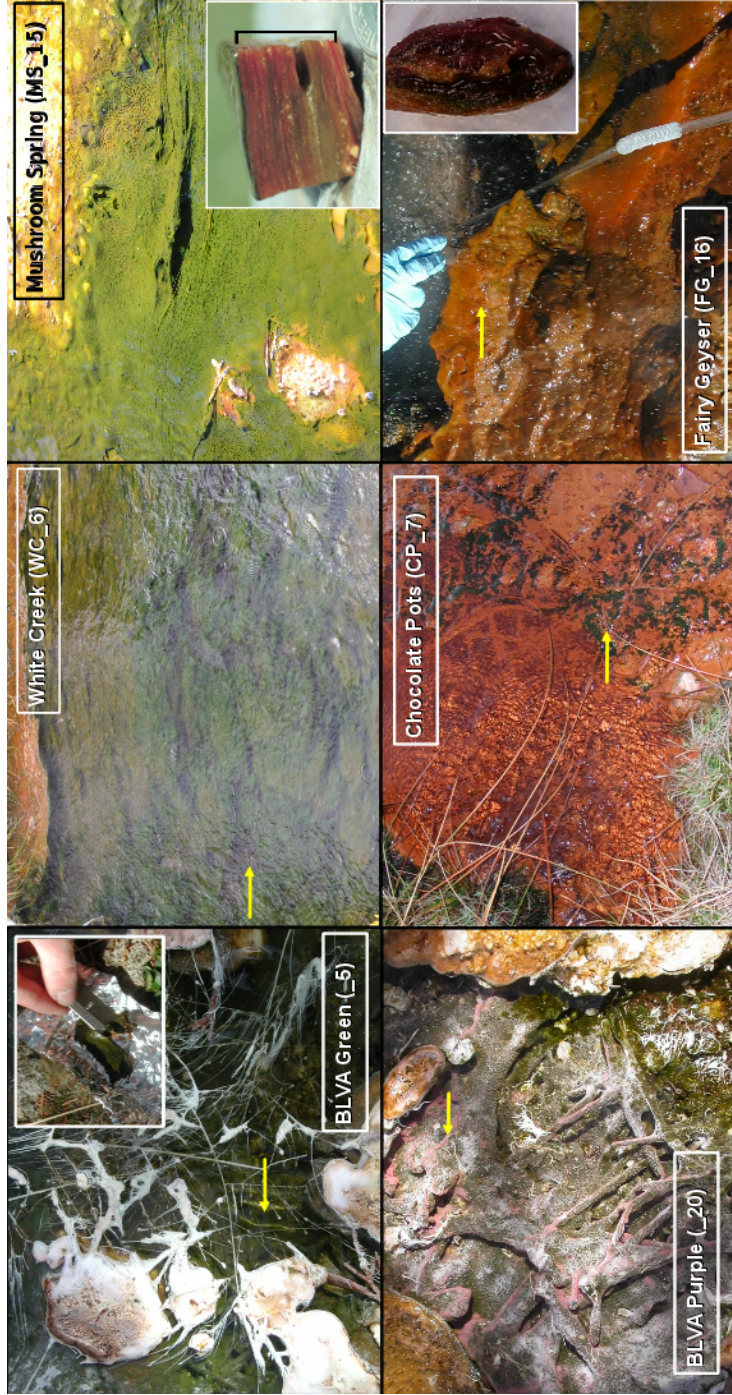


Figure 4.1: Site Photographs of the Microbial Mats Selected for Metagenome Sequencing in the Current Study. The sites cover a range in geochemical conditions including oxygenic phototrophic communities at White Creek (WC_6) and Chocolate Pots (CP_7), deeper-mat positions at Mushroom Spring (MS_15) and Fairy Geyser (FG_16) (also oxygenic systems), as well as anoxygenic phototrophic communities at Bath Lake Vista Annex (BLVA), sampled at two different time points to compare green *Chloroflexus* mats in the absence (BLVA_5) and presence (BLVA_20) of purple-bacteria (arrows indicate approximate sample locations and types; inset at BLVA_5 shows mat dissection at sampling). Insets for (MS_15) and (FG_16) illustrate subsurface mats of the type that were sampled from these springs.

by cyanobacteria. These non-sulfidic environments have been shown in prior work to contain greater numbers of various members of the Chloroflexi relative to communities found in top mat-layers (Nübel *et al.*, 2002; Boomer *et al.*, 2002). The phototrophic mats at FG_16 are referred to as 'splash-mats' due to the fact that these communities receive constant inputs of geothermal water emanating from the main source pool (85-88 °C) (Figure 4.1). The 'splash-mats' surrounding FG_16 are reasonably thick (e.g., 3 - 5 cm), but the target sample discussed here is a 2 - 4 mm 'red-layer' usually found at a temperature range of 35-50 °C and pH approaching 9 (Boomer *et al.*, 2000, 2002). Although the two subsurface mat samples (MS_15 and FG_16) are less oxic than their respective near-surface layers, no significant DS is present in the bulk aqueous phase (Table 4.1).

Analysis of Metagenome Sequences

Individual sequences (average length ~800 bp) were analyzed using two complementary approaches: an alignment-based comparison to reference databases, and an evaluation of the guanine and cysteine content (% G+C) of each sequence read. Comparison of all sequences to the NCBI nr database (blastx) was accomplished using MEGAN (Huson *et al.* 2008). The most highly represented phyla across all sites included the Chloroflexi (28%), Cyanobacteria (12%), Proteobacteria (8%) and Cytophaga/Flavobacteria/Bacteroidetes (CFB, 6%). Many sequences (27%) did not match those available in NCBI ('no hits'), indicating that some members of these communities are not represented in current genome databases.

Taxonomic assignment of individual sequences was combined with % G+C distribution to obtain a profile of community composition (Figure 4.2). Each site contained populations similar to *Chloroflexus* and/or *Roseiflexus* spp, with average G+C contents of 55 and 61%, respectively. The two sulfidic samples (BLVA_5 and BLVA_20)

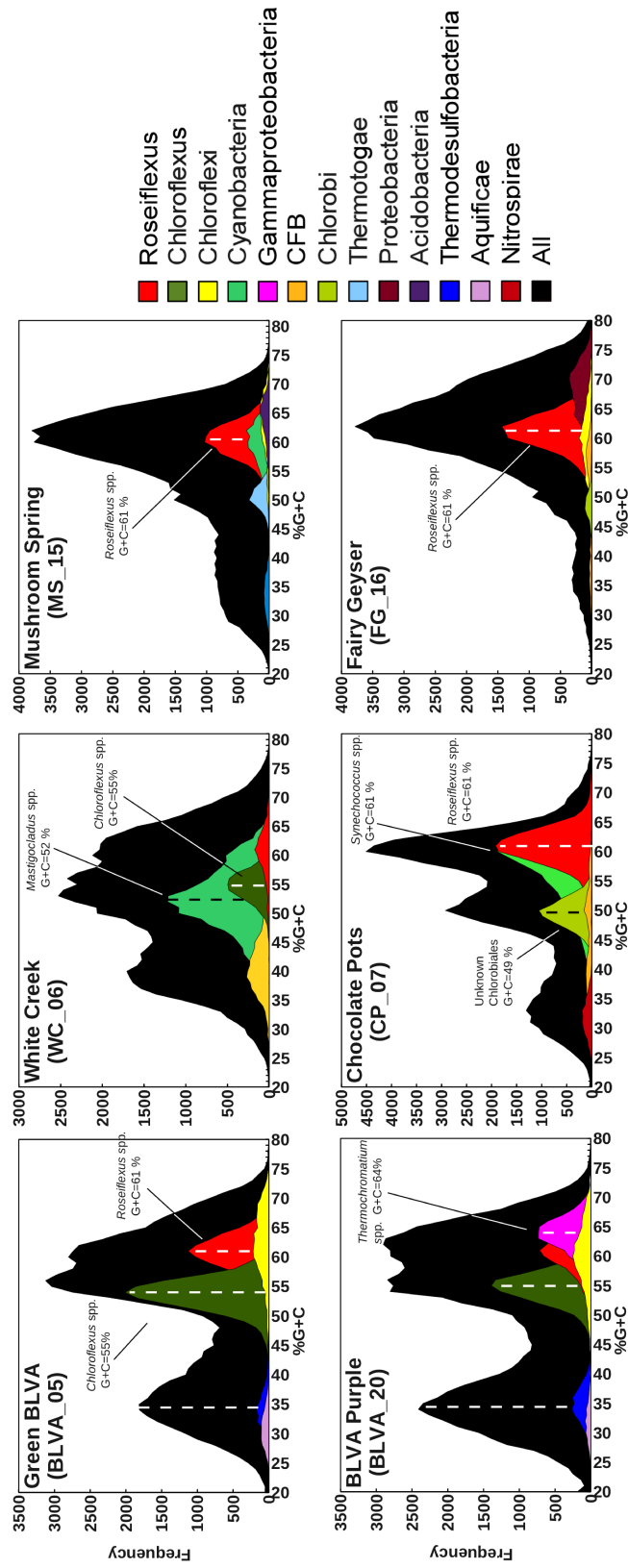


Figure 4.2: Percent G+C Content of Individual Metagenome Sequences. Subsets of sequences from each community that exhibited taxonomic calls above thresholds (determined by MEGAN-BLASTX) are indicated by the color key.

clearly show contributions from both the *Chloroflexus*-like (average=55%) and *Roseiflexus*-like (average=61%) populations (Figure 4.2). The phototrophic community from White Creek (WC_6) also contains significant contributions from *Chloroflexus*-like organisms, while CP_7, MS_15 and FG_16 are more enriched in *Roseiflexus*-like sequences (Figure 4.2). All sites contain a significant number of sequences contributed from novel Chloroflexi populations that have not been adequately characterized, and for which appropriate reference organisms have not been cultivated or sequenced.

The phototrophic mat communities from WC_6 and CP_7 contain a significant fraction of sequences (23 and 25%, respectively) corresponding to cyanobacteria. Both sites contain expected targets related to *Synechococcus* spp. strains A and B' that exhibit a mean G+C content of 60%. The WC_6 community also contains a large proportion (73%) of cyanobacteria that could not be classified beyond the kingdom-level. These sequences exhibit a large range in G+C content (40 to 65%, with a major peak at 51.5%). *Mastigocladus*-like organisms (Order Stigonematales) have been shown to be important community members at the WC_6 site (Miller *et al.*, 2009), but no reference genomes are currently available from this group of cyanobacteria.

The G+C frequency plots also reveal major contributions from organisms within the Chlorobi (CP_7 and to a lesser extent FG_16), *Thermotoga* (MS_15), as well as the targeted population of γ -proteobacteria (purple-sulfur bacteria) in BLVA_20 with an average G+C content of 64%. Moreover, all sites contained sequences with G+C contents ranging from 20-40%; however, the lack of reference genomes precludes phylogenetic identity beyond the level of Bacteria.

Phylogenetic Analysis of Metagenome Assemblies

The assembly of individual sequences into large contigs and scaffolds provides a powerful tool for linking functional attributes and gene assignment with specific

phylotypes. Sequence data from each site was assembled independently (both Celera and PGA assemblies are available at the Joint Genome Institute's IMG/M website, <http://img.jgi.doe.gov/cgi-bin/m/main.cgi>), resulting in an average scaffold size of 2,330 bp, ranging from small contigs of 1 kb to large scaffolds approaching 126 kb. The largest assemblies were obtained from CP_7, and represented 42% of the larger scaffolds (≥ 10 kb) obtained across all six sites. Long assemblies were also obtained from the anoxygenic mats at BLVA sampled eight months apart (BLVA_5, BLVA_20). Sequences from sub-surface mat communities (MS_15 and FG_16) did not result in long assemblies, and only two scaffolds ≥ 10 kb were obtained from each site. The difficulty generating longer assemblies from these lower mat layers reflects the greater diversity of operational taxonomic units (demarcated at 1% difference in nucleotide identity at the 16S rRNA locus) observed relative to other samples; both MS_15 and FG_16 exhibited greater species richness estimates from the PCR-based 16S rRNA surveys (see Supplementary Table 1 in Appendix C). Sequence assemblies were examined using principle components analysis of nucleotide word frequencies (NWF PCA) in conjunction with a taxonomic classification algorithm of average scaffold identity (APIS; Rusch *et al.* 2007), providing a mechanism for visualizing the dominant community members inferred from genome coverage and subsequent assembly. For example, NWF PCA plots of the sulfidic system at BLVA sampled 8 months apart reveal the major differences in community composition associated with a visible bloom of purple-sulfur bacteria in BLVA_20 (Figures 4.1 and 4.3). The major change in community composition between the two samples was the appearance of the *Chromatiaceae*-like population in BLVA_20, which corresponded with a decrease in *Roseiflexus*-like sequences (Figure 4.3). Both samples reveal a dominant *Chloroflexus*-like population that corresponds to the G+C peak at 55% (Figure 4.2, and was an expected target population in these sulfidic habitats at 56 °C. Similar NWF

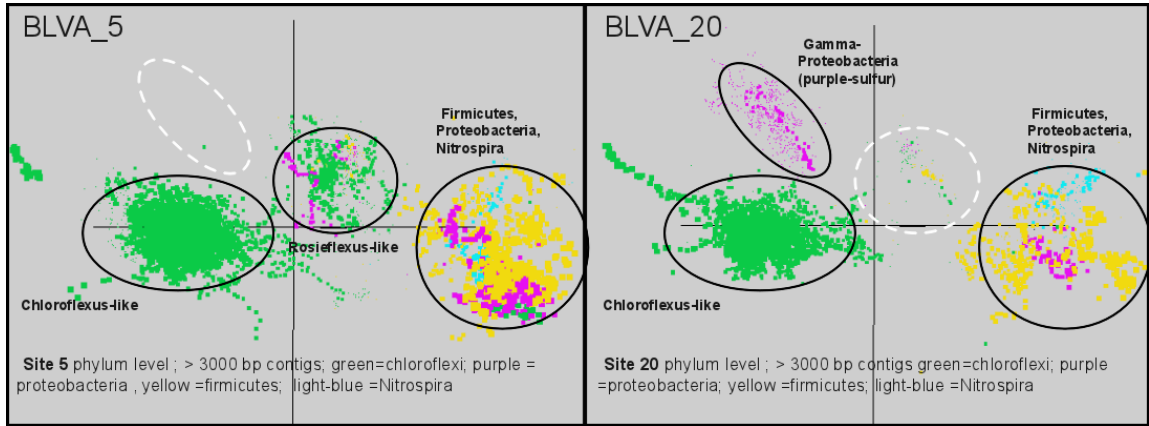


Figure 4.3: Oligonucleotide Frequency Principal Components Ordination of Assemblies from BLVA_5 and BLVA_20. BLVA_20 was sampled to capture a bloom of purple-sulfur bacteria shown in prior work to be related to *Thermochromatium tepidum*. Both sites contained scaffolds from dominant populations of *Chloroflexus* spp., and BLVA_5 contained scaffolds corresponding to *Roseiflexus* spp. BLVA_20 contained numerous scaffolds from purple-sulfur bacteria (γ -proteobacteria, family *Chromatiaceae*, average G+C ~6%).

PCA analyses of assemblies from CP_7 revealed three major populations (*Roseiflexus*, *Synecochoccus*, and Chlorobiales), as well as sub-dominant community members distantly related to members of the phyla Firmicutes, Bacteroidetes and Spirochaetes (Supplementary Figure 3 in Appendix C).

A Monte-Carlo approach was also used to compare normalized oligonucleotide frequencies from all sites, which clustered scaffolds that originated from phylogenetically related organisms. A minimum scaffold length of 10 kbp was used to focus the analysis on dominant assemblies with maximal phylogenetic signal; however, smaller scaffolds from sub-surface mat communities (MS_15 and FG_16) were not well-represented in this analysis. Twelve scaffold clusters corresponding to the consensus of 100 replicated k-means groupings were observed, and these clusters were found to correspond to dominant community members when examined further (Table 4.2, Figure 4.4).

Table 4.2: Properties of Metagenomic Scaffold Clusters as Demarcated with Oligonucleotide Composition.

Metagenome Scaffold Cluster	Taxonomic Affiliation	Sites	Number of scaffolds	Median scaffold size (Kbp)	Average G+C(%)	Total amount of assembled sequence (Kbp)	Estimated Depth of Coverage (mean read depth)
1	<i>Roseiflexus</i> spp.	BLVA_5 CP_7 MS_15 FG_16	112	12.5	60.0 $\pm$ 1.2	1554	2.6x $\pm$ 0.4
2	<i>Chloroflexus</i> spp.	BLVA_5 WC_6 CP_7 BLVA_20	211	13.5	54.3 $\pm$ 1.2	3205	2.9x $\pm$ 0.7
3	Order Chlorobiales	CP_7	73	14.8	49.5 $\pm$ 0.8	1128	2.7x $\pm$ 0.5
4	<i>Thermochromatium</i> spp.	BLVA_20	29	12.5	63.0 $\pm$ 1.3	374	2.1x $\pm$ 0.4
5	<i>Synechococcus</i> spp.	WC_6 CP_7	78	26.2	58.7 $\pm$ 1.1	2589	4.0x $\pm$ 0.7
6	Cyanobacteria	WC_6 CP_7	26	11.7	49.8 $\pm$ 1.2	319	2.4x $\pm$ 0.5
7	Cytophaga-Flavobacterium -Bacteroidetes (CFB) group	WC_6	30	11.1	37.7 $\pm$ 0.9	368	2.4x $\pm$ 0.4
8	Unknown	BLVA_5 MS_15 BLVA_20	37	10.6	63.9 $\pm$ 2.3	441	2.5x $\pm$ 0.5
9	Unknown	BLVA_5 CP_7 BLVA_20	47	14.2	36.0 $\pm$ 1.5	790	2.7x $\pm$ 0.4
10	Unknown	CP_7	21	11.8	30.5 $\pm$ 1.4	249	2.3x $\pm$ 0.4
11	Unknown	BLVA_5 BLVA_20	11	12.7	29.0 $\pm$ 1.4	162	2.6x $\pm$ 0.6
12	Unknown	BLVA_5 CP_7 BLVA_20	6	9.21	32.6 $\pm$ 1.5	70	2.0x $\pm$ 0.4
13	Unknown	CP_7	5	12.8	29.2 $\pm$ 1.5	67	2.3x $\pm$ 0.3

Clustering by oligonucleotide frequency afforded greater discrimination among organism groups that exhibit similar G+C composition. For example, *Roseiflexus*-like populations have similar G+C content (61%) to the dominant cyanobacterial population related to *Synechococcus* sp. strains A and B' (Supplementary Figure 3), yet the differences in sequence character of these different genera are clearly separated using oligonucleotide clustering analysis (Figure 4.4). Site-specific oligonucleotide clusters were observed in several cases corresponding to major populations identified using G+C% frequency analysis. A γ -proteobacterial cluster related to *Thermochromatium* spp. contains sequences solely from BLVA_20, and is consistent with the visual evidence of this targeted population when this site was sampled in May 2008 (Figure 4.1). Other site-specific clusters include the Chlorobiales-like population from CP_7

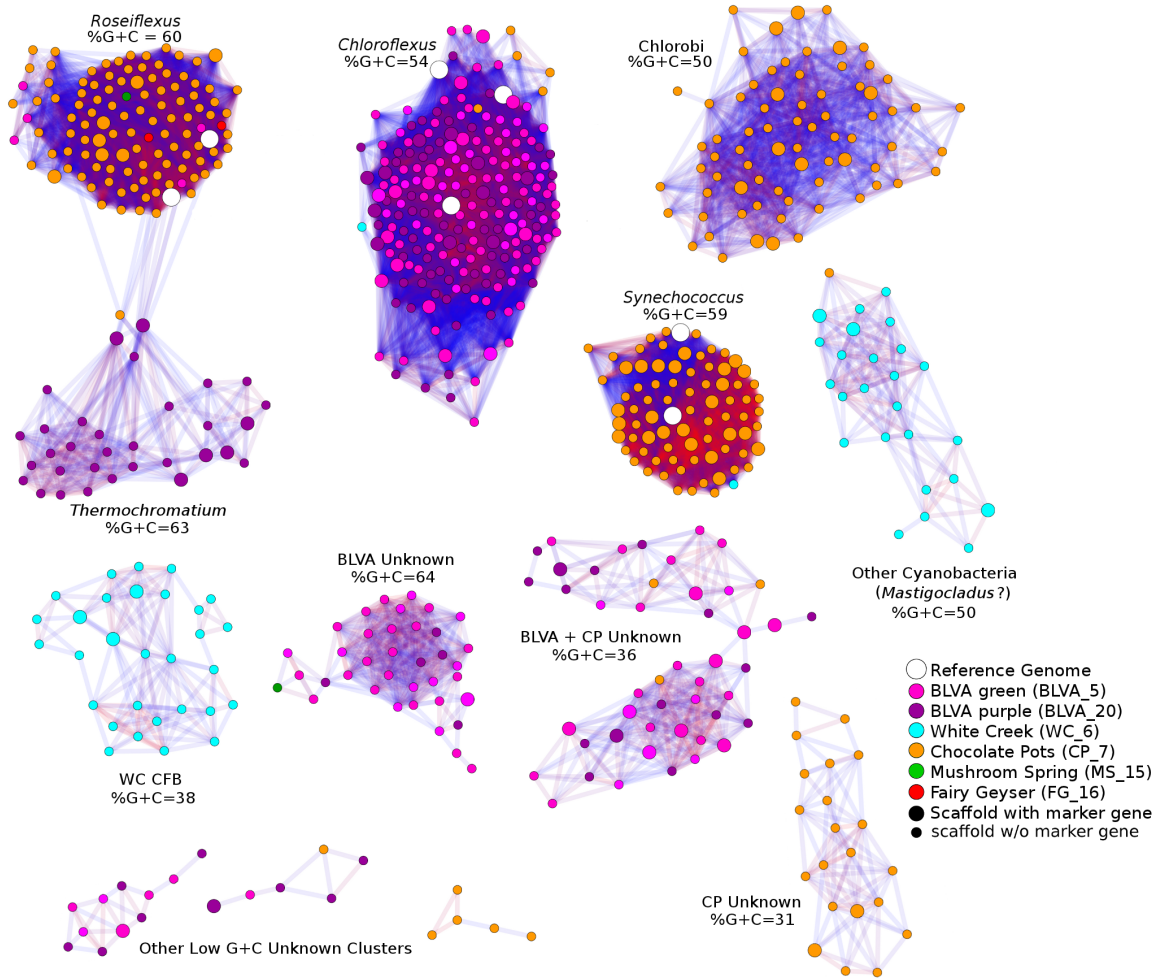


Figure 4.4: Scaffold Oligonucleotide Frequency Similarity Network. Oligonucleotide (tri-, tetra-, penta-, and hexa-nucleotide) counts were normalized to scaffold length and subject to k-means clustering ($k=8$, 100 trials). The scaffolds that group together in $\geq 90\%$ trials are shown, with lines connecting scaffolds ranging from blue (90%) to red (100%). Scaffolds that contain phylogenetic or functional marker genes are indicated by larger nodes, and colors correspond to the sampling site. CFB = Cytophaga-Flavobacterium-Bacteroidetes.

as well as smaller clusters from WC_6 corresponding to members of the Cytophaga-Flavobacterium-Bacteroidetes (CFB) group. The coverage of community members belonging to the Cytophaga-Flavobacterium-Bacteroidetes group was greater in the WC_6 community, resulting in larger assemblies (Figure 4.4), although relatives of

the Bacteroidetes were found to occupy all sites (Figure 4.5A). Three scaffold clusters with comparatively low G+C content (<40%) were observed, but both AMPHORA (based on phylogenetic analysis) and MEGAN (based on BLASTX alignments) were unable to classify the sequences in these groups, suggesting that they originate from organisms currently unrepresented in public databases. Phylogenetically informative single-copy genes were identified among the metagenome assemblies using AMPHORA (Wu and Eisen, 2008), and these sequences were examined further to predict the predominant taxa represented in the six metagenome samples. The distribution of dominant phylotypes predicted using AMPHORA (Figure 4.5A) was similar to that observed using the combined BLASTX and G+C analyses of individual sequences (Figure 4.2), and corresponded to the taxonomic distributions of PCR-based 16S rRNA gene libraries from these same sites (Figure 4.5B). Results from 16S rRNA gene surveys are consistent with results obtained using random shotgun sequencing, and support the major phylotypes observed using shotgun metagenome sequencing. All three approaches supported the observation that members of the Chloroflexi are ubiquitous across all sites (Figures 4.2 and 4.5). The distribution of sub-kingdom lineages of this group, with particular focus on the relative contribution of *Chloroflexus*-versus *Roseiflexus*-like organisms, as well as identification of novel lineages within this kingdom, are discussed below. Cyanobacteria were highly abundant in WC_6 and CP_7, and as expected, were not as important in sub-surface communities from MS_15 and FG_16 (Figure 4.5). A γ -proteobacterial population most closely related to the purple-sulfur bacterium *Thermochromatium tepidum* (Madigan, 1984; Imhoff *et al.*, 1998), was one of three dominant community members observed in BLVA_20. Other major contributions from anoxygenic phototrophs included populations of purple non-sulfur α -proteobacteria (Family *Hyphomicrobiaceae*) in FG_16, *Candidatus* Chloracidobacterium thermophilum (Bryant *et al.*, 2007) in WC_6, and novel bacteria

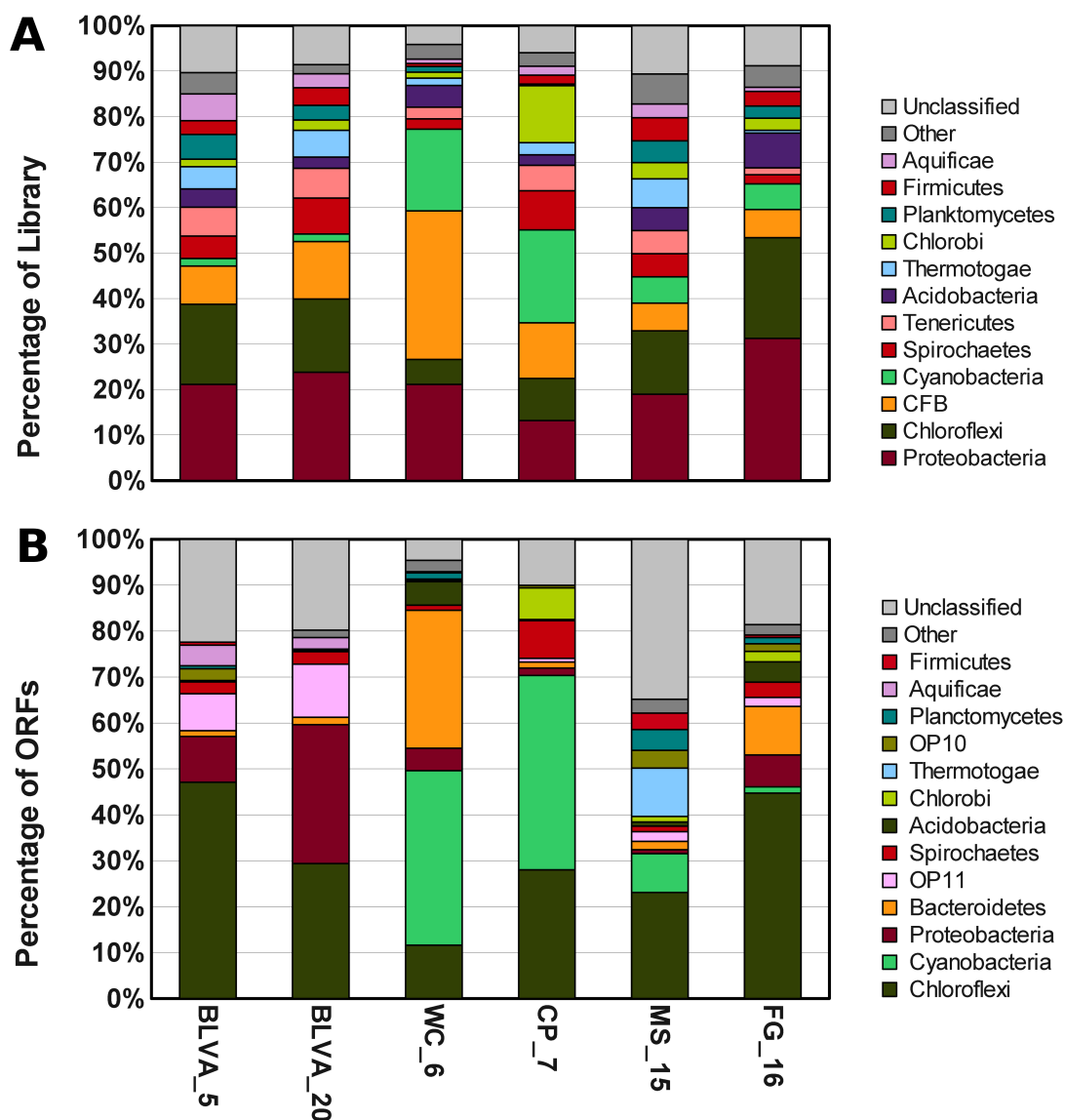


Figure 4.5: Comparison of the Distribution of Phylogenetic Marker Genes from Metagenomes and from 16S rRNA Clones. (A) displays the phylogenetic marker genes in the metagenome classified at the level of kingdom by AMPHORA. (B) 16S rRNA sequences from clone libraries were classified to kingdoms by the RDP Bayesian Classifier at a confidence threshold of 80%.

within the order Chlorobiales in MS_15, FG_16 and especially CP_7 (Figure 4.5B). The MS_15 community contains a *Thermotoga*-like population as well as several low %G+C organisms that have not yet been characterized. FG_16 contains a significant Chlorobiales population as well as a novel high %G+C proteobacterial population not seen in the other sites. The Chlorobiales population in CP_7 is distantly related to *Chloroherpeton thalassium*, (BLASTN alignments had 79% NT ID on average), and uncultivated *Candidatus Thermochlorobacter* spp. (average NT ID = 91%) observed in metagenomes from the phototrophic mat communities of Octopus Spring and Mushroom Spring (Chapter 3, Liu *et al.* 2011a). The possible roles of these novel populations are discussed below.

Chloroflexi Diversity and Distribution

The phylogenetic diversity of Chloroflexi 16S rRNA gene sequences among sites was compared to the abundance of Chloroflexi marker genes in the metagenome assemblies identified using AMPHORA (Figure 4.6A). The majority of Chloroflexi-like 16S sequences were most similar to either *Chloroflexus* or *Roseiflexus* spp.; however, many sequences were more closely related to Chloroflexi that fall outside of the family *Chloroflexaceae*, clading with organisms not known to exhibit phototrophy (Figure 4.6B). Additionally, *Roseiflexus*-like populations from MS_15, CP_7 and FG_16 each formed monophyletic groups that excluded sequences from any other springs, suggesting that each of these clades is specific to its corresponding spring (Figure 4.7). Interestingly, the predominant sequences from *Chloroflexus* spp. originating from the two BLVA sites and WC_6 were closely related (Figure 4.7), despite the very different geochemical context of these environments (Table 4.1); a similar phenomenon was observed with sequences from *Roseiflexus* spp. from BLVA and CP_7. Other spring-specific clades were observed for *Chloroflexus* spp. sequences from FG_16 within the

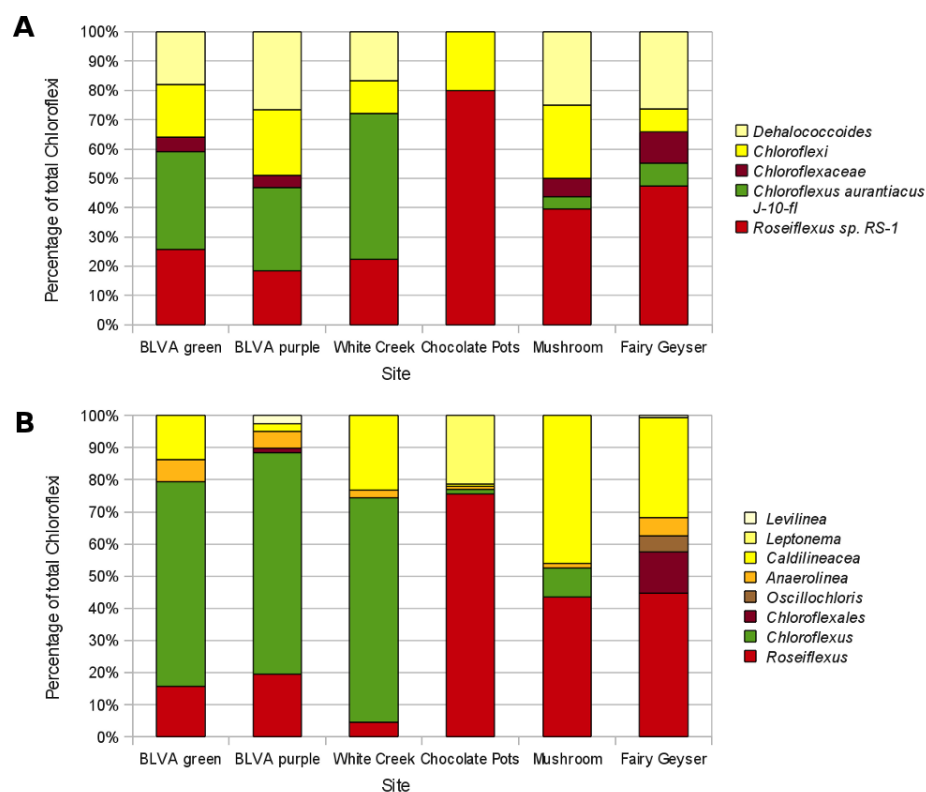


Figure 4.6: Comparison of Chloroflexi Phylogenetic Marker Genes from Metagenomes and Chloroflexi 16S rRNA Clones. (A) Phylogenetic marker genes in the metagenome classified as Chloroflexi by AMPHORA. (B) 16S rRNA composition of the Chloroflexi kingdom classified by the RDP at a confidence threshold of 80%. Colors correspond to similar taxonomic groupings of Chloroflexi as follows: red = *Roseiflexus* spp., green = *Chloroflexus* spp., shades of brown = other taxa within Order Chloroflexales, and shades of yellow = other taxa within kingdom Chloroflexi.

Chloroflexi class Anaerolineae, a group that until recently was not known to contain phototrophic members (Chapter 3). The presence of these 16S rRNA gene sequences combined with observed photosynthesis genes most similar to the *Chloroflexaceae* suggests that currently unknown and uncultured phototrophic Chloroflexi exist in many of these mat communities.

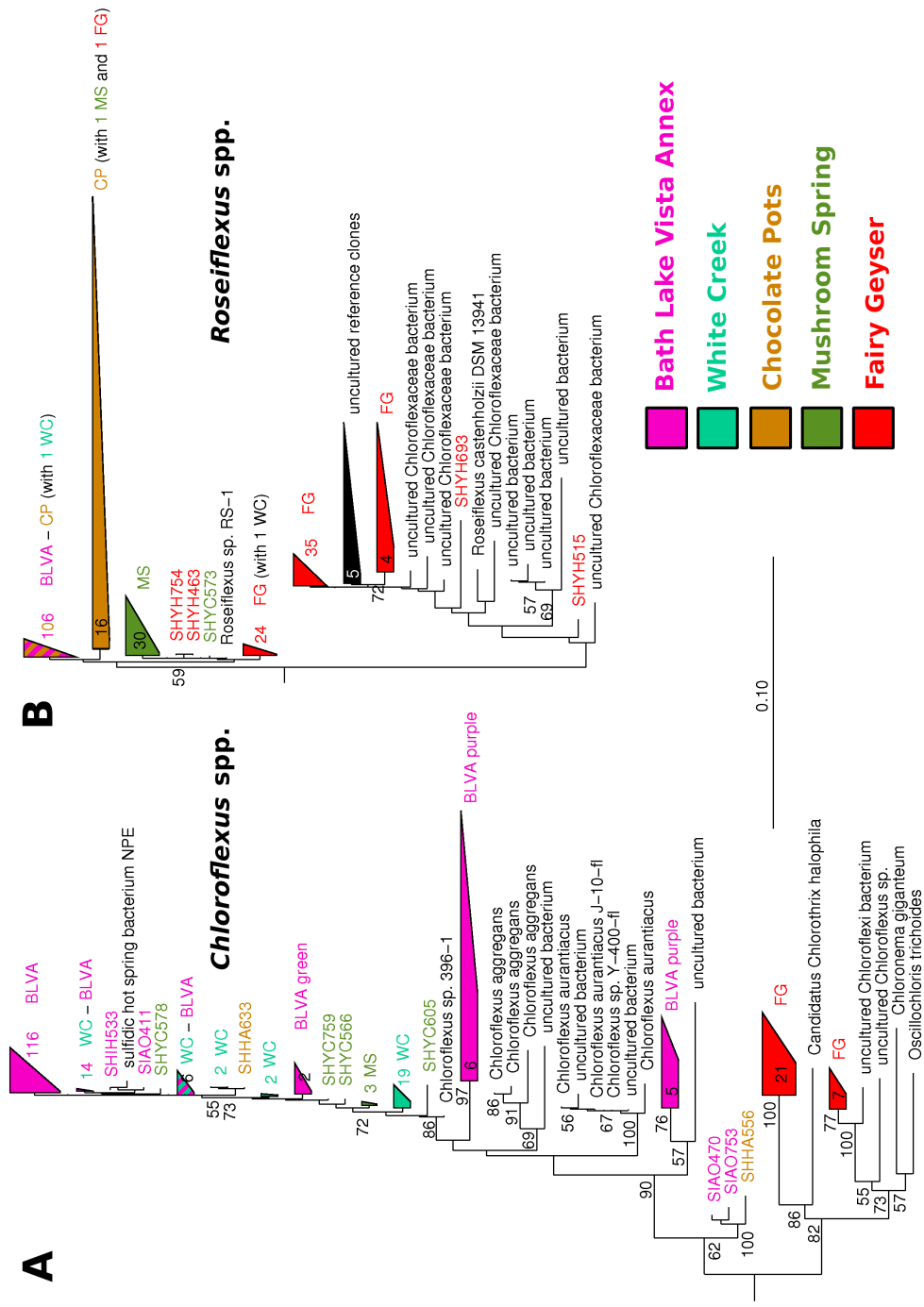


Figure 4.7: Unrooted Neighbor-joining Phylogenetic Trees of Chloroflexi 16S rRNA Sequences from PCR Clone Libraries. (A) Sub-branch of tree corresponding to *Chloroflexus* spp. and other FAPs capable of producing BChl *c*. (B) Sub-branch of tree corresponding to FAPs related to *Roseiflexus* spp. Sequences are color coded according to spring origin, and numbers adjacent to or within polygons indicate the number of clones in each clade. Bootstrap support for $\geq 50\%$ of 1000 replicate trees are shown at nodes. BLVA refers to both sites BLVA_5 and BLVA_20 unless indicated otherwise.

Geochemical Influences on Community Composition

Community composition differences among Chloroflexi were analyzed to determine whether there was evidence that geochemistry influenced the spring-specificity of clades observed in the phylogenetic analysis (Figure 4.7). To compare the environmental characteristics of the sites, a distance matrix of all geochemical variables was constructed, and ordination was used to visualize the similarity of measured environmental variables among the sites (Figure 4.8A). The two sampling times at BLVA were geochemically similar in contrast to the other sites given their high sulfide and NH_4^+ concentrations, whereas the MS_15 and FG_16 geochemical profiles showed similarity contributed by their higher pH, and elevated Na^+ concentrations. The patterns apparent from the differences in geochemistry were also reflected in differences of Chloroflexi community compositions within each site. A comparison of the phylogenetic makeup of the Chloroflexi communities across all sites was visualized with an ordination of the weighted Unifrac distance matrix of pairwise comparisons for all sites (Figure 4.8B). Consistent with the geochemistry, the BLVA sites exhibited similar communities, as did the under-layer communities, which both contained closely related *Roseiflexus* spp. (Figure 4.7B). Despite the difference in sulfide concentrations between WC_6 and the BLVA sites, there was notable similarity in the Chloroflexi community compositions among these samples (Figure 4.8B), which was largely due to the occurrence of closely related *Chloroflexus* spp. in all three sites (Figure 4.7A).

Functional Analysis of Predominant Sequence Assemblies

Genes Involved in Autotrophy and Phototrophy: The gene content of each scaffold cluster provides a basis for inferring the functional roles of the dominant commu-

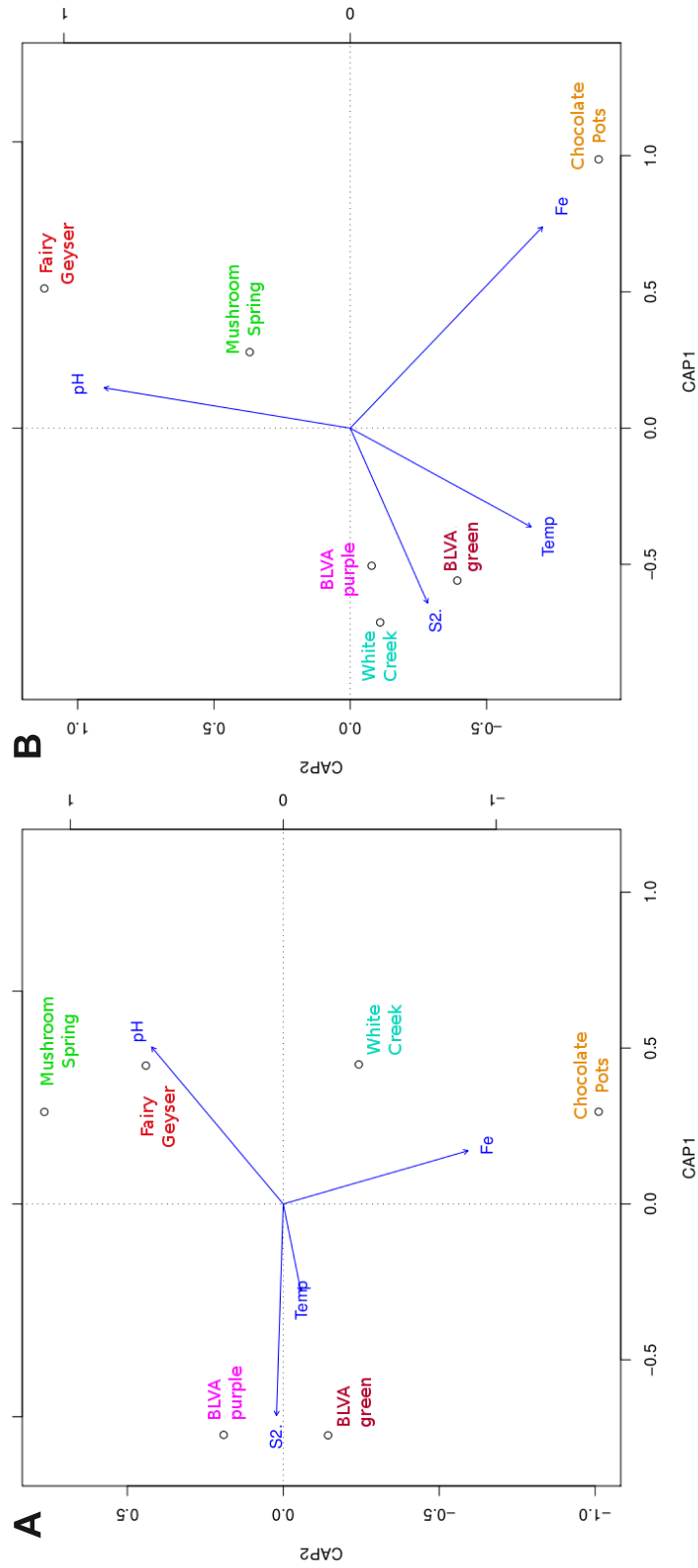


Figure 4.8: Ordination of Geochemical and Community Distance Matrices. (A) Constrained analysis of principal coordinates (CAP) for the environmental dissimilarity matrix, with vectors indicating the direction of constrained environmental variables pH, temperature, sulfide (HS^-), and Fe. (B) CAP analysis of the Weighted UniFrac community dissimilarity matrix based upon the Chloroflexi 16S rRNA neighbor-joining tree.

nity members represented in these metagenomes. For example, genes encoding key enzymes involved in the 3-hydroxypropionate (3-OHP) pathway of inorganic carbon fixation were present in metagenomes from all six sites (Table 4.3), and were associated with the predominant *Chloroflexus* and *Roseiflexus* like populations present across these respective habitats. Genes coding for subunits of ribulose biphosphate carboxylase-oxygenase (RuBisCO), a key enzyme in the reductive pentose phosphate pathway (i. e., Calvin-Benson-Bassham cycle) were observed in cyanobacterial (in WC_6 and CP_7) or proteobacterial (in FG_16 and BLVA_20) sequences. No CO₂ fixation genes were found in the Chlorobiales-like populations from CP_7, despite the fact that other cultivated members of this kingdom are capable of fixing CO₂ via the reductive tricarboxylic acid (rTCA) cycle. While the relative depths of coverage of these metagenomes were not sufficient to conclude that these Chlorobiales organisms lack the capacity to fix inorganic carbon, metatranscriptomic studies with deeper coverage have demonstrated that there is an absence of rTCA cycle genes in the *Candidatus* Thermochlorobacter spp. populations in Mushroom Spring (Liu *et al.*, 2011a). Genes involved in bacteriochlorophyll biosynthesis and the production of photosynthetic reaction centers were present in scaffold clusters corresponding to *Roseiflexus*, *Chloroflexus*, *Thermochromatium* and *Synechococcus* spp., as well as from undescribed Chlorobi and Cyanobacteria (Figure 4.4). Consequently, all dominant phototrophs in each community showed genomic evidence for chlorophototrophic metabolism. Examination of shorter (<10 kbp) scaffolds revealed additional genes involved in chlorophototrophy, and these could be assigned to distinct phylogenetic groups (Table 4.3). For example, phototrophy genes from *Ca. Chloracidobacterium* spp. were present in WC_6, and sequences from uncultivated proteobacteria were present in the FG_16 subsurface mat community. Phototrophy genes most closely related to members of the Chloroflexi, but too distant (~70% amino acid identity)

Table 4.3: Phylogenetic Distribution of Phototrophic, Autotrophic, and Sulfur Cycling Genes in Metagenomes. Bacteriochlorophyll/chlorophyll biosynthesis genes included *acsF*, *chlGILP*, and *bchBCDEFGHIJKLMNPRSUXYZ*. Photosynthetic reaction center genes included *pufLMC*, *psaA*, and *pscA*. Genes for carbon fixation included those involved in the 3-OHP pathway (*ccl*, *mch*, *mcl*, *mcr*, *mct*, *meh*, *pcs*, *sct*, and *smtAB*) and the Calvin-Benson-Bassham cycle (*cbbQX*, PRK, *rbcSLX*). Sulfur-cycling genes included *aprABM*, *dsrACEFHKNORS*, *fccAB*, and *sqr*.

SPRING	<i>Roseiflexus</i> sp.	<i>Chloroflexus</i> sp.	Other Chloroflexi	Chlorobiales	<i>Chloracidobacteria</i>	Cyanobacteria	Proteobacteria
Bacteriochlorophyll/Chlorophyll Biosynthesis Genes							
BLVAgreen	X	X					
BLVApurple	X	X					X
White Creek	X	X	X	X	X	X	
Choc Pots	X		X	X		X	
MS 60 undermat	X	X	X	X	X	X	
Fairy Geyser	X		X	X			X
Photosystem Reaction Center Genes							
BLVAgreen	X	X					
BLVApurple	X	X					
White Creek					X	X	
Choc Pots	X			X		X	
MS 60 undermat							
Fairy Geyser	X					X	
Autotrophic Pathway Diagnostic Genes							
BLVAgreen	X	X					
BLVApurple	X	X					X
White Creek	X	X				X	
Choc Pots	X	X				X	
MS 60 undermat	X						
Fairy Geyser	X		X				X
Sulfur-cycling Genes							
BLVAgreen							
BLVApurple		X					X
White Creek							
Choc Pots	X						
MS 60 undermat							
Fairy Geyser							

to originate from either *Chloroflexus* or *Roseiflexus* spp. were present in all non-sulfidic sites, and were especially prevalent in FG_16. The translated peptide sequences of three novel phototrophy genes from MS_15 were highly similar (96-100% amino acid identity) to sequences observed in a recent metagenomic and metatranscriptomic study of the Mushroom Spring top-layer mat (Liu *et al.*, 2011b), and which linked these genes to a group within the Chloroflexi not previously known to contain chlorophototrophic organisms. Novel chlorophototrophy genes from FG_16 were distinct from previously described metagenome sequences (<70% amino acid

identity) and any phototrophic peptide sequences residing in public databases as of July 2011.

This study targeted anoxygenic photosynthesis as an important process in the sulfidic community at BLVA and possibly in the high ferrous iron system at CP_7. The potential for sulfide and ferrous Fe to serve as electron donors for phototrophy was examined using query genes for both sulfur oxidation and Fe-oxidation, respectively (Frigaard and Dahl, 2009; Bryant *et al.*, 2012; Grimm *et al.*, 2011). Interestingly, no genes with significant similarity to those experimentally characterized to be involved in the phototrophic oxidation of ferrous iron in *Rhodopseudomonas* spp. (*pioAB*; Jiao and Newman 2007) were observed in CP_7, or any site described here with the exception of one sequence in FG_16, a site that contains below detectable levels of iron. Genes involved in sulfide oxidation (*dsr*) that are used by some anoxygenic phototrophs (such as those characterized in the γ -proteobacterium *Allochromatium vinosum*) were identified in the *Thermochromatium*-like population present in BLVA_20, providing a definitive linkage with the high dissolved sulfide levels measured in situ. The dominant Chloroflexi populations observed in both BLVA samples do not contain known genes for dissimilatory oxidation of reduced-sulfur compounds, such as *dsr* or *sox*, which is consistent with the lack of these genes in representative genomes (Tang *et al.*, 2011) and the idea that sulfide-oxidation occurs via an unknown mechanism in these organisms (Frigaard and Dahl, 2009). Both *Chloroflexus* and *Roseiflexus* spp. genomes and the BLVA metagenomes contain *sqr* genes encoding potential sulfide-quinone oxidoreductases, suggesting that these genes enable FAPs to obtain electrons from reduced-sulfur compounds (Frigaard and Dahl, 2009; Bryant *et al.*, 2012).

The scaffold clusters corresponding to undescribed CFB organisms and those with low G+C did not contain genes indicative of chlorophototrophy, but they do con-

tain genes involved in anaerobic metabolism. These genes allow for the oxidation or fermentation of organic acids, such as acyl-CoA synthetase in the BLVA-specific (G+C=64%) and CP-specific (G+C=31%) unknown clusters, or lactate dehydrogenase in the mixed BLVA and CP unknown cluster (G+C=36%). Also included were genes that encode integral components of anaerobic carbon metabolism and electron transfer, such as subunits of pyruvate ferredoxin:oxidoreductase (PFOR), which were found in both unknown BLVA clusters. While the CP-specific cluster showed evidence of anaerobic metabolisms, metagenomic coverage was insufficient for the detection of genes involved in aerobic metabolisms, most importantly those encoding terminal cytochrome c oxidases. While the organisms represented by this cluster co-inhabit the CP_7 site with cyanobacteria and presumably live in oxic conditions during the day, it is possible that fermentative metabolisms are more important at this anoxic, Fe(II)-rich spring compared to more-oxic downstream communities.

Discussion

The six sites investigated in this study are representative of three types of geothermal springs that support bacterial phototrophic communities in Yellowstone National Park, namely (i) alkaline-siliceous chloride springs (pH 7.5-8), (ii) sulfidic-carbonate springs (pH 6), and (iii) mildly acidic (pH 6) non-sulfidic springs high in Fe(II) and Mn(II) (Rowe *et al.*, 1973; McClesky *et al.*, 2005). The major physical and geochemical constraints that have been postulated to control the distribution of phototrophs (and photosynthesis) in these thermal springs are pH, temperature, sulfide concentration, and gradients in light and/or other chemicals existing as a function of mat depth (Cox *et al.*, 2011). Most springs that support prokaryotic phototrophic mats occur at pH >5, with rare exceptions (such as the purple phototrophic bacterial communities

comprised of organisms related to *Rhodopila* sp. observed in Nymph Lake (YNP) and in small sulfidic, acidic (pH 3.5-4.5) springs near the Gibbon River; Pfennig 1974; Madigan *et al.* 2005). The bulk aqueous pH levels at CP_7 and BLVA_5 and _20 are near the lower limit observed for thermophilic cyanobacteria (Brock, 1973); however, consumption of dissolved $\text{CO}_2/\text{HCO}_3^-$ by cyanobacteria results in significant pH increases of interstitial aqueous environments. Specifically, previous microelectrode studies of pH profiles at CP_7 and MS_15 reveal daytime pH maxima to be as high as 9 to 10 in the top 1 mm (Revsbech and Ward, 1984; Pierson *et al.*, 1999; Jensen *et al.*, 2011). Consequently, CP_7 supports an active community of cyanobacteria that are similar to *Synechococcus*-like populations observed in Mushroom Spring and Octopus Spring phototrophic mats.

Anoxygenic phototrophs have long been known to colonize sulfidic springs of YNP (van Niel and Thayer, 1930; Madigan, 1984; Giovannoni *et al.*, 1987), and this was confirmed in samples from BLVA where sulfide levels exceed $100\ \mu\text{M}$. However, the only population in the BLVA samples with genes similar to the sulfide-oxidizing pathway identified in other anoxygenic phototrophs was that composed of the *Thermochromatium*-like organisms observed in BLVA_20. The other prominent anoxygenic phototrophs identified across sites include the *Chloroflexus*, *Roseiflexus* and Chlorobiales-like populations. The abundance of phototrophic Chloroflexi across sites is reflective of their previously established physiological diversity including photoheterotrophy on organic acids such as acetate and propionate, photoautotrophy, and aerobic chemoorganotrophy (Pierson and Castenholz, 1974a; Madigan *et al.*, 1974; Hanada *et al.*, 2002; van der Meer *et al.*, 2003, 2010). While these organisms generally grow in culture as photoheterotrophs, their metabolic flexibility and ability to produce diverse electron and carbon storage compounds such as polyhydroxyalkanoic acids, polyglucose and wax esters may, in part, be why these organisms colonize a

broad spectrum of phototrophic environments (Castenholz and Pierson, 1995). Highly similar (>98% average nucleotide identity) *Roseiflexus*-like organisms were abundant populations in nearly all sites, while *Chloroflexus*-like populations were limited to BLVA (sulfidic) and WC_6 (oxic), which indicates that other ecological factors aside from O₂ and sulfide are important for controlling the relative abundance of *Chloroflexus* and *Roseiflexus* spp. in YNP phototrophic mat environments.

Trophic interactions between FAPs and cyanobacteria have been studied in phototrophic geothermal mats, where it has been shown that FAP photoheterotrophs utilize organic acids and or storage compounds produced by autotrophic cyanobacteria (Anderson *et al.*, 1987; Nold and Ward, 1996; van der Meer *et al.*, 2003; Bauld and Brock, 1974). Moreover, it has been proposed that *Thermochromatium* spp. (purple-sulfur bacteria) are primary producers in sulfidic springs and cross-feed FAP populations with low-molecular weight organic acids (Madigan *et al.*, 1989, 2005) analogous to the cyanobacterial primary production and trophic interactions that have been documented in Octopus Spring and Mushroom Spring mats (Anderson *et al.*, 1987; van der Meer *et al.*, 2005). This hypothesis has been challenged by the relatively heavy carbon isotope compositions of *Chloroflexaceae*-specific lipid biomarkers, which can also be explained by *Chloroflexus* and *Roseiflexus* spp. autotrophy via the 3-OHP pathway (Strauss and Fuchs, 1993; Holo and Sirevåg, 1986; van der Meer *et al.*, 2000; Klatt *et al.*, 2007). The isotope values have been interpreted as too heavy to have originated from compounds originally fixed by Calvin-Benson-Bassham cycle autotrophy (from *Thermochromatium* spp.) and subsequently cross-fed to *Chloroflexus* (van der Meer *et al.*, 2000). Metagenome sequence obtained in the current study shows that *Chloroflexus* and *Roseiflexus* spp. both contain genes necessary for CO₂ fixation via the 3-OHP pathway, supporting the hypothesis that all three groups contribute to primary productivity in sulfidic-carbonate springs (Table 4.3).

It remains to be determined whether FAPs augment their carbon metabolism utilizing the 3-OHP autotrophic pathway in springs where they coexist with cyanobacteria, and whether their primary productivity is supported in sulfidic springs that contain higher concentrations of reductants than alkaline siliceous springs.

Conclusion

This study highlights some of the major differences in phototrophic bacterial community composition and metagenomic gene content from representative geothermal springs that support chlorophototrophic metabolism. The degree to which these community composition differences reflect differences in overall process rates (e. g. primary productivity or biologically mediated sulfur cycling) is currently unknown. Regardless, the observation of genes involved in these processes (e. g. autotrophy or sulfide oxidation) provide an initial step necessary for assigning the appropriate members of each community to corresponding functional groups capable of mediating the geochemical transformations of interest.

Materials and Methods

Sample Collection and Geochemical Analyses

Six different samples were taken from five hot springs from August 2007 to May 2008 (Table 4.1) and immediately frozen in liquid N₂. These springs were sampled at different distances down the effluent channels from the source of each respective spring, and two of these samplings are from the subsurface communities in Mushroom Spring and Fairy Geyser. Geochemical characterizations were done with bulk spring water at the sampling locations after filtration (0.2 μ m polycarbonate filter), and they include temperature, pH, total dissolved sulfide, dissolved gasses (O₂, CO₂,

CH₄, and H₂), and a survey of total dissolved ions. Techniques for determining total dissolved sulfide and dissolved gasses have been published elsewhere (Clesceri *et al.*, 1998; Inskeep *et al.*, 2004; Macur *et al.*, 2004; Inskeep *et al.*, 2005), and total dissolved ions were determined using ion chromatography and inductively coupled plasma spectrometry as previously described (Inskeep *et al.*, 2005).

DNA Extraction and Preparation

DNA extractions were carried out on mat samples using a previously published protocol (Inskeep *et al.*, 2010). Briefly, 0.5-1 g of frozen mat samples were processed for parallel DNA extractions using both enzymatic (Proteinase K (1 mg/ml) with sodium dodecyl sulfate (SDS) (0.3% w/v) for 0.5 hour at 37 °C) and mechanical (bead-beating with 2% w/v SDS and 15% v/v TRIS-equilibrated phenol, shaken at 5.5 m/s for 30 s) treatments, then both lysates were pooled, and subsequent extractions were done with phenol:chloroform:isoamyl alcohol (25:24:1), and chloroform:isoamyl alcohol (24:1). All samples were treated with RNase I (Promega, Madison WI USA) and DNA was precipitated with ethanol and sodium acetate. Small insert metagenome libraries were constructed as described previously (Inskeep *et al.*, 2010). DNA was randomly sheared via nebulization, end-polished with consecutive BAL31 nuclease and T4 DNA polymerase treatments, and size-selected using gel electrophoresis on 1% low-melting-point agarose. After ligation to BstXI adapters, DNA fragments were purified, then inserted into BstXI-linearized, medium-copy pBR322 plasmid vectors. The resulting library was electroporated into *Escherichia coli* resulting in high-quality random plasmid libraries with few clones without inserts, and no clones with chimeric inserts (Rusch *et al.*, 2007). Clones were sequenced from both ends to produce pairs of linked sequences representing ~820 bp at the end of each insert, and resulted in a total of 320.6 Mbp in 424,982 sequences. 16S rRNA sequence PCR

amplicons were produced with universal primers targeting domains Archaea (4aF, TCCGGTTGATCCTGCCRG; 1391R, GACGGGCRGTGWGTRCA) and Bacteria (27F, AGAGTTTGATCCTGGCTCAG and 1391R). Amplicons were cloned using the TOPO TA Cloning Kit (Invitrogen, Carlsbad CA USA) and sequenced using Big Dye v3.1 chemistry.

Pre-Assembly Metagenomic Sequence Analyses

All metagenomic sequences were used as queries in an NCBI BLAST+ (Camacho *et al.*, 2009) BLASTX search against the NCBI nr database (accessed 22 March 2011) with default parameters. The results were parsed and visualized with the MEGAN software version 2.3.2 (Huson *et al.*, 2007) with the default parameters (MinScore = 35.0, TopPercent=10.0, MinSupport=5,) and taxonomic assignments of the top BLASTX matches were extracted. A customized perl script was used to determine the %G+C of all sequences.

Sequence Assembly and Annotation

Metagenomic scaffolds of overlapping end sequences were constructed separately for each of the six samples using the Celera assembler (Miller *et al.*, 2008). This resulted in 206,469 scaffolds containing 183.2 Mbp (27 to 33 Mbp per site) of assembled sequence, or a 57% compression of the raw sequence data. The JCVI annotation pipeline including open reading frame (ORF) prediction, BLAST alignments, and hidden Markov model analysis (Tanenbaum *et al.*, 2010) was used as an initial step for inferring functions for predicted ORFs on metagenomic scaffolds. Translated peptide sequences from predicted ORFs were analyzed with the AMPHORA package (Wu and Eisen, 2008), which identified homologs to 31 different genes (mostly predicted to encode ribosomal proteins or enzymes with housekeeping functions) that could be

used as phylogenetic markers in comparison to 16S rRNA sequences. Genes encoding particular functions were identified by BLASTP using reference sequences as queries, with the additional requirement that candidate sequences had a top BLASTP match to a sequence with the same annotated function in NCBI's nr database.

Ribosomal RNA Sequence Analyses

All bacterial 16S rRNA sequences from the 16S rRNA-specific PCR clone libraries were aligned and screened for chimeras with Bellerophon (Huber *et al.*, 2004) with subsequent manual curation. OTUs were determined using the CAP3 assembler (Huang and Madan, 1999) at the 99% demarcation level. Rarefaction curves were determined (Supplementary Figure 1), the Chao1 and ACE richness indexes and the Fisher's alpha, Shannon-Weaver, and Simpson's diversity indexes were calculated for each library. The RDP Bayesian Classifier (Wang *et al.*, 2007) was used to assign taxonomy to 16S rRNA sequences at the 80% confidence level, and all sequences belonging to Kingdom Chloroflexi were aligned with reference sequences corresponding with *E. coli* positions 29 to 1349 (1321 positions) in ARB (Ludwig *et al.*, 2004). A phylogenetic tree was produced using the BioNJ algorithm (Gascuel, 1997) and bootstrapped with 1000 replicates. Reference sequences shorter than the initial alignment were subsequently added to the tree using the ARB parsimony tool. RaxML (Stamatakis, 2006) was used to produce a consensus maximum likelihood tree from 1000 replicate trees, which were masked with bacterial complexity filters. Reference sequences were removed and a second neighbor-joining phylogenetic tree was produced as an input tree for community composition analysis using weighted Unifrac (Lozupone *et al.*, 2007). A pairwise distance matrix of weighted Unifrac dissimilarity coefficients was constructed from these data.

Statistical Analyses

A distance matrix of environmental variables was constructed by first eliminating columns containing missing values, then Gower coefficients were calculated using the R Statistical Package (R Core Development Team, 2011). The Gower coefficient allows for different data types (qualitative presence/absence vs. quantitative numerical) with different dimensional scales to be combined into a general dissimilarity metric (Gower, 1971). Ordinations of the community composition and the geochemical distance matrices with respect to geochemical variables were done using constrained analysis of principal coordinates with the `capscale` function of the `vegan` package (URL = <http://vegan.r-forge.r-project.org/>) (R Core Development Team, 2011). This constrained analysis focused on environmental variables found to be significant in the Pearson correlation analysis. Mantel tests using environmental distance matrix and community composition distance matrix were performed using the `mantel` function in `vegan` (Legendre, 1998). Metagenomic scaffolds that were 10 kb or larger were analyzed in terms of their oligonucleotide composition. All possible tri-, tetra-, penta-, and hexanucleotides were counted with custom perl scripts, and counts were normalized by the length of the scaffold. Normalized oligonucleotide composition matrices were subjected to k-means clustering with a range of $k = 4$ to 12 with 100 trials each. The composite summary of these k-means trials was displayed as an interaction network using the program Cytoscape 2.8.1 (Shannon *et al.*, 2003).

Sequence Availability

All individual sequences and assembled contigs have been deposited with NCBI under the GenomeProject database (ID #41119) and are assigned a registered locus tag prefix of YNPJCVI.

CHAPTER 5

TEMPORAL PATTERNING OF IN SITU GENE EXPRESSION IN
UNCULTIVATED PHOTOTROPHIC CHLOROFLEXI INHABITING AN
ALKALINE SILICEOUS GEOTHERMAL SPRING.Contribution of Authors and Co-Authors

Manuscript in Chapter 5

Author: Christian G. Klatt

Contributions: Designed the study, conducted the experiments, collected and analyzed output data and wrote the manuscript.

Co-author: Zhenfeng Liu

Contributions: Assisted with experimental design, assisted in data analysis, and edited the manuscript.

Co-author: Marcus Ludwig

Contributions: Assisted with experimental design, assisted in data analysis, and edited the manuscript.

Co-author: Donald A. Bryant

Contributions: Obtained funding, assisted with experimental design, assisted in data analysis, discussed the results and edited the manuscript at all stages.

Co-author: David M. Ward

Contributions: Obtained funding, assisted with experimental design, assisted in conducting field experiments, assisted in data analysis, discussed the results and edited the manuscript at all stages.

Manuscript Information Page

Christian G. Klatt, Zhenfeng Liu, Marcus Ludwig, Donald A. Bryant, and David M. Ward

Journal Name: The ISME Journal

Status of Manuscript:

☒ Prepared for submission to a peer-reviewed journal

☐ Officially submitted to a peer-reviewed journal

☐ Accepted by a peer-reviewed journal

☐ Published in a peer-reviewed journal

Published by the International Society for Microbial Ecology.

Abstract

Filamentous anoxygenic phototrophs (FAPs) are dominant members of microbial communities inhabiting neutral and alkaline geothermal springs in Yellowstone National Park. Natural populations of FAPs related to *Chloroflexus* and *Roseiflexus* spp. have been particularly well characterized in Mushroom Spring mats, where they co-inhabit the mats with unicellular cyanobacteria related to *Synechococcus* spp. strains A and B'. Metatranscriptomic sequencing was applied to the microbial community over a diel period to determine how FAPs regulate their gene expression in response to fluctuating environmental conditions and resource availability. Both *Roseiflexus* and *Chloroflexus* spp. were found to express key genes involved in the 3-hydroxypropionate carbon fixation pathway during the day, when these organisms were thought to primarily use photoheterotrophic and/or aerobic chemoorganotrophic metabolisms. Transcripts for genes involved in phototrophic metabolism such as the biosynthesis of bacteriochlorophylls and photosynthetic reaction centers, were much more abundant at night; this suggests that these organisms prepare at night for phototrophic activity in the early morning. The expression of genes involved in the synthesis and degradation of storage polymers, such as glycogen, polyhydroxyalkanoates (PHAs), and wax esters, suggests that these organisms produce and utilize these compounds at different times during the diel cycle. From these data, we infer that *Chloroflexus* and *Roseiflexus* spp. primarily produce polyglucose during the day, and ferment this to intermediates that are used to construct polyhydroxyalkanoates and possibly and possibly wax esters as forms of energy storage during the night. We summarize these results by proposing a conceptual model for temporal changes in central carbon metabolism and energy production for FAPs living in a natural environment.

Introduction

Molecular characterization of the thermophilic microbial communities in Octopus Spring and Mushroom Spring revealed that the most dominant community members consist of cyanobacteria related to cultivated *Synechococcus* spp. strains A and B' (Ward *et al.*, 1990; Ferris *et al.*, 1996a; Allewalt *et al.*, 2006; Bhaya *et al.*, 2007), in addition to filamentous anoxygenic phototrophs (FAPs) related to *Chloroflexus* and *Roseiflexus* spp. (Nübel *et al.*, 2002). Past work has suggested that *Synechococcus* spp. are the primary producers responsible for most inorganic carbon fixation, while they also produce low-molecular organic compounds as byproducts of their metabolism, and it has been shown that FAPs assimilate these compounds photoheterotrophically (Figure 1.2; Anderson *et al.* 1987; Bateson and Ward 1988; Nold and Ward 1996). Metabolites excreted by cyanobacteria in these mats fluctuate between daytime production of glycolate (a byproduct of photorespiration under conditions of oxygen supersaturation during the day; Bateson and Ward 1988) and nighttime production of acetate and propionate (both produced in part by cyanobacterial or other bacterial fermentation under anoxic conditions; Anderson *et al.*, 1987, Nold and Ward 1996, van der Meer *et al.*, 2005). FAPs are thought to perform photoheterotrophic metabolism for the uptake of low-molecular weight carbon sources both in culture and in situ (Pierson and Castenholz, 1974a; Madigan *et al.*, 1974; Sandbeck and Ward, 1981; Anderson *et al.*, 1987; van der Meer *et al.*, 2003; Hanada *et al.*, 2002; van der Meer *et al.*, 2005, 2010). However, *Chloroflexus aurantiacus* strain OK-70-fl can be grown photoautotrophically on a minimal medium gassed with H₂ and CO₂ as the sole source of carbon (Holo and Sirevåg, 1986; Strauss *et al.*, 1992), and there was also evidence that *Chloroflexus* and *Roseiflexus* spp. might fix inorganic carbon *in situ* when electron donors such as H₂ and H₂S as well as light

are available at dawn and dusk (van der Meer *et al.*, 2003; Klatt *et al.*, 2007). Furthermore, the 3-hydroxypropionate (3-OHP) carbon fixation pathway that has been described for these organisms (Strauss and Fuchs, 1993; Zarzycki *et al.*, 2009) can also operate mixotrophically, in which these organisms simultaneously incorporate both CO₂ and organic compounds as carbon sources, such as acetate (by way of acetyl-CoA synthetase) and glycolate (by way of glycolate dehydrogenase) (Bryant *et al.*, 2012; Zarzycki and Fuchs, 2011).

The recent metagenomic characterizations of phototrophic microbial mat communities in Octopus Spring and Mushroom Spring have revealed three additional and abundant photoheterotrophic groups of organisms: Acidobacteria related to "Candidatus *Chloracidobacterium thermophilum*" (Bryant *et al.*, 2007), Chlorobi related to "Candidatus *Thermochlorobacter aerophilum*" (Chapter 3; Liu *et al.* 2011a,b); and a novel clade of organisms related to Chloroflexi of the Class Anaerolineae (Chapter 3). These organisms are predicted to be photoheterotrophs and utilize some of the same resources as FAPs, and photoheterotrophic community members could escape competition for resources by temporally partitioning their nutrient uptake. The abundance of these organic carbon compounds, combined with the availability of inorganic carbon, light as an energy source and hydrogen or sulfide as a source of electrons, are factors that shape the relative degree to which FAPs use heterotrophic, mixotrophic, or autotrophic metabolisms.

This study utilized metatranscriptomic sequencing from hourly samples taken over the course of a diel period to obtain a more complete view of how chlorophototrophic members of the Chloroflexi temporally transcribe their genes in relation to environmental conditions and the metabolisms of other community members. This experiment enabled high-resolution temporal transcription profiles of genes involved in photosynthesis, central carbon metabolism and energy production of uncultivated

FAPs in their natural habitat. From these transcriptional analyses, we infer a model of how members of the Chloroflexi regulate their metabolism and contribute to the food webs of these microbial mats. The metatranscriptomic analysis of cyanobacterial *Synechococcus* spp. and the photoheterotrophic "*Ca. C. thermophilum*" and "*Ca. T. aerophilum*" in this mat will be reported elsewhere (Liu *et al.*, 2011a).

Materials and Methods

Metagenomic Analyses

The sequencing and assembly of the metagenome scaffolds of the entire mat community and the clustering of scaffolds associated with various bacterial populations were described previously (Chapter 3). In order to identify many of the transcripts originating from FAPs, it was first necessary to expand the database of metagenomic scaffolds to which these transcripts could be assigned. Uncultivated *Roseiflexus* and *Chloroflexus* spp. were represented by two distinct clusters of scaffolds larger than 20 kb (Figure 3.1), which contained signature genes characteristic of members of these genera. Metagenomic scaffolds that were smaller than the 20-kb cutoff and (thus were not included in clusters), but which were still greater than 5 kb (and thus were included in the bioinformatic annotation workflow as previously described; (Tanenbaum *et al.* 2010; Chapter 3), also contained genes that were highly similar to *Chloroflexus* and *Roseiflexus* spp. reference genomes. This larger grouping of scaffolds of length 5 kb and greater is referred to as the 'expanded set' below. Open reading frames (ORFs) on all scaffolds were demarcated and annotated as previously described (Chapter 3). All scaffolds containing ORFs that had at least 90% amino acid identity (% AA ID) to the *Roseiflexus* sp. strain RS-1 genome and 80% AA ID to the *Chloroflexus* sp. strain 396-1 genomes (TBLASTN of translated metagenomic

ORFs used as queries against the genome databases with default parameters) were categorized as *Roseiflexus* spp. and *Chloroflexus* spp., respectively. The alignment cutoffs were determined based upon previous work, which established the level of relatedness between metagenomic sequence derived from uncultivated FAPs and the genomes of corresponding reference isolates (Appendix B). The genomes of these isolates have been shown in past analyses to be most closely related to the dominant uncultivated populations in the mat (Chapter 3). Scaffolds meeting the %AA ID criteria to both the *Roseiflexus* sp. RS-1 and the *Chloroflexus* sp. 396-1 genomes were manually assigned to either genus while also considering their guanine and cytosine content (*Roseiflexus* spp. scaffolds contained an average of 60% G+C, while *Chloroflexus* spp. contained an average of 54% G+C; see Chapter 4). ORFs on scaffolds demarcated as either *Roseiflexus* or *Chloroflexus* spp. which aligned to the *Roseiflexus* sp. RS-1 genome or the *Chloroflexus* sp. 396-1 genome above the 90% or 80% AA ID cutoffs, respectively, were reciprocally aligned to the database of total metagenomic ORFs. Pairs of genomic and metagenomic ORFs that exhibited reciprocal top BLAST matches were determined to be orthologous.

Collection and Preparation of Microbial Mat Samples

The microbial mat community inhabiting the effluent channel of Mushroom Spring at 60 °C was sampled hourly beginning at 5:00 PM September 11, 2009 and ending at 4:00 PM on the following day. Mat cores were collected in the following manner: two #4-sized cores (each 9 mm in diameter, resulting in a total area of 1.26 cm² sampled per timepoint) were randomly taken from the this region of the mat, and a razor blade was used to remove mat material below the top ~2 mm. These top-mat subsamples were subsequently split in half through the vertical aspect of the mat. All samples were immediately frozen in liquid N₂ and were stored at -80 °C until further

processing. Light data were collected simultaneously using a LI-1400 light meter equipped with a LI-192 irradiance sensor (LI-COR, Lincoln, NE). Depth profiles of oxygen concentrations were measured in situ using microelectrodes as had been done in a previous study (Jensen *et al.*, 2011).

Nucleic Acid Extraction and Analysis

Prior to RNA extraction, the halved samples from the two different cores were combined to account for heterogeneity in the mat community within the sampling region. Diethyl pyrocarbonate (DEPC)-treated 10 mM sodium acetate, pH 4.5 (250 μ l), and 500 mM Na₂-EDTA, pH 8.0 (37.5 μ l) were added to tubes containing the combined half-core mat samples and the samples were subsequently homogenized by bead-beating with a velocity of 6.5 m s⁻¹ for 10 s (Fastprep-24 Instrument, MP Biomedicals, Solon, OH). DEPC- treated lysis buffer (375 μ l) containing 10 mM sodium acetate and 10% (w/v) sodium dodecyl sulfate (pH 4.5) was added to the mat homogenate, which was incubated at 65 °C for 3 min. Acidic phenol equilibrated with DEPC-treated H₂O (700 μ l) was added, and the samples were incubated at 65 °C for an additional 3 min. Two subsequent organic extractions were performed, the first with Tris-HCl-equilibrated phenol (pH 8) and the second with equal parts of Tris-HCl-equilibrated phenol and chloroform (1:1). Nucleic acids were precipitated by adding 0.1 volume of 10 M LiCl₂ and 2.5 volumes of absolute ethanol; after a 30-min incubation at -20 °C, the solutions were centrifuged at 17,000 $\times$ g for 30 min at 0 °C. The resulting pellets were resuspended in DEPC-treated H₂O (88 μ l), and two successive DNase treatments were performed using Ambion Turbo DNase™ (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. A final extraction with chloroform:isoamyl alcohol (24:1, v/v) was performed on the DNase-treated solution to remove protein and residual phenol, and RNA was precipitated

from the aqueous phase with 10 M LiCl₂ and absolute ethanol as described above. The RNA was pelleted by centrifugation, washed, and resuspended in DEPC-treated H₂O (60 μ l). RNA concentrations and purity were estimated by absorbance at 260 nm and 280 nm with a NanoDrop Spectrophotometer ND-1000 (Thermo Fisher Scientific, Wilmington DE), and RNA integrity was verified by analyzing aliquots on an RNA NanoChip with the Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Samples had RNA integrity numbers averaging 5.5 (range 4.5 to 6.2), indicating that these RNA extractions were acceptable for further analyses (Schroeder *et al.*, 2006).

cDNA Synthesis

The 11:00 AM sample from 12 September was omitted from further analysis due to a processing error. The remaining 23 hourly samples were subjected to cDNA synthesis and sequencing at the Genomics Core Facility at The Pennsylvania State University (University Park, PA). The cDNA libraries were constructed from 0.5 μ g RNA samples according to the "Whole Transcriptome Library Preparation for SOLiD Sequencing" protocol (Applied Biosystems, Foster City, CA), and samples were barcoded using multiplexing barcode set B (Applied Biosystems, Foster City, CA). The SOLiD ePCR and SOLiD Bead Enrichment Kits (Applied Biosystems, Foster City, CA) were used for processing the samples, and the SOLiDTM-3.5 System (Applied Biosystems, Foster City, CA) was used for sequencing.

Alignment and Statistical Analyses of cDNA Sequences

Sequences from the cDNA libraries were assigned to metagenomic ORFs as previously described (Liu *et al.*, 2011b). Briefly, the sequences from the SOLiDTM3.5 were aligned to the metagenomic scaffold database in color space using the BWA algorithm (Li and Durbin, 2009) allowing a maximum of 5 mismatches per sequence

($\geq 90\%$ nucleotide sequence identity). A sequence was assigned to a specific gene if at least half of the sequence was aligned to the coding region of the gene. Uniquely assigned cDNA sequences were then counted for each ORF, and Fisher's exact tests were performed to determine if the transcript counts for a given gene for at least one pairwise combination of timepoints were significantly different, compared to the difference in the total number of cDNA sequences for those respective timepoints. Relative expression values were determined for each gene that was determined to have statistically significant differences in gene expression, and these relative expression values were calculated as in the previous study (Liu *et al.*, 2011b). The expression values E for a given timepoint i (E_i) were calculated using the following formula:

$$E_i = n_i / (N_i * p_i).$$

Here, n_i denotes the number of mRNA sequences assigned to a gene for a given timepoint, N_i denotes the number of total mapped sequences at that timepoint, and p_i denotes the percentage of mRNA sequences that were assigned to that particular taxonomic cluster at that timepoint. Each expression value from this formula was then normalized by the mean of the expression values for all 23-timepoints for that particular gene. This calculation was a slight departure from the previous study (Liu *et al.*, 2011b), in that p originally represented the percentage of rRNA sequences of the taxonomic cluster to which the gene had been assigned. The pilot study had employed 454 pyrosequencing as well as SOLiD sequencing platforms, and the former platform had produced sequences that were long enough (~ 250 bp) for accurate taxonomic assignments of SSU and LSU rRNAs (Liu *et al.*, 2011b). In contrast, the sequences produced by the SOLiDTM-3.5 platform averaged ~ 50 bp in length, and thus sequences that were mapped to rRNA were not analyzed further in this study due to the inability to classify them accurately.

Clustering and Visualization of Gene Expression Patterns

Normalized expression levels of genes were $\log_2$ transformed, centered by mean, and then clustered using the k-means algorithm with the program Cluster (Eisen *et al.*, 1998) ($k = 10$ and $k = 5$ for *Roseiflexus* spp. transcripts; $k = 10$ and $k = 6$ for *Chloroflexus* spp. transcripts; runs = 1000). A conservative k-means clustering approach ($k = 5$ clusters for *Roseiflexus* spp. and $k = 6$ clusters for *Chloroflexus* spp.) was chosen for further analysis. The resulting gene expression patterns in each cluster were visualized using Java Treeview (Saldanha, 2004). These k-means clusters were then assigned temporal transcription categories, such as "diurnal" patterns when they exhibited higher expression levels during the day (typically 8:00 AM to 6:00 PM), "nocturnal" patterns with higher expression levels between 6:00 PM and 8:00 AM, and "constitutive" patterns when genes in the cluster exhibited expression levels that could not be unambiguously assigned to diurnal or nocturnal groups (Figure 5.1).

Results and Discussion

Metagenomes of FAP Populations

The number of transcripts that could be assigned to *Chloroflexus* and *Roseiflexus* spp. significantly increased when sequences were mapped to the expanded sets of reference metagenomic scaffolds (Table 5.1). The 78 scaffolds of the *Roseiflexus*-like metagenomic cluster that were demarcated on the basis of oligonucleotide frequency constituted approximately 69% of the length of the *Roseiflexus* sp. strain RS-1 genome and exhibited very similar G+C content and average nucleotide identity to the genome (Table 5.1, Chapter 3). The expanded set of *Roseiflexus* spp. metagenomic scaffolds significantly increased the amount of metagenomic sequence,

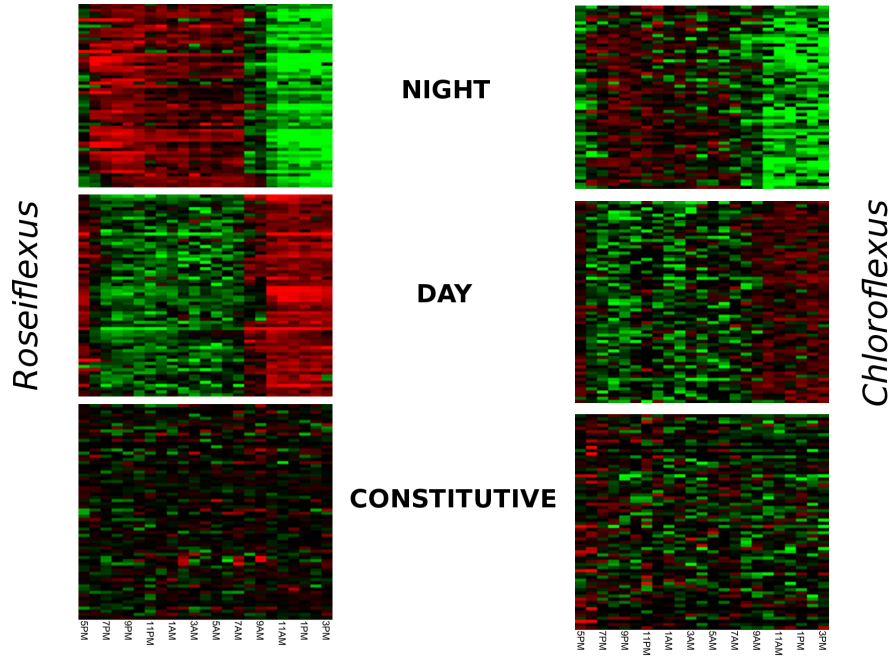


Figure 5.1: Major Transcription Categories. The normalized relative expression levels of selected genes (along the vertical axis) are indicated for multiple timepoints throughout the diel cycle (along the horizontal axis) and are colored to indicate higher (red) or lower (green) relative expression.

such that the summed length of these scaffolds was 88.6% of the total length of the *Roseiflexus* sp. strain RS-1 genome (Table 5.1). These scaffolds contained 50 novel ORF sequences that did not have reciprocal orthologs in the reference organism (Supplementary Table 1 in Appendix D). These gene differences may impart phenotypic differences to members of the *in situ* populations and the reference strain, but most are annotated as hypothetical proteins, which precludes inferences regarding their function.

While the metagenomic scaffolds associated with uncultivated *Chloroflexus* spp. are similar to the *C. aurantiacus* J-10-fl genome (Tang *et al.*, 2011), they are less related to this reference genome than to the unfinished draft genome of *Chloroflexus* sp. strain 396-1, which is sufficiently distant from the *C. aurantiacus* isolates to be

Table 5.1: Genome and Metagenome Scaffolds Used in the Analysis of Metatranscriptomes.

Organism Category	Number of Scaffolds	Amount of Sequence (Mbp)	Number of protein encoding ORFs	%G + C
<i>Roseiflexus</i> spp. cluster scaffolds	78	4.02	3688	60.2
<i>Roseiflexus</i> spp. expanded scaffolds	349	5.14	4844	59.8
<i>Roseiflexus</i> sp. strain RS-1 genome	1	5.8	4621	60.4
<i>Chloroflexus</i> spp. cluster scaffolds	18	0.4	373	55.0
<i>Chloroflexus</i> spp. expanded scaffolds	320	2.64	2561	54.4
<i>Chloroflexus</i> spp. strain 396-1 genome	81	4.86	N/A ¹	55.2
Anaerolineae-like cluster scaffolds	45	2.07	282	63.2

¹This genome has not been closed, and the equivalent bioinformatic analysis has not yet been completed.

considered a separate species (5% difference in the full-length 16S rRNA sequence) (Nübel *et al.* 2002; Bryant *et al.* 2007; Chapter 3). There were relatively fewer scaffolds attributed to *Chloroflexus* spp. that were demarcated by oligonucleotide clustering compared to the more dominant *Roseiflexus* spp. This was problematic from the perspective of mapping metatranscriptomic sequences from *Chloroflexus* spp., as the cluster scaffolds represented only $\sim 8\%$ of the length of a typical *Chloroflexus* spp. genome (which are on average 5.0 Mbp). The expanded set of scaffolds increased the summed metagenomic scaffold length nearly 7-fold (Table 5.1). *Chloroflexus* spp. scaffolds contained 74 ORFs that were unique (i.e., not reciprocal orthologs) compared to *Chloroflexus* spp. genomes (Supplementary Table 2 in Appendix D). The annotations for these genes were commonly hypothetical proteins and transposases, making it difficult to discern what ecological differences these in situ populations may exhibit with respect to cultivated *Chloroflexus* spp. isolates.

The composition of the metagenome scaffolds from *Roseiflexus* spp. was complete with respect to the presence of homologs for all of the known genes involved in phototrophy, central carbon metabolism, and electron transport that are conserved in both the *Roseiflexus* sp. RS-1 and *R. castenholzii* genomes. In contrast, metagenome scaffolds from *Chloroflexus* populations lacked many homologs that were expected due to their universal presence in other *Chloroflexus* spp. genomes (Table 5.2). Given their essential role in common metabolic processes shared by *Chloroflexus* spp., it is unlikely that these genes are missing in environmental populations, but rather that the relative level of metagenomic coverage is not high enough to assemble metagenomic scaffolds containing these genes for this group of organisms.

The scaffolds contributed from Anaerolineae-like organisms are relatively distantly related to reference genomes, which prevented the discovery of additional scaffolds beyond those of the oligonucleotide-demarcated cluster described in Chapter 3. For

Table 5.2: Expected *Chloroflexus* spp. Genes Absent in the *Chloroflexus* Metagenome Scaffolds.

Function	<i>Chloroflexus auratiacus</i> J-10-fl homolog
Polyglucose biosynthesis/degradation	
Cellulose synthase	Caur_1954
cellulase	Caur_1697
β -glucosidase	Caur_0360
	Caur_1073
Glycogen synthesis/degradation	Caur_3107
Glycolysis	
<i>pgi</i>	Caur_2179
<i>pfk</i>	Caur_2662
<i>tpiA</i>	Caur_3825
<i>gapA</i>	Caur_0010
<i>gapA</i>	Caur_3729
<i>gpm</i>	Caur_0353
<i>gpm</i>	Caur_1199
<i>eno</i>	Caur_3808
<i>pyk</i>	Caur_3128
Non-oxidative pentose phosphate pathway	
<i>rpiA</i>	Caur_3198
<i>rpe</i>	Caur_3197
<i>rbsK</i>	Caur_1720
<i>rbsK</i>	Caur_2197
Anapleurotic reactions	
<i>ppc</i>	Caur_3888
Oxidative tricarboxylic acid cycle	
<i>korD</i>	Caur_1567
<i>korB</i>	Caur_0250
<i>sucA</i>	Caur_3727
<i>sucA</i>	Caur_3726
<i>sucD</i>	Caur_0702
<i>fumC</i>	Caur_1443

continued on next page

continued from previous page

Function	<i>Chloroflexus auratiacus</i> J-10-fl homolog
Polyhydroxyalkanoate synthesis/degradation	
PHB $\leftrightarrow$ 3-hydroxybutanoyl-CoA	Caur_3263
3-hydroxybutanoyl-CoA $\leftrightarrow$ acetoacetyl-CoA	Caur_1462
Acetoacetyl-CoA $\leftrightarrow$ acetyl-CoA	Caur_1461
Acetoacetyl-CoA $\leftrightarrow$ acetoacetate	Caur_3394
Branched-chain amino acid biosynthesis	
branched-chain amino acid transferase	Caur_0488
branched-chain amino acid transferase	Caur_1435
2-isopropylmalate synthase	Caur_0166
isopropylmalate isomerase small subunit (<i>leuD</i>)	Caur_0169
3-Hydroxypropionate pathway	
Acetyl-CoA carboxylase (<i>accB</i>)	Caur_3739
Propionyl-CoA carboxylase (<i>pcc</i>)	Caur_3433
Methylmalonyl-CoA epimerase	Caur_3037
Methylmalonyl-CoA mutase	Caur_1844
succinyl-CoA-malate-CoA transferase (<i>smtA</i>)	Caur_0179
succinyl-CoA-malate-CoA transferase (<i>smtB</i>)	Caur_0178
Mesaconyl-CoA C1-C4 CoA transferase (<i>mct</i>)	Caur_0175
malyl-CoA lyase (<i>mcl</i>)	Caur_0174
mesaconyl-C1-CoA hydratase (<i>mch</i>)	Caur_0173
mesaconyl-C4-CoA hydratase (<i>meh</i>)	Caur_0180
succinyl-CoA:D-malate CoA transferase (<i>sct</i>)	Caur_2266
Glyoxylate bypass	
malate synthase	Caur_2969
Acetate metabolism	
Acetyl-CoA synthetase	Caur_0003
Alcohol dehydrogenase	Caur_2809
Alcohol dehydrogenase	Caur_0032

continued on next page

continued from previous page

Fatty Acid Metabolism

Biosynthesis

<i>fabH</i>	Caur_2406
<i>fabG</i>	Caur_3773
<i>fabG</i>	Caur_2362
<i>fabG</i>	Caur_1462
<i>fabG</i>	Caur_3262
<i>fabZ</i>	Caur_1433

β -oxidation

3-hydroxyacyl-CoA hydrolase	Caur_1346
-----------------------------	-----------

Phototrophy

<i>pufL</i>	Caur_1052
<i>pufB</i>	Caur_2091
<i>pufA</i>	Caur_2090
<i>bchG</i>	Caur_2088
<i>bchZ</i>	Caur_3806
<i>bchU</i>	Caur_0137

Oxidative stress

superoxide dismutase	Caur_1176
----------------------	-----------

Electron transport

NADH menaquinone oxidoreductase subunit (<i>nuoE</i>)	Caur_1184
alternative complex III quinone oxidoreductase subunit <i>actG</i> (Cp)	Caur_0627
cytochrome c oxidase subunit <i>cyoE</i>	Caur_0029
cytochrome c oxidase subunit <i>coxB</i>	Caur_2141

example, these scaffolds contained conserved housekeeping genes (e.g., *recA*, *rpoB* and ribosomal proteins) that exhibited 50-70% amino acid identity with the genomes of other Chloroflexi (including the genomes of *Chloroflexus* or *Roseiflexus* spp., *Oscillochloris trichoides*, *Anaerolinea thermophila*, and *Dehalococcoides* spp.).

Metatranscriptomes of FAP Populations

The transcripts detected from FAPs at hourly timepoints provide insights into how these organisms temporally regulate their gene expression, which in turn informs how these organisms respond to changing environmental conditions over a diel cycle. In the discussion that follows, it is acknowledged that transcript abundance does not imply physiological function, and all statements regarding the timing of particular metabolisms are put forward as hypotheses. The total number of transcripts that uniquely mapped to ORFs on *Roseiflexus* scaffolds (11,159,969) was 30-fold higher than the total number of *Chloroflexus* transcripts (365,812), which was notable in comparison to the 2-fold difference in metagenomic scaffold sequence contributed between these two groups (Table 5.1). While it is acknowledged that many *Chloroflexus* spp. transcripts cannot be detected due to incomplete metagenomic coverage for these organisms, it is improbable that the remaining undetected transcripts for *Chloroflexus* spp. genes could account for the 30-fold difference in transcript abundance between *Roseiflexus* and *Chloroflexus* spp. Alternatively, it is proposed that there are fewer transcripts from *Chloroflexus* spp. at the temperature at which this study was conducted (60 °C). By comparison, the metagenomic scaffolds from *Chloroflexus* spp. included more sequences that were constructed from samples taken at 65 °C, a temperature at which *Chloroflexus* spp. have been shown to be more abundant (Nübel *et al.* 2002; Chapters 2 and 3). After transcript abundance was normalized to these unique mRNA totals, it was observed that both FAP genera exhibited their

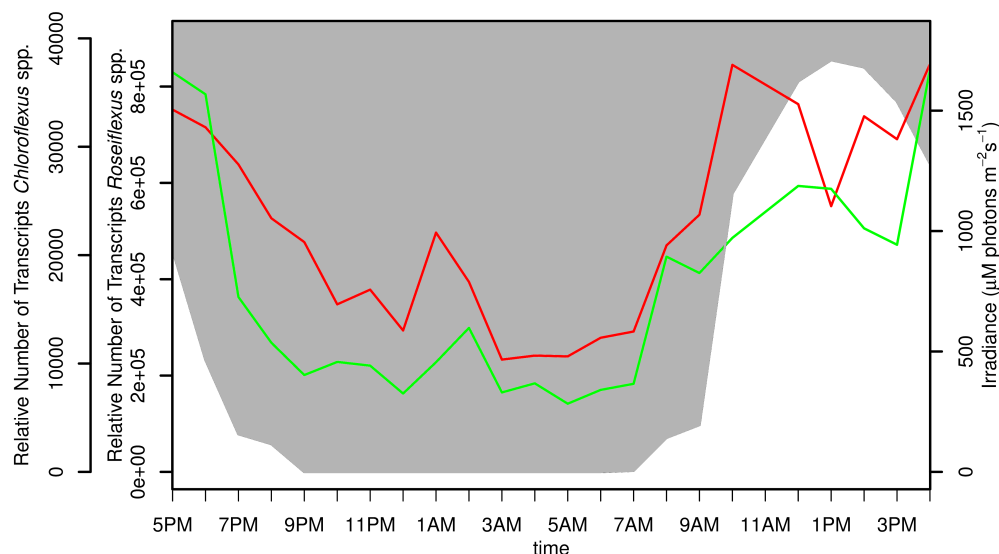


Figure 5.2: Total transcript Abundance Levels of *Roseiflexus* (red) and *Chloroflexus* (green) Transcripts. Light intensity is indicated in white.

lowest transcript levels at 7:00 AM and their highest levels at 6:00 PM (Figure 5.2). Despite differences in metagenomic coverage, *Chloroflexus* and *Roseiflexus* spp. metatranscriptomes were similar in that 97.7% of the *Roseiflexus*-like metagenomic ORFs and 97.6% of the *Chloroflexus*-like ORFs had at least one metatranscriptomic sequence uniquely mapped to them.

Three major transcription patterns, diurnal, nocturnal and constitutive, were observed in the metatranscriptomes of members off the Chloroflexi after the normalized relative expression values were subjected to k-means clustering. K-means clusters exhibiting diurnal or nocturnal patterns were more finely categorized into "strong" and "weak" patterns (dependent upon the relative difference in day and night expression levels), or into other subcategories that may be physiologically meaningful (e.g., a cluster of diurnal genes from *Roseiflexus* spp. that had increased transcript levels into the evening). Anaerolineae-like organisms had the highest proportion of diurnally expressed genes ($\sim 14:1$ diurnal:nocturnal ratio, or D:N), which supported

the hypothesis that this phototrophic bacterium is most transcriptionally active when light is available (Liu *et al.*, 2011b). While the majority of *Chloroflexus*-like genes were diurnal ($\sim 8:1$ D:N), most *Roseiflexus*-like genes had constitutive expression patterns and there was a relatively higher proportion of genes with nocturnal expression, thus the ratio of genes with diurnal to nocturnal patterns was lower for *Roseiflexus* spp. ($\sim 2:1$). While these organisms must be able to cope with both oxic and anoxic conditions in these mats, the relative degree to which they utilize aerobic or anaerobic metabolism is currently unknown.

Photosynthesis

Consistent with the prediction that FAPs perform photoautotrophy during low-light periods in the evening and early morning (Revsbech and Ward, 1984; van der Meer *et al.*, 2005), initial metatranscriptomic investigations suggested that members of the Chloroflexi transcribe genes encoding type-2 photosynthetic reaction centers (i.e., *pufLM*, homologs of Rose_3268, Caur_1052, and Caur_1051; *pufC*, RoseRS_3269 and Caur_2089) during these times (Liu *et al.*, 2011b). The higher temporal resolution afforded by the hourly sampling in the present study revealed that transcripts for the *pufLM* genes of both *Chloroflexus* and *Roseiflexus* spp. are highly abundant at night (Figure 5.3). The *pufLMC* homologs from the more distantly related Anaerolineae-like population also showed highest transcript levels during the night (Figure 5.3). These results are consistent with the patterning of transcript abundance of type-1 reaction center genes from the other anoxygenic photoheterotrophs in this mat, namely "*Ca. C. thermophilum*" and "*Ca. T. aerophilum*" (Liu *et al.*, 2011a), and are opposite of the diurnal expression of cyanobacterial photosynthesis genes (Steunou *et al.*, 2006; Liu *et al.*, 2011b). Transcripts for genes encoding proteins for chlorosomes in *Chloroflexus* spp. (*csmA*, Caur_0126; *csmM*, Caur_0139; *csmN*, Caur_0140) were

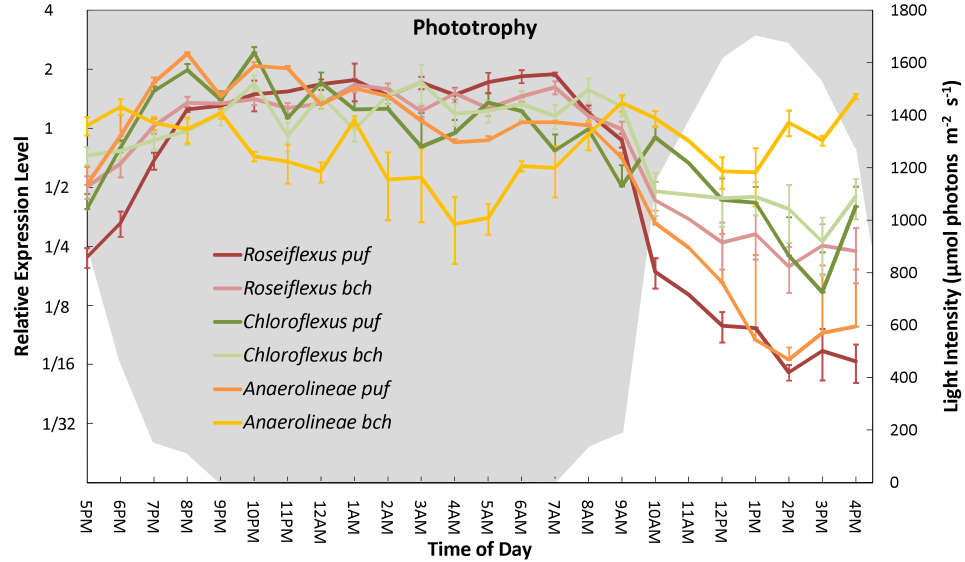


Figure 5.3: Expression of Phototrophy Genes. The mean relative expression level ($\pm$ standard error) is displayed for photosynthetic reaction center genes *pufLMC* (dark) and BChl biosynthesis genes (light) for *Roseiflexus* spp. (red), *Chloroflexus* spp. (green), and Anaerolineae-like (orange) Chloroflexi. BChl biosynthesis gene expression was the mean expression level of all BChl biosynthesis genes known in *Roseiflexus* and *Chloroflexus* genomes, while for Anaerolineae-like Chloroflexi, the mean expression was taken from *bchH*, *bchX*, *bchY*, and *bchZ* identified in previous metagenomic analyses.

also more abundant at night, which is consistent with observations of chlorosomes in cells grown anoxically in light (Sprague *et al.*, 1981).

Bacteriochlorophyll Biosynthesis

With a few exceptions, transcripts for genes involved in the biosynthesis of bacteriochlorophyll (BChl) pigments were most abundant in FAPs at night (Figure 5.3); this temporal linkage with *pufLMC* transcription is logical, as these pigment molecules are required to assemble functional photosynthetic reaction centers. Likewise, the incomplete set of Anaerolineae-like bacteriochlorophyll biosynthesis genes (*bchXYZ*, *bchD*, *bchF*, *bchH*, *bchI*) were most highly expressed at night.

While this expression of BChl biosynthesis genes under anoxic conditions is consistent with findings from anoxygenic phototrophic proteobacteria (Gregor and Klug, 1999), *Chloroflexus* and *Roseiflexus* spp. genomes lack some of the transcriptional regulatory mechanisms present in proteobacteria, such as a photosynthetic gene cluster superoperon, or homologs to the oxygen-activated transcriptional repressor *ppsR*. Despite the lack of a single photosynthesis gene cluster, some BChl biosynthesis genes are co-localized in *Chloroflexus* spp. and *Roseiflexus* spp. genomes (van der Meer *et al.* 2010; Chapter 3), and the coordinated expression exhibited in the metatranscriptome suggests that there is an undiscovered, oxygen- or redox-sensitive regulatory mechanism that is common to these organisms. Both *Chloroflexus* and *Roseiflexus* spp. have two genes predicted to be involved in the same step of BChl biosynthesis, the oxygen-dependent Mg-protoporphyrin IX monomethylester oxidative cyclase (encoded by *acsF*, Caur_2590 and RoseRS_1905) and the oxygen-independent oxidative cyclase (*bchE*, Caur_3676 and RoseRS_0942). In the purple sulfur bacterium *Rubrivivax gelatinosus*, AcsF is required for the production of BChl *a* under oxic growth conditions, while BchE is required under anoxic conditions, although the *bchE* gene is transcribed in both the presence and absence of O₂ (Ouchane *et al.*, 2004). The *acsF* and *bchE* homologs in the metatranscriptome for both *Roseiflexus* and *Chloroflexus* spp. exhibited a nocturnal expression pattern, which suggested that, unlike *R. gelatinosus*, the transcription of *bchE* in FAPs may be inhibited by oxygen. There were a few BChl biosynthesis genes that showed either diurnal (*bchY*, Caur_0417, RoseRS_3260) or constitutive (paralogs of *bchI*, Caur_1255 and RoseRS_0883; *bchH*, Caur_2591) expression patterns. The *bchY* gene, along with *bchX* and *bchZ* (which had a nocturnal expression pattern), are subunits of the light-independent protochlorophyllide reductase that reduces tetrapyrrole ring B of chlorophyllide *a*, an essential step leading to the production of BChl *a* (Nomata *et al.*,

2006). This enzyme is labile and generates superoxide in the presence of oxygen (Kim *et al.*, 2008). This observation suggests that the actual translation of this enzyme is likely to occur in anoxic conditions coordinately with the presence of transcripts for the *bchX* and *bchZ* subunits.

Electron Transport Complexes

Because the data could provide information about metabolic modes employed by FAPs at different periods throughout the diel cycle, the transcript abundances for genes encoding various proteins involved in electron transport were of particular interest. Different components of the electron transport chain may become more important at different times. For example, the need for an external source of electrons might increase when FAPs couple phototrophy with carbon fixation. *Roseiflexus* spp. contain a NiFe hydrogenase that could function to oxidize H₂ as a source of reductant for carbon fixation, and homologs of these genes (*hoxABCD*, RoseRS_2319 - RoseRS_2322) had nocturnal expression patterns, similar to the patterns observed for the *puf* and *bch* genes mentioned previously (also see discussion below regarding nitrogen metabolism).

Given the environmental fluctuations in oxygen concentration that these organisms experience, it is intuitive that they would maintain different sets of enzymes for some of the same reactions, specialized for either oxic or anoxic conditions (Bryant *et al.*, 2012; Tang *et al.*, 2011). FAP genomes contain paralogous genes encoding some of the major enzyme complexes involved in the electron transport chain, namely NADH:menaquinone oxidoreductase (Complex I) in *Chloroflexus* and *Roseiflexus* spp. (van der Meer *et al.*, 2010; Tang *et al.*, 2011), and both the Alternative Complex III, or ACIII (Yanyushin *et al.*, 2005; Gao *et al.*, 2009) and the soluble electron carrier auracyanin in *Chloroflexus* spp. (McManus *et al.*, 1992; van Driessche *et al.*, 1999;

Tsukatani *et al.*, 2007). The expression patterns of these genes are shown in Table 5.3 and are discussed below.

Respiratory Electron Transport Complexes: There were similar expression patterns within paralogous gene groups that encode different forms of NADH: menaquinone oxidoreductase in both *Chloroflexus* and *Roseiflexus* spp. (Table 5.3), with the exception of a few genes that were categorized as having constitutive expression patterns due to a weaker day or night pattern.

Chloroflexus spp. also contain two paralogous groups of genes encoding for subunits of ACIII, which function to oxidize menaquinol and donate electrons to soluble carriers such as the blue-copper protein auracyanin on the periplasmic side of the cytoplasmic membrane. In the past, these two gene sets have been named Cp (for the ACIII predicted to operate primarily for cyclical phototrophic electron transfer) and Cr (the ACIII predicted for linear respiratory electron transfer to a terminal electron acceptor such as O₂) (Yanyushin *et al.*, 2005). Interestingly, the transcript levels for both Cp and Cr genes did not show much temporal variation, and there was no evidence to suggest that *Chloroflexus* spp. modulate the transcriptional activity of their paralogous ACIII complexes in order to specialize in either phototrophic or respiratory electron transfer. *Roseiflexus* spp. contain only one set of genes encoding a Cp-like ACIII, which thus are likely to function in both phototrophic and respiratory electron transfer. Because of this predicted dual function, it was unexpected that the corresponding genes of *Roseiflexus* spp. ACIII would exhibit temporal expression patterns; however transcripts for these genes (*actABCDEFG*) were most abundant at night (Table 5.3).

Table 5.3: Expression Categories of Genes Involved in Electron Transport. Locus ID names are marked if they are specific to *Roseiflexus* spp. (*) or specific to *Chloroflexus* spp. (**). Genes that were expected but not found in either the metagenome scaffolds or did not have a significant level of uniquely mapped transcripts are also indicated (***). Expression categories were determined by labelling the dominant trends shared by clusters of genes demarcated by k-means analysis.

Gene	<i>Roseiflexus</i> sp. RS1 homolog	<i>Roseiflexus</i> expression category	<i>Chloroflexus</i> <i>aurantiacus</i> J-10-fl homolog	<i>Chloroflexus</i> expression category
NADH menaquinone oxidoreductase (Complex I)				
<i>nuoA</i>	RoseRS_2089*	weak night		
<i>nuoB</i>	RoseRS_2090*	***		
<i>nuoC</i>	RoseRS_2091*	strong day		
<i>nuoD</i>	RoseRS_2092*	***		
<i>nuoE</i>	RoseRS_3543	strong night	Caur_1184	***
<i>nuoF</i>	RoseRS_3542	***	Caur_1185	night
<i>nuoA</i>	RoseRS_2989	constitutive	Caur_1987	constitutive
<i>nuoB</i>	RoseRS_2990	weak night	Caur_1986	night
<i>nuoC</i>	RoseRS_2991	weak night	Caur_1985	night
<i>nuoD</i>	RoseRS_2992	weak night	Caur_1984	constitutive
<i>nuoI</i>	RoseRS_2993	weak night	Caur_1983	***
<i>nuoH</i>	RoseRS_2994	constitutive	Caur_1982	constitutive
<i>nuoJ</i>	RoseRS_2995	weak day	Caur_1981	constitutive
<i>nuoK</i>	RoseRS_2996	constitutive	Caur_1980	constitutive
<i>nuoL</i>	RoseRS_2997	constitutive	Caur_1979	day
<i>nuoM</i>	RoseRS_2998	constitutive	Caur_1978	night
<i>nuoM</i>	RoseRS_2999	***	Caur_1977	constitutive
<i>nuoN</i>	RoseRS_3000	***	Caur_1976	4:00 PM spike
<i>nuoA</i>	RoseRS_3678	strong day	Caur_2896	day
<i>nuoB</i>	RoseRS_3677	strong day	Caur_2897	day
<i>nuoC</i>	RoseRS_3676	strong day	Caur_2898	constitutive
<i>nuoD</i>	RoseRS_3675	strong day	Caur_2899	day

continued on next page

continued from previous page

Gene	<i>Roseiflexus</i> sp. RS1 homolog	<i>Roseiflexus</i> expression category	<i>Chloroflexus</i> <i>aurantiacus</i> J-10-fl homolog	<i>Chloroflexus</i> expression category
NADH menaquinone oxidoreductase (Complex I) <i>continued</i>				
<i>nuoE</i>	RoseRS_2238	strong day	Caur_2900	day
<i>nuoF</i>	RoseRS_2237	strong day	Caur_2901	day
<i>nuoG</i>	RoseRS_2236	strong day	Caur_2902	4:00 PM spike
<i>nuoH</i>	RoseRS_2235	strong day	Caur_2904	4:00 PM spike
<i>nuoI</i>	RoseRS_2234	constitutive	Caur_2905	4:00 PM spike
<i>nuoJ</i>	RoseRS_2233	weak day	Caur_2906	4:00 PM spike
<i>nuoK</i>	RoseRS_2232	strong day	Caur_2907	4:00 PM spike
<i>nuoL</i>	RoseRS_2231	strong day	Caur_2908	day
<i>nuoM</i>	RoseRS_2230	weak day	Caur_2909	day
Alternative complex III menaquinol/auracyanin oxidoreductase				
<i>actA</i> (Cp)	RoseRS_4139	weak night	Caur_0621	constitutive
<i>actB</i> (Cp)	RoseRS_4140	weak night	Caur_0622	constitutive
<i>actC</i> (Cp)	RoseRS_4141	weak night	Caur_0623	constitutive
<i>actD</i> (Cp)	RoseRS_4142	weak night	Caur_0624	***
<i>actE</i> (Cp)	RoseRS_4143	weak night	Caur_0625	constitutive
<i>actF</i> (Cp)	RoseRS_4144	weak night	Caur_0626	day
<i>actG</i> (Cp)			Caur_0627**	***
<i>actB</i> (Cr)			Caur_2136**	day
<i>actE</i> (Cr)			Caur_2137**	4:00 PM spike
<i>actA</i> (Cr)			Caur_2138**	constitutive
<i>actG</i> (Cr)			Caur_2139**	constitutive
SC01/SenC e-transport?			Caur_2140**	day
Auracyanin				
auracyanin A	RoseRS_2366	weak day	Caur_3248	constitutive
auracyanin B			Caur_1950**	constitutive
Cytochrome c oxidase (Complex IV)				
<i>cyoE</i>	RoseRS_0224	weak day	Caur_0029	***
COX II	RoseRS_2263	weak day	Caur_2141	***
COX I	RoseRS_2264	strong day	Caur_2142	day
COX III	RoseRS_2265	***	Caur_2143	day
COX IV (cyoD)	RoseRS_2266	constitutive	Caur_2144	***
COX I	RoseRS_0934	strong night	Caur_2426	***
COX II	RoseRS_0933	strong night	Caur_2425	***

Aerobic respiration in these organisms requires that they use a terminal cytochrome *c* oxidase. *Chloroflexus* spp. contain two paralogs of cytochrome *c* oxidase (COX III and COX IV, homologous to Caur_2143 and Caur_2144), which showed a diurnal transcription pattern (Table 5.3). Transcripts were detected for more genes encoding subunits of cytochrome *c* oxidase from *Roseiflexus* spp., and different paralogs exhibited either diurnal and nocturnal patterns (Table 5.3).

Soluble Electron Carriers: *Chloroflexus* spp. genomes contain two paralogs of the soluble blue-copper protein auracyanin, which have been labeled auracyanin A and B (*aurA*, Caur_3248 and *aurB*, Caur_1950). Similar to the Cp and Cr paralogs of ACIII, these proteins have been hypothesized to function during phototrophic (AurA) and respiratory (AurB) electron transfer, based upon the absence of AurA in cultures grown aerobically in the dark (Lee *et al.*, 2009). The transcript levels for the *aurA* and *aurB* genes of *Chloroflexus* spp. were relatively constant over a diel cycle (Table 5.3); additional work is needed to verify whether there are differences in the expression of AurA in situ. *Roseiflexus* spp. contain only one gene for auracyanin, and its transcript levels were highest during the day. Very little is currently known about the regulation of electron transport in FAPs, and continued proteomic characterization of this community could indicate whether the abundance of proteins correlates with the observed transcription patterns (Steinke *et al.*, 2011).

Mixotrophy and the TCA/3-OHP Cycles

The 3-OHP bi-cycle was discovered and characterized as an autotrophic pathway in *C. aurantiacus* cultures (Holo and Sirevåg, 1986; Strauss and Fuchs, 1993), and studies utilizing isotopic labeling have suggested that FAPs in these mats incorporate inorganic carbon in the morning (van der Meer *et al.*, 2005). In contrast, the oxidative

TCA cycle is of importance in chemoorganoheterotrophic metabolism, and cultures of FAPs have all shown the capacity to respire organic compounds under dark anoxic conditions. Thus, it was thought that FAPs in these natural environments primarily fix inorganic carbon during low light conditions when H_2 is available, then switch to photoheterotrophic metabolism during the day, and aerobically respire organic compounds at night when O_2 is available near the mat surface (van der Meer *et al.*, 2005). It had even been proposed that FAPs migrate to the surface of the mat at night when O_2 is only available via diffusion from the overlying water (Brock, 1978). Contrary to these previous models, we have suggested that FAPs in these natural environments are more likely utilizing both the 3-OHP and the TCA cycles as mixotrophic pathways, which results in the simultaneous incorporation of organic and inorganic carbon (Chapter 2, Bryant *et al.* 2012; Zarzycki and Fuchs 2011).

The TCA cycle is intimately linked with the 3-hydroxypropionate bi-cycle; two enzymes (succinyl-CoA dehydrogenase and fumarate hydratase) and three metabolites (succinyl-CoA, fumarate, and malate) are shared by these cycles, and glyoxylate forms an intermediate of both the glyoxylate bypass of the TCA cycle and the 3-hydroxypropionate bi-cycle (Figure 5.4). Transcripts for genes encoding enzymes of the TCA cycle and the glyoxylate bypass were higher during the day for both *Roseiflexus* and *Chloroflexus* spp. populations; likewise, genes for key steps in the 3-hydroxypropionate bi-cycle had diurnal expression patterns for *Roseiflexus* spp. A putative operon occurs in *Roseiflexus* spp., which contains genes encoding the enzymes acetyl-CoA carboxylase, malonyl-CoA reductase, and propionyl-CoA synthase (RoseRS_3199 - RoseRS_3203, see Chapter 2). Transcripts for these genes, which are involved in the first three steps of the 3-OHP pathway, were all more abundant during the day (Figure 5.5A).

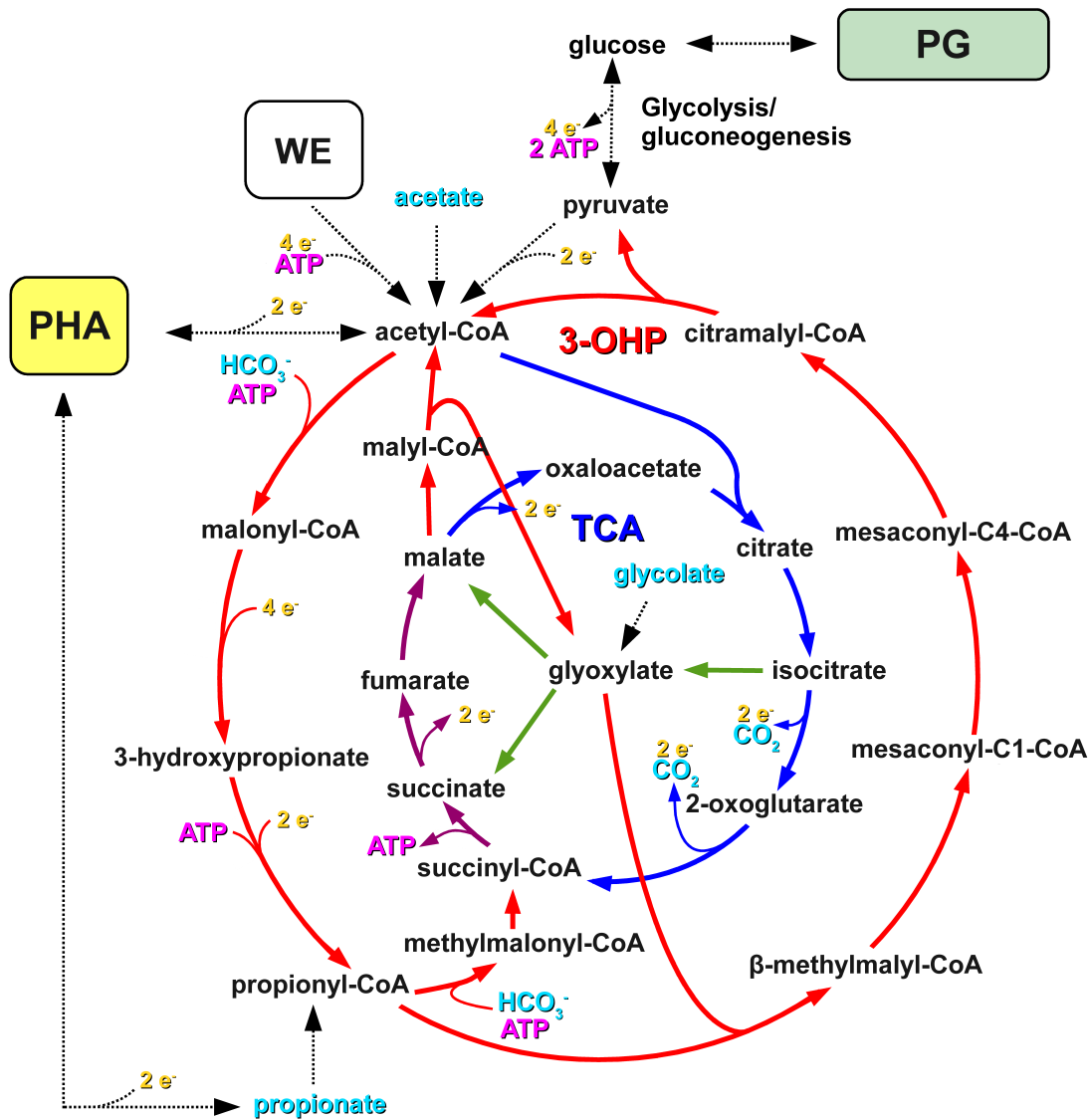


Figure 5.4: The Integrated TCA and 3-OHP Pathways for Mixotrophic Metabolism. The TCA cycle (blue) operates in the oxidative direction, while the 3-OHP cycle (red) reduces inorganic carbon. Shared steps are in purple, and the glyoxylate bypass is indicated in green. Metabolites indicated in light blue are substrates that can be obtained from outside the cell. PHA = polyhydroxyalkanoates, PG = polyglucose, WE = wax esters.

Chloroflexus spp. homologs of genes encoding acetyl-CoA carboxylase also showed diurnal or constitutive expression patterns, but malonyl-CoA reductase exhibited a nocturnal pattern. The coordinated transcript patterns of key genes in the 3-OHP and TCA cycles indicate that this may be a way in which *Roseiflexus* spp. incorporate organic acids (glycolate $\rightarrow$ glyoxylate, acetate $\rightarrow$ acetyl-CoA, and propionate $\rightarrow$ propionyl-CoA) while they simultaneously produce key substrates for anabolic pathways (i.e., 2-oxoglutarate, succinyl-CoA, and oxaloacetate) and reduce the loss of carbon as CO₂ or the need for an external electron acceptor (Figure 5.5).

Alternative Reactions Involving CO₂

Many other enzymes that are not involved in the 3-hydroxypropionate pathway have the potential to either incorporate or release inorganic carbon, depending upon the direction of the reaction. One such enzyme is pyruvate:ferredoxin oxidoreductase (PFOR, EC 1.2.7.1), which has the potential to convert acetyl-CoA and bicarbonate to pyruvate; however, it more typically operates in the reverse (oxidative) direction. Two different enzymes catalyze the reaction converting pyruvate to acetyl-CoA. Pyruvate dehydrogenase (PDH, ECs 1.2.4.1, 2.3.1.12, and 1.8.1.4) is an enzyme complex typically found in aerobic organisms, and PFOR (EC 1.2.7.1) is typically observed in organisms with anaerobic metabolism (Buckel and Golding, 2006; Tang *et al.*, 2011). Consistent with the presence or absence of oxygen, the transcripts for *nifJ/por* (PFOR) genes of both *Chloroflexus* and *Roseiflexus* spp. were most abundant at night, whereas transcripts for the PDH genes were highest during the day. While PFOR is hypothetically a reversible enzyme, if there is not a source of reduced ferredoxin available, it is energetically unfavorable for this reaction to operate in the direction of pyruvate synthesis (and CO₂ incorporation). Thus, without additional information regarding how FAPs produce reduced ferredoxin, we assume that

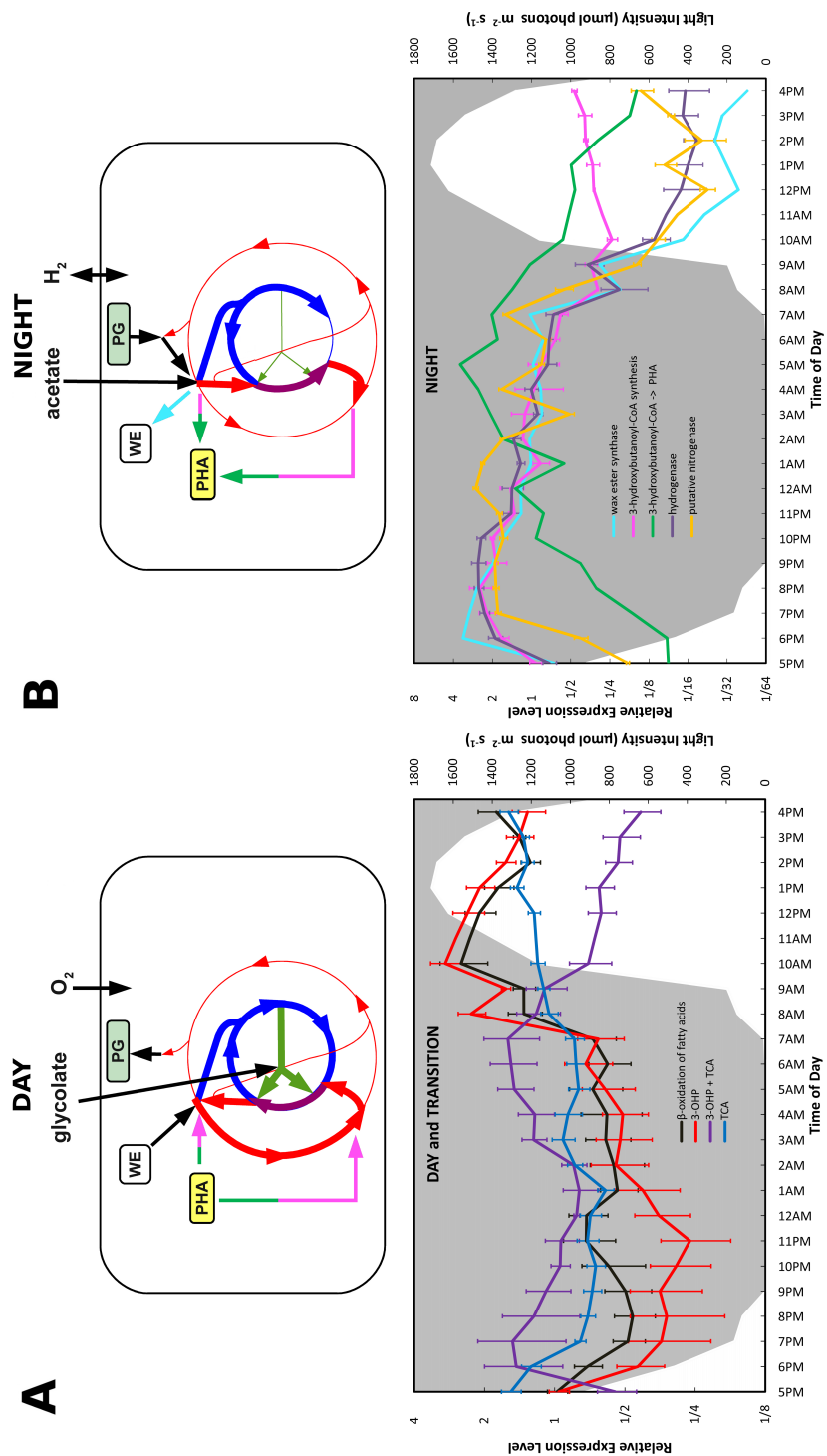


Figure 5.5: A Diel Model of Central Carbon Metabolism in *Roseiflexus* spp. The top panel displays a simplified diagram of Figure 5.4, where bold arrows indicate the predicted flow of carbon through the 3-OHP/TCA cycles and related pathways. The bottom panel shows transcription patterns for relevant genes for these pathways. A) Genes with diurnal transcription patterns such as malonyl-CoA reductase (*mcr*) and propionyl-CoA synthase (*pcs*) were averaged for the 3-OHP bi-cycle (red), malonyl-CoA mutase and malonyl-CoA epimerase were averaged to indicate the expression of shared components of the TCA and 3-OHP cycles (purple), and the remaining genes of the TCA cycle were averaged (blue). B) Nocturnally expressed genes are shown as the mean expression values of those encoding subunits of hydrogenase (*hoxABCD*) and the putative nitrogenase (*nifHBDK*). Genes involved in PHB synthesis/degradation (including multiple paralogs of β -ketothiolase and acetoacetyl-CoA reductase) are represented by 3-hydroxybutanoyl-CoA synthase. Normalized relative expression for wax ester synthase and PHA synthase are displayed individually.

PFOR likely operates in the direction of pyruvate decarboxylation in these organisms. Another anaplerotic carboxylation reaction catalyzed by phosphoenolpyruvate (PEP) carboxylase (*ppc*, E.C. 4.1.1.31) is predicted to occur in *Chloroflexus* and *Roseiflexus* spp. genomes and may provide an additional way in which inorganic carbon is fixed in these organisms. The transcripts for genes encoding PEP carboxylase (homologs of RoseRS_2753 and Caur_3161) are more abundant in the day for *Chloroflexus* spp. and have a constitutive pattern in *Roseiflexus* spp., concomitant with the daytime transcript abundance of genes involved in the 3-hydroxypropionate pathway and TCA cycle. If this reversible enzyme is primarily operating in the PEP-producing direction, which is highly plausible given the co-transcription of genes involved in glycogen and cellulose synthesis (see below), this may also be an important step to consider when estimating CO₂-fixing potential. Transcripts for a gene encoding a third potential anaplerotic reaction catalyzed by PEP carboxykinase (*pckA*, E.C. 4.1.1.32, Rose_2496 and Caur_2331) were highest at night in both of these organisms. This implies that *Roseiflexus* spp. may direct carbon flux through an oxaloacetate intermediate resulting in CO₂ release during the night.

Glycolysis/Gluconeogenesis

Past work has revealed that polyglucose levels fluctuate in mat organisms over a diel cycle, such that mat samples enriched in either *Synechococcus* spp. or FAPs accumulate glycogen during the day, and subsequently degrade it at night (van der Meer *et al.*, 2007). *Chloroflexus* and *Roseiflexus* spp. scaffolds both contain genes involved in glycogen storage and utilization. Consistent with observations of fluctuating polyglucose levels in the mat, the nocturnal expression of the gene encoding pyruvate kinase (RoseRS_1428), which catalyzes the unidirectional ATP-generating, substrate-level phosphorylation step in glycolysis, was taken as evidence that *Rosei-*

flexus spp. route carbon through glycolysis at night or in the early morning. The metagenome of *Chloroflexus* spp. did not contain a homolog of this gene (Table 5.2), presumably due to lower sequencing depth-of-coverage for this group. Other genes encoding steps in glycolysis/gluconeogenesis were bidirectional, used in both pathways, or they did not exhibit strictly diurnal or nocturnal transcription patterns. For example, *Roseiflexus* spp. contain a novel bifunctional fructose 1,6 bisphosphate phosphatase/aldolase (RoseRS_2049; Say and Fuchs 2010) which catalyzes key steps in both gluconeogenesis (phosphatase, E.C. 3.1.3.11) and glycolysis (aldolase, E.C. 4.1.2.13). Transcripts for this gene were found to be more abundant at night; however, the dual function of the corresponding enzyme precludes predictions regarding the potential effect upon temporal flux to or from stored glycogen.

Heterotrophic Carbon Assimilation and Storage

FAPs take up low-molecular weight organic compounds, such as acetate and propionate, during either photoheterotrophic or chemoorganotrophic metabolism, and these acids must be converted to acyl-CoA derivatives in order to be utilized by other metabolic reactions. Genes catalyzing the conversion of acetate to acetyl-CoA (acetyl-CoA synthetase, EC 6.2.1.1, RoseRS_2003) had constitutive expression patterns for *Roseiflexus* spp. (a *Chloroflexus* spp. homolog was not detected, Table 5.2). This observation suggests that this enzyme may allow acetate to be used to replenish acetyl-CoA throughout the diel cycle.

Acetyl-CoA and other acyl-CoA derivatives also serve as crucial intermediary metabolites for the biosynthesis of polyhydroxyalkanoic acid (PHA), a common carbon and electron storage compound, which is known to be produced by FAPs. Transcripts for one paralog of 3-ketothiolase from *Roseiflexus* spp. were more abundant at night (RoseRS_4348; Figure 5.5), as were homologs for the two remaining steps

in PHA biosynthesis: acetoacetyl-CoA reductase (EC 1.1.1.36, RoseRS_4347), and polyhydroxyalkanoate synthase (EC 2.3.1.-, RoseRS_4553). The transcripts for these three genes were temporally offset such that there was an increase in transcripts for 3-ketothiolase and acetoacetyl-CoA reductase in the evening (5:00 PM to 10:00 PM; pink line in 5.5B) followed by an increase in transcripts for PHA synthase in the (peak at 5:00 AM, green line in 5.5B). These transcript patterns are consistent with the hypothesis that *Roseiflexus* spp. are building PHA at night. Metagenomic coverage was limited in the case of *Chloroflexus* spp., and homologs of genes encoding the enzymes for these latter steps were not observed (Table 5.2).

The production of PHAs at night could potentially be commensurate with the breakdown of glycogen in these organisms. It has been proposed that some anaerobic bacteria produce PHA by incorporating acetate and reducing it as described above, but they also obtain supplemental acetyl-CoA, ATP, and reducing power from stored polyglucose via glycolysis (Hesselmann *et al.*, 2000). Pyruvate that is produced from glycolysis then enters a branched TCA cycle, in which it is converted to 2-oxoglutarate via the first three steps of the oxidative TCA cycle. This 2-oxoglutarate could then be used as a precursor for BChl biosynthesis, and thus very little 2-oxoglutarate dehydrogenase activity would be expected. Phosphoenolpyruvate produced from glycolysis could simultaneously be converted to oxaloacetate from the reaction catalyzed by PEP carboxylase mentioned above, and it could then be reduced on the opposite branch of the branched TCA cycle. The reversible steps catalyzed by malate dehydrogenase, fumarate hydratase, succinate dehydrogenase, and succinyl-CoA synthase could reductively convert oxaloacetate to succinyl-CoA. This intermediate might then enter the methylmalonyl pathway to form propionyl-CoA, a precursor for PHA biosynthesis. Finally, some of the acetyl-CoA produced from PFOR could be directly incorporated into PHA. Using this pathway, FAPs can produce both polyhydroxybutarate (from

2 acetyl-CoA + 2 e⁻) or polyhydroxyvalerate (1 acetyl-CoA + 1 propionyl-CoA + 2e⁻). This proposed pathway would allow FAPs to build PHA at night for carbon and energy storage (Figure 5.5A, with electrons and acetyl-CoA released from the fermentation of stored polyglucose. *Roseiflexus* spp. could also obtain acetate from cyanobacterial fermentation (Nold and Ward, 1996). Below the surface of the mat, oxygen levels are below detection limits (<1 μ M) at night, and this metabolic strategy would allow *Roseiflexus* spp. to regenerate NADP⁺ obviating the need for an external electron acceptor and retaining most of the carbon from glucose, while simultaneously building BChl molecules from 2-oxoglutarate. When O₂ is plentiful during the day, the subsequent degradation of PHA (possibly by the paralogs of 3-ketothiolase that exhibited diurnal expression) would release carbon and electrons for use in the operation of the combined TCA and 3-OHP cycles, when acetate and electron donors are more scarce due to the lack of cyanobacterial fermentation, and competition with aerobic chemoorganoheterotrophs for these compounds.

Wax esters represent another potential class of carbon and electron storage compounds that FAPs produce (Shiea *et al.*, 1991), and isotopic labeling studies have shown that inorganic carbon is incorporated into FAP wax esters in the morning (either indirectly via cross-feeding from cyanobacteria, or directly by FAPs) but not in the afternoon (van der Meer *et al.*, 2005). These wax esters could be utilized by FAPs as a carbon and energy source, and the degradation of this storage compound would be favorable under conditions when O₂ can be used as a terminal electron acceptor. Consistent with this prediction, transcripts for genes encoding enzymes for the β -oxidation of fatty acids were universally most highly abundant during the day for both *Roseiflexus* and *Chloroflexus* spp. (Figure 5.5B). Both *Roseiflexus* and *Chloroflexus* spp. exhibited constitutive expression of fatty acid biosynthesis genes, and transcripts for a *Roseiflexus* sp. gene homologous to wax ester synthase

(RoseRS_2456) were most highly abundant at night (Figure 5.5B). A corresponding ortholog of this wax ester synthase was not detected in any *Chloroflexus* spp. genome, and it is currently unknown how *Chloroflexus* spp. produce wax esters. Caution is warranted regarding inferences of photoautotrophy in the morning based upon the incorporation of labeled bicarbonate into wax esters (van der Meer *et al.*, 2005), because these observations are dependent upon when wax esters are produced in these cells. Bicarbonate labeling studies that did not involve compound-specific labeling suggested significant incorporation by *Roseiflexus* spp. during the day (van der Meer *et al.*, 2007), and this could potentially be driven by daytime photomixotrophy.

During the day, glycolate is produced from photorespiration by cyanobacteria when the mat is highly oxic (Bassham and Kirk, 1962; Bateson and Ward, 1988), and past work has demonstrated that FAPs assimilate this organic compound (Bateson and Ward, 1988). Both *Roseiflexus* and *Chloroflexus* spp. encode homologs to glycolate oxidase (*glcD*, RoseRS_3360, Caur_2132) that had diurnal expression patterns, and this gene would convert glycolate to glyoxylate, a key intermediate in the central metabolism of FAPs.

Nitrogen and Hydrogen Metabolism

Chloroflexus and *Roseiflexus* spp. differ in their acquisition of nitrogen, which is an essential nutrient for the biosynthesis of proteins, nucleic acids, and BChls. *Chloroflexus* spp. does not possess the genes for dinitrogen fixation, but *Roseiflexus* spp. contain homologs of the nitrogenase genes (*nifHBDK*, RoseRS_1201 - RoseRS_1198), and transcripts for these genes were more abundant at night (Figure 5.5B). Transcript levels for *nif* genes in *Synechococcus* spp. were also most abundant at night, and nitrogenase activity has been detected during the night and in the early morning (Steunou *et al.*, 2006). Hydrogen generation by cyanobacterial nitrogen fixation would

be an important source of electrons for hypothesized photoautotrophic metabolism in FAPs. Genes encoding subunits of a [Ni-Fe] hydrogenase in *Roseiflexus* spp. exhibited a nocturnal pattern (Figure 5.5B).

Roseiflexus spp. also contain a homolog for an ammonium transporter (*amtB*), which had a diurnal expression pattern. The expanded set of metagenomic scaffolds for *Chloroflexus* spp. contains two homologs of *amtB* (Caur_1002), and transcript levels for both were also highest during the day. *Chloroflexus* spp. may also assimilate nitrate as a nitrogen source, which is reduced to nitrite by a putative *narG* homolog (Caur_3201) that had a nocturnal pattern. The transcripts of a *Roseiflexus* spp. homolog of *narG* (RoseRS_1793) were highest during the day, which may indicate that *Roseiflexus* and *Chloroflexus* spp. do not compete directly for nitrate. The temporal-offset of transcript abundance for the same activity (nitrate reduction) coupled with the potential for *Roseiflexus* spp. nitrogen fixation both illustrate ways in which *Chloroflexus* and *Roseiflexus* spp. could acquire the same resource using different ecological strategies.

Conclusions

While the functions of uncultivated Chloroflexi inhabiting Mushroom Spring cannot explicitly be inferred from gene expression patterns, such patterns nevertheless provide evidence for the regulation of metabolic functions at the transcriptional level and provide the basis for modeling the metabolic responses of these organisms to environmental stimuli over a diel cycle. The results presented here lead to the following hypotheses about FAP metabolism during the diel cycle. FAPs utilize photomixotrophy during the day, either by degrading internal carbon storage polymers such as wax esters and PHAs to obtain metabolic intermediates and electrons, or by incorporating

and metabolizing glycolate crossed from cyanobacteria. Both the TCA and the 3-OHP pathways are predicted to function for central carbon metabolism; the resulting ATP and electrons can be applied to gluconeogenesis for polyglucose storage, and some CO₂ produced from the TCA cycle can be reduced by the 3-hydroxypropionate cycles. During the transition between light and dark periods, FAPs are predicted to utilize photomixotrophic metabolism; they reduce their need for external electron acceptor (O₂) by using cyclical phototrophic electron flow. As light and oxygen levels decrease and phototrophy diminishes, FAPs couple fermentation of their stored polyglucose to the synthesis of PHAs and possibly wax esters. Hydrogenase activity during the night may act as an electron valve for the disposal of excess reductant, however many electrons may be retained in the production of PHAs. The branched TCA cycle operates during this time to simultaneously produce both succinyl-CoA, which can be converted to propionyl-CoA and thus PHAs, and 2-oxoglutarate for BChl biosynthesis, which in turn is applied to the production of photosynthetic reaction centers and antennae structures. At dawn, oxygen concentrations in the illuminated portions of the mat are low (Jensen *et al.*, 2011) and FAPs are predicted to utilize photomixotrophic growth. During this time, fermentation and nitrogen fixation from *Synechococcus* spp. produce H₂ that can be cross-fed to FAPs, thus providing an external supply electrons for the reduction steps in the 3-hydroxypropionate pathway. The differential timing of genes encoding hydrogenase and enzymes of the 3-OHP bi-cycle seem to contrast with the hypothesis that FAPs are photoautotrophic in the early morning. The production and degradation of carbon and energy storage polymers are predicted to be a central component of FAP physiology, in which they provide metabolic resources such as carbon and electrons at times when these resources are not available externally.

CHAPTER 6

CONCLUSIONS AND RELATION TO OTHER COLLABORATIVE WORK

The work presented in this dissertation has advanced the knowledge of the community context and physiological ecology of phototrophic Chloroflexi bacteria in two major ways. First, the co-inhabiting community members and their functional potential have been described for multiple FAP-dominated communities under different geochemical contexts. Second, detailed genomic, metagenomic, and transcriptional data have revealed the genetic potential and temporal regulation of key physiological functions in these organisms. The major findings from this work include *i)* there are novel Chloroflexi and Chlorobi phototrophic bacteria in some of these environments, *ii)* that the detection of *Chloroflexus* and *Roseiflexus* spp. across a wide variety of geothermal environments (e.g., *Roseiflexus* spp. had not previously been reported in mats from Bath Lake Vista Annex or Chocolate Pots springs), and *iii)* that *Roseiflexus* spp. may be significant catalysts of inorganic carbon fixation with their capacity for mixotrophic metabolism discovered herein.

It is still poorly understood how the relative abundances of either *Chloroflexus* or *Roseiflexus* spp. are affected by changes in geochemical conditions, or by the presence of co-inhabiting community members. Based on the presence of bacteiochlorophyll *c* and early cultivation efforts (Giovannoni *et al.*, 1987), it was previously thought that *Chloroflexus* spp. were dominant in geothermal sites with sulfide concentrations ≥ 30 μM (Castenholz, 1977; Ward *et al.*, 1989b; Castenholz and Pierson, 1995). Cultivation studies have shown that *Roseiflexus* spp. can grow at sulfide concentrations up to 100 μM (van der Meer *et al.*, 2010), and the application of molecular sequencing approaches has revealed that *Roseiflexus* spp. are dominant members of these high-

sulfide communities as well (Chapter 4). It is not yet clear whether *Roseiflexus* spp. utilize sulfide as an electron donor for carbon fixation, and by which physiological mechanism if they do.

Metagenomic and metatranscriptomic approaches have given unparalleled insight into the presence and regulation of key metabolisms for uncultivated bacteria in these springs. Metagenomics enabled the discovery of three new phototrophic bacteria, namely *Ca. Chloracidobacterium thermophilum* (Bryant *et al.*, 2007), *Ca. Thermochlorobacter aerophilum* (Liu *et al.*, 2011a), and novel phototrophic *Chloroflexi* (Chapter 3). Additionally, this technique was used to describe other dominant chemoorganotrophic community members that had not been previously characterized. As coverage depth and annotations for the genes contained by these chemorganotrophic groups are improved, their trophic roles can be determined in more detail. The metagenomic characterization of the communities studied has shown that *Roseiflexus* spp. contain the key genes involved in the 3-hydroxypropionate pathway in all the sites they inhabited (Chapters 2 and 3).

Metatranscriptomes of the Mushroom Spring community could not have been interpreted without initial metagenomic sequencing, such that transcripts could not have been assigned to many of the dominant community members without the contextual information that metagenomic scaffolds contained. Once transcripts were properly assigned, the expression of *Roseiflexus* spp. 3-OHP pathway genes provided an additional level of evidence that these organisms utilize this pathway. It was unexpected that two of the genes encoding key enzymes of this pathway, namely, propionyl-CoA synthase and malonyl-CoA reductase, were most highly expressed during the day.

While it is acknowledged that metatranscriptomics cannot determine whether a particular function is occurring at the same time that a corresponding gene is

transcribed, initial work using metaproteomic techniques with samples from Octopus Spring and Mushroom Spring done in collaboration with Dr. Laurey Steinke (University of Nebraska Medical Center) has provided information about which enzymes are present at a given time, thus overcoming interpretive limitations due to temporal incongruence between transcription and translation (Schaffert *et al.*, 2011). Examples of discontinuity between the metatranscriptome and metaproteome exist, such as the daytime lack of peptides from malonyl-CoA reductase and propionyl-CoA synthase (Mcr and Pcs), as these enzymes have not been detected in the daytime top green layer proteomic samples. Although it is unlikely that these peptides are completely absent, they are significantly underrepresented compared to the presence of chaperones such as GroEL or those of the cpn10 family. Interestingly, peptides were detected for malonyl-CoA reductase during the day in a proteome constructed from subsurface layers (samples in which the top 2mm of mat material was removed). These proteomes were sampled with similar depth of coverage, where cyanobacteria are not as active (i.e., fewer cyanobacterial peptides were detected in subsurface samples) and less oxic conditions persist due to steep light gradients in the mat, even during periods of high light at the mat surface (Jensen *et al.*, 2011). It is possible that these enzymes are translated for later afternoon/evening activity in the top layers of the mat, when oxygen concentrations are diminished and more reductant is available for the incorporation of inorganic carbon. Ongoing analysis of deeper-coverage proteomes taken during the entire diel cycle will provide a more definitive answer as to when this pathway is active.

The transcriptional regulation of the 3-hydroxypropionate pathway concomitant with genes involved in heterotrophic assimilation of organic compounds (Chapter 5) suggests that *Chloroflexus* and *Roseiflexus* spp. are never strictly autotrophic in their natural habitats, as a source of low molecular weight dissolved organic carbon

is likely to be available to these organisms most of the time. Mixotrophy complicates the inferences that have been made regarding the stable isotopic composition of lipid biomarkers (van der Meer *et al.*, 2003), as the reference used to interpret the heavier isotopic composition of FAP biomarkers were cultures of autotrophically grown *Chloroflexus aurantiacus* (Holo and Sirevåg, 1986; van der Meer *et al.*, 2001). One other possible contribution to the heavier composition of FAP biomarkers could be the cross-feeding of acetate from fermenting cyanobacteria. Bulk polyglucose has been shown to exhibit relatively heavy isotopic composition ($\delta^{13}\text{C} \sim -10 \text{ ‰}$), and the fermentation of this storage compound by *Synechococcus* spp. would not exhibit a noticeable fractionation pattern compared to DIC (van der Meer *et al.*, 2007). It remains to be determined if the isotopic composition remains unchanged if *Chloroflexus aurantiacus* are grown mixotrophically, with varying amounts of organic carbon and HCO_3^- . Furthermore, many of the FAP biomarkers previously studied such as C_{32} - C_{35} wax esters are shared between both *Chloroflexus* and *Roseiflexus* spp. (van der Meer *et al.*, 2002, 2010), such that nucleic acid biomarkers (or the proteins they encode) could provide a more informative basis from which to discern the relative contributions to inorganic carbon fixation among organisms in these genera.

Inorganic carbon assimilation can be determined directly with stable isotope probe (SIP)-labeling experiments, as was done previously with lipid biomarkers (van der Meer *et al.*, 2005, 2007). In the last decade, both RNA-SIP and DNA-SIP have become standard approaches for identifying the constituents of particular functional groups within microbial communities (Manefield *et al.*, 2002; Lueders *et al.*, 2004; Buckley *et al.*, 2007). During an Integrative Graduate Education and Research Traineeship (IGERT)-sponsored collaboration with Dr. James Prosser (University of Aberdeen), I investigated these approaches to characterize the mixotrophic FAPs capable of HCO_3^- assimilation. A series of experiments was done at Mushroom Spring,

in which mat cores were incubated in vials with or without $\text{H}^{13}\text{CO}_3^-$. These cores were processed to extract RNA; the ^{12}C - and ^{13}C -RNAs were separated using isopycnic centrifugation on a cesium trifluoroacetate (CsTFA) gradient, and the fractions collected over the gradient were subsequently reverse transcribed. The cDNAs from these heavy RNA fractions were used as templates for PCR amplifications of 16S rRNA sequences using general primers targeting domain Bacteria, as well as those specific for *Chloroflexi*. Unfortunately, insufficient quantities of RNA were recovered from these CsTFA fractions, and the PCR amplifications were not successful despite successful PCR amplification of cDNA from uncentrifuged RNA. Another promising application of SIP has been achieved with the detection of ^{13}C -labeled peptides (Steinke *et al.*, 2011). This technique employed a similar approach to incubate mat cores and obtain labeled peptides. Protein-SIP allows for us to simultaneously probe which organisms are autotrophic, and to which enzymes they are allocating most of this carbon. Preliminary analysis of these protein-SIP data indicate that *Roseiflexus* spp. were the most active community members to take up $\text{H}^{13}\text{CO}_3^-$ in the low-light morning incubations. No labeled peptides were detected that were indicative of 3-OHP pathway activity, but the presence of peptides from chaperonins and enzymes with housekeeping functions (e.g., RpoB, ribosomal proteins) indicate that *Roseiflexus* spp. are taking up significant levels of HCO_3^- in the morning. These definitive links between community phylogeny and physiological functions provide the basis for linking functions to the taxa performing them, therefore allowing us to model community interactions more precisely.

Linking community structure and function has been a primary research aim in microbial ecology (Fuhrman, 2009), and studies concerning the trophic structure and dynamics of microbial communities have recently become more prominent in the literature (e.g., Lueders *et al.* 2006; Ruan *et al.* 2006; Fuhrman and Steele 2008;

Langenheder *et al.* 2010). Regardless, microbial ecology as a discipline is still far behind macroorganismal ecology with respect to the development of theory regarding food webs and linkages between community structure and function (Prosser *et al.*, 2007). This discrepancy can be attributed to the gaps in knowledge regarding the functional attributes of most of the constituents of any given microbial community. As illustrated in Chapters 2, 3, and 4), the application of metagenomics is a promising approach for closing gaps in understanding of microbial communities. Systems biological approaches that evaluate metabolic interaction networks from metagenomic data can approximate food webs in very simple communities (Röling *et al.*, 2010), such as has been done with syntrophic co-cultures (Stolyar *et al.*, 2007).

As a fellow for the IGERT program, I participated in a project which used a systems-level approach to investigate the potential interactions in a simplified model of the phototroph communities in alkaline siliceous springs. The attributes of three functional guilds of cyanobacteria, FAPs, and sulfate reducing bacteria were represented by the genomes of *Synechococcus* spp. A and B', *Roseiflexus* sp. RS-1, and *Thermodesulfovibrio yellowstonii* (Taffs *et al.*, 2009). We then utilized metabolic flux modeling as a means to understand how the flows of materials and energy are partitioned among these three interacting guilds. We found support for the cross-feeding of glycolate and acetate from *Synechococcus* spp. to FAPs and gained insight into the temporal use or production of hydrogen by all three of these members of the mat community (Figure 6.1). These metabolic flux simulations comprise the first quantitative models for the trophic interactions occurring between community members in these alkaline-silicious spring mats. Now that metagenomic data have revealed that sulfate reducing bacteria are not dominant community members, but instead three other photoheterotrophic bacteria are (Chapter 3), future work could be aimed toward constructing models that integrate these more abundant community

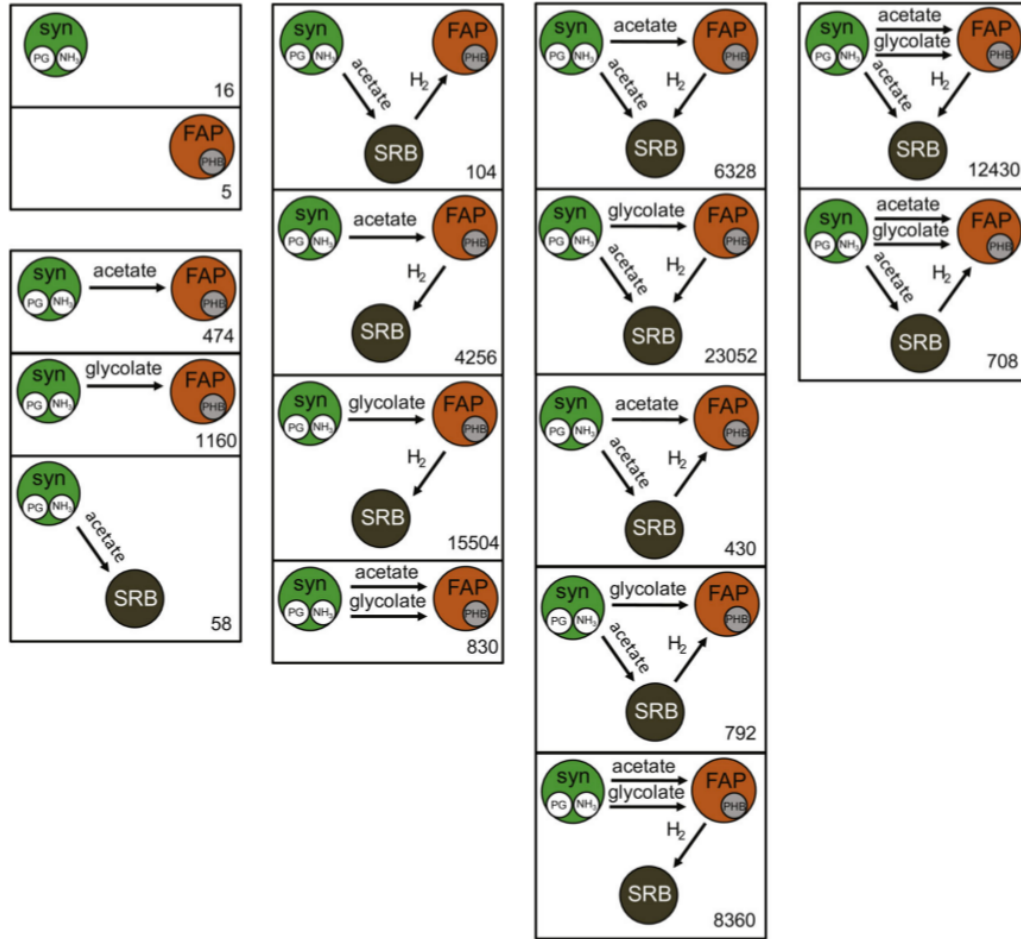


Figure 6.1: Daytime Guild Interactions Derived from Flux Models. Elementary flux mode analysis was done with compartmentalized metabolic networks for each of the three guilds, such that they could exchange external metabolites with the others while maximizing biomass production. Each box represents a grouping of models that exhibited the displayed exchange of external metabolites. Numbers in each box indicate the number of elementary modes (i.e., unique metabolic pathway combinations) in each category. Storage compounds are abbreviated with the following labels; PG = polyglucose, PHB = polyhydroxybutyrate, NH₃ = cyanophycin. This figure was originally published in Taffs *et al.* 2009.

members, which will enable more accurate predictions of their functions in the food webs of these and similar communities.

APPENDICES

APPENDIX A

CHAPTER 2 APPENDIX

Enzymatic lysis and DNA extraction. Frozen samples were thawed and resuspended in 100 μ l Medium DH (Castenholz's Medium D; Castenholz, 1969a), then homogenized with sterile mini-pestle in 2 ml screw cap tubes. 900 l Medium DH was added to the homogenized sample, then lysozyme (ICN Biomedicals, Irvine, CA) was added at approximately 200 μ g ml⁻¹, and the mixture was incubated for 45 minutes at 37 °C. 110 μ l of 10% (w/v) SDS and 200 μ g ml⁻¹ Proteinase K (Qiagen, Valencia, CA) was added, and the mixture was incubated on a shaker for 50 minutes at 50 °C. Lysis was verified by microscopy. DNA was purified using a phenol/chloroform extraction.

Mechanical lysis and DNA extraction. Frozen samples were processed with a MoBio UltraClean Soil DNA extraction kit (catalog #12800, MO BIO Laboratories, Inc. Carlsbad, CA) according to the manufacturer's instructions.

Metagenome clone library construction. DNA from both extraction procedures resulted in DNA of ~2 - 12 kb in length. Various insert sizes (See Supplementary Figure 1) were separated in gel analysis and ligated into HT plasmid vectors. End sequencing of inserts was performed using BigDye Terminator chemistry and sequences were determined with an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA).

Sample	Temperature	Layer	Lysis Procedure	Clone Insert Size	# of Reads
Octopus Spring High Temperature	58-67°C	0-1 mm	Enzymatic	2-3 kb	4216
		0-1 mm	Enzymatic	10-12 kb	3838
Octopus Spring Low Temperature	53.5-63.4°C	0-1 mm	Enzymatic	2-3 kb	19142
		0-1 mm	Enzymatic	10-12 kb	79674
Octopus Spring Total (collected November 5 th 2004)					106870
Sample	Temperature	Layer	Lysis Procedure	Clone Insert Size	# of Reads
Mushroom Spring High Temperature	65°C	0-1 mm	Enzymatic	2-3 kb	8001
		0-1 mm	Enzymatic	10-12 kb	7280
		0-1 mm	Mechanical	2-3 kb	689
		0-1 mm	Mechanical	10-12 kb	722
		1-2 mm	Mechanical	3-4 kb	5759
		1-2 mm	Mechanical	5-6 kb	5427
		7-8 mm	Mechanical	4-5 kb	5772
Mushroom Spring Low Temperature	60°C	0-1 mm	Enzymatic	3-4 kb	15837
		0-1 mm	Enzymatic	8-9 kb	23341
		0-1 mm	Mechanical	3-4 kb	5650
		0-1 mm	Mechanical	5-6 kb	5661
		1-2 mm	Mechanical	5-6 kb	5679
		4-5 mm	Mechanical	5-6 kb	5643
Mushroom Spring Total (collected October 2 nd 2003)					95461
Metagenome Total					202331

Supplementary Figure 1. Metagenome Libraries used in Chapter 2.

APPENDIX B

CHAPTER 3 APPENDIX

1. Photographs of Octopus and Mushroom Spring. See Supplementary Figure 1.

2. Reference genomes used in this study. See Supplementary Table 1.

3. Detailed Materials and Methods.

DNA extraction. The uppermost 1 mm-thick green layer from each microbial mat core was physically removed using a razor blade and DNA was extracted using either enzymatic or mechanical bead-beating lysis protocols. The two methods resulted in different abundances of community members (see below) (Bhaya et al., 2007; Klatt et al. 2007). For enzymatic lysis and DNA extraction, frozen mat samples were thawed, resuspended in 100 μ l Medium DH (Castenholz's Medium D with 5 mM HEPES, pH = 8.2; Castenholz, 1988), and homogenized with a sterile mini-pestle in 2 ml screw cap tubes. Medium DH (900 μ l) was added to the homogenized sample, then lysozyme (ICN Biomedicals, Irvine, CA) was added to $\sim 200 \mu\text{g ml}^{-1}$, and the mixture was incubated for 45 min at 37 °C. Sodium dodecyl sulfate (110 μ l of 10% (w/v) solution) and Proteinase K (Qiagen, Valencia, CA) (to $200 \mu\text{g ml}^{-1}$) were added, and the mixture was incubated on a shaker for 50 min at 50 °C. Microscopic analysis suggested efficient lysis of *Synechococcus* spp. cells, but a possible bias against some filamentous community members (Supplementary Figure 2). Phase contrast micrographs were obtained with a Zeiss Axioskop 2 Plus (Carl Zeiss Inc., Thornwood NY, USA) using a Plan NeoFluar magnification objective, and autofluorescence was detected using a HBO 100 mercury arc lamp as excitation

source and a standard epifluorescence filter set (Leistungselektronik Jena GmbH, Jena, Germany). DNA was purified using a series of organic extractions, the first using Tris-HCl-equilibrated phenol (pH=8.0) and three subsequent extractions using phenol:chloroform:isoamyl alcohol (25:24:1). Nucleic acids were precipitated at -20°C by adding 2.5 volumes ethanol and 0.1 volume 3.0 M sodium acetate (pH=5.2). The mechanical bead-beating extraction was performed on frozen mat samples with a MoBio UltraClean Soil DNA extraction kit (catalog #12800, MO BIO Laboratories, Inc. Carlsbad, CA) according to the manufacturer's instructions.

16S rRNA analysis of samples used in construction of metagenomic libraries. Denaturing gradient gel electrophoresis analysis of PCR-amplified 16S rRNA genes in DNA extracted using the enzymatic protocol was analyzed by denaturing gradient gel electrophoresis according to methods previously described (Ferris and Ward, 1997), and confirmed a familiar distribution pattern (Ferris and Ward, 1997; Ward et al., 2006) of *Synechococcus* spp. A/B genotypes along the effluent channel of Mushroom Spring and Octopus Spring, as shown in Supplementary Figure 3.

Pyrosequencing of 16S rDNA. A pyrosequencing test plate (Roche 454 FLX) was completed at JCVI using DNA extracted from a #15 core sampled at Mushroom Spring 60°C on 17 December 2007. Four different protocols were followed for the extraction of DNA; (i) the enzymatic protocol detailed above, (ii) an enzymatic and mechanical method used to construct metagenome libraries at

the US DOE Joint Genome Institute (see Inskeep et al., 2010 for details), (iii) a MoBio UltraClean Soil DNA extraction kit as above, and (iv) a pressure based lysis procedure. For this procedure, mat samples were resuspended into the Epicentre gram positive lysis buffer supplemented with Epicentre Ready-lyse at 1ug/ml and proteinase K 1 ug/ml (Epicentre Biotechnologies, Madison, WI) and samples processed in the PCT Barocycler NEP2320 (Pressure BioSciences, South Easton, MA). Briefly, resuspended samples were added to PCT tubes with shredder disk. Samples were homogenized in the shredder tube for 20 seconds. Homogenized samples were processed further in the Barocycler for 45 cycles at 65°C. Cycles were as follow: 5 seconds at 35K p.s.i. followed by 5 second at 0 p.s.i. After 45 cycles in the Barocycler, nucleic acids were extracted as per Epicentre protocol. Pyrosequencing was conducted using the sequencing primers V3-V5F: 5'-CCTACGGGAGGCAGCAG-3', and V3-V5R: 5'-CCGTCAATTCMTTTRAGT-3'. Taxonomic calls were determined using the Ribosomal Database Project Bayesian Classifier (Wang et al., 2007). The taxonomic distribution of these sequences is shown in Supplementary Figure 4.

Metagenome construction and sequencing. DNA from both extraction procedures was size-fractionated using agarose gel electrophoresis, and fragments between ~2-3 kb and ~10-12 kb (Supplementary Table 2) were ligated into HT plasmid vectors. Paired-end sequencing of inserts was done at the J. Craig Venter Institute (JCVI) using BigDye Terminator chemistry and an ABI 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA). Metagenomic assemblies were deposited in GenBank (Project number 20953).

BLASTN recruitment by reference genomes. The 202 331 paired-end sequences derived from the plasmid insert libraries contain approximately 167 Mbp of sequence with an average sequence length of 817 nucleotides. Due to concerns of lysis bias and lower cyanobacterial representation in mechanical lysis protocols, we used only the 161 976 sequences that were produced from the enzymatic lysis protocol for further analysis (see Supplementary Figures 5 and 6). These sequences were used as a query in a preliminary WU-BLASTX (Altschul et al., 1990) (default parameters) comparison to NCBI's protein database of bacterial and archaeal genomes (obtained on 26 February 2008) to identify publicly available genomes that recruited numerous metagenome sequences at an amino acid identity above ~70%. In addition, the metagenomic sequences were subjected to BLASTN recruitment by all 1 414 genomes available at NCBI (May 2nd, 2009). These results guided the selection of twenty isolate genomes (Supplementary Table 1) to be used as a reference set. These genomes were selected on the basis of whether the isolates containing them were (i) known to be genetically representative of populations inhabiting these mat communities based on prior molecular analysis (*e. g.*, 16S rRNA or 16S-23S internal transcribed spacer region analyses), (ii) cultivated from these or similar Yellowstone alkaline siliceous hot spring cyanobacterial mats, (iii) cultivated from another kind of Yellowstone geothermal feature; (iv) cultivated from geothermal features outside Yellowstone, (v) representative of physiological groups whose activities are known to occur in the mat (*e. g.*, oxygenic photosynthesis, anoxygenic photosynthesis, aerobic respiration, fermentation, sulfate reduction and methanogenesis), or (vi) representative of relevant phylogenetic groups that

were not otherwise included in the set of reference genomes. WU-BLASTN was used to align the metagenome sequences to the concatenated twenty-genome database with the parameters $M=3$, $N=-2$, $E=1e-10$, and $wordmask=dust$. Recruitment plots to these and a large number of other genomes can be produced using <http://gos.jcvi.org/users/FIBR/advancedReferenceViewer.html> found at <http://gos.jcvi.org/users/FIBR/advancedReferenceViewer.html>). These parameters were designed using Karlin-Altschul statistics (Karlin and Altschul, 1990) to obtain significant alignments as low as 50% identity with a target length of approximately 100 bp. Sequences that did not meet these criteria were labeled “null”, which indicates a lack of sufficient sequence similarity from which to assign phylogeny. Supplementary Figures 4 and 5 show recruitment results metagenomes obtained using different lysis protocols and samples.

Taxonomic resolution of recruited sequences. To estimate the taxonomic resolution offered by the recruitment of metagenomic sequences to reference genomes, cyanobacterial and FAP genomes of differing relatedness were aligned to a reference genome (Supplementary Figure 7). The distributions of % NT ID for each genome in comparison to the reference genome determined the level of % NT ID that corresponded to strains within the same named genus, within different genera within the same kingdom (*i.e.*, sub-Domain lineage) or within different kingdoms. We used these % NT ID ranges to inform decisions as to the % NT ID distributions that could be confidently associated with the respective reference genome, as indicated in Table 3.2 in the main text. Specifically, we examined the relationships between homologs in genomes from cyanobacteria and

Chloroflexi (and relatives) with different levels of relatedness (Supplementary Figure 7). *Synechococcus* spp. strain A and B' homologs range from ~75 to 100% NT ID (mean $\pm$ standard deviation = 85.0 ± 6.5 %). To ensure that the metagenomic sequences recruited by the *Synechococcus* spp. A and B' genomes were more closely related to the genome that recruited them than to the other genome, these sequences were separately queried against the *Synechococcus* spp. A and B' genomes in two independent BLASTN experiments. Results indicating efficient separation are shown in Supplementary Figure 8.

Genes of more distantly related cyanobacteria (*Thermosynechococcus elongatus*, *Nostoc* sp. strain PCC 7120 and *Gloeobacter violaceus*) range from 50-75% NT ID (with means 61 to 64%) to homologs in *Synechococcus* sp. strain A. Similarly, *Roseiflexus* sp. strain RS1 and *R. castenholzii* homologs range from ~70 to ~90% NT ID (mean 78.3 ± 7.1 %), but genes in more distantly related members of the kingdom (*Chloroflexus* and *Herpetosiphon*) range from 50-75% NT ID (means 58.3 to 64.1 %) with *Roseiflexus* sp. strain RS1 homologs. According to a one-way analysis of variance, there is a statistically significant difference between the distributions of % NT ID in these pairwise genome comparisons ($F_{4,7021} = 6179.2$, $P < 10^{-10}$ for comparisons to *Synechococcus* sp. strain A; $F_{3,8283} = 4352.3$, $P < 10^{-10}$ for comparisons to *Roseiflexus* sp. strain RS1). A Tukey HSD post hoc test indicated that homologs between organisms as divergent as *Synechococcus* sp. strain A vs. sp. strain B' (Supplementary Figure 7A) and between *Roseiflexus* sp. strain RS1 vs. *R. castenholzii* in (Supplementary Figure 7B) can be significantly distinguished from comparisons of more distant taxonomic pairings, supporting inferences about the differences observed in

metagenomic recruitment. Furthermore, the differences in distribution of % NT ID between *Synechococcus* sp. strain A and more distantly-related cyanobacteria were significantly greater than were those between *Synechococcus* sp. strain A and the *Chloroflexi* outgroup (Supplementary Figure 7A), just as more distantly-related *Chloroflexi* were significantly greater than the cyanobacterial outgroup in the comparison to *Roseiflexus* sp. strain RS1 (Supplementary Figure 7B).

Synteny determination of clones. When both end sequences of a particular clone insert had most significant WU-BLASTN high-scoring pairs (HSPs, or alignments) to the same isolate genome, these end sequences were considered "jointly recruited." When paired-end sequences had best BLAST HSPs to different genomes, these sequences were considered "disjointly recruited." Jointly recruited sequences were analyzed further to determine their degree of synteny with the reference genomes, based on both the separation and orientation of end sequences, as described below (Rusch et al., 2007; Bhaya et al., 2007).

i) Length component. "Jointly recruited" sequences were mapped to the genome recruiting them by the locations of the alignments on each end. The size estimated *in silico* was then compared to the expected size of the DNA fragments used to construct the library from which the sequence was derived (Supplementary Table 2), and paired-end sequences were considered "syntenous" with respect to length if the genome-mapped size was within 30% of the expected size. Those pairs that mapped to sizes $\geq 30\%$ greater or less than the expected size were considered "nonsyntenous". The 30% tolerance value was determined for jointly recruited sequences by comparing the expected size of each metagenome library to the positions that these recruited sequences aligned to for eight

different reference genomes. When the stringency of the distance requirement is relaxed, larger numbers of sequences are considered to be jointly recruited and syntenous. However, 30% is the level at which a further relaxation of divergence from the expected size does not further increase the percentage of syntenous sequences (Supplementary Figure 9). The 30% cutoff is thus a very conservative estimate and may obscure fine-scale loss in synteny amongst the lineages studied. As an example, a jointly recruited pair of sequences from the largest expected insert-size library of 10-12 kbp was considered "syntenous" with the 30 % error rate if the two end sequences were within 7 to 15.6 kbp of each other when aligned to the recruiting reference genome, and thus the hypothetical loss of a gene ~1 kb in length would not be detected. This method ensured that significant changes in gene order had occurred in cases where sequences were considered non-syntenous. While we acknowledge that much of the sequence data analyzed would likely be syntenous by the classic definition of being located on the same chromosome (Passarge et al., 1999), our use of this term (*sensu* Bhaya et al., 2007) refers more specifically to changes in local genome architecture based upon the hypothesized separation distance of loci on a chromosome compared to a reference chromosome (Dempsey et al., 2006).

ii) Orientation component. A second criterion for synteny was the correct orientation of jointly recruited end sequences (Rusch et al., 2007). A jointly recruited pair of sequences was considered syntenous only if both end sequences aligned to the reference genome in 5' to 3' orientations on their respective opposite strands, in addition to the alignments being the expected distance apart on the genome as described above.

***In silico* analysis of synteny among genomes.** The conservation of synteny of metagenomic sequences in comparison to the reference genomes of *Synechococcus* spp. A and B' was determined by querying these sequences in a WU-BLASTN alignment to each genome independently in a “forced” comparison (i.e. “forced” to align to a single genome as opposed to allowing a sequence to be recruited by one of many genomes). To establish the relationship of how gene order conservation changes with increasing evolutionary distance, control experiments were performed in which *in silico* “metagenomes” were created by randomly fractionating five cyanobacterial genomes (*Synechococcus* sp. strain B', *Thermosynechococcus elongatus* BP-1, *Gloeobacter violaceus*, *Nostoc* sp. strain PCC 7120, and *Synechococcus* sp. strain WH8102) and one outgroup *Chloroflexi* genome (*Roseiflexus* sp. strain RS1) each into 10 000 jointly recruited metagenomic sequences 800 bp long and clone mates 2 000 bp apart on their respective genomes with custom Perl scripts (Supplementary Table 3). This initial control metagenome simulates an artificial community in which organisms are represented by equal fractions of a particular metagenome library (but with varying degrees of coverage, depending on genome size), given a uniform clone-insert size for this metagenome library. Synteny relationships for these pairwise genome comparisons declined as the relationship between genomes decreased (Supplementary Table 3), and also with increasing clone insert lengths (data not shown), and this complicated direct comparisons of metagenome recruitment content and pairwise genome comparisons due to differences in clone insert lengths used to construct the environmental metagenome libraries. To overcome this limitation, an *in silico* metagenome was created to reflect the distribution of

clone insert sizes observed for those sequences recruited to the *Synechococcus* sp. strain A genome, enabling direct comparison of synteny between the *in silico* and the observed metagenome recruitment. This consisted of an *in silico* metagenome containing 1 936 clones with a 2 000 bp insert size, 978 clones with 3 000 bp insert size, 1 441 clones with 8 000 insert size, and 5 645 clones with a 10 000 bp insert size. These *in silico* metagenomes were used as queries in a BLASTN alignment to the *Synechococcus* sp. strain A genome with the same parameters described above (M=3 N=-2 E=1e-10 workmask=dust) and were subjected to the same length and orientation analyses to determine synteny (Figure 3.5 in the main text).

This method of analyzing and comparing synteny of metagenomic sequences is specialized for datasets produced by end-sequencing of clone inserts, and differs from a previous method that analyzed the predicted genes that are co-localized on a single metagenomic sequence and determined if the homologs of these genes were also co-localized on a reference genome (Wilhelm et al., 2007). Many of the metagenomic sequences in this dataset contained regions with sequence similarity to more than one gene on the genome of interest. Our method of aligning sequences against entire genomic scaffolds encompassed both multiple genes and intergenic regions, which increased the probability of correctly identifying homologous regions to isolate chromosomes given these stringent BLAST criteria.

Scaffold Clustering and Annotation. The oligonucleotide frequencies of all scaffolds $\geq 20\,000$ bp in length in addition to the genomes of *Synechococcus* sp.

strain A and B', *Roseiflexus* sp. strain RS1, *Chloroflexus* sp. strain 396-1, *Cand. C. thermophilum*, and *Chloroherpeton thalassium* were subjected to k-means analysis using the stats R package (The R Core Development Team, 2011) and custom perl scripts with multiple *a priori* values of k ranging from 5 to 12. For each value of k, the clustering analysis was simulated 100 times with random starting points to obtain “core clusters” that grouped together in $\geq 90\%$ runs. Eight clusters of scaffolds that grouped together in at least 90% of the monte carlo simulations were consistently observed across the range of initial k values, thus k=8 was chosen for final analysis. To determine gene annotations for the metagenome scaffolds, the DNA sequences were submitted to the JCVI Annotation Service, where they were analyzed using JCVI's prokaryotic annotation pipeline. This pipeline includes open reading frame prediction using Glimmer (Delcher et al., 1999), and comparative annotation using hidden markov models, (Haft et al., 2001; Finn et al., 2008), TMHMM searches (Krogh et al., 2001), and SignalP predictions (Bendtsen et al., 2004) to assign names, functions, and Gene Ontology terms to the predicted peptide sequences (Tanenbaum et al., 2010).

Recovery of phylogenetic marker sequences from metagenomes. Known 16S rRNA and *recA* sequences were used in WU-BLASTN analyses (default parameters) against the metagenomic sequences to identify putative 16S rRNA and *recA* homologs. Phylogenetic assignments of the 16S rRNA sequences were made by sequence alignment with sequences from past studies of these springs (Ward et al., 2006). If 16S rRNA sequences could not be unambiguously

classified in this way, they were classified taxonomically with the Ribosomal Database Project Classifier (Wang et al., 2007). Putative *recA* metagenome sequences were translated and analyzed against the NCBI non-redundant protein database using WU-BLASTX with default parameters to identify the best BLAST HSPs to known RecA sequences. Alignments of RecA sequences were verified by comparison to the curated alignment used to construct the PFAM hidden Markov model PF00154 (Finn et al., 2008). Phylogenetic assignments of the RecA sequences were based on taxonomic affiliations of the organisms with homologs identified by best matches in BLAST analyses (Supplementary Table 4), sequence alignments and in some cases by phylogenetic analysis. A Neighbor-Joining phylogenetic tree of partial translated metagenomic RecA sequences consisting of 103 amino acid positions was constructed with evolutionary distances calculated using the Poisson correction method of the MEGA 4 software package (Tamura et al., 2007) (Supplementary Figure 10). The program AMPHORA was used to detect and phylogenetically assign homologs to 31 phylogenetic marker genes from Domain Bacteria on the translated sequences of predicted ORFs on metagenomic scaffolds (Wu and Eisen, 2008) (see Supplementary Table 5). Phylogenetic analysis in reference to 578 genome sequences was done with the maximum likelihood method implemented by RAxML (Stamatakis, 2006). Many sequences exhibiting sequence similarity to these 31 marker genes could not be assigned to a more specific taxonomic level than Domain, and therefore Archaea might contribute some of these sequences. The relative abundances of 16S rRNA and RecA sequences for different phylogenetic groups is compared in Supplementary Table 6.

Comparative Analyses. With the exception of the programs specifically mentioned above, all comparative data analyses were performed and images were created using custom Perl scripts developed by J. M. Wood. These scripts are available from the corresponding author by request.

4. Phylogeny of *Chloroflexi* sequences.

A full-length 16S rRNA sequence from scaffold scf1113211797825 was imported into ARB (Ludwig et al., 2004) and aligned with other representative environmental clone sequences and isolates from Kingdom Chloroflexi. All columns in the resulting alignment containing gaps were removed from analysis. A neighbor-joining tree (Supplementary Figure 11) was constructed using 1 128 nucleotide positions with the Jukes-Cantor model using the BioNJ algorithm (Gascuel et al., 1997). A more detailed version of the neighbor-joining PufL and PufM tree (Figure 3.3 in the main text) which supports the basal position of these Chloroflexi sequences is shown in Supplementary Figure 12.

5. Genomes recruiting low-quality homologs from metagenomic samples.

Many genomes recruited mostly distantly related metagenomic sequences that were disjointly recruited as shown in Supplementary Figure 13.

Oxygenic phototrophs. The *Thermosynechococcus elongatus* strain BP-1 genome recruited less than 1% (n=1 419) of the total metagenomic sequences,

most of which were disjointly recruited (72% of the sequences recruited by the *T. elongatus* genome) and had low % NT ID (mean $63.3 \pm 6.6\%$). When these sequences were aligned to the *Synechococcus* sp. strain A genome in a separate experiment, the % NT IDs of these alignments were not discernibly different from the alignments of genome fragments from *Roseiflexus* sp. strain RS1, used as a taxonomic outgroup to the cyanobacteria (see Supplementary Figure 7). *T. elongatus* strain BP-1 was cultivated from a Japanese geothermal system (Nakamura et al., 2002). While this isolate is typical of cyanobacteria found in Japanese hot springs (Papke et al., 2003), *Synechococcus* spp. strains whose 16S rRNA sequences are 96% identical in the 16S rRNA V9 region (157 positions) to that of *T. elongatus* strain BP-1 have been cultivated from the Octopus Spring mat (Ferris et al. 1996b). However, dilution cultivation (Ferris et al., 1996b), and oligonucleotide probing (Papke et al., 2003; Ruff-Roberts et al., 1994) suggest that these cyanobacteria are present at very low abundance compared to A/B-like *Synechococcus* spp.

Aerobic non-phototrophic organisms. The metagenomic sequences recruited by the *Herpetosiphon aurantiacus* and *Candidatus* Koribacter versatilis strain Ellin345 genomes were mainly disjointly recruited sequences of very low % NT ID and cannot be confidently associated with these organisms or their close relatives. Aerobic chemolithotrophy, mediated by communities of filamentous organisms belonging to the bacterial Order *Aquificales*, also occurs in these springs in higher temperature waters upstream of the cyanobacterial mats (Reysenbach et al. 1994). We included the *Aquifex aeolicus* strain VF5 genome to represent this

group and to evaluate possible immigration of organisms from upstream communities due to transport. The small number of low % NT ID matches with this genome suggests that contributions from *Aquificales* are rare in these mat metagenomes.

Anaerobic non-phototrophic organisms. Fermentation and other anaerobic decomposition processes occur during the night when the oxygen level in the mat is low (Anderson et al., 1987; Nold and Ward, 1996; van der Meer et al., 2007). Organisms driving fermentation processes were queried using the reference genome of *Thermoanaerobacter pseudethanolicus*, which was originally cultivated from the Octopus Spring mat (Zeikus et al., 1980); this genome recruited less than 0.2% (n=278) of all metagenome sequences, most of which were disjointly recruited and aligned to this reference genome with a low % NT ID (mean $58.9 \pm 6.0\%$ NT ID, 92% disjointly recruited). The genome of *Carboxydotherrmus hydrogeniformans*, which was used to probe for sequences from related organisms involved in anaerobic carbon monoxide oxidation, recruited even fewer sequences than did the *T. yellowstonii* genome (n = 368), mean $60.6 \pm 7.7\%$ NT ID, 97% disjointly recruited). A phylogenetically distinct sulfate reducer, *Thermodesulfobacterium commune*, was also originally cultivated from the Octopus Spring mat, but dissimilatory sulfite reductase (*dsrAB*) genes related to this isolate were not detected in the Mushroom Spring mat (Dillon et al., 2007). The genomes of *Methanothermobacter thermoautotrophicus* strain delta H and *Thermoproteus neutrophilus* served as taxonomic representatives of the Euryarchaeota and Crenarchaeota, respectively, but both recruited few sequences

of low % NT ID (means < 60%). *M. thermoautotrophicus* represented another terminal anaerobic metabolic group known to occur these mats (Ward, 1978; Sandbeck and Ward, 1981). The lower contributions of anaerobic nonphototrophic community members might have been due to our focus on the uppermost photosynthetic layers of the mat and/or to trophic structure, as inferred from lipid biomarker abundances (Ward et al., 1989a).

6. Comparison of metagenomes for evidence of *Synechococcus* sp. A'-like sequences.

To ensure that the sequences recruited to the *Synechococcus* sp. strain A genome with 83-92% NT ID from the Mushroom Spring 65 °C metagenome were indeed originating from A'-like organisms, we compared this subset of sequences to a random shotgun Titanium 454 pyrosequencing library constructed from a sample taken from Mushroom Spring at 68 °C (ED Becraft, CG Klatt, DB Rusch and DM Ward, unpublished). This comparison indicated that this subset of Sanger sequences are more closely related to native *Synechococcus* spp. from higher temperatures (Supplementary Figure 14) where A'-like *Synechococcus* spp. are dominant (Supplementary Figure 3).

7. Taxonomic resolution of assembled *Synechococcus* populations.

We compared the sequence content of assembled scaffolds to their respective recruitment by reference genomes to assess whether assembly put together rational combinations of sequences. A compilation of the recruitment results for the metagenomic sequences in each scaffold cluster is presented in Supplementary

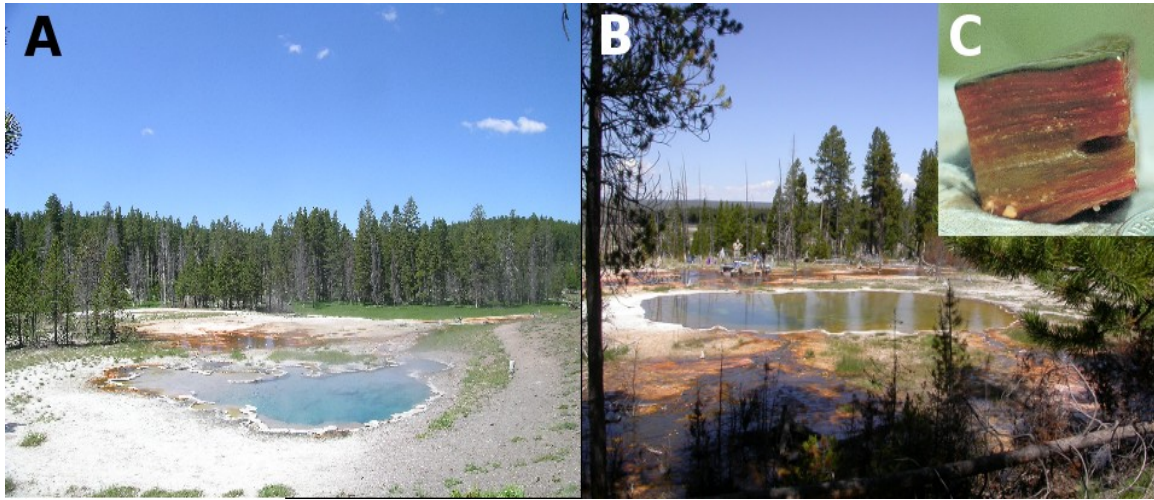
Table 7. Of the 1 472 scaffolds that contained sequences that were recruited by the *Synechococcus* spp. A and B' genomes in the recruitment analysis, 63.1% (n=930) consist exclusively of sequences recruited by these two reference genomes (*i. e.*, they contained sequences recruited to no other genomes). Of these exclusively cyanobacterial scaffolds, 35% (n=321) are “pure” in that they are made entirely of sequences recruited by the *Synechococcus* sp. strain A genome, 39% (n=364) are pure with respect to recruitment by the *Synechococcus* sp. strain B' genome, and 26% (n=245) are mixed scaffolds, which consist of sequences recruited by both the *Synechococcus* spp. A and B' genomes (Supplementary Table 8). These mixed scaffolds had a mean % NT ID that was significantly different than the pure A and B' scaffolds with respect to both genomes (Supplementary Table 8), suggesting that these scaffolds are derived from organisms more distantly related to both the A and B' reference organisms. Without comparison to a closely related representative genome, we could not verify whether these scaffolds were representative of uncultivated cyanobacterial genomes, or whether they were artifacts of assembly. After scaffolds were characterized and compared with respect to oligonucleotide frequency, scaffolds that clustered together >90% were analyzed to determine how the individual sequences underlying these scaffolds were recruited by reference genomes (Supplementary Table 7).

In our analysis of scaffolds containing sequences that were exclusively recruited by the two *Synechococcus* reference genomes, we excluded subsets of cyanobacteria that have genes that the reference genomes do not and were thus recruited to different genomes or the “null” bin. There are 36 mixed scaffolds of

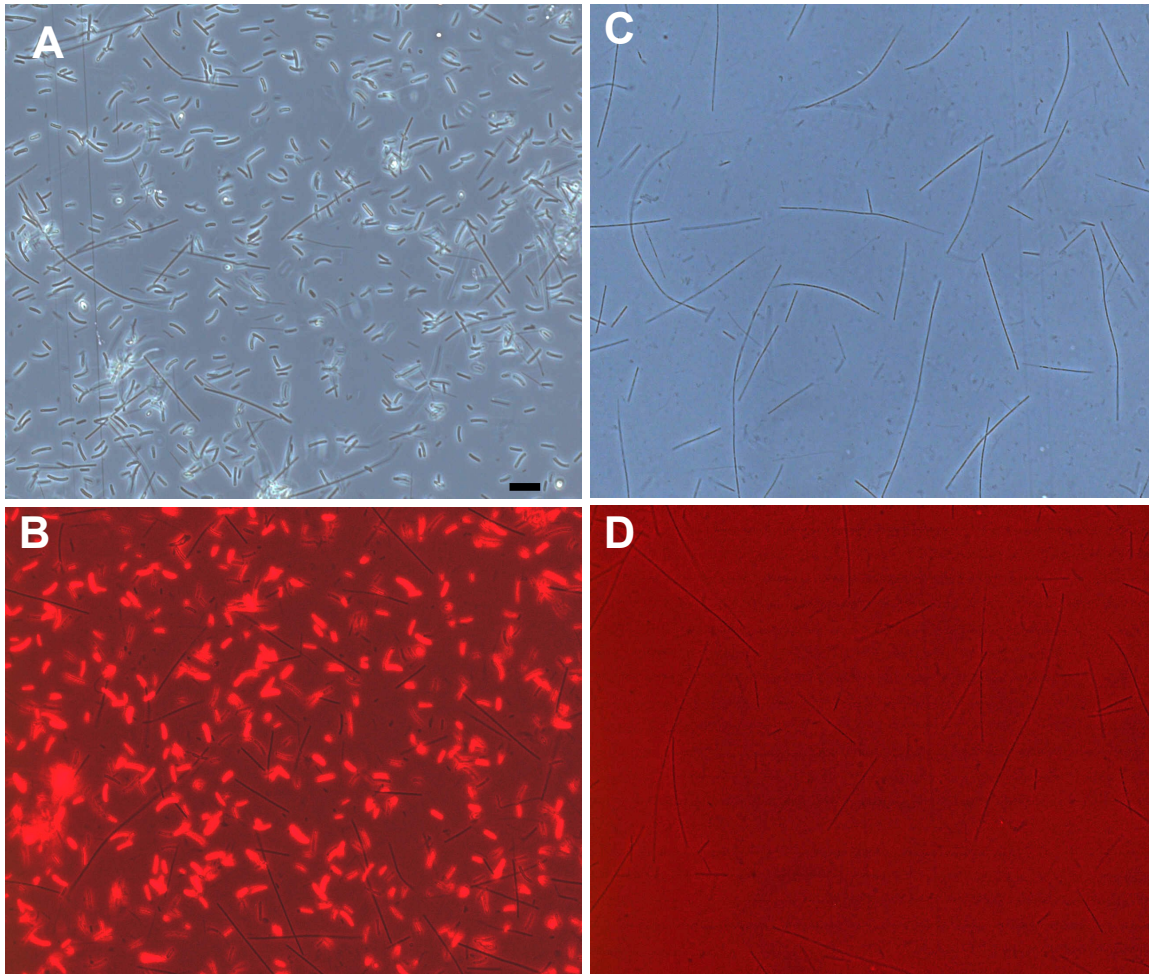
which 80% of sequences are recruited to either the *Synechococcus* sp. strain A or B' genomes, and the remaining sequences typically fall into the null bin. These assemblies may reflect the existence of environmental cyanobacterial genomes that contain genes not present in the *Synechococcus* spp. reference genomes, such as those that contain homologs to *feoA* and *feoB* genes that may confer the ability to use ferrous iron in the mat (Bhaya et al., 2007).

8. Metagenomic sequences possibly found in native *Synechococcus* spp. populations but not in *Synechococcus* spp. A and B' isolates.

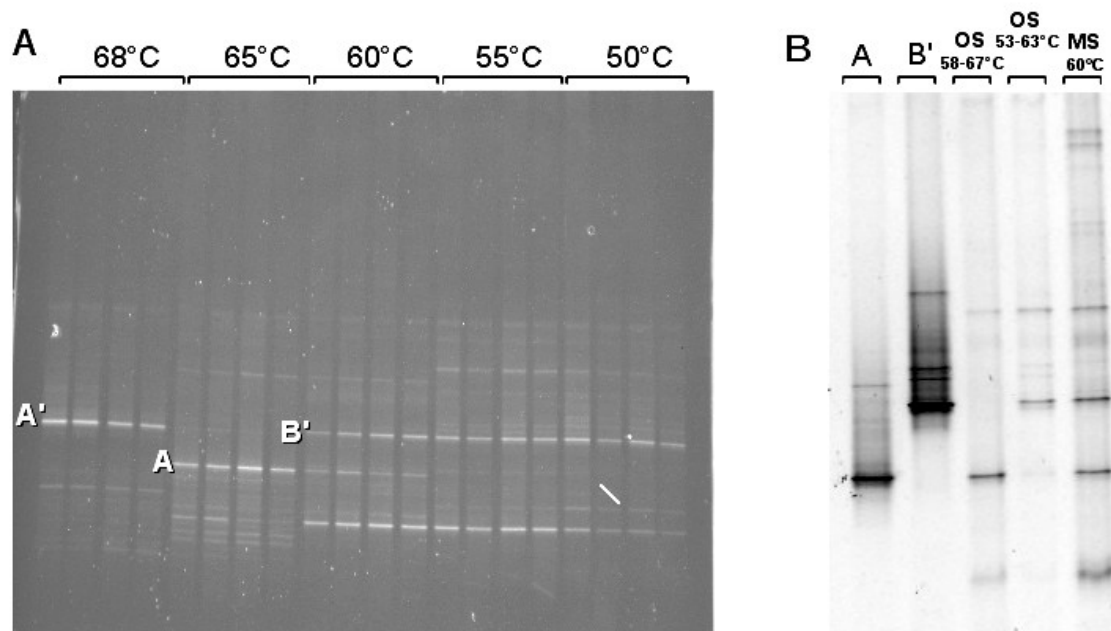
Disjointly recruited metagenomic clones with only one end sequence that can be confidently associated with a reference genome may contain sequences on the other end that are present in native populations, though absent in the isolates whose genomes are used in recruitment experiments (Bhaya et al., 2007). Metagenomic clones that had one end sequence that aligned with greater than 93% NT ID to the *Synechococcus* sp. B' genome or greater than 95% NT ID to the *Synechococcus* sp. A genome and whose paired-end sequence did not align to either *Synechococcus* spp. genomes were further analyzed. Supplementary Table 9 lists the recruitment of these paired-end sequences and their corresponding best matches in BLASTX searches (default parameters) against NCBI's nr database.



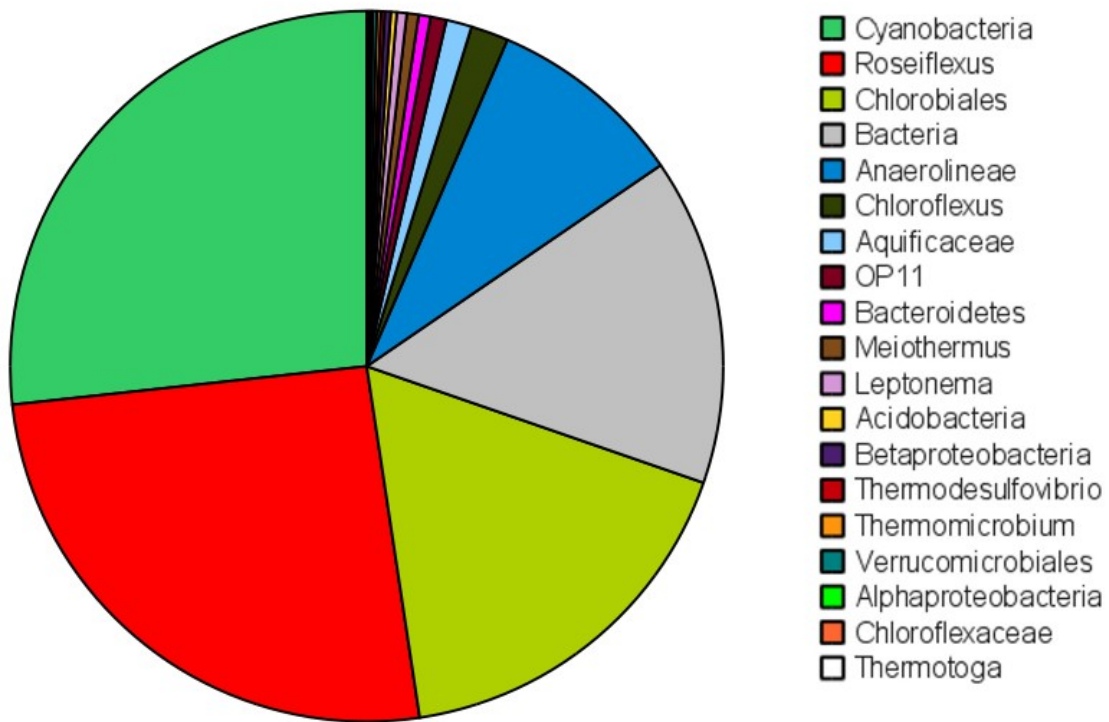
Supplementary Figure 1. Hot spring microbial mats sampled. (A) Octopus Spring, (B) Mushroom Spring, (C) mat sample ~2 X 2 cm, showing top green *Synechococcus* layer used to make metagenomic libraries used in this study.



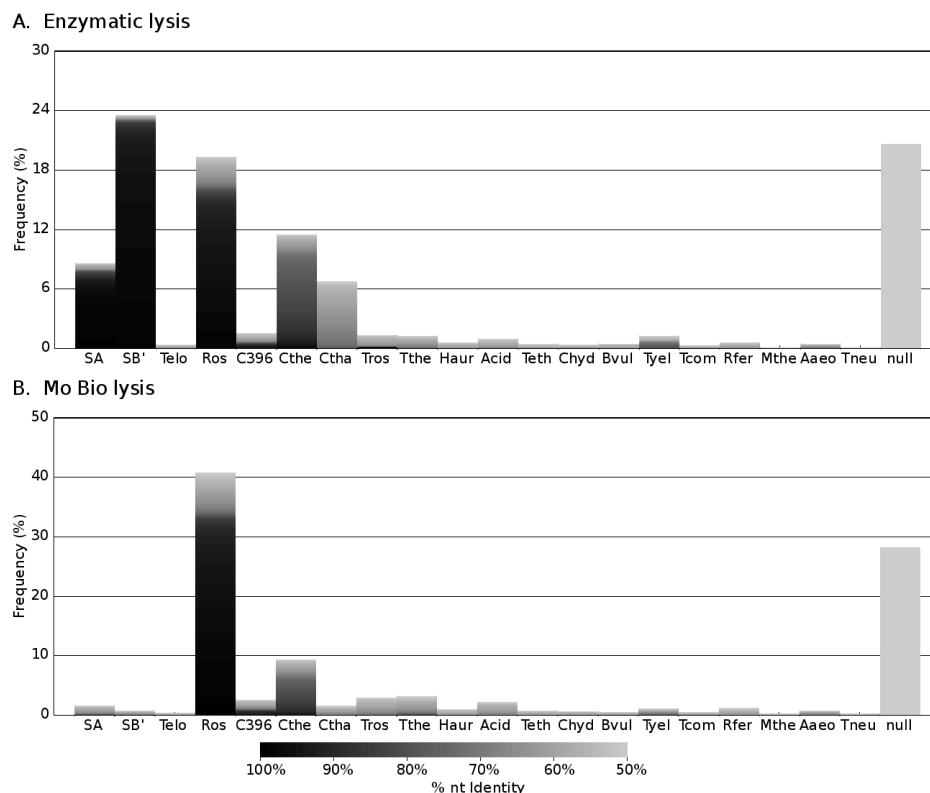
Supplementary Figure 2. Microscopic evidence of the efficiency of the enzymatic protocol in lysing *Synechococcus* spp. cells. (A) and (B) before and (C) and (D) after lysis. (A and C) phase contrast. (B and D) fluorescence with phase contrast dimmed. The scale bar in Panel A corresponds to 10 μm .



Supplementary Figure 3. Denaturing gradient gel electrophoresis analysis of PCR-amplified 16S rRNA genes in replicate samples used to produce metagenomes. (A) Mushroom Spring. (B) Comparison of *Synechococcus* spp. strains A and B' unicyanobacterial cultures with Octopus Spring and Mushroom Spring samples.

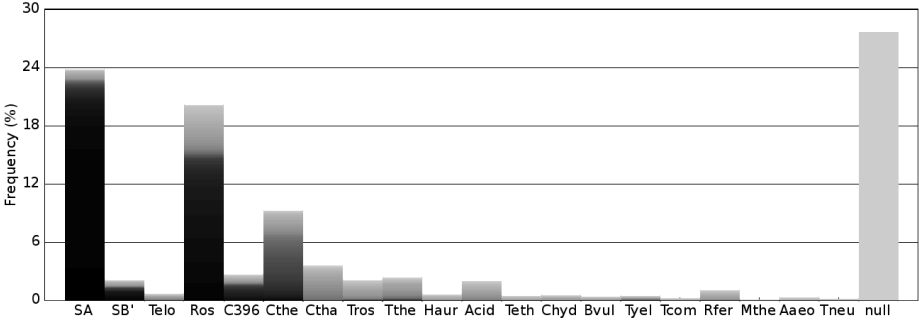


Supplementary Figure 4. Fractional contribution of taxa to 16S rDNA sequences detected by pyrosequencing. The samples correspond to the pooled results of four different DNA extraction protocols. The most specific taxonomic level determined from the R is shown.

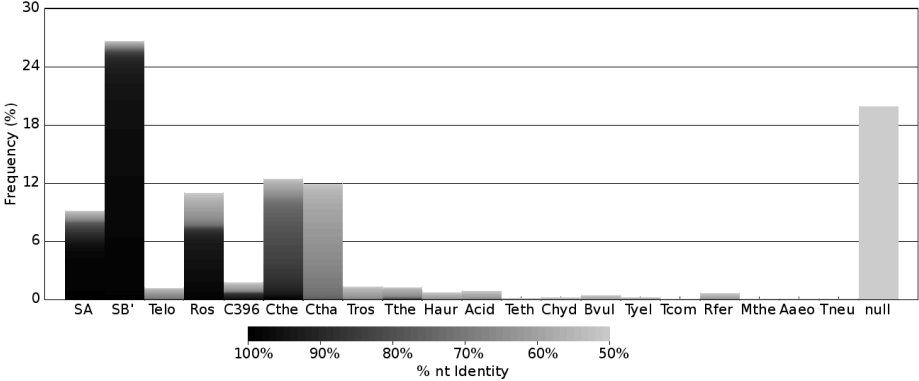


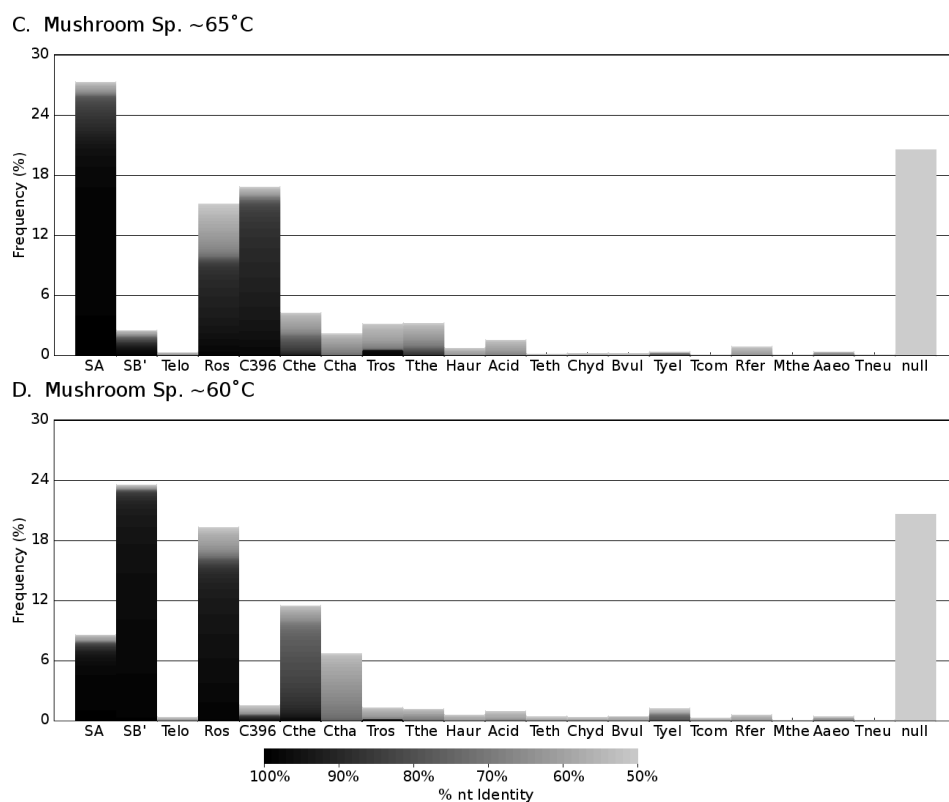
Supplementary Figure 5. Evidence of lysis bias. BLASTN-based recruitment of metagenomic sequences from libraries prepared from top green (0-1 mm) mat layers from sequences produced from DNA isolated using (A) an enzymatic lysis protocol, and (B) the MoBio soil extraction kit. Sequences were recruited by genomes of 20 microorganisms using BLASTN. SA, *Synechococcus* sp. strain A; SB', *Synechococcus* sp. strain B'; Telo, *Thermosynechococcus elongatus* strain BP-1; Ros, *Roseiflexus* sp. strain RS1; Caur, *Chloroflexus* sp. strain 396-1; Cthe, *Candidatus Chloracidobacterium thermophilum*; Ctha, *Chloroherpeton thalassium*; Tros, *Thermomicrobium roseum*; The, *Thermus thermophilus*; Haur, *Herpetosiphon aurantiacus*; Acid, *Acidobacterium* sp. strain; Tpse, *Thermoanaerobacter pseudoethanolicus*; Chyd, *Carboxydotherrmus hydrogenoformans*; Bvul, *Bacteroides vulgatus*; Tyel, *Thermodesulfobacterium yellowstonii*; Tcom, *Thermodesulfobacterium commune*; Rfer, *Rhodoferrax ferrireducens*; Mthe, *Methanothermobacter thermoautotrophicum*; Aaeo, *Aquifex aeolicus*; and Tneu, *Thermoproteus neutrophilus*. Shading indicates % NT ID of sequences recruited to each genome.

A. Octopus Sp. 58-67°C



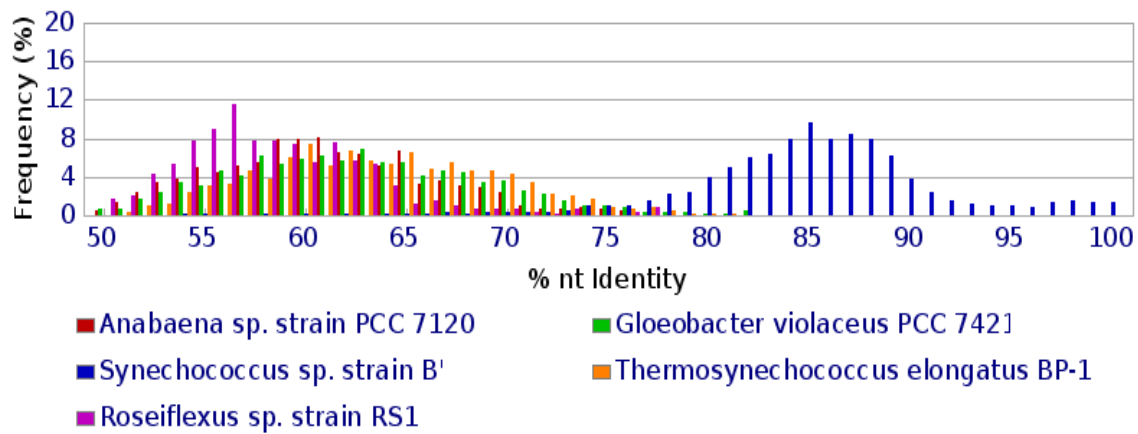
B. Octopus Sp. 53-63°C



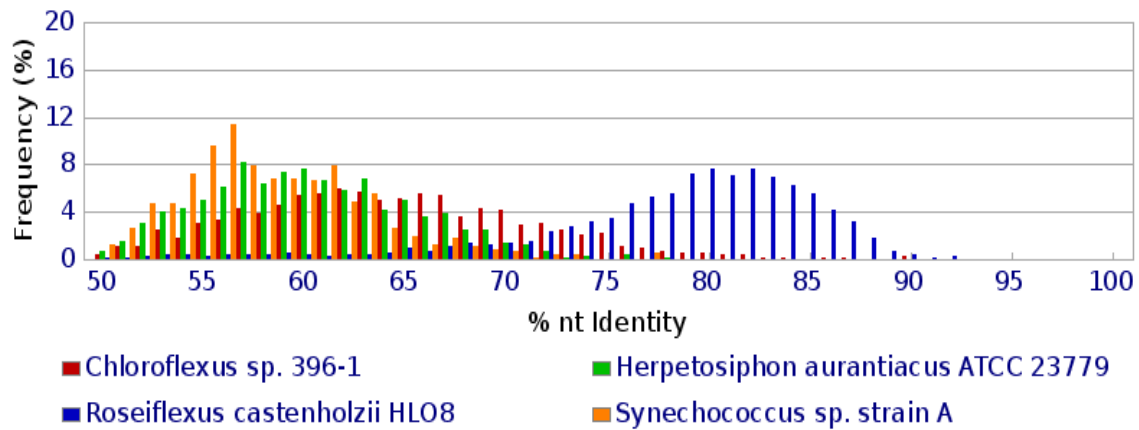


Supplementary Figure 6. BLASTN-based recruitment of metagenomic reads from libraries prepared from DNA obtained by enzymatic lysis of the top green (0-1 mm) mat layers from (A) Octopus Sp. 58-67°C, (B) Octopus Sp. 53-63°C, (C) Mushroom Sp. ~65°C and (D) Mushroom Sp. ~60°C by genomes of 20 microorganisms of possible relevance to these mats. The frequency of sequences recruited by each genome (unnormalized to genome size) displayed with the relative degree of shading indicating the % NT ID of the alignments between metagenomic and isolate homologs are indicated by the degree of shading. SA, *Synechococcus* sp. strain A; SB', *Synechococcus* sp. strain B'; Telo, *Thermosynechococcus elongatus*; Ros, *Roseiflexus* sp. strain RS-1; C396, *Chloroflexus* sp. strain 396-1; Cthe, *Candidatus Chloracidobacterium thermophilum*; Ctha, *Chloroherpeton thalassium*; Tros, *Thermomicrobium roseum*; Tthe, *Thermus thermophilus*; Haur, *Herpetosiphon aurantiacus*; Acid, *Candidatus Koribacter versatilis* strain Ellin 345; Tpse, *Thermoanaerobacter pseudoethanolicus*; Chyd, *Carboxydotherrhus hydrogenoformans*; Bvul, *Bacteroides vulgatus*; Tyel, *Thermodesulfovibrio yellowstonii*; Tcom, *Thermodesulfobacterium commune*; Rfer, *Rhodoferrax ferrireducens*; Mthe, *Methanothermobacter thermoautotrophicum*; Aao, *Aquifex aeolicus*; and Tneu, *Thermoproteus neutrophilus*.

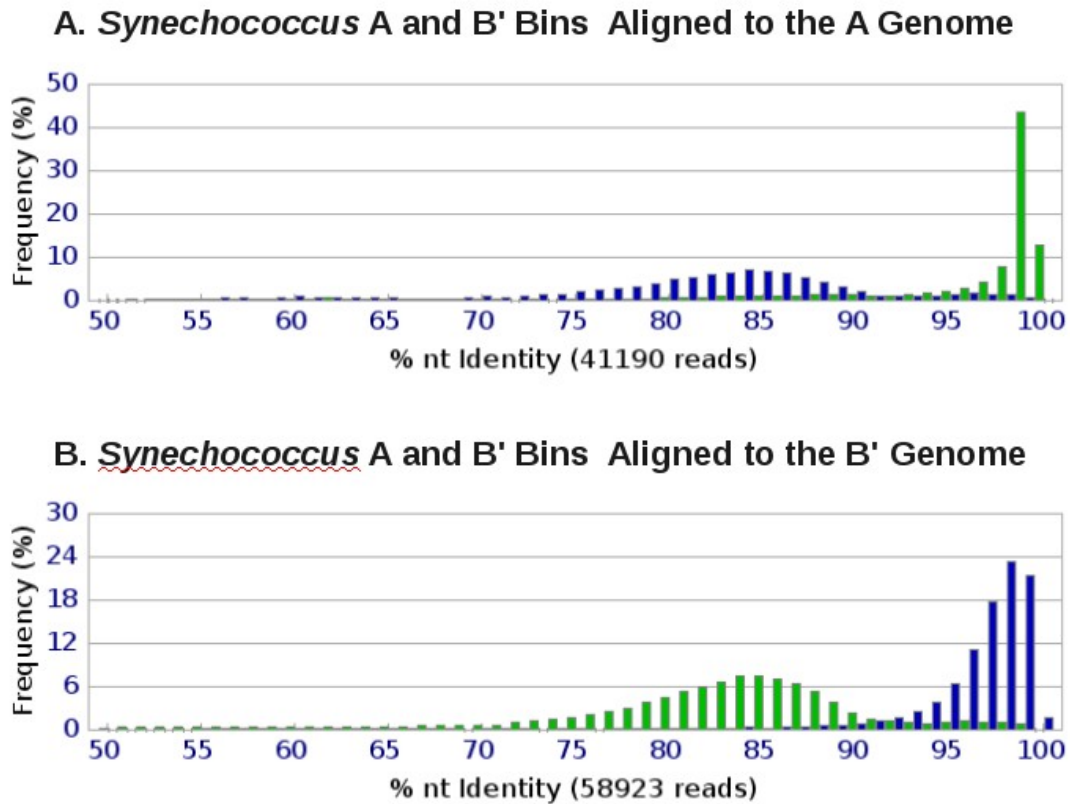
A. Cyanobacteria



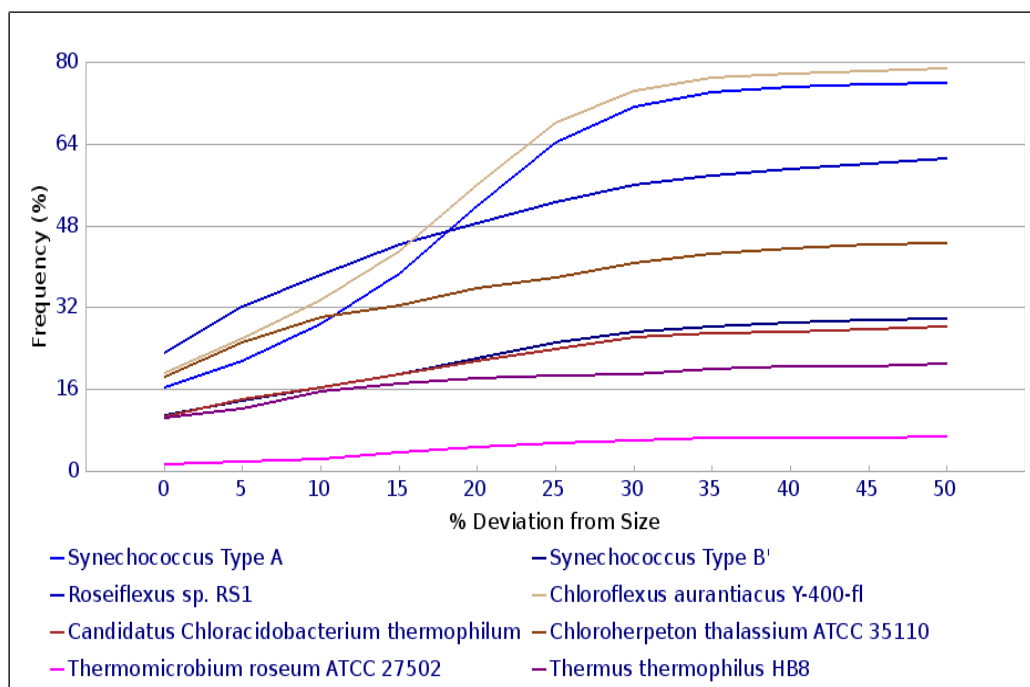
B. Chloroflexi



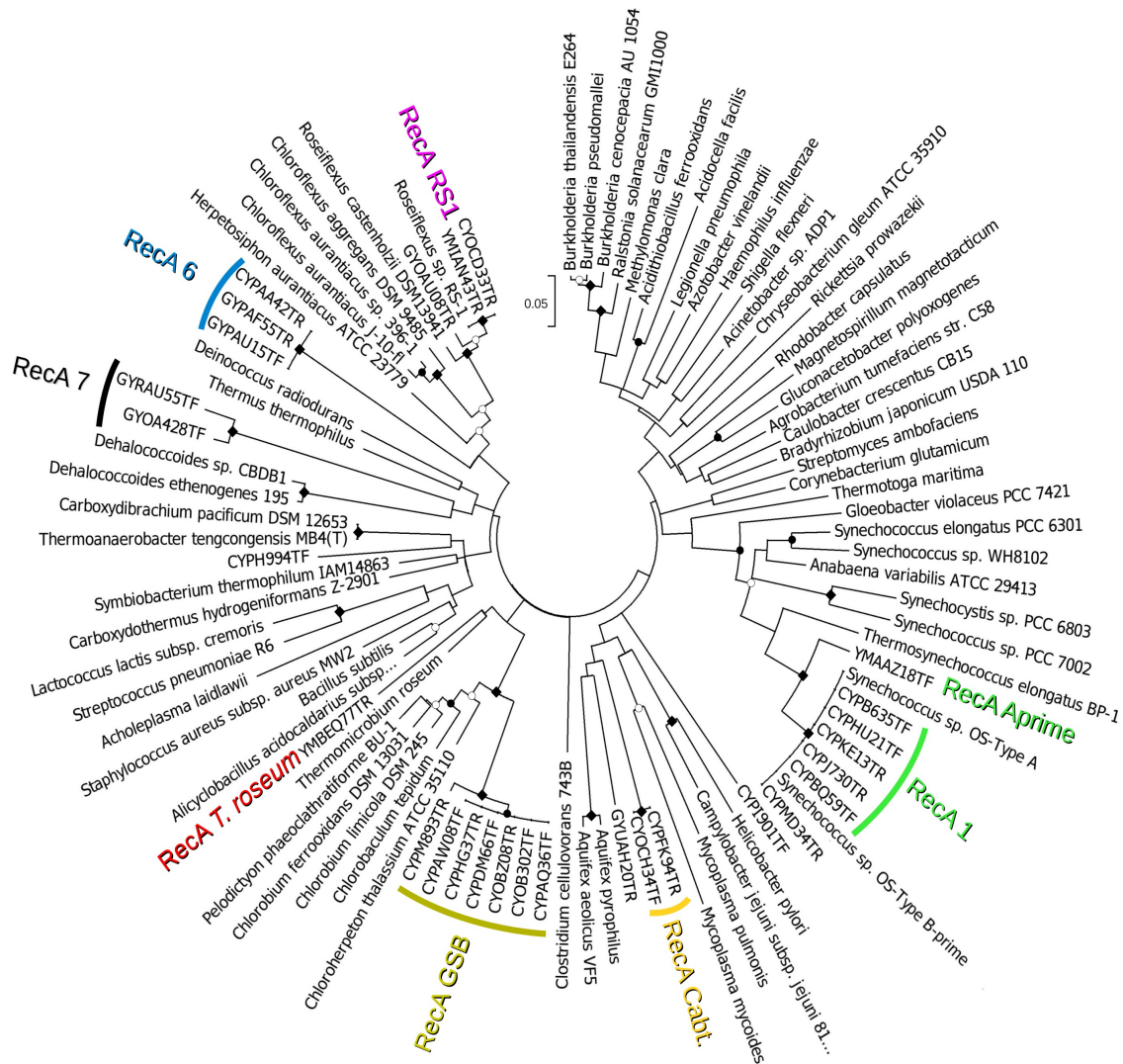
Supplementary Figure 7. Histograms of % NT ID of homologs in different genomes of (A) cyanobacteria compared to the *Synechococcus* sp. strain A genome (*Roseiflexus* sp. strain RS1 as outgroup) and (B) Chloroflexi and relatives compared to the *Roseiflexus* sp. strain RS1 genome (*Synechococcus* sp. strain A as outgroup).



Supplementary Figure 8. Histograms of % NT ID of metagenomic sequences from all libraries recruited by either the *Synechococcus* sp. strain A (green) or *Synechococcus* sp. strain B' genome (blue) aligned to the (A) *Synechococcus* sp. strain A genome, and (B) aligned to the *Synechococcus* sp. strain B' genome.

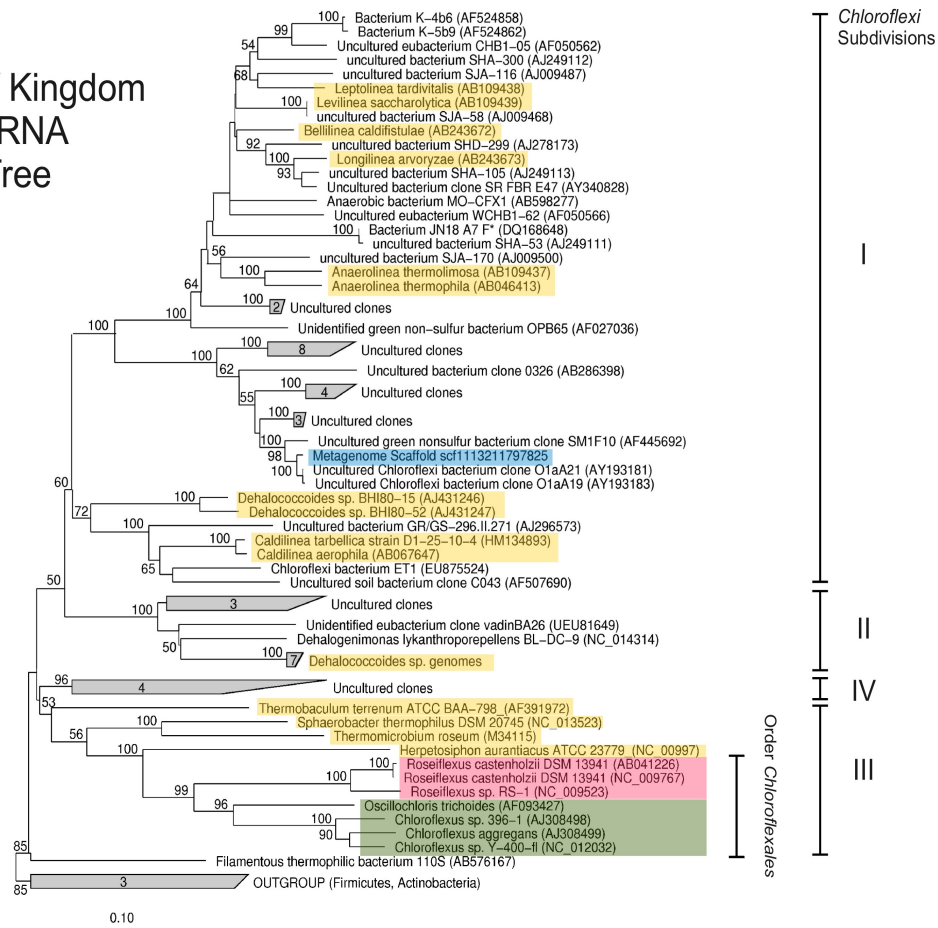


Supplementary Figure 9. Synteny as a function of deviation from estimated clone length.

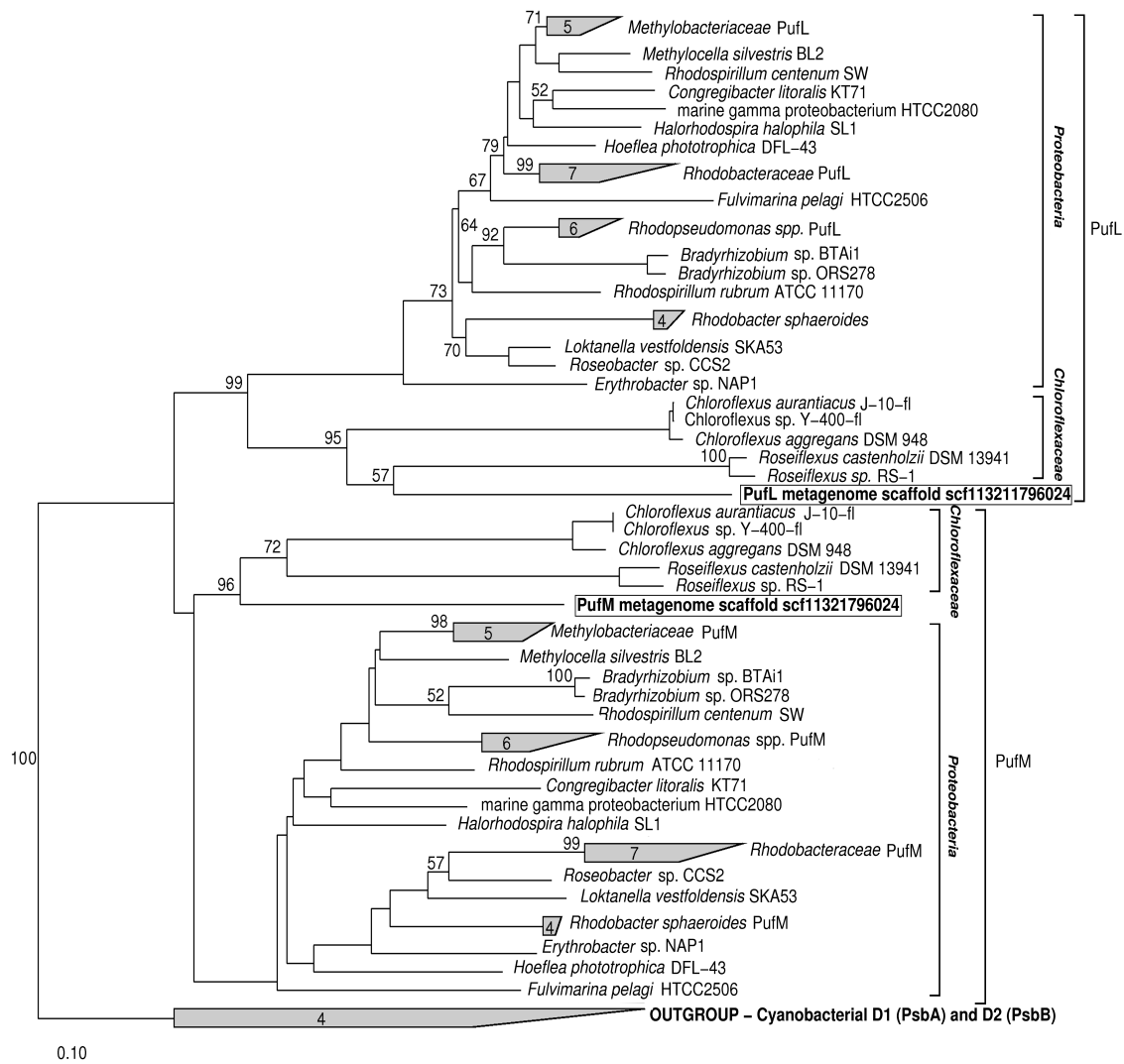


Supplementary Figure 10. Phylogenetic analysis of metagenomic RecA sequences using the Neighbor Joining method. The percentage of replicate trees in which associated taxa clustered together with bootstrapping (1000 replicates) are indicated at the nodes with the following symbols: ○ 50 to 75%, ● 75 to 90%, and ◆ >90%. Labeled RecA sequences were located in assemblies 20 kbp or greater in length and correspond to labels in Figure 3.4.

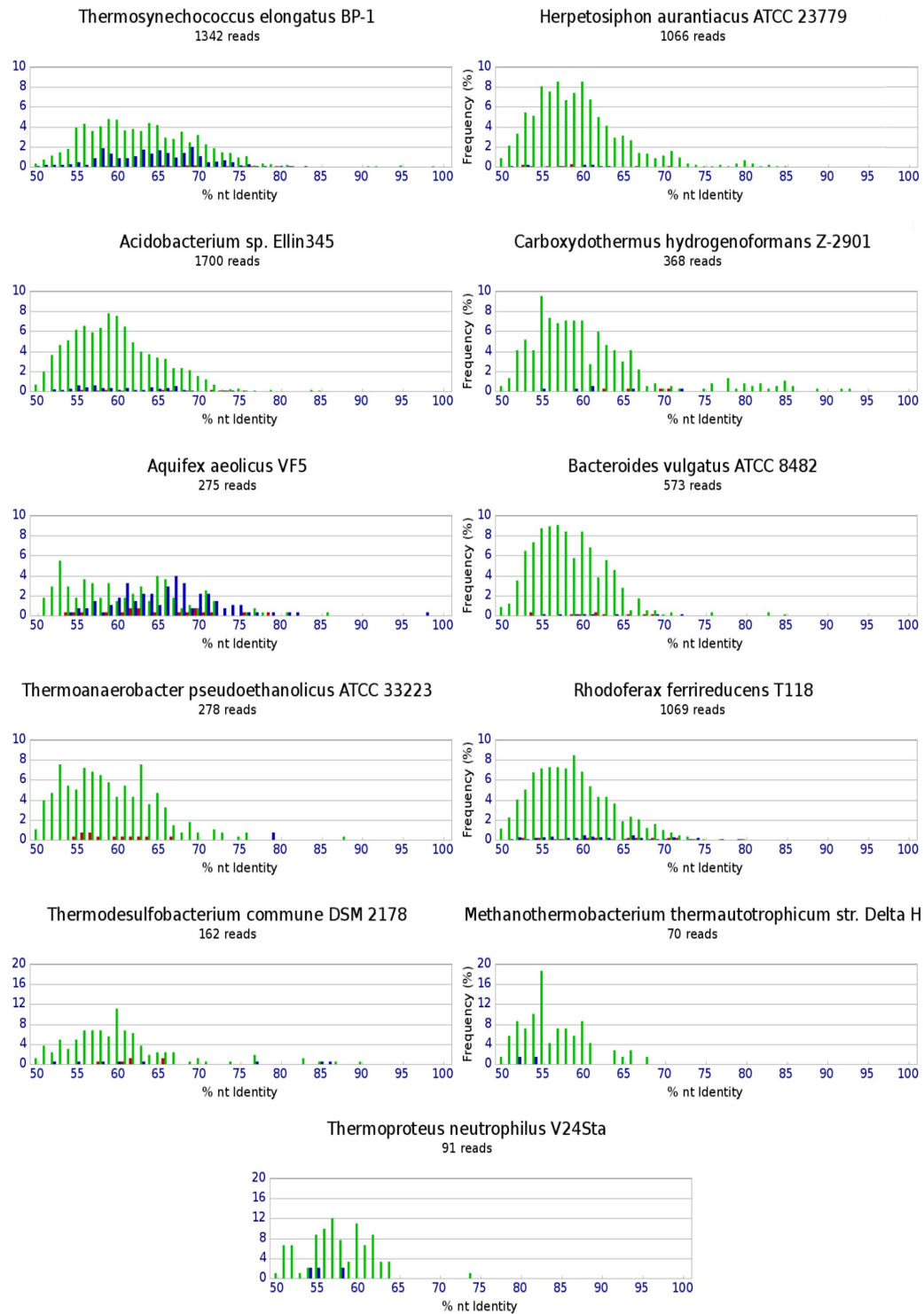
Chloroflexi Kingdom
16S rRNA
NJ Tree



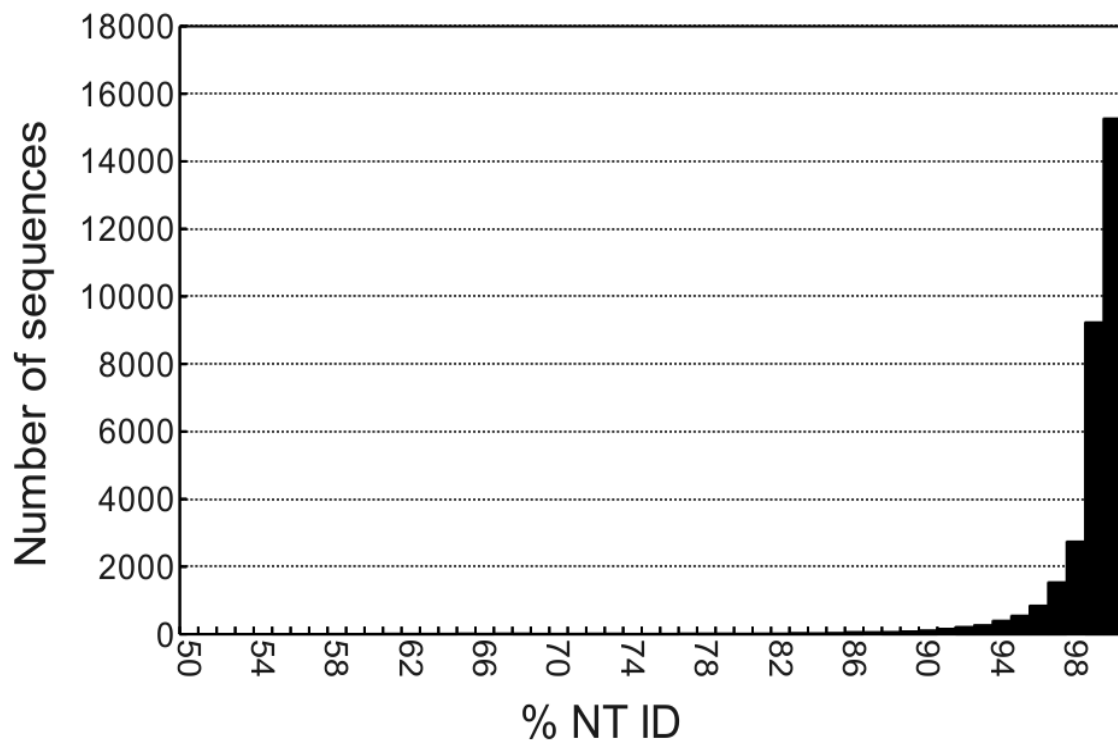
Supplementary Figure 11. Neighbor-joining 16S rRNA phylogenetic tree of novel chlorophototrophic Chloroflexi. Highlighting indicates sequences from chlorophototrophic isolates that contain chlorosomes (green) or do not contain chlorosomes (red). Yellow highlighting indicates isolates that are non-phototrophic chemoorganoheterotrophs, and blue indicates the metagenomic sequence from Cluster 6 in this study. Subdivisions are labeled *sensu* Sekiguchi et al. 2003.



Supplementary Figure 12. Detailed neighbor-joining phylogenetic tree based on PufL and PufM sequences from a novel Chloroflexi metagenomic scaffold from Cluster 6 (boxed) and from sequenced genomes. Numbers at nodes reflect bootstrap support after 1000 replications.



Supplementary Figure 13. Histograms of disjointly recruited (green), jointly recruited syntenous (red) and jointly recruited non-syntenous (blue) metagenomic sequences than cannot be associated confidently with a reference genome.



Supplementary Figure 14. Comparison of Mushroom Spring high temperature metagenomes. The suspected *Synechococcus* sp. A' Sanger metagenome sequences from Mushroom 65 °C were used as queries in a BLASTN to a database consisting of a random shotgun Titanium 454 pyrosequencing metagenome constructed from a Mushroom Spring 68 °C sample.

Supplementary Table 1. Genomes used as references in this study.

	Genome	Source of genome	Source of isolate	Reference	Rationale
1	<i>Synechococcus</i> sp. strain A [JA-3-3Ab]	FIBR; JCVI	58-65 °C Octopus Sp. mat; 7-25-2002	Allewalt et al., 2006; Bhaya et al., 2007	Oxygenic phototroph; known genetic relevance to mat
2	<i>Synechococcus</i> sp. strain B' [JA-2-3B'a(2-13)]	FIBR; JCVI	51-61 °C Octopus Spring mat; 7-10-2002	Allewalt et al., 2006; Bhaya et al., 2007	Oxygenic phototroph; known genetic relevance to mat
3	<i>Thermosynechococcus elongatus</i> BP-1	Kazusa DNA Research Institute	Beppu hot spring in Japan	Nakamura et al. 2002	Oxygenic phototroph; suspected low population density community member
4	<i>Roseiflexus</i> sp. strain RS1	JGI/Don Bryant	60°C Octopus Sp. mat; 7-27-2002	van der Meer et al., 2010; Klatt et al., 2007	FAP; known genetic relevance to mat
5	<i>Chloroflexus</i> sp. strain 396-1	JGI/Don Bryant	30-40°C Conophyton Pool, Fairy Springs Meadow, YNP	Bauld, 1973; Nübel et al., 2002	FAP; distant relative of mat <i>Chloroflexus</i> , but from YNP (unfinished)
6	<i>Candidatus</i> Chloracidobacterium thermophilum	JGI/Don Bryant	51-61°C Octopus Spring mat; 7-10-2002; cultivated from enrichment in 2	Bryant et al., 2007	Anoxygenic phototroph; known genetic relevance to mat (unfinished)
7	<i>Chloroherpeton</i>	PSU/Don	25°C,	Gibson	Anoxygenic

	<i>thalassium</i> ATCC 35110	Bryant	Sippowisset Salt Marsh, Woods Hole, MA	et al., 1984	phototroph; closest known relative to mat GSB (unfinished)
8	<i>Thermomicrobium roseum</i> DSM 5159	Jonathan Eisen	YNP; 74°C Toadstool sp. mat beneath wax paper	Jackson et al., 1973; Wu et al., 2009	Aerobic heterotroph; cultivated from similar YNP mat; recruits some high-quality hits
9	<i>Thermus thermophilus</i> HB8	JCVI CMR	Japanese hot spring; 80°C, pH 6.3	Oshima and Imahori, 1974	Aerobic heterotroph; similar strains commonly isolated from mats
10	<i>Herpetosiphon aurantiacus</i> DSMZ 785	JGI/Don Bryant	Slime coat of green alga (<i>Chara</i> sp.); Birch Lake, MN	Holt and Lewin, 1968	Filamentous aerobic heterotrophic <i>Chloroflexi</i> strain; recruits some reads in test BLASTX
11	<i>Aquifex aeolicus</i> VF5	JCVI	Hydrothermal system, Porto di Levante, Vulcano, Italy (102°C)	Eder and Huber 2002; Deckert et al., 1998	Representative of <i>Aquificales</i> known to inhabit Octopus Spring upstream sampling sites
12	<i>Acidobacterium</i> sp. Ellin345	JGI/Cheryl Kuske	Soil core from mixed rye grass and clover pasture	Davis et al., 2005; Ward et al., 2009	Acidobacterium kingdom representative
13	<i>Thermoanaerobacter pseudoethanolicus</i> 39E	JGI	65°C Octopus Sp. mat, YNP;	Zeikus et al., 1980	Anaerobic fermentor; cultivated from Octopus Spring
14	<i>Carboxydotherrmus hydrogenoformans</i> strain Z-2901		hot swamp from Kunashir Island,	Wu et al., 2005	CO metabolizing anaerobe isolated from hot springs

			Russia; 78°C opt		
15	<i>Bacteroides vulgatus</i> ATCC 8482	Washington Univ. Genome Sequencing Center	Human gut	Xu et al., 2007	CFB representative; several CFBs recruit some hits moderate-quality hits in test BLASTX
16	<i>Rhodoferrax</i> <i>ferrireducans</i> T118 ^T (DSM 15236)	JGI/Derek Lovely	Subsurface sediments; Oyster Bay, VA	Finnera n et al., 2003	Anaerobe Fe reducer; recruits some moderate- quality hits in test BLASTX
17	<i>Thermoproteus</i> <i>neutrophilus</i> V24Sta	JGI/Todd Lowe	Iceland hot spring, 85°C, pH 6.5	Fischer et al., 1983	Crenarchaeota representative; anaerobic fermentor
18	<i>Thermodesulfobacteriu</i> <i>m commune</i> DSM 2178	Jonathan Eisen	YNP spring isolate YSRA-1 from Inkpot Sp., 70°C edge sediment water, pH 6.6	Zeikus et al., 1983; Dillon et al., 2007	YNP isolate whose lipids resemble those found in these mats; not found in <i>dsrA</i> study
19	<i>Thermodesulfovibrio</i> <i>yellowstonii</i> YP87 (ATCC51303)	JCVI	YNP lake thermal vent water	Dillon et al., 2007, Kunisaw a et al., 2010	YNP isolate with <i>dsrA</i> 85-95% NT ID to cloned mat sequences
20	<i>Methanothermobacter</i> <i>thermautotrophicus</i> Δ H		fermenting sludge from Urbana, IL sewage treatment plant	Zeikus & Wolfe, 1972; Smith et al. 1997.	Euryarchaeota representative; other M. thermo strains cultivated from this mat

Supplementary Table 2. Metagenomic libraries produced from DNA obtained after lysis of top green 0-1 mm layer of alkaline siliceous hot spring microbial mats analyzed in this study.¹

Metagenomic library	Clone Insert Size	Number of sequences
Octopus Sp. 58-67°C	2-3 kb	4 216
	10-12 kb	3 838
Octopus Sp. 53-56°C	2-3 kb	19 142
	10-12 kb	80 321
Mushroom Sp. ~65°C	3-4 kb	15 837
	8-9 kb	23 341
Mushroom sp. ~60°C	2-3 kb	8 001
	10-12 kb	7 280
TOTAL		161 976

¹ Additional libraries were produced for both Mushroom Spring samples using DNA obtained by mechanical means (see Klatt et al., 2007; Bhaya et al., 2007).

Supplementary Table 3. Synteny conservation between the *Synechococcus* sp. A and genomes as a function of relatedness. Genomes were fractionated in silico and aligned to the *Synechococcus* sp. A genome to simulate a single 2kb-insert metagenome library of jointly recruited end-sequences.

Genome origin	16S % NT ID to A ¹	% syntenous ²	n	Mean $\pm$ SD % NT ID of syntenous	statistical significance ³
<i>Synechococcus</i> sp. strain B'	96.4	62.20%	12422	84.76 $\pm$ 6.42	mean greater than all other genomes ($p < 10^{-7}$)
<i>Thermosynechococcus elongatus</i>	87.1	8.40%	1680	66.27 $\pm$ 5.76	mean not significantly different from <i>G. violaceus</i> but greater than <i>Synechococcus</i> sp. WH8203 ($p < 0.005$) and <i>Anabaena</i> sp. PCC 7120 & <i>Roseiflexus</i> sp. RS1 ($p < 10^{-7}$)
<i>Gloeobacter violaceus</i>	87.1	5.60%	1112	66.48 $\pm$ 6.16	mean greater than WH8102 ($p < 0.001$) <i>Anabaena</i> sp. PCC 7120 and <i>Roseiflexus</i> sp. RS1 ($p < 10^{-7}$)
<i>Synechococcus</i> sp. WH8102	84.3	8.80%	1752	65.58 $\pm$ 6.05	mean greater than <i>Anabaena</i> sp. PCC 7120 ($p < 10^{-7}$)
<i>Anabaena</i> sp. strain PCC 7120	83.2	3.30%	650	64.74 $\pm$ 5.15	mean greater than <i>Roseiflexus</i> sp. RS1 ($p < 10^{-7}$)
<i>Roseiflexus</i> sp. strain RS1	69.7	1.50%	296	62.14 $\pm$ 5.60	mean less than all other genomes ($p < 10^{-7}$)

¹ pairwise distance matrix of 1284 ungapped positions in the 16S rRNA gene computed using MEGA.

² % Synteny = No. jointly recruited syntenous sequences / No. syntenous and non-syntenous sequences (within range) * 100%.

³ ANOVA with Tukey's HSD post hoc test, unequal sample sizes (conservative), $\alpha = 0.05$. Adjusted p-value from Tukey's HSD reported.

Supplementary Table 4. Top BLASTX matches of metagenomic RecA sequences to the NCBI nr database. Sequences matching *Candidatus Chloracidobacterium thermophilum* were determined by BLASTN to metagenomic scaffolds later identified to originate to relatives of this organism.

Phylogeny	Metagenome Sequence	Library	% AA ID	Top BLASTX match in nr	E-value
cy,A-recA	CYPMD34TR	OS Low	99.9	<i>Synechococcus</i> sp. strain strain A	4.20E-146
cy,A-recA	YMBA716TR	MS High	100.0	<i>Synechococcus</i> sp. strain strain A	1.20E-189
cy,A'orB-recA	YMAAK22TF	MS High	85.0	<i>Synechococcus</i> sp. strain strain A	2.00E-153
cy,A'orB-recA	YMAAZ18TF	MS High	84.3	<i>Synechococcus</i> sp. strain strain A	3.10E-140
cy,A'orB-recA	YMBBJ95TR	MS High	78.9	<i>Synechococcus</i> sp. strain strain B'	3.30E-37
cy,A'orB-recA	YMBBN34TF	MS High	78.8	<i>Synechococcus</i> sp. strain strain B'	1.20E-48
cy,A'orB-recA	YMBCI39TR	MS High	82.7	<i>Synechococcus</i> sp. strain strain B'	9.50E-127
cy,A'orB-recA	YMJB173TR	MS Low	82.3	<i>Synechococcus</i> sp. strain strain A	9.80E-103
cy,B'-recA	CYOAR93TF	OS Low	99.9	<i>Synechococcus</i> sp. strain strain B'	5.70E-152
cy,B'-recA	CYPAQ25TR	OS Low	99.3	<i>Synechococcus</i> sp. strain strain B'	2.30E-183
cy,B'-recA	CYPB635TF	OS Low	98.0	<i>Synechococcus</i> sp. strain strain B'	7.60E-172
cy,B'-recA	CYPBE81TF	OS Low	88.4	<i>Synechococcus</i> sp. strain strain B'	1.40E-129
cy,B'-recA	CYPBQ59TF	OS Low	98.5	<i>Synechococcus</i> sp. strain strain B'	2.30E-188
cy,B'-recA	CYPD180TR	OS Low	99.2	<i>Synechococcus</i> sp. strain strain B'	3.00E-201
cy,B'-recA	CYPED65TF	OS Low	99.0	<i>Synechococcus</i> sp. strain strain B'	1.40E-179
cy,B'-recA	CYPHU21TF	OS Low	97.9	<i>Synechococcus</i> sp. strain strain B'	1.80E-173
cy,B'-recA	CYPIT19TF	OS Low	99.8	<i>Synechococcus</i> sp. strain strain B'	7.40E-177
cy,B'-recA	CYPJ730TR	OS Low	98.4	<i>Synechococcus</i> sp. strain strain B'	9.00E-169
cy,B'-recA	CYPKE13TR	OS Low	97.9	<i>Synechococcus</i> sp. strain strain B'	4.70E-153
cy,B'-recA	YMIA963TF	MS Low	98.7	<i>Synechococcus</i> sp. strain strain B'	4.00E-200
cy,B'-recA	YMJAL81TR	MS Low	99.0	<i>Synechococcus</i> sp. strain strain B'	1.20E-188
cy,other-recA	CYPM011TR	OS Low	72.9	<i>Synechococcus</i> sp. strain strain B'	8.50E-48
cfx3-rs	CYOB093TF	OS Low	96.2	<i>Roseiflexus</i> RS1	1.90E-176
cfx3-rs	CYOC33TR	OS Low	97.4	<i>Roseiflexus</i> RS1	4.40E-177
cfx3-rs	YMIAN43TR	MS Low	98.6	<i>Roseiflexus</i> RS1	5.40E-156
cfx-1	GYOAU08TR	MS Low	89.6	<i>Roseiflexus</i> RS1	8.50E-158
cfx-1	YMAB934TF	MS High	89.3	<i>Roseiflexus</i> RS1	2.00E-139

cfx2	CYPAA42TR	OS Low	73.6	<i>Symbiobacterium thermophilum</i> IAM14863	2.40E-81
cfx2	CYPJ232TF	OS Low	66.6	<i>Symbiobacterium thermophilum</i> IAM14863	4.30E-55
cfx2	GYPAF55TR	MS Low	73.0	<i>Symbiobacterium thermophilum</i> IAM14863	5.50E-57
cfx2	GYPAU15TF	MS Low	73.3	<i>Symbiobacterium thermophilum</i> IAM14863	1.00E-83
cfx2	YMABV46TF	MS High	65.8	<i>Symbiobacterium thermophilum</i> IAM14863	3.40E-52
cfx2	YMBBH30TF	MS High	68.6	<i>Symbiobacterium thermophilum</i> IAM14863	7.50E-71
chlorobi	YMJA487TR	MS Low	68.9	<i>Chlorobium tepidum</i> TLS	2.20E-50
chlorobi	YMJA904TR	MS Low	69.1	<i>Chlorobium tepidum</i> TLS	1.30E-51
chlorobi	CYOA050TR	OS Low	68.0	<i>Chlorobium tepidum</i> TLS	1.20E-10
chlorobi	CYOB302TF	OS Low	69.4	<i>Chlorobium tepidum</i> TLS	3.60E-70
chlorobi	CYOBZ08TR	OS Low	68.7	<i>Chlorobium tepidum</i> TLS	5.30E-69
chlorobi	CYOBZ28TR	OS Low	69.4	<i>Chlorobium tepidum</i> TLS	3.10E-46
chlorobi	CYOC922TR	OS Low	68.9	<i>Chlorobium tepidum</i> TLS	2.60E-48
chlorobi	CYPAQ36TF	OS Low	70.0	<i>Chlorobium tepidum</i> TLS	5.10E-80
chlorobi	CYPAW08TF	OS Low	69.7	<i>Chlorobium tepidum</i> TLS	1.20E-69
chlorobi	CYPBL73TF	OS Low	68.2	<i>Chlorobium tepidum</i> TLS	3.10E-33
chlorobi	CYPC421TF	OS Low	67.7	<i>Chlorobium tepidum</i> TLS	9.80E-25
chlorobi	CYPC505TR	OS Low	66.6	<i>Chlorobium tepidum</i> TLS	2.70E-22
chlorobi	CYPDM66TF	OS Low	69.0	<i>Chlorobium tepidum</i> TLS	2.30E-76
chlorobi	CYPEE96TR	OS Low	60.4	<i>Chloroflexis aurantiacus</i> J-10-fl	0.06
chlorobi	CYPEH75TR	OS Low	66.0	<i>Chlorobium tepidum</i> TLS	5.90E-36
chlorobi	CYPHG37TR	OS Low	69.1	<i>Chlorobium tepidum</i> TLS	3.10E-78
chlorobi	CYPM893TR	OS Low	68.4	<i>Chlorobium tepidum</i> TLS	2.30E-75
chlorobi	CYPME37TF	OS Low	65.8	<i>Chlorobium tepidum</i> TLS	1.70E-20
firmicuti	CYPH994TF	OS Low	65.7	<i>Caldicellulosiruptor saccharolyticus</i>	1.00E-49
firmicuti	CYPJZ78TF	OS Low	64.2	<i>Symbiobacterium thermophilum</i> IAM14863	5.90E-23
firmicuti	CYPL354TR	OS Low	66.1	<i>Acidobacterium sp. strain</i> Ellin6076	1.60E-39
firmicuti	GYOA428TF	MS Low	70.3	<i>Symbiobacterium thermophilum</i> IAM14863	3.10E-79
firmicuti	GYRAU55TF	MS High	69.9	<i>Symbiobacterium thermophilum</i> IAM14863	2.30E-68
firmicuti	GYSA222TF	MS High	67.6	<i>Symbiobacterium thermophilum</i> IAM14863	1.50E-50
firmicuti	GYTA875TR	MS High	66.0	<i>Symbiobacterium thermophilum</i> IAM14863	8.80E-57

firmicuti	GYUAD41TF	MS High	67.7	<i>Roseiflexus</i> RS1	8.40E-29
firmicuti	YMABG37TF	MS High	67.4	<i>Symbiobacterium thermophilum</i> IAM14863	2.60E-42
firmicuti	YMBBP66TF	MS High	67.6	<i>Symbiobacterium thermophilum</i> IAM14863	1.90E-24
firmicuti	YMBCJ32TF	MS High	66.0	<i>Symbiobacterium thermophilum</i> IAM14863	9.80E-47
firmicuti	YMBEQ77TR	MS High	71.2	<i>Symbiobacterium thermophilum</i> IAM14863	9.10E-70
firmicuti	YMBER53TF	MS High	63.2	<i>Symbiobacterium thermophilum</i> IAM14863	2.30E-40
firmicuti	YMIA184TF	MS Low	67.1	<i>Symbiobacterium thermophilum</i> IAM14863	3.20E-63
gfp-recA	CYMAF31TF	OS High	100.0	<i>Chloracidobacterium thermophilum</i>	1.70E-173
gfp-recA	CYCH34TF	OS Low	100.0	<i>Chloracidobacterium thermophilum</i>	2.00E-202
gfp-recA	CYPEZ61TF	OS Low	99.9	<i>Chloracidobacterium thermophilum</i>	4.70E-171
gfp-recA	CYPFK94TR	OS Low	99.9	<i>Chloracidobacterium thermophilum</i>	1.40E-182
gfp-recA	CYPIC44TF	OS Low	100.0	<i>Chloracidobacterium thermophilum</i>	2.40E-191
gfp-recA	CYPKS71TF	OS Low	86.3	<i>Chloracidobacterium thermophilum</i>	3.10E-128
gfp-recA	CYPLM15TF	OS Low	98.9	<i>Chloracidobacterium thermophilum</i>	3.50E-53
gfp-recA	CYPLX42TR	OS Low	99.9	<i>Chloracidobacterium thermophilum</i>	1.60E-171
gfp-recA	YMJB724TF	MS Low	100.0	<i>Chloracidobacterium thermophilum</i>	5.70E-195
proteo-recA	CYPH352TF	OS Low	66.8	<i>Thermoanaerobacter ethanolicus</i> strain 39E	1.10E-37
proteo-recA	CYPI901TF	OS Low	66.8	<i>Thermoanaerobacter ethanolicus</i> strain 39E	4.50E-66
proteo-recA	YMIAU71TF	MS Low	67.8	<i>Thermoanaerobacter ethanolicus</i> strain 39E	1.80E-40
proteo-recA	GYUAH20TR	MS High	69.8	<i>Symbiobacterium thermophilum</i> IAM14863	3.80E-74
other-recA	GYRA005TF	MS High	65.5	<i>Thermus thermophilus</i> HB8	5.10E-09
other-recA	GYOA442TF	MS Low	70.0	<i>Symbiobacterium thermophilum</i> IAM14863	1.40E-54
other-recA	YMAAU07TR	MS High	75.8	<i>Gemmata obscuriglobus</i> UQM 2246	2.00E-065

Supplementary Table 5. AMPHORA identification of 31 different phylogenetic marker genes and their associated taxonomic calls. Taxonomic ranks indicate the most specific (Rank 2) and next-most specific (Rank 1) taxonomic level that these sequences could be assigned above a 70% bootstrap cutoff.

Putative metagenomic ORF	Rank 1	Rank 2
JCVI_PEP_metagenomic.orf.21162558.1	<i>Acidobacteria</i>	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21461737.1	<i>Acidobacteria</i>	<i>Acidobacteria</i> bacterium Ellin345
JCVI_PEP_metagenomic.orf.20810374.1	<i>Acidobacteria</i>	<i>Acidobacteria</i> bacterium Ellin345
JCVI_PEP_metagenomic.orf.20824390.1	<i>Acidobacteria</i>	<i>Solibacter usitatus</i> Ellin6076
JCVI_PEP_metagenomic.orf.20932260.1	<i>Acidobacteria</i>	<i>Solibacter usitatus</i> Ellin6076
JCVI_PEP_metagenomic.orf.21074597.1	<i>Alphaproteobacteria</i>	<i>Orientia tsutsugamushi</i> Boryong
JCVI_PEP_metagenomic.orf.21523186.1	<i>Alphaproteobacteria</i>	<i>Orientia tsutsugamushi</i> Boryong
JCVI_PEP_metagenomic.orf.21071750.1	<i>Aquifex aeolicus</i>	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21319792.1	<i>Aquifex aeolicus</i>	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21010294.1	<i>Aquifex aeolicus</i>	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21409163.1	<i>Aquifex aeolicus</i>	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.20920732.1	<i>Aquifex aeolicus</i>	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21526199.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21526695.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21572994.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.20938253.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21407097.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21460848.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21453812.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21453449.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21537268.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21158376.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.21132746.1	Bacteria	<i>Acidobacteria</i>
JCVI_PEP_metagenomic.orf.20801436.1	Bacteria	<i>Acidobacteria</i> bacterium Ellin345
JCVI_PEP_metagenomic.orf.21453551.1	Bacteria	<i>Actinobacteria</i>
JCVI_PEP_metagenomic.orf.20840483.1	Bacteria	<i>Actinobacteria</i>
JCVI_PEP_metagenomic.orf.20930790.1	Bacteria	<i>Actinobacteria</i>

JCVI_PEP_metagenomic.orf.21569090.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.21179659.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.21358671.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.21330466.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.21359781.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.21206699.1	Bacteria	<i>Actinobacteridae</i>
JCVI_PEP_metagenomic.orf.20933632.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21458889.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21317712.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21100457.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21383095.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21320407.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.20919892.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.20824065.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.20804555.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21034594.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.21459128.1	Bacteria	<i>Aquifex aeolicus</i> VF5
JCVI_PEP_metagenomic.orf.20815561.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21199224.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21036241.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21290807.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20968313.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20879377.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21102520.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20942391.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21519377.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20949884.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20924335.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21324945.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20814215.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21314654.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20938965.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21459216.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20780591.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20989192.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21519362.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20901504.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20872036.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20784203.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20851993.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21306373.1	Bacteria	Bacteria

JCVI_PEP_metagenomic.orf.20975679.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21529158.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.21273117.1	Bacteria	Bacteria
JCVI_PEP_metagenomic.orf.20912906.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21260236.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21346610.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20774295.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21194420.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21245345.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20898942.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20793661.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20808486.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21026213.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21072816.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20994853.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21081500.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.20911265.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21192055.1	Bacteria	<i>Bacteroidetes/Chlorobi</i> group
JCVI_PEP_metagenomic.orf.21296930.1	Bacteria	<i>Bdellovibrio bacteriovorus</i> HD100
JCVI_PEP_metagenomic.orf.20819148.1	Bacteria	<i>Borrelia burgdorferi</i> group
JCVI_PEP_metagenomic.orf.20962537.1	Bacteria	<i>Campylobacteriales</i>
JCVI_PEP_metagenomic.orf.21529129.1	Bacteria	<i>Candidatus Pelagibacter ubique</i> HTCC1062
JCVI_PEP_metagenomic.orf.21245353.1	Bacteria	<i>Candidatus Pelagibacter ubique</i> HTCC1062
JCVI_PEP_metagenomic.orf.21214568.1	Bacteria	<i>Candidatus Sulcia muelleri</i> GWSS
JCVI_PEP_metagenomic.orf.21480770.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.21079280.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.20988791.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.21022832.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.21529448.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.20918451.1	Bacteria	<i>Chlamydiales</i>
JCVI_PEP_metagenomic.orf.21303636.1	Bacteria	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21082550.1	Bacteria	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.20954524.1	Bacteria	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.20868803.1	Bacteria	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21321292.1	Bacteria	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21094829.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21036381.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21205210.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21528989.1	Bacteria	<i>Chloroflexi</i>

JCVI_PEP_metagenomic.orf.20924839.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21517280.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21187897.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21292768.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21159321.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.20908197.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21200677.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21196120.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21529044.1	Bacteria	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21459391.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20781204.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21074298.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20872896.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21155277.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21276587.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20776314.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21529055.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20957978.1	Bacteria	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20868799.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21358004.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21409399.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21528932.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21091317.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21200392.1	Bacteria	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.20989736.1	Bacteria	<i>Desulfococcus oleovorans</i> Hxd3
JCVI_PEP_metagenomic.orf.20784658.1	Bacteria	<i>Desulfovibrionaceae</i>
JCVI_PEP_metagenomic.orf.20920368.1	Bacteria	<i>Epsilonproteobacteria</i>
JCVI_PEP_metagenomic.orf.21375401.1	Bacteria	<i>Flavobacteriaceae</i>
JCVI_PEP_metagenomic.orf.21153260.1	Bacteria	<i>Fusobacterium nucleatum</i> subsp. <i>nucleatum</i> ATCC 25586
JCVI_PEP_metagenomic.orf.21144108.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21111304.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21458602.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21221840.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.20777017.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.20854335.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21221353.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21317541.1	Bacteria	<i>Leptospira</i>
JCVI_PEP_metagenomic.orf.21028950.1	Bacteria	<i>Mollicutes</i>
JCVI_PEP_metagenomic.orf.20924569.1	Bacteria	<i>Mollicutes</i>
JCVI_PEP_metagenomic.orf.21028614.1	Bacteria	<i>Mycoplasma</i>

JCVI_PEP_metagenomic.orf.21297364.1	Bacteria	<i>Mycoplasma</i>
JCVI_PEP_metagenomic.orf.20784353.1	Bacteria	<i>Mycoplasma</i>
JCVI_PEP_metagenomic.orf.21365464.1	Bacteria	<i>Mycoplasma</i>
JCVI_PEP_metagenomic.orf.20855263.1	Bacteria	<i>Mycoplasma</i>
JCVI_PEP_metagenomic.orf.20838881.1	Bacteria	<i>Mycoplasma hyopneumoniae</i>
JCVI_PEP_metagenomic.orf.21139435.1	Bacteria	<i>Mycoplasma penetrans</i> HF-2
JCVI_PEP_metagenomic.orf.20920133.1	Bacteria	<i>Myxococcales</i>
JCVI_PEP_metagenomic.orf.20938562.1	Bacteria	<i>Nitrosococcus oceani</i> ATCC 19707
JCVI_PEP_metagenomic.orf.21320314.1	Bacteria	<i>Novosphingobium aromaticivorans</i> DSM 12444
JCVI_PEP_metagenomic.orf.20858256.1	Bacteria	<i>Orientia tsutsugamushi</i> Boryong
JCVI_PEP_metagenomic.orf.21362477.1	Bacteria	<i>Pelotomaculum thermopropionicum</i> SI
JCVI_PEP_metagenomic.orf.21104271.1	Bacteria	<i>Peptococcaceae</i>
JCVI_PEP_metagenomic.orf.21478868.1	Bacteria	<i>Petrogoga mobilis</i> SJ95
JCVI_PEP_metagenomic.orf.21016285.1	Bacteria	<i>Proteobacteria</i>
JCVI_PEP_metagenomic.orf.21504562.1	Bacteria	<i>Proteobacteria</i>
JCVI_PEP_metagenomic.orf.21012197.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21117197.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21240118.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21121086.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21003034.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.20834448.1	Bacteria	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21251814.1	Bacteria	<i>Rickettsia</i>
JCVI_PEP_metagenomic.orf.20905428.1	Bacteria	<i>Rickettsia</i>
JCVI_PEP_metagenomic.orf.21487014.1	Bacteria	<i>Rickettsia</i>
JCVI_PEP_metagenomic.orf.21458512.1	Bacteria	<i>Rickettsia</i>
JCVI_PEP_metagenomic.orf.20927832.1	Bacteria	<i>Rickettsiales</i>
JCVI_PEP_metagenomic.orf.21561058.1	Bacteria	<i>Rickettsiales</i>
JCVI_PEP_metagenomic.orf.21123815.1	Bacteria	<i>Rubrobacter xylophilus</i> DSM 9941
JCVI_PEP_metagenomic.orf.21362765.1	Bacteria	<i>Rubrobacter xylophilus</i> DSM 9941
JCVI_PEP_metagenomic.orf.20890232.1	Bacteria	<i>Rubrobacter xylophilus</i> DSM 9941
JCVI_PEP_metagenomic.orf.20967497.1	Bacteria	<i>Rubrobacter xylophilus</i> DSM 9941
JCVI_PEP_metagenomic.orf.21246109.1	Bacteria	<i>Rubrobacter xylophilus</i> DSM 9941
JCVI_PEP_metagenomic.orf.20821160.1	Bacteria	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.21321134.1	Bacteria	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.20819782.1	Bacteria	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.20939865.1	Bacteria	<i>Solibacter usitatus</i> Ellin6076
JCVI_PEP_metagenomic.orf.20995039.1	Bacteria	<i>Solibacter usitatus</i> Ellin6076
JCVI_PEP_metagenomic.orf.21560300.1	Bacteria	<i>Solibacter usitatus</i> Ellin6076
JCVI_PEP_metagenomic.orf.21453415.1	Bacteria	<i>Spirochaetaceae</i>
JCVI_PEP_metagenomic.orf.21479159.1	Bacteria	<i>Spirochaetales</i>

JCVI_PEP_metagenomic.orf.20857581.1	Bacteria	<i>Spirochaetales</i>
JCVI_PEP_metagenomic.orf.21304401.1	Bacteria	<i>Spirochaetales</i>
JCVI_PEP_metagenomic.orf.20885735.1	Bacteria	<i>Spirochaetales</i>
JCVI_PEP_metagenomic.orf.21072517.1	Bacteria	<i>Spirochaetales</i>
JCVI_PEP_metagenomic.orf.21305898.1	Bacteria	<i>Symbiobacterium thermophilum</i> IAM 14863
JCVI_PEP_metagenomic.orf.21382847.1	Bacteria	<i>Symbiobacterium thermophilum</i> IAM 14863
JCVI_PEP_metagenomic.orf.20840930.1	Bacteria	<i>Syntrophus aciditrophicus</i> SB
JCVI_PEP_metagenomic.orf.20806281.1	Bacteria	<i>Syntrophus aciditrophicus</i> SB
JCVI_PEP_metagenomic.orf.21086467.1	Bacteria	<i>Syntrophus aciditrophicus</i> SB
JCVI_PEP_metagenomic.orf.20840144.1	Bacteria	<i>Thermosiphon melanesiensis</i> BI429
JCVI_PEP_metagenomic.orf.20878775.1	Bacteria	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.21166258.1	Bacteria	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.20868399.1	Bacteria	<i>Thermotogaceae</i>
JCVI_PEP_metagenomic.orf.20821605.1	Bacteria	<i>Thermotogaceae</i>
JCVI_PEP_metagenomic.orf.21537248.1	Bacteria	<i>Thermotogaceae</i>
JCVI_PEP_metagenomic.orf.21137644.1	Bacteria	<i>Thermotogaceae</i>
JCVI_PEP_metagenomic.orf.21139632.1	Bacteria	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.20959128.1	Bacteria	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21223408.1	Bacteria	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21169968.1	Bacteria	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21269687.1	Bacteria	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21023707.1	Bacteria	<i>Treponema</i>
JCVI_PEP_metagenomic.orf.20914997.1	Bacteria	<i>Treponema</i>
JCVI_PEP_metagenomic.orf.20877458.1	Bacteria	<i>Tropheryma whipplei</i>
JCVI_PEP_metagenomic.orf.20845195.1	Bacteria	<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700970
JCVI_PEP_metagenomic.orf.20845703.1	Bacteria	<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700970
JCVI_PEP_metagenomic.orf.20901408.1	Bacteria	<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700970
JCVI_PEP_metagenomic.orf.20832135.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.20800567.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.21181541.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.21014944.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.21296525.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.20913455.1	<i>Bacteroidetes</i>	<i>Bacteroidetes</i>
JCVI_PEP_metagenomic.orf.21244858.1	<i>Bacteroidetes</i>	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.20888786.1	<i>Bacteroidetes</i>	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.20830705.1	<i>Borrelia</i>	<i>Borrelia burgdorferi</i> group
JCVI_PEP_metagenomic.orf.21055845.1	<i>Caldicellulosirum</i>	<i>Caldicellulosiruptor saccharolyticus</i>

	<i>ptor saccharolyticus</i>	DSM 8903
JCVI_PEP_metagenomic.orf.20800317.1	<i>Chlorobi</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21478931.1	<i>Chlorobi</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21086604.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21479461.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21458039.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21479751.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21355077.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21283734.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21050474.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21480065.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21193815.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21478970.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21273913.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21467880.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21391959.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21479323.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21480160.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21479651.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.20853327.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21392184.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21467991.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21320079.1	<i>Chlorobiaceae</i>	<i>Chlorobiaceae</i>
JCVI_PEP_metagenomic.orf.21338132.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21352272.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21369456.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21378223.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21353085.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21113622.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21352028.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21378122.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21360339.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21352864.1	<i>Chloroflexaceae</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.20915475.1	<i>Chloroflexaceae</i>	<i>Roseiflexus</i>
JCVI_PEP_metagenomic.orf.20920295.1	<i>Chloroflexaceae</i>	<i>Roseiflexus</i>
JCVI_PEP_metagenomic.orf.20916343.1	<i>Chloroflexaceae</i>	<i>Roseiflexus</i>
JCVI_PEP_metagenomic.orf.21250843.1	<i>Chloroflexaceae</i>	<i>Roseiflexus</i>
JCVI_PEP_metagenomic.orf.20918336.1	<i>Chloroflexaceae</i>	<i>Roseiflexus</i>
JCVI_PEP_metagenomic.orf.21529098.1	<i>Chloroflexi</i>	<i>Chloroflexaceae</i>
JCVI_PEP_metagenomic.orf.21353743.1	<i>Chloroflexi</i>	<i>Chloroflexaceae</i>

JCVI_PEP_metagenomic.orf.20944654.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21193400.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.20926054.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21484749.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21306408.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21529438.1	<i>Chloroflexi</i>	<i>Chloroflexi</i>
JCVI_PEP_metagenomic.orf.21432603.1	<i>Chloroflexi</i>	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.20937505.1	<i>Chloroflexi</i>	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.21127801.1	<i>Chloroflexus aurantiacus</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21252220.1	<i>Chloroflexus aurantiacus</i>	<i>Chloroflexus aurantiacus</i> J-10-fl
JCVI_PEP_metagenomic.orf.21430555.1	<i>Chroococcales</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21014528.1	<i>Chroococcales</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21495503.1	<i>Chroococcales</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21320249.1	<i>Cyanobacteria</i>	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21357323.1	<i>Cyanobacteria</i>	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20960197.1	<i>Cyanobacteria</i>	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21361995.1	<i>Cyanobacteria</i>	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.20785980.1	<i>Cyanobacteria</i>	<i>Cyanobacteria</i>
JCVI_PEP_metagenomic.orf.21183812.1	<i>Cyanobacteria</i>	<i>Nostocaceae</i>
JCVI_PEP_metagenomic.orf.21495622.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21495846.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21002342.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20891998.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20882732.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20990243.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21065389.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21495769.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20828793.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21160673.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20915255.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21243447.1	<i>Cyanobacteria</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21393958.1	<i>Cyanobacteria</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21255393.1	<i>Cyanobacteria</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21529004.1	<i>Dehalococcoides</i>	<i>Dehalococcoides</i>
JCVI_PEP_metagenomic.orf.20931361.1	<i>Deinococci</i>	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21491290.1	<i>Deinococcus-Thermus</i>	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21283254.1	<i>Deinococcus-Thermus</i>	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21027325.1	<i>Deltaproteobact</i>	<i>Desulfuromonadales</i>

	<i>eria</i>	
JCVI_PEP_metagenomic.orf.20952189.1	<i>Deltaproteobact eria</i>	<i>Syntrophus aciditrophicus</i> SB
JCVI_PEP_metagenomic.orf.21490837.1	<i>Desulfuromona dales</i>	<i>Pelobacter carbinolicus</i> DSM 2380
JCVI_PEP_metagenomic.orf.20854100.1	<i>Mollicutes</i>	<i>Mycoplasmataceae</i>
JCVI_PEP_metagenomic.orf.20937749.1	<i>Mycoplasmatac eae</i>	<i>Mycoplasma gallisepticum</i> R
JCVI_PEP_metagenomic.orf.21320148.1	<i>Mycoplasmatac eae</i>	<i>Ureaplasma parvum</i> serovar 3 str. ATCC 700970
JCVI_PEP_metagenomic.orf.21495895.1	<i>Proteobacteria</i>	<i>Buchnera aphidicola</i>
JCVI_PEP_metagenomic.orf.20921425.1	<i>Proteobacteria</i>	<i>Candidatus Pelagibacter ubique</i> HTCC1062
JCVI_PEP_metagenomic.orf.20883551.1	<i>Proteobacteria</i>	<i>Deltaproteobacteria</i>
JCVI_PEP_metagenomic.orf.21320150.1	<i>Proteobacteria</i>	<i>Proteobacteria</i>
JCVI_PEP_metagenomic.orf.21022556.1	<i>Rhodopirellula baltica</i>	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21204324.1	<i>Rhodopirellula baltica</i>	<i>Rhodopirellula baltica</i> SH 1
JCVI_PEP_metagenomic.orf.21006065.1	<i>Rickettsiales</i>	<i>Rickettsiales</i>
JCVI_PEP_metagenomic.orf.20919174.1	<i>Roseiflexus</i>	<i>Roseiflexus castenholzii</i> DSM 13941
JCVI_PEP_metagenomic.orf.21527565.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.21057935.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20890429.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.21251664.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20913152.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20779084.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20773306.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20793956.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20846400.1	<i>Roseiflexus</i>	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.21328284.1	<i>Roseiflexus</i> sp. RS-1	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.20911660.1	<i>Roseiflexus</i> sp. RS-1	<i>Roseiflexus</i> sp. RS-1
JCVI_PEP_metagenomic.orf.21039100.1	<i>Salinibacter ruber</i>	<i>Salinibacter ruber</i> DSM 13855
JCVI_PEP_metagenomic.orf.21126128.1	<i>Sphingobacteria les</i>	<i>Cytophaga hutchinsonii</i> ATCC 33406
JCVI_PEP_metagenomic.orf.21430387.1	<i>Synechococcus</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.21126862.1	<i>Synechococcus</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20925562.1	<i>Synechococcus</i>	<i>Synechococcus</i>
JCVI_PEP_metagenomic.orf.20962497.1	<i>Synechococcus</i>	<i>Synechococcus</i>

JCVI_PEP_metagenomic.orf.21068513.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21256465.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20978440.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21513105.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21254805.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21244105.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21243955.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21058422.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21390991.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20833897.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21257613.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21347094.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21180175.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20810453.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21254152.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21394656.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21376275.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21101614.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21256008.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20781587.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21350917.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20791093.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21092388.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21180528.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21384207.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21111842.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21375545.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21007810.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21376207.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21257234.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21365622.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20806901.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.21495105.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21021783.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20827679.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20907307.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20860436.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21384670.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21430107.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21390764.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21244570.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A

JCVI_PEP_metagenomic.orf.21495545.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21316567.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20840724.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21230198.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21538252.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20791624.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21026008.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20881549.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21085342.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21495965.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21297553.1	<i>Synechococcus</i>	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21362913.1	<i>Synechococcus</i> sp. strain B'	<i>Synechococcus</i> sp. strain B'
JCVI_PEP_metagenomic.orf.20829181.1	<i>Synechococcus</i> sp. strain A	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21495374.1	<i>Synechococcus</i> sp. strain A	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.20822899.1	<i>Synechococcus</i> sp. strain A	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21495447.1	<i>Synechococcus</i> sp. strain A	<i>Synechococcus</i> sp. strain A
JCVI_PEP_metagenomic.orf.21223210.1	<i>Thermales</i>	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.20772473.1	<i>Thermales</i>	<i>Thermus thermophilus</i>
JCVI_PEP_metagenomic.orf.21162240.1	<i>Thermotoga</i>	<i>Thermotoga</i>
JCVI_PEP_metagenomic.orf.21053607.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.20909029.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.20946999.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.21098502.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.20865826.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.20864887.1	<i>Thermotoga</i>	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.21059200.1	<i>Thermotogacea</i> e	<i>Thermotoga lettingae</i> TMO
JCVI_PEP_metagenomic.orf.21270101.1	<i>Thermus</i> <i>thermophilus</i>	<i>Thermus thermophilus</i> HB27
JCVI_PEP_metagenomic.orf.21269829.1	<i>Thermus</i> <i>thermophilus</i>	<i>Thermus thermophilus</i> HB27

Supplementary Table 6. 16S rRNA and RecA sequences detected in the metagenomes

Reference genome	No. 16S rRNA genes in reference genome	16S rRNA % of total ¹		RecA % of total ²
		Raw	normalized	
<i>Synechococcus</i> sp. strain A	2	13.5	6.75	2.4
<i>Synechococcus</i> A' ³	-	3.68	(3.68)	2.4
<i>Synechococcus</i> sp. strain B'	2	19	9.51	15.8
<i>Roseiflexus</i> sp. RS1	2	6.75	3.37	6.1
<i>Chloroflexus</i> sp. strain 396-1	?	6.75	(6.75)	
<i>Cand.</i> <i>Chloracidobacterium thermophilum</i>	1	1.84	1.84	11.0
<i>Chloroherpeton thalassium</i>	1	9.82	9.82	22.0
<i>Thermomicrobium roseum</i>	3	1.23	0.41	
<i>Thermus thermophilus</i>	2	1.84	1.84	
<i>Thermodesulfovibrio yellowstonii</i>	3	ND	ND	
Firmicutes (OS-L)	-	11.6	(11.6)	6.1
Planctomyces	-	0.61	(0.61)	
CFG OPB88	2	3.1	(3.1)	
OP99	-	0.61	(0.61)	
<i>Synechococcus</i> sp. strain C9/other cyano	2	1.23	(1.23)	6.1
Spirochete	2	0.61	(0.61)	
Unknown.	-	17.8	(17.8)	

¹ number of 16S rRNA matches / (total number of 16S rRNA matches * number of 16S rRNA copies per genome); low percentages are suspect due to low numbers of matches.

² percentage of RecA with top matches to sequenced genomes from total RecA sequences in metagenome. Sequences with top matches below 70% identity to sequenced genomes using NCBI BLASTX were categorized as “Unknown”. Normalizing corrections were not used due to most genomes containing recA in single copy.

³ values in parentheses were not normalized for 16S rRNA copy number, which is unknown.

Supplementary Table 7. Relationship between sequences in clusters and recruitment bins.

	% <i>Synechococcus</i> sp. strain A	% <i>Synechococcus</i> sp. strain B	% <i>T. elongatus</i> BP-1	% <i>Roseiflexus</i> sp. strain RS1	% <i>Chloroflexus</i> sp. 396-1	% <i>Cand. C. thermophilum</i>	% <i>C. thalassium</i>	% <i>T. roseum</i>	% <i>T. thermophilus</i>	% <i>H. aurantiacus</i>	% <i>Cand. K. versatilis</i>	% <i>T. ethanolicus</i>	% <i>C. hydrogenoformans</i>	% <i>B. vulgatus</i>	% <i>T. yellowstonii</i>	% <i>T. commune</i>	% <i>R. ferrireducens</i>	% <i>M. thermautotrophicus</i>	% <i>A. aeolicus</i>	% <i>T. neutrophilus</i>	% Null	Total No. of Sequences
Cluster 1	59.3	39.6	0	0	0	0.2	0	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0.6	19452
Cluster 2	0	0	0	97.9	0.8	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.1	18203
Cluster 3	2	0.9	0	1.9	81.9	1.1	0.3	0.7	0.8	0.5	0.2	0	0.4	0.3	0	0	0.4	0.2	0	0	8.3	1080
Cluster 4	0.1	0.1	0	0.1	0.1	98.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.2	13381
Cluster 5	0.5	0.3	0.6	1.5	1.1	3.9	52.4	0.4	0.1	1.5	0.9	0.1	0.3	0.9	0.3	0.1	0.5	0	0.1	0	34.4	17358
Cluster 6	2.6	1.3	0.4	26.3	6.3	3.4	0.8	11.3	3.6	3.6	3.4	0.1	0.2	0.2	0.1	0	3	0	0	0.1	33.3	8650
Cluster 7	3.8	1.5	0.7	13.8	1.8	6.9	0.5	7.5	6.2	1.2	6.3	0.1	0.5	0.2	0.1	0.1	2.8	0.1	0.3	0.4	45.1	8354
Cluster 8	2.6	1.6	1	2.3	1.4	6.4	4.1	1.5	5.8	0.9	2.3	0.3	0.9	4	0.3	0.3	2.5	0.1	0.4	0.1	61	3512

Supplementary Table 8. Celera assembly statistics of scaffolds consisting entirely of sequences recruited by either the *Synechococcus* sp. strain A or B' genome in metagenome recruitment. All % NT ID values were obtained from alignments made using BLASTN against the *Synechococcus* spp. strain A or B' genomes separately (*i. e.*, “forced” alignment, see Methods).

Recruitment bins	number of scaffolds	Mean $\pm$ S.D. % NT ID with respect to <i>Synechococcus</i> sp. A	Mean $\pm$ S.D. % NT ID with respect to <i>Synechococcus</i> sp. B'	statistical significance
Exclusively <i>Synechococcus</i> sp. strain A	321	94.8 $\pm$ 7.96	82.2 $\pm$ 5.98	mean to A is greater than mean to B' ($p < 10^{-15}$), and is greater than the exclusively B' scaffold mean to A ($p < 10^{-15}$)
Exclusively <i>Synechococcus</i> sp. strain B'	364	82.9 $\pm$ 6.21	96.8 $\pm$ 4.48	mean to B' is greater than mean to A ($p < 10^{-15}$), and is greater than the exclusively A scaffold mean to B' ($p < 10^{-15}$)
mixture of <i>Synechococcus</i> spp. A and B'	244	90.4 $\pm$ 9.31	90.0 $\pm$ 8.66	Mean to A is greater than mean to B' ($p < 0.001$), means to A and B' genomes are less than exclusive scaffolds to their respective genomes ($p < 10^{-15}$)

Supplementary Table 9. List and annotation of disjunctly recruited metagenomic sequences that can be confidently assigned to the *Synechococcus* sp. strain A or B' reference genome on one end. Sequences that were split between these two genomes are not reported here. The % NT ID cutoffs used to be considered a putative horizontal gene transfer event between *Synechococcus* spp. strain A or B' and another organism were as follows: $\geq 80\%$ for both *Chloroflexus* sp. 396-1 and *Roseiflexus* sp. RSI, $\geq 70\%$ for Cab. thermophilum. No cutoff was used for the *Thermosynechococcus elongatus* genome, as matches to this genome may represent distantly related cyanobacteria.

Metagenomic Sequence ID Recruited to A	%NT ID to A	Library	Clone-mate Metagenomic Sequence	% NT ID to Other Genome	Other Reference Genome	Top BLASTX match in nr	% AA ID to nr
1041025354856	100	oslow	1041025153962	69.21	thermosynechococcus.dlongatus.bp-1	3-methyl-2-oxobutanoate hydroxymethyltransferase [Anabaena variabilis ATCC 29413].	69.27
1099477830904	100	mslow	1099474235500	0	Null	ABC transporter membrane spanning protein (spermidine/putrescine) [Agrobacterium tumefaciens str. C58].	68.95
1047284316719	96.88	mslow	1047284094146	56.57	chloroflexus.sp.-396-1	ABC transporter nucleotide binding/ATPase protein (spermidine/putrescine) [Agrobacterium tumefaciens str. C58].	63.09
1041023504250	99.47	mslow	1041024576912	78.81	thermosynechococcus.dlongatus.bp-1	AGPSU1 [Ostreococcus tauri].	60.53
1041024440482	99.58	mslow	1041024232340	0	Null	aliphatic sulfonates family ABC transporter periplasmic ligand-binding protein [Cyanotheca sp. PCC 7425].	72
1047280758777	98	oslow	1047280758776	0	Null	allophanate hydrolase [Cyanotheca sp. PCC 7425].	58.93
1041023395436	98.41	oslow	1041024575930	0	Null	amino acid or sugar ABC transport system permease protein putative [Synchococcus sp. PCC 7335].	77.5
1041023157071	99.65	mslow	1041025157972	0	Null	aminoglycoside phosphotransferase [Xanthobacter autotrophicus Px2].	62.96
1047292926291	95.39	mslow	1047292935551	86.36	thermus.thermophilus.bb8	AMP-dependent synthetase and ligase [Thermus aquaticus Y51MC23].	93.27
1041025467236	99.42	mslow	1041024903422	50.9	thermus.thermophilus.bb8	basic proline-rich protein [Sus scrofa].	35.61
1041024851061	99.88	mslow	1041024468747	0	Null	binding-protein-dependent transport systems inner membrane component [Cyanotheca sp. PCC 7425].	74.58
1041083547885	97.74	mslow	1041083547884	0	Null	binding-protein-dependent transport systems inner membrane component [Cyanotheca sp. PCC 7425].	59.73
1041023347728	98.59	mslow	1041025158534	58.57	thermosynechococcus.dlongatus.bp-1	biotin/acetyl-CoA-carboxylase ligase [Cyanotheca sp. PCC 7425].	50.45
1041025125061	100	mslow	1041024232346	0	Null	cell division protein [Rhizobium cfr. CIAT 894].	51.85
1041025286867	99.86	mslow	1041025158356	0	Null	collagen alpha 1(xviii) chain [Aedes aegypti].	31.22
1041024830336	100	oslow	1041024830337	0	Null	conserved hypothetical protein [Nectoc azollae' 0708].	50
1047182015206	99.86	mslow	1047181731328	0	Null	conserved hypothetical protein [Actinomyces urogenitalis DSM 15434].	32.31
1041025274576	100	oslow	1041025343850	0	Null	conserved hypothetical protein [Thermus aquaticus Y51MC23].	63.64
1047293934911	97.89	oslow	1047296121885	79.92	thermus.thermophilus.bb8	conserved hypothetical protein [Thermus aquaticus Y51MC23].	95.65
1041025346494	97.89	mslow	1041024882384	58.07	thermus.thermophilus.bb8	conserved hypothetical protein [Thermus aquaticus Y51MC23].	76.83
1047292896340	99.58	mslow	1047292888069	57.14	thermus.thermophilus.bb8	conserved hypothetical protein [Thermus aquaticus Y51MC23].	59.14
1047296173752	99.49	oslow	1047296996717	51.27	thermus.thermophilus.bb8	DNA polymerase III beta subunit [Desulfotomaculum reducens MI-1].	36.84
1047296308883	100	oslow	1047296230968	61.96	thermomicrobium.roseum	extracellular solute-binding protein [Anabaena variabilis ATCC 29413].	67.25
1041025152056	99.78	mslow	1041025276449	58.03	rosedflexus.sp._rs1	extracellular solute-binding protein family 5 [Crocospaera watsonii WH 8501].	58.79
1041025125024	100	oslow	1041025241892	56.07	thermomicrobium.roseum	extracellular solute-binding protein family 5 [Crocospaera watsonii WH 8501].	56.99
1047280780264	99.89	oslow	1047280780265	0	Null	ferrous iron transport protein A [uncultured bacterium].	96.36
1041024576464	96.67	mslow	1041024850811	67.95	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein A [uncultured bacterium].	95.85
1047284301153	97.21	mslow	1047283931060	70.76	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein B [uncultured bacterium].	95.29
1047280785127	98.71	oslow	1047280785126	68.64	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein B [uncultured bacterium].	96.47
1041025125315	100	mslow	1041024834447	67.95	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein B [uncultured bacterium].	98.29
1041024232410	99.67	mslow	1041024430517	66.29	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein B [uncultured bacterium].	94.92
1041025158452	99.64	mslow	1041025347087	61.49	thermosynechococcus.dlongatus.bp-1	ferrous iron transport protein B [uncultured bacterium].	96.43
1041025274622	99.87	oslow	1041025466106	0	Null	FkbM family methyltransferase [Synchococcus sp. strain B'].	35.71
1041024917594	99.77	mslow	1041025174025	0	Null	FkbM family methyltransferase [Synchococcus sp. strain B'].	36.42
1041023347127	99.7	mslow	1041025296465	0	Null	FkbM family methyltransferase [Synchococcus sp. strain A].	52.94
1099474232849	100	mslow	109947703159	50.31	thermus.thermophilus.bb8	GTP-binding protein Obg/CgtA [Ammonifex degensii KC4].	44.54
1047296030835	95.57	oslow	1047296997323	0	Null	head-tail adaptor putative [Roseovarius nubinhibens ISM].	47.37
1041025276774	96.31	mslow	1041025347055	0	Null	hypothetical protein ABC0569 [Bacillus clausii KSM-K16].	55.96

1041025473064	96.94	oslow	1041025464775	74.74	chloracidobacterium_thermophilum	hypothetical protein Acid345_0630 [Candidatus Koribacter versatilis Ellin345].	36.5
1041025296648	99.78	mshigh	1041025175106	55.42	thermosynechococcus_elongatus_bp-1	hypothetical protein ANACOL_03340 [Anaerotruncus colliominis DSM 17241].	53.85
1041025163876	98.47	oslow	1041023785660	56.65	thermomicrobium_roseum	hypothetical protein Cagg_2700 [Chloroflexus aggregans DSM 9485].	66.21
1041025283379	99.09	oslow	1041025334905	55.03	thermomicrobium_roseum	hypothetical protein Cagg_2700 [Chloroflexus aggregans DSM 9485].	67.1
1041024600400	98.64	mshigh	1041025313830	0	Null	hypothetical protein Cagg_2701 [Chloroflexus aggregans DSM 9485].	73.71
1041025240008	99.64	oslow	1041025304906	91.98	chloroflexus_sp._396-1	hypothetical protein Caur_0093 [Chloroflexus aurantiacus J-10-fl].	64.43
1041025466899	98.74	mshigh	1041024624548	94.38	chloroflexus_sp._396-1	hypothetical protein Caur_0621 [Chloroflexus aurantiacus J-10-fl].	95.89
1041024853284	99.84	mshigh	1041024624326	0	Null	hypothetical protein CFE428DRAFT_0450 [Chthoniobacter flavus Ellin428].	42.18
1047280759058	99.69	oslow	1047280759059	0	Null	hypothetical protein CYB_0691 [Synecococcus sp. strain B].	81.82
1041026333968	100	oslow	1041025285098	0	Null	hypothetical protein Pac1_07176 [Ferropasma acidamarum fer1].	37.66
1047284179626	99.55	mshigh	1047284180736	0	Null	hypothetical protein L8106_04981 [Lyngbya sp. PCC 8106].	31.78
1041025337536	99.88	oslow	1041025337535	0	Null	hypothetical protein L8106_12830 [Lyngbya sp. PCC 8106].	34.65
1047182015284	99.64	mshigh	1047181731484	0	Null	hypothetical protein L8106_12830 [Lyngbya sp. PCC 8106].	39.47
1041025295871	99.88	mshigh	1041024576820	0	Null	hypothetical protein MAE_01000 [Microcystis aeruginosa NIES-843].	46.81
1041025297376	99.86	mshigh	1041025347288	0	Null	hypothetical protein MAE_01000 [Microcystis aeruginosa NIES-843].	46.75
1041025297258	99.75	mshigh	1041025243354	0	Null	hypothetical protein MAE_01000 [Microcystis aeruginosa NIES-843].	46.41
1047296997359	99.58	oslow	1047296030907	86.84	chloracidobacterium_thermophilum	hypothetical protein RoseRS_0299 [Roseiflexus sp. RS-1].	91.49
1047280758989	99.01	oslow	1047280758990	78.44	roseiflexus_sp._rs1	hypothetical protein RoseRS_1882 [Roseiflexus sp. RS-1].	64.63
1041025156821	97.38	mshigh	1041024469145	78.1	roseiflexus_sp._rs1	hypothetical protein RoseRS_1882 [Roseiflexus sp. RS-1].	82.61
1041025145528	99.13	mshigh	1041024371894	89.55	roseiflexus_sp._rs1	hypothetical protein RoseRS_2488 [Roseiflexus sp. RS-1].	90.32
1099474214539	99.44	mslow	1099474235133	0	Null	hypothetical protein S7335_905 [Synecococcus sp. PCC 7335].	26.09
1041035353867	99.77	mshigh	1041025158602	0	Null	hypothetical protein Sden_1914 [Shewanella denitrificans OS217].	26.67
1047284302339	99.11	mshigh	1047284307096	0	Null	hypothetical protein Sden_1914 [Shewanella denitrificans OS217].	33.99
1041025158350	97.44	mshigh	1041025286864	58.92	thermus_thermophilus_hb8	Kelch repeat-containing protein [Thermus aquaticus Y51MC23].	57.55
1041025467523	95.92	mshigh	1041025278049	67.08	roseiflexus_sp._rs1	M_EsaWC21 [uncultured bacterium].	100
1041032391906	99.31	oslow	1041032391907	0	Null	major ampullate spideroin 2-like [Nephila inaurata madagascariensis].	33.59
1099474227051	98.64	mslow	1099474004023	0	Null	methyltransferase FkbM family [Geobacter bemidjiensis Ben].	46.19
1047296192966	98.22	oslow	1047296192965	71.6	chloracidobacterium_thermophilum	novel kinesin motor domain containing protein [Danio rerio].	41.18
1041025167098	96.24	mshigh	1041025242732	92.94	chloroflexus_sp._396-1	nucleotidyl transferase [Chloroflexus aurantiacus J-10-fl].	92.09
1041024847580	100	oslow	1041024370752	0	Null	null	
1041025152047	100	mshigh	1041024856671	0	Null	null	
1041025166779	100	mshigh	1041024856839	0	Null	null	
1041024850885	100	mshigh	1041024624144	0	Null	null	
1047176444077	99.89	mshigh	1047176444076	0	Null	null	
1041025166756	99.89	mshigh	1041024856793	0	Null	null	
1041025165997	99.89	mshigh	1041025125570	0	Null	null	
1041025354965	99.88	oslow	1041025338049	0	Null	null	
1041025242634	99.88	mshigh	1041024856981	0	Null	null	
1041024644087	99.78	mshigh	1041024644086	0	Null	null	
1041024469643	99.76	mshigh	1041025156878	0	Null	null	

1041025276416	99.63	mshigh	1041024857197	0	Null	Null	Null	
1041024596648	99.56	oslow	1041024807766	0	Null	Null	Null	
1041024430620	99.55	mshigh	1041024917414	0	Null	Null	Null	
1041025346486	98.44	mshigh	1041024882368	0	Null	Null	Null	
1041025355877	95.41	oslow	1041025143504	0	Null	Null	Null	
1047284181624	99.58	mshigh	1047283951366	88.89	roseiflexus.sp._rs1	roseiflexus.sp._rs1	Null	
1041025157848	100	mshigh	1041025167339	63.78	chloroflexus.sp._396-1	chloroflexus.sp._396-1	Null	
1047176988464	99.6	mshigh	1047176826037	67.41	thermomicrobium.roseum	thermomicrobium.roseum	oligopeptide ABC transporter ATP-binding protein [Lyngbya sp. PCC 8106].	75.72
1041025239276	99.89	oslow	1041025338901	59.09	thermomicrobium.roseum	thermomicrobium.roseum	oligopeptide binding protein of ABC transporter [Lyngbya sp. PCC 8106].	71.13
1041025242329	99.88	mshigh	1041025145392	67.02	thermomicrobium.roseum	thermomicrobium.roseum	oligopeptide/dipeptide ABC transporter ATPase subunit [Chloroflexus aggregans DSM 9485].	60.81
1047169476010	100	mshigh	1047169468147	68.4	thermomicrobium.roseum	thermomicrobium.roseum	Oligopeptide/dipeptide transporter domain family protein [Synecococcus sp. PCC 7335].	70.87
1047176671098	100	mshigh	1047176345489	0	Null	Null	ORF 73 [Human herpesvirus 8].	73.97
1041024902608	100	oslow	1041024908506	0	Null	Null	ORF73 [Human herpesvirus 8].	26.15
1041024849319	96.13	oslow	1041024881607	84.35	chloracidobacterium.thermophilum	chloracidobacterium.thermophilum	Pantothenate synthetase [Thermotoga neapolitana DSM 4359].	24.44
1041024821657	99.73	oslow	1041025238728	52.57	chloroflexus.sp._396-1	chloroflexus.sp._396-1	Pentapeptide repeat protein [Microcoleus chthonoplastes PCC 7420].	56.29
1047296997104	99.13	oslow	1047296015153	64.93	thermosynechococcus.elongatus.bp-1	thermosynechococcus.elongatus.bp-1	Pentapeptide repeat protein [Microcoleus chthonoplastes PCC 7420].	41.67
1041025150086	97.69	oslow	1041024090080	64.93	thermosynechococcus.elongatus.bp-1	thermosynechococcus.elongatus.bp-1	Pentapeptide repeat protein [Microcoleus chthonoplastes PCC 7420].	58.33
1041024621678	98.59	oslow	1041024643517	61.44	roseiflexus.sp._rs1	roseiflexus.sp._rs1	periplasmic sugar binding protein-like protein [Rubrobacter xylanophilus DSM 9941].	52.4
1041025277262	99.89	mshigh	1041025277261	64.82	thermomicrobium.roseum	thermomicrobium.roseum	permease protein of ABC transporter [Lyngbya sp. PCC 8106].	77.97
1041025338447	99.65	oslow	1041025338448	63.72	thermomicrobium.roseum	thermomicrobium.roseum	permease protein of ABC transporter [Nostoc sp. PCC 7120].	73.93
1099477822261	99.34	mslow	1099474238503	58.57	thermosynechococcus.elongatus.bp-1	thermosynechococcus.elongatus.bp-1	Phycobilisome protein [Synecococcus sp. PCC 7335].	71.62
1041024621490	99.88	oslow	1041024907435	0	Null	Null	polymorphic outer membrane protein [Roseiflexus castenholzii DSM 13941].	43.88
1047292896503	99.3	mshigh	1047292926170	0	Null	Null	PREDICTED: hypothetical protein isoform 1 [Vitis vinifera].	41.28
1047284174511	98.61	mshigh	1047284299257	0	Null	Null	protein of unknown function DU990 [Chloroflexus aggregans DSM 9485].	45.45
1047292896371	99.76	mshigh	1047292926104	0	Null	Null	proteophosphoglycan ppg4 [Leishmania braziliensis MHOM/BR/75/M2904].	35.34
1047292926437	99.51	mshigh	1047292926436	0	Null	Null	putative hydroxyproline-rich protein [Micrococcus sp. 28].	31.3
1041025297758	99.88	mshigh	1041025347863	59.87	chloroflexus.sp._396-1	chloroflexus.sp._396-1	putative transposase [Thermosynechococcus elongatus BP-1].	59.07
1041025286750	98.17	mshigh	1041025307455	59.32	chloroflexus.sp._396-1	chloroflexus.sp._396-1	putative transposase [Thermosynechococcus elongatus BP-1].	58.84
1041025462588	97.97	oslow	1041024900208	52.92	chloroflexus.sp._396-1	chloroflexus.sp._396-1	putative transposase [Thermosynechococcus elongatus BP-1].	57.81
1041025243086	99.32	mshigh	1041025146138	0	Null	Null	subtilisin-like serine protease [Rhodothermus marinus DSM 4252].	26.84
1041024841021	97.73	oslow	1041024841022	0	Null	Null	Tetratricopeptide TPR.2 repeat protein [Geobacter sp. M21].	45.27
1041024835898	98.7	oslow	1041024835899	0	Null	Null	TPR domain/SecC motif-containing domain protein [Geobacter sulfurreducens PCA].	49.81
1041024600412	100	mshigh	1041025313836	0	Null	Null	TPR repeat-containing protein [Cyanotherce sp. PCC 8801].	40.26
1041025157019	99.71	mshigh	1041025126334	81.49	chloroflexus.sp._396-1	chloroflexus.sp._396-1	transcriptional regulator domain-containing protein [Chloroflexus aurantiacus J-10-fl].	30.67
1047284115553	98.37	mshigh	1047284181705	0	Null	Null	translation initiation factor IF-2 [Frankia sp. EAN1pecc].	33.98
1041025125297	99	mshigh	1041024853411	95.42	roseiflexus.sp._rs1	roseiflexus.sp._rs1	transporter DMT superfamily protein [Roseiflexus sp. RS-1].	94.62
1041025158618	97.97	mshigh	1041035333875	60.86	thermosynechococcus.elongatus.bp-1	thermosynechococcus.elongatus.bp-1	transposase [Nostoc sp. PCC 7120].	58.63
1047284308388	98.85	mshigh	1047284178143	60.17	thermosynechococcus.elongatus.bp-1	thermosynechococcus.elongatus.bp-1	transposase [Synecocystis sp. PCC 6803].	57.89
1047284173703	98.39	mshigh	1047284176441	84.04	thermus.thermophilus.bhs8	thermus.thermophilus.bhs8	transposase [SI116/IS110/IS902 family protein [Thermus aquaticus Y51MC23].	85.84

1041024468001	98.46	oslow	1041023957426	59.9	chloracidobacterium_thermophilum	twin-arginine translocation pathway signal [Anabaena variabilis ATCC 29413].	65.65
1047182014828	97.76	mshigh	1047181731148	58.96	chloracidobacterium_thermophilum	twin-arginine translocation pathway signal [Anabaena variabilis ATCC 29413].	60.89
1041024623256	100	oslow	1041025142830	0	Null	uncharacterized conserved protein [Spirosoma linguale DSM 74].	49.81
1041025287449	99.77	mshigh	1041025287448	0	Null	unknown [Myxococcus xanthus].	34.25
1047181891082	100	mshigh	1047181968611	0	Null	urea carboxylase [Cyanotherce sp. PCC 7425].	49.73
1041024855667	100	mshigh	1041024910539	51.21	rhodoferrax_ferrireducens.t118	urea carboxylase [Cyanotherce sp. PCC 7425].	65.59

Metagenomic Sequence ID Recruited to B'	%NT ID to B'	Library	Clone-mate Metagenomic Sequence	%NT ID to Other Genome	Other Genome	Top BLASTX match in nr	%AA ID to nr
1047283951022	96.09	ushigh	1047284301134	63.89	thermus_thermophilus.lb8	2-phosphoglycerate kinase [Methotermus ruber DSM 1279].	87.21
1099474205197	96.28	uslow	1099474238401	62.97	thermonitrobium_roseum	AAA ATPase [Chloroflexus aggregans DSM 9485].	72.3
1041025123383	99.26	oslow	1041024907646	51.66	roselliflexus.sp.jrsl	ABC transporter periplasmic substrate-binding protein [Slicibacter sp. TrichCH4B].	53.53
1041024839919	97.6	oslow	1041024598342	0	Null	ABC-type spermidine/putrescine transport system permease component II [NoCARDopsis dassonvillei subsp. dassonvillei DSM 43111].	45.16
1041024429502	98.71	oslow	1041024843365	61.95	thermosynechococcus_elongatus_bp-1	ABC-type transporter ATPase component [Ralstonia eutropha H16].	36.29
1047296365345	96.19	oslow	1047296031907	62.6	chloroflexus.sp._396-1	acetamidase/formamidase [Nostoc punctiforme PCC 73102].	72.56
1041025304351	93.44	oslow	1041025164388	97.57	chloroflexus.sp._396-1	alpha/beta hydrolase fold-containing protein [Chloroflexus aurantiacus J-10-8].	87.1
1041025343511	97.01	oslow	1041025343510	64.21	acidobacterium_bacterium_ellin345	AMP-dependent synthetase and ligase [Candidatus Korarchaeum versatilis Ellin345].	42.54
1041025304973	97.17	oslow	1041025240142	0	Null	AraM [Thermomicrobium roseum DSM 5159].	27.67
1047281677062	99.67	oslow	1047281677063	0	Null	ATP-binding cassette transporter putative [Richius communis].	40
1047283984220	93.28	ushigh	1047284312537	50.87	chlorocyclobacterium_thermophilum	ATPase component of ABC transporters with duplicated ATPase domain [Methotermus ruber DSM 1279].	81.72
1041024834767	99.43	oslow	1041024834768	59.22	rhodoflexus.ferritducens.4118	Basic membrane protein [Synchococcus sp. PCC 7335].	63.56
1041025465632	97.73	oslow	1041025143892	58.7	rhodoflexus.ferritducens.4118	Basic membrane protein [Synchococcus sp. PCC 7335].	66.9
1099474162414	96.22	uslow	1099474247358	0	Null	BimA [Burkholderia pseudomallei].	45
1041025124160	99.87	oslow	1041024908720	0	Null	binding-protein-dependent transport system inner membrane component [Cyanodorea sp. PCC 7425].	77.17
1041025344927	98.86	oslow	1041025344926	60.49	chlorocyclobacterium_thermophilum	Carboxymethylhydrobutenolide [Cyanodorea sp. PCC 7425].	66.97
1041025122576	97.95	oslow	1041025335457	59.14	chlorocyclobacterium_thermophilum	Carboxymethylhydrobutenolide [Methylobacterium populi B1001].	66.13
1041024572138	96.66	oslow	1041024620015	0	Null	Carboxymethylhydrobutenolide [Methylobacterium populi B1001].	68.35
1041024552364	98.59	oslow	1041025238318	60.51	chlorocyclobacterium_thermophilum	carboxymethylhydrobutenolide [Synchococcus elongatus PCC 6301].	65.47
1041024908608	95.97	oslow	1041025124104	61.11	chlorocyclobacterium_thermophilum	carboxymethylhydrobutenolide [Synchococcus elongatus PCC 6301].	72.73
1041024645534	99.03	oslow	1041024598422	0	Null	CG14021 [Drosophila melanogaster].	30.61
1041024231796	97.7	oslow	1041024916123	49.68	herpetosiphon_aurantiacus_atcc_23779	chlorohydrolase [Butyrivibrio crossotus DSM 2876].	54.33
1041025535915	98.54	oslow	1041025143580	66.34	chlorocyclobacterium_thermophilum	conserved hypothetical protein [Arthrosira maxima CS-328].	69.26
1047283966426	99.42	ushigh	1047284308969	0	Null	conserved hypothetical protein [Arthrosira maxima CS-328].	53.45
1041024819781	99.79	oslow	1041025283807	0	Null	conserved hypothetical protein [Chthoniobacter flavus Ellin428].	45.1
1099474177603	97.86	uslow	1099474202754	0	Null	conserved hypothetical protein [Chthoniobacter flavus Ellin428].	50
1041025240545	97.67	oslow	1041025343054	0	Null	conserved hypothetical protein [Chthoniobacter flavus Ellin428].	45.21
1041025143828	98.93	oslow	1041025341568	80.89	chloroflexus.sp._396-1	conserved hypothetical protein [Gramicicella adiacens ATCC 49175].	41.38
1041024807577	96.61	oslow	1041024807576	0	Null	conserved hypothetical protein [Halothiobacillus neopoltanus c2].	30.46
1041024805395	98.69	oslow	1041025149139	60.45	thermus_thermophilus.lb8	conserved hypothetical protein [Thermus aquaticus Y51MC29].	66
1047296099931	97.4	oslow	1047296016127	0	Null	Conserved protein/domain typically associated with flavoprotein oxygenases DIM6/NTAB family [Vibrio angustum S14].	45.31
1041024847505	100	oslow	1041024902484	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	53.09
1041024817583	99.25	oslow	1041024817582	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	43.96
1041024231498	98.8	oslow	1041024916359	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	44.94
1041024841578	98.37	oslow	1041024370748	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	48.94
1041023804932	97.7	oslow	1041025123160	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	48.15
1041025535854	97.36	oslow	1041025306132	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	43.61
1041025354551	97.22	oslow	1041025141328	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	65.46
1041024834130	96.51	oslow	1041024834129	0	Null	CRISPR-associated helicase Cas3 domain protein [Microcoleus chthonoplastes PCC 7420].	42.86

1041025305608	98.8	oslow	1041025341865	54.73	chloracidobacterium_thermophilum	CRISPR-associated protein Cas1	Cyanothecae sp. PCC 7424].	78.66
1041024620938	93.2	oslow	1041024573208	54.38	chloracidobacterium_thermophilum	CRISPR-associated protein Cas1	Cyanothecae sp. PCC 7424].	77.14
10410252242371	96.59	mshigh	1041025145476	0	Null	CRISPR-associated protein Cas1	[Fibrobacter succinogenes subsp. succinogenes S85].	36.17
10410253343033	98.37	oslow	1041025165032	0	Null	CRISPR-associated protein Cas1	putative [Microcoleus chthonoplastes PCC 7420].	83.93
1041024835596	97.84	oslow	1041024467897	0	Null	CRISPR-associated protein Cas1	putative [Microcoleus chthonoplastes PCC 7420].	75.19
1041024843374	99.56	oslow	1041024429610	56.86	chloracidobacterium_thermophilum	CRISPR-associated protein DevS	[Microcoleus chthonoplastes PCC 7420].	60.59
1041025463252	99.79	oslow	1041025292331	0	Null	CRISPR-associated protein DevS	[Microcoleus chthonoplastes PCC 7420].	61.08
1041025336678	98.18	oslow	1041024823635	0	Null	CRISPR-associated protein DevS	[Microcoleus chthonoplastes PCC 7420].	51.85
1041023304182	97.82	oslow	1041025338172	0	Null	CRISPR-associated protein DevS	[Microcoleus chthonoplastes PCC 7420].	60.59
1041025338924	99.45	oslow	1041025273122	0	Null	CRISPR-associated protein Csm2	family [Arthrospira maxima CS-328].	39.31
1041024090278	98.9	oslow	1041024838322	0	Null	CRISPR-associated protein Csm2	family [Arthrospira maxima CS-328].	39.72
1041024427796	96.45	oslow	1041025141684	0	Null	CRISPR-associated RAMP Csm2	family protein [Synecococcus sp. strain B'].	37.44
1099474157150	93.67	mslow	1099474243520	0	Null	CRISPR-associated regulatory protein DevR	family [Microcoleus chthonoplastes PCC 7420].	65.97
1041025355100	99.12	oslow	1041025338703	97.36	chloroflexus.sp._396-1	cyclopropane fatty acyl phospholipid synthase	[Synecococcus sp. strain B'].	94.25
1041025473141	99.55	oslow	1041025464929	0	Null	dipeptidase	[Thermoanaerobacter italicus Ab9].	43
1041024572496	95.25	oslow	1041024816255	0	Null	dTDP-6-deoxy-L-hexose 3-O-methyltransferase	[Planctomyces maris DSM 8797].	59.32
1041025354608	99.18	oslow	1041025141442	56.06	thermomicrobium_roseum	extracellular solute-binding protein family 5	[Crocospira watsonii WH 8501].	57.2
1041025123683	96.15	oslow	1041024902058	0	Null	ferrous iron transport protein A	[uncultured bacterium].	84.38
1101131329510	95.78	oslow	1101131329511	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.49
1101131329519	95.74	oslow	1101131329520	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.47
1101131329649	95.65	oslow	1101131329648	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.46
1101131329489	95.63	oslow	1101131329490	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.46
1101131329589	94.97	oslow	1101131329588	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.49
1101131329441	94.54	oslow	1101131329442	0	Null	ferrous iron transport protein A	[uncultured bacterium].	99.49
1041025356251	99.2	oslow	1041025356252	67.64	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein A	[uncultured bacterium].	94.76
1041024837974	96.98	oslow	1041024622458	71.35	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein A	[uncultured bacterium].	95.48
1041025465197	93.17	oslow	1041025239388	66.05	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein A	[uncultured bacterium].	94.41
1041024802091	99.65	oslow	1041025122025	58.18	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	92.02
1041025141966	99.01	oslow	1041024574288	59.1	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	96.34
1047297000173	98.46	oslow	1047296309186	66.04	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	91.09
1041024623298	98.16	oslow	1041025142851	69.12	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	95.86
1047176345611	97.68	mshigh	1047176345610	69.68	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	97.41
1041024230686	97.64	oslow	1041025293220	72.24	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	97.34
1041024231124	97.46	oslow	1041024552462	70.51	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	98.48
1041025165939	97.14	mshigh	1041025125454	58.61	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	97.05
1041083861584	96.4	mslow	1041083861583	68.51	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	95.99
10410244468151	95.4	oslow	1041024839803	59.23	thermosynechococcus_elongatus_bp-1	ferrous iron transport protein B	[uncultured bacterium].	94.44
1041025295276	98.09	oslow	1041025173522	93.68	roseiflexus.sp._rs1	GHP kinase	[Roseiflexus sp. RS-1].	97.02
1041025335949	97.67	oslow	1041025238556	0	Null	glycosyl transferase group 1	[Nostoc azollae' 0708].	35.58
1041025163575	98.41	oslow	1041024367680	0	Null	GntR family transcriptional regulator	[Roseiflexus castenholzii DSM 13941].	42.57

1041024916289	96.73	oslow	1041024880478	52.5	thermomicrobium_roseum	HAD family hydrolase [Rhodospirillum rubrum ATCC 11170].	51.23
1047297000126	98.14	oslow	1047296309092	65.35	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	76.53
10410253338910	97.51	oslow	1041025273094	59.81	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	70.49
1041024849160	96.96	oslow	1041024623966	68.25	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	86.38
1099474247822	96.18	mslow	1099474224204	61.55	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	81.67
10410253555748	95.57	oslow	1041025172972	63.31	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	82.07
1099474168786	93.08	mslow	1099474191051	59.95	chloracidobacterium_thermophilum	helicase domain protein [Cyanotheca sp. PCC 7425].	76.84
1041025313262	98.69	oslow	1041025170973	0	Null	helic-turn-helix domain-containing protein [Geobacter uranireducens Rf4].	53.23
1041025337692	95.81	oslow	1041025337691	0	Null	Henolysin activation/secretion protein [Magnetospirillum gryphiswaldense MSR-1].	35.16
1041025340199	99.15	oslow	1041025304684	0	Null	hydrolase carbon-nitrogen family [Synecococcus sp. PCC 7335].	78.57
1041025355783	98.57	oslow	1041025304990	62.6	chloroflexus.sp..396-1	hypothetical protein all0706 [Nostoc sp. PCC 7120].	73.3
1041024907320	94.7	oslow	1041024428280	0	Null	hypothetical protein all8519 [Nostoc sp. PCC 7120].	36.52
1041025355786	98.01	oslow	1041025304996	53.29	thermosynechococcus_elongatus_bp-1	hypothetical protein AM1.4519 [Acaryochloris marina MBIC11017].	53.99
1099474138324	99.3	mslow	1099474238236	0	Null	hypothetical protein AmaxDRAFT_3735 [Arthrospira maxima CS-328].	51.88
1041024835022	97.08	oslow	1041024835023	0	Null	hypothetical protein AmaxDRAFT_3735 [Arthrospira maxima CS-328].	52.86
1041025277851	98.83	mshigh	1041025347249	0	Null	hypothetical protein An08g03930 [Aspergillus niger].	33.93
1041024794705	98.26	oslow	1041024900360	0	Null	hypothetical protein An08g03930 [Aspergillus niger].	31.74
1099474199257	95.68	mslow	1099471728576	55.19	thermosynechococcus_elongatus_bp-1	hypothetical protein ANACOL_03340 [Anaerotruncus colihominis DSM 17241].	53.13
1041025336191	97.36	oslow	1041025464112	80.59	chloroflexus.sp..396-1	hypothetical protein Apar.0219 [Atopobium parvulum DSM 20469].	41.71
1041024810784	93.86	oslow	1041024810785	81.2	chloroflexus.sp..396-1	hypothetical protein Apar.0219 [Atopobium parvulum DSM 20469].	43.93
1041024810924	96.33	oslow	1041024810923	67.97	chloracidobacterium_thermophilum	hypothetical protein Ava.2190 [Anabaena variabilis ATCC 29413].	64.06
1041024843222	99.38	oslow	1041024231914	0	Null	hypothetical protein BanaMEX5DRAFT_6929 [Burkholderia ambifaria MEX-5].	56.41
1041024370884	98.89	oslow	1041024847646	0	Null	hypothetical protein BanaMEX5DRAFT_6929 [Burkholderia ambifaria MEX-5].	52.38
1041024917312	97.63	oslow	1041024917311	0	Null	hypothetical protein Cagg.2700 [Chloroflexus aggregans DSM 9485].	33.9
1041024812058	96.14	oslow	1041024572342	54.04	thermomicrobium_roseum	hypothetical protein Caur.0093 [Chloroflexus aurantiacus J-10-fl].	63.27
1041024815315	98.86	oslow	1041025292808	80	chloroflexus.sp..396-1	hypothetical protein Caur.2700 [Chloroflexus aurantiacus J-10-fl].	61.11
1041024845329	95.47	oslow	1041024430136	65.29	chloracidobacterium_thermophilum	hypothetical protein cce.0356 [Cyanotheca sp. ATCC 51142].	55.67
1041024901661	99.66	oslow	1041024880110	0	Null	hypothetical protein CHE428DRAFT_0450 [Clithionobacter flavus Elin428].	47.18
1041025303922	96.9	oslow	1041024429308	0	Null	hypothetical protein CY0110.30950 [Cyanotheca sp. CCY0110].	46.41
1041024574100	99.49	oslow	1041024574101	0	Null	hypothetical protein CY0110.30950 [Cyanotheca sp. CCY0110].	62.02
1041025141320	99.1	oslow	1041025354547	0	Null	hypothetical protein CYA.0321 [Synecococcus sp. strain A].	60.91
1041025285982	99.3	oslow	1041025174100	81.59	roseiflexus.sp..rs1	hypothetical protein Cyan7425.2444 [Cyanotheca sp. PCC 7425].	82.61
1047296388134	98.31	oslow	1047297001072	0	Null	hypothetical protein CYB.1700 [Synecococcus sp. strain B].	39.26
1041025336237	96.69	oslow	1041025464135	83.85	roseiflexus.sp..rs1	hypothetical protein DDB.G0280701 [Dictyostelium discoideum AX4].	67.29
1041025313200	97.86	oslow	1041025170849	0	Null	hypothetical protein DDB.G0280701 [Dictyostelium discoideum AX4].	33.33
1041024849907	98.29	oslow	1041024908815	0	Null	hypothetical protein DDB.G0295727 [Dictyostelium discoideum AX4].	31.13
1041025150233	94.67	oslow	1041025142444	85.75	chloroflexus.sp..396-1	hypothetical protein GCWU000182.00560 [Albortrophia defectiva ATCC 49176].	41.6
1041024790653	97.34	oslow	1041024790652	0	Null	hypothetical protein gfr4333 [Gloeobacter violaceus PCC 7421].	32.14
1047281102649	99.51	oslow	1047281102650	0	Null	hypothetical protein L8106.12830 [Lyngbya sp. PCC 8106].	32.23
1041024847931	99.39	oslow	1041024847930	0	Null	hypothetical protein L8106.30020 [Lyngbya sp. PCC 8106].	49.81

1041024850145	97.55	oslow	1041025124242	0	Null	hypothetical protein L8106.30020 [Lyngbya sp. PCC 8106].	42.53
1041025345709	98.36	oslow	1041025345710	0	Null	hypothetical protein L8106.30025 [Lyngbya sp. PCC 8106].	53.1
1041024090636	97.72	oslow	1041024623716	0	Null	hypothetical protein L8106.30025 [Lyngbya sp. PCC 8106].	60.62
1041025342102	97.24	oslow	1041025342103	0	Null	hypothetical protein L8106.30025 [Lyngbya sp. PCC 8106].	56.3
1041024807383	95.65	oslow	1041024807384	0	Null	hypothetical protein L8106.30025 [Lyngbya sp. PCC 8106].	61.41
1041025466380	95.63	oslow	1041025466379	0	Null	hypothetical protein L8106.30025 [Lyngbya sp. PCC 8106].	62.5
1041023784008	98.05	oslow	1041024831863	0	Null	hypothetical protein L8106.30030 [Lyngbya sp. PCC 8106].	52.15
1041024575094	97.69	oslow	1041024575093	0	Null	hypothetical protein L8106.30035 [Lyngbya sp. PCC 8106].	57.07
1041024596888	93.7	oslow	1041024572656	0	Null	hypothetical protein L8106.30055 [Lyngbya sp. PCC 8106].	52.09
1041025335842	96.57	oslow	1041025163164	0	Null	hypothetical protein LA3189 [Leptospira interrogans serovar Lai str. 56601].	54.4
1041025285352	99.88	oslow	1041025294743	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	55.96
1041025463646	97.65	oslow	1041024820653	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	58.73
1041024846274	95.37	oslow	1041024846273	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	57.65
1099474199421	95.19	mslow	1099474235271	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	56.25
104102504805	94.98	oslow	1041025143340	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	55.82
1041024901157	93.8	oslow	1041024819459	0	Null	hypothetical protein MC7420.3829 [Microcoleus chthonoplastes PCC 7420].	53.8
1041025466344	95.61	oslow	1041025466343	0	Null	hypothetical protein MGG.12193 [Magnetococcus marinus DSM 11131].	31.85
1041024839904	99.77	oslow	1041024598312	0	Null	hypothetical protein MSMEG.5916 [Mycobacterium smegmatis str. MC2 155].	45
1047295935273	99.43	oslow	1047296999776	0	Null	hypothetical protein Npun.R5419 [Nostoc punctiforme PCC 73102].	49.21
1041025356183	97.62	oslow	1041025356182	0	Null	hypothetical protein PCC7424.3103 [Cyanobacterium sp. PCC 7424].	46.9
1047295934063	99.26	oslow	1047296348047	0	Null	hypothetical protein PM8797T.07829 [Planctomyces maris DSM 8797].	46.48
1041024832938	96.6	oslow	1041024428518	0	Null	hypothetical protein PROVRETT.01298 [Providencia rettgeri DSM 1131].	36.36
1041024907960	97.73	oslow	1041025123780	0	Null	hypothetical protein RmarDRAFT.16570 [Rhodothermus marinus DSM 4252].	41.27
1041026333973	97.85	oslow	1041025285108	71.71	roseiflexus.sp..rs1	hypothetical protein RoseRS.0296 [Roseiflexus sp. RS-1].	73.83
1041024840301	98.3	oslow	1041024231256	90.42	roseiflexus.sp..rs1	hypothetical protein RoseRS.1409 [Roseiflexus sp. RS-1].	84.46
1041024848596	99.52	oslow	1041024848597	81.94	roseiflexus.sp..rs1	hypothetical protein RoseRS.1882 [Roseiflexus sp. RS-1].	72.62
1041024819749	99.1	oslow	1041024915744	77.23	roseiflexus.sp..rs1	hypothetical protein RoseRS.1882 [Roseiflexus sp. RS-1].	80.43
1041023956804	96.75	oslow	1041024832354	83.44	roseiflexus.sp..rs1	hypothetical protein RoseRS.1882 [Roseiflexus sp. RS-1].	91.67
1041025155124	93.26	oslow	1041025478216	84.56	roseiflexus.sp..rs1	hypothetical protein RoseRS.1882 [Roseiflexus sp. RS-1].	91.67
1041024427718	99.65	oslow	1041025238495	0	Null	hypothetical protein Rru.A1723 [Rhodospirillum rubrum ATCC 11170].	63.12
1041025340424	94.9	oslow	1041025340425	0	Null	hypothetical protein Rru.A1723 [Rhodospirillum rubrum ATCC 11170].	59.09
1041025306319	99.87	oslow	1041025275286	0	Null	hypothetical protein slr1815 [Synecococcus sp. PCC 6803].	57.56
1041025124364	97.85	oslow	1041025339423	53.36	chloroflexus.sp..396-1	hypothetical protein SUN_0884 [Sulfurovum sp. NBC37-1].	41.94
1041025463653	94.85	oslow	1041024820667	81.44	chloroflexus.sp..396-1	hypothetical protein SUN_0885 [Sulfurovum sp. NBC37-1].	40.65
1041024468549	95.4	oslow	1041024623005	0	Null	hypothetical protein syc1447.d [Synecococcus elongatus PCC 6301].	52.91
1041024816999	96.21	oslow	1041024467275	0	Null	hypothetical protein Tery.1283 [Trichodesmium erythraeum IMS101].	53.85
1041025305673	93.74	oslow	1041025155272	61.79	thermus.thermophilus.hb8	hypothetical protein Thu.1317 [Thermobifida fusca YX].	34.46
1041024642656	99.56	oslow	1041024807977	0	Null	hypothetical protein TTC1429 [Thermus thermophilus HB27].	80
1041024880494	98.82	oslow	1041024916297	62.89	thermomicrobium.roseum	hypothetical protein TTC1430 [Thermus thermophilus HB27].	69.43
1041025305440	97.64	oslow	1041024917056	0	Null	hypothetical protein VEIDISOL_00231 [Veillonella dispar ATCC 17748].	29.66

1041024855094	99.63	mslhigh	1041023958540	0	Null	chloracidobacterium.thermophilum	integral membrane protein MviN [Desulfotomaculum reducens MI-1].	39.14
1041025339263	94.89	oslow	1041025273416	83.13	Null		ISSoc9 transposase [Synecococcus sp. strain B ¹].	84.92
1041025143702	98.07	oslow	1041025341505	0	Null		methyltransferase FkbM family [Geobacter bemidjensis Bem].	46.11
1041024621858	97.65	oslow	1041024835320	0	Null		nucleoside ABC transporter membrane protein [Meiothermus ruber DSM 1279].	47.06
1041025142859	97.3	oslow	1041024623314	0	Null		nucleoside ABC transporter membrane protein [Meiothermus ruber DSM 1279].	52.52
1041025336770	100	oslow	1041024824119	0	Null		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	48.26
1041025336559	98.94	oslow	1041024823197	0	Null		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	45.56
1041032377192	98.03	oslow	1041032377191	0	Null		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	48.53
1041025156102	97.92	oslow	1041025285836	0	Null		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	48.26
1041024907369	96.57	oslow	1041024621358	0	Null		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	48.04
1041025283855	97.99	oslow	1041024819977	60.77	thermus.thermophilus.hb8		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	70.09
1041025336029	96.91	oslow	1041024901168	64.32	thermus.thermophilus.hb8		nucleoside ABC transporter membrane protein [Meiothermus silvanus DSM 9946].	69.44
1041024090238	100	oslow	1041025150115	0	Null		null	
1099474235217	99.68	mslow	1099474214707	0	Null		null	
1041025238161	99.57	oslow	1041024814427	0	Null		null	
1041024814975	99.48	oslow	1041025292738	0	Null		null	
1041024468443	99.47	oslow	1041024840141	0	Null		null	
1041024907726	99.41	oslow	1041025123423	0	Null		null	
1041024089628	99.37	oslow	1041024834199	0	Null		null	
1041023784326	99.36	oslow	1041024428703	0	Null		null	
1041024575932	99.29	oslow	1041023395440	0	Null		null	
1041024909052	99.1	oslow	1041025339373	0	Null		null	
1041025163784	99.04	oslow	1041024836517	0	Null		null	
1041024847574	98.78	oslow	1041024370740	0	Null		null	
1041025164146	98.71	oslow	1041025465261	0	Null		null	
1041024468215	98.68	oslow	1041024839835	0	Null		null	
1041024839866	98.63	oslow	1041024598236	0	Null		null	
1041023078943	98.63	oslow	1041024367948	0	Null		null	
1041025336880	98.54	oslow	1041025239074	0	Null		null	
1041032354190	98.48	oslow	1041032354191	0	Null		null	
1041024367562	98.42	oslow	1041024827087	0	Null		null	
1041026445946	98.29	oslow	1041025305204	0	Null		null	
1041024622786	98.06	oslow	1041024840158	0	Null		null	
1099474245907	97.14	mslow	1099474237927	0	Null		null	
1041024089670	97.08	oslow	1041024834320	0	Null		null	
1041024846513	96.99	oslow	1041024846514	0	Null		null	
1099474247753	96.97	mslow	1099474224066	0	Null		null	
1041024817137	96.85	oslow	1041024817136	0	Null		null	
1099474238225	96.83	mslow	1099474138302	0	Null		null	
1041023784138	96.77	oslow	1041024428659	0	Null		null	

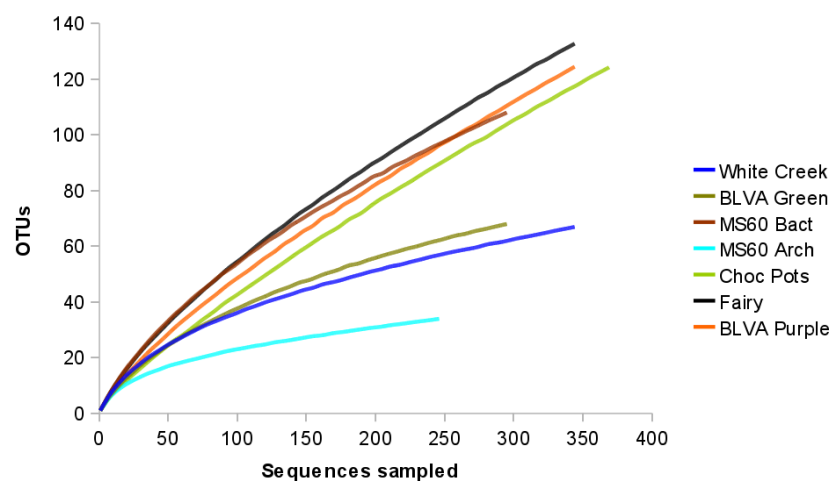
1041025339325	96.73	oslow	1041024908956	0	Null		null	
1041024830303	96.62	oslow	1041024830304	0	Null		null	
1041023784394	96.49	oslow	1041025141867	0	Null		null	
1099474247763	96.25	mslow	1099474224086	0	Null		null	
1041025285936	96.21	oslow	1041025174008	0	Null		null	
1099474241153	96.04	mslow	1099474220605	0	Null		null	
1041024596624	95.72	oslow	1041024807754	0	Null		null	
1099474177543	95.71	mslow	1099474202724	0	Null		null	
1047297000462	95.71	oslow	1047296348548	0	Null		null	
1041025238994	95.13	oslow	1041025336840	0	Null		null	
1041025164250	94.79	oslow	1041025465313	0	Null		null	
1041024807320	94.71	oslow	1041024807321	0	Null		null	
1041025241940	94.02	oslow	1041025151154	0	Null		null	
1099474293704	97.41	mslow	1099474174322	61.41	thermomicrobium.roseum		oligopeptide binding protein of ABC transporter [Nostoc sp. PCC 7120].	66.29
10410253336940	99.28	oslow	1041025143212	0	Null		ORF73 [Human herpesvirus 8].	26.3
1041024844137	99.27	oslow	1041024844136	0	Null		ORF73 [Human herpesvirus 8].	28.11
1041024642922	99.03	oslow	1041025163135	0	Null		ORF73 [Human herpesvirus 8].	28.09
1041025155890	97.48	oslow	1041025241315	0	Null		ORF73 [Human herpesvirus 8].	28.29
1041024903044	99.09	mslow	1041025346056	0	Null		outer membrane autotransporter barrel domain [Burkholderia ubonensis Bu].	27.45
1041025173486	99.16	oslow	1041025295258	0	Null		oxidoreductase FAD-dependent [Methylobacterium ruber DSM 1279].	95.45
104728111410	99.45	oslow	104728111411	94.88	chloracidobacterium.thermophilum		PAS domain S-box protein [Methylobacterium ruber DSM 1279].	39.81
1041025122996	97.98	oslow	1041024366826	60	thermosynechococcus.elongatus_bp-1		Peptidase M23B [Lyngbya sp. PCC 8106].	53.36
1099474159779	98.9	mslow	1099474246333	64.36	thermomicrobium.roseum		permease protein of ABC transporter [Lyngbya sp. PCC 8106].	77.78
1099471703455	99.19	mslow	1099474247042	0	Null		phage integrase [Synecococcus sp. PCC 7002].	38.81
10410253336877	97.16	oslow	1041025239068	64.4	thermosynechococcus.elongatus_bp-1		Phycobilisome protein [Synecococcus sp. PCC 7335].	71.88
1041024828574	98.74	oslow	1041024828573	0	Null		predicted protein [Coprinopsis cinerea okayama7#130].	34.71
1041024907440	97.97	oslow	1041024621500	0	Null		predicted protein [Coprinopsis cinerea okayama7#130].	35.65
1099474247543	98.78	mslow	1099474212034	86.73	chloracidobacterium.thermophilum		PREDICTED: hypothetical protein kinase [Halogemetricum borinquense DSM 11551].	37.78
1041025341518	97.42	oslow	1041025143728	0	Null		PREDICTED: similar to guanylate binding protein 1 [Vitis vinifera].	41.35
1047296016155	99.43	oslow	1047296999945	0	Null		PREDICTED: similar to guanylate binding protein 1 [Gallus gallus].	33.05
1041024231526	98.72	oslow	1041024916373	50.55	roseiflexus.sp.-rs1		probable transport system permease transmembrane abc transporter protein [Vibrio shilonii AK1].	40.71
1041025144128	96.51	oslow	1041025294770	49.73	roseiflexus.sp.-rs1		probable transport system permease transmembrane abc transporter protein [Vibrio shilonii AK1].	40.4
1041025337707	96.4	oslow	1041025337706	57.2	chloracidobacterium.thermophilum		protein of unknown function DUF1156 [Arthrospira maxima CS-328].	54.18
1041024643484	97.15	oslow	1041024621612	0	Null		protein of unknown function DUF1156 [Arthrospira maxima CS-328].	59.55
1041025339083	95.83	oslow	1041025293655	65.22	roseiflexus.sp.-rs1		protein of unknown function DUF1156 [Arthrospira maxima CS-328].	56.72
1041024834900	99.76	oslow	1041024834899	69.58	chloracidobacterium.thermophilum		protein of unknown function DUF1156 [Cyanotheca sp. PCC 7425].	79.86
1041025122788	99.45	oslow	1041025354642	62.62	chloracidobacterium.thermophilum		protein of unknown function DUF1156 [Cyanotheca sp. PCC 7425].	66.8
1041025172592	97.71	oslow	1041025339167	68.88	chloracidobacterium.thermophilum		protein of unknown function DUF1156 [Cyanotheca sp. PCC 7425].	72.99
1041025466349	95.83	oslow	1041025466350	64.07	chloracidobacterium.thermophilum		protein of unknown function DUF1156 [Cyanotheca sp. PCC 7425].	66.22

1041024815391	94.63	oslow	1041025283726	63.39	chloracidobacterium_thermophilum	protein of unknown function DUF1156 [Cyanotherce sp. PCC 7425].	72.22
1041025465398	97.59	oslow	1041025340657	60.29	roseiflexus_sp._rs1	protein of unknown function DUF1156 [Cyanotherce sp. PCC 7425].	60
1041025466367	96.23	oslow	1041025466368	0	Null	Protein of unknown function DUF1963 [Paenibacillus sp. JDR-2].	44.69
1099474247265	100	mslow	1099474132511	0	Null	protein of unknown function DUF820 [Cyanotherce sp. PCC 7425].	76.89
1041025143362	99.69	oslow	1041025144361	0	Null	protein of unknown function DUF820 [Cyanotherce sp. PCC 7425].	76.89
1041024823661	99.79	oslow	1041025336691	53.4	thermus_thermophilus_hb8	putative ABC transporter permease component [Rhizobium leguminosarum bv. viciae 3841].	41.98
1041025465471	94.77	oslow	1041025355562	0	Null	putative CHRSPR-associated protein [Synecococcus sp. PCC 7002].	67.29
1041024850203	98.75	oslow	1041025124271	0	Null	putative periplasmic solute-binding protein [Xanthobacter autotrophicus Py2].	52.7
1041025141801	98.84	oslow	1041024826395	0	Null	putative transposase [Cyanotherce sp. ATCC 51142].	63.52
1041025462929	97.74	oslow	1041025237558	0	Null	putative transposase [Cyanotherce sp. ATCC 51142].	62.72
1041024827808	97.44	oslow	1041024827807	0	Null	putative transposase [Cyanotherce sp. ATCC 51142].	61.46
1041025165663	97.44	oslow	1041025144780	0	Null	putative transposase [Cyanotherce sp. ATCC 51142].	64.21
1041024800018	93.39	oslow	1041024800017	0	Null	putative transposase [Cyanotherce sp. ATCC 51142].	50.94
1047296340491	96.15	oslow	1047296007291	69.18	thermosynechococcus_elongatus_bp-1	putative transposase [Thermosynechococcus elongatus BP-1].	72.69
1041025142373	94.3	oslow	1041025123874	72.84	thermosynechococcus_elongatus_bp-1	putative transposase [Thermosynechococcus elongatus BP-1].	77.64
1041025463033	94.18	oslow	1041025303282	72.86	thermosynechococcus_elongatus_bp-1	putative transposase [Thermosynechococcus elongatus BP-1].	78.88
1041024623254	98.65	oslow	1041025142829	0	Null	putative transposase [S891/JS136/JS1341 family [Cyanotherce sp. PCC 8802].	46.36
1047297000192	96.81	oslow	1047296309224	0	Null	response regulator receiver protein [Cyanotherce sp. PCC 7425].	33.52
1041024824791	96.72	oslow	104102428028	0	Null	ribosomal protein S12 methylthiotransferase rimO [Synecococcus sp. strain B']	
1099477832215	96.69	mslow	1099474238411	0	Null	Serine/Threonine protein kinase [Sagittula stellata E-37].	32.48
10410253334800	99.73	oslow	1041025162616	50.73	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.33
1041025344524	99.4	oslow	1041025344525	58.24	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1101131329381	99.29	oslow	1101131329382	54.52	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.5
1101131329517	99.29	oslow	1101131329516	54.52	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.53
1101131329391	99.27	oslow	1101131329390	54.49	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.52
1101131329399	99.26	oslow	1101131329400	54.43	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.53
1101131329624	99.26	oslow	1101131329625	54.52	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.53
1041024468335	98.87	oslow	1041024840087	58.82	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1101131329553	98.27	oslow	1101131329552	57.93	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1101131329501	98.06	oslow	1101131329502	57.69	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1041025340654	97.51	oslow	1041025150964	52.4	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1041024806379	97.35	oslow	1041024900857	59.08	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	100
1041024847137	97.27	oslow	1041024509122	55.23	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.48
1041025122748	94.15	oslow	1041025355543	58.96	thermosynechococcus_elongatus_bp-1	SRA-YDG domain protein [uncultured bacterium].	99.32
1041024798129	99.74	oslow	1041026740267	0	Null	Sugar transport system permease protein [Bacillus thuringiensis serovar monterrey BGSC 4A.J1].	38
1047284179366	96.88	mshigh	1047284178271	0	Null	Tetrapeptide TPR-2 repeat protein [Geobacter bemiidjensis Ben].	50
1041024838631	98.99	oslow	1041024880327	56.59	herpetosiphon_aurantiacus_atcc.23779	TM1410 hypothetical-related protein [Chloroflexus aggregans DSM 9485].	58.29
1041024852820	95.37	mshigh	1041024852819	0	Null	TPR domain/SecC motif-containing domain protein [Geobacter sulfurreducens PCA].	46.67
1041024572790	93.21	oslow	1041024506955	0	Null	TPR domain/SecC motif-containing domain protein [Geobacter sulfurreducens PCA].	48.85
1041025354791	96.84	oslow	1041024824507	49.6	methanothermobacter_thermautotrophicus_str._delta.h	TPR repeat-containing protein [Cyanotherce sp. PCC 8801].	36.07

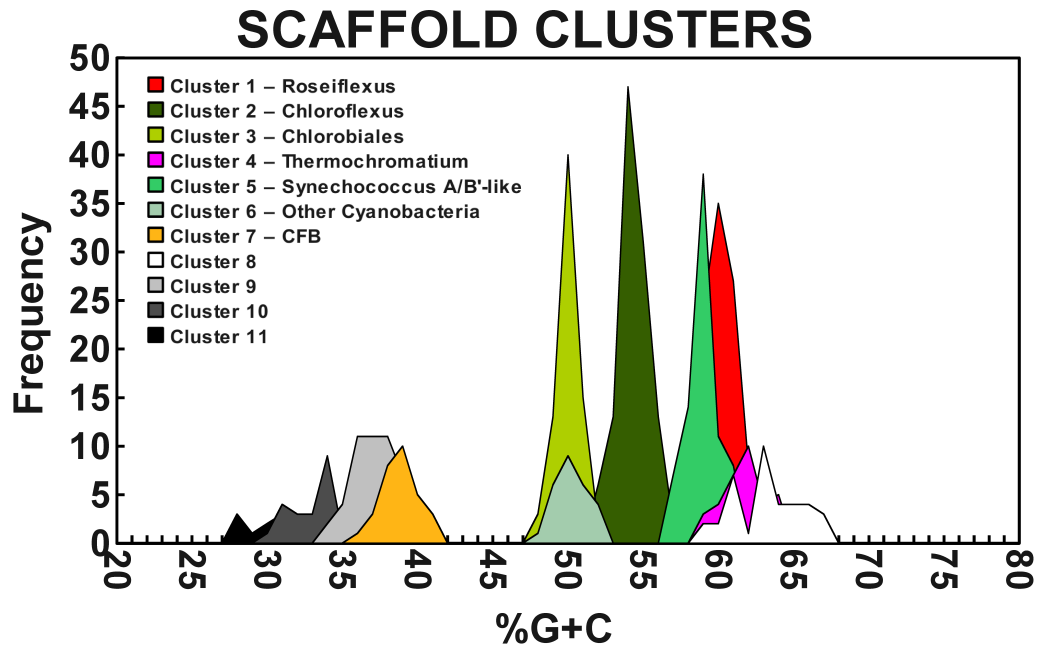
1041025286464	95.62	mshigh	1041025296163	0	Null	TPR repeat-containing protein [Pelobacter propionicus DSM 2379].	60.58
1041025464230	97.14	oslow	1041025336427	0	Null	transcriptional regulator [Stappia aggregata IAM 12614].	55.71
1041024881604	98.26	oslow	1041024849313	0	Null	Transposase (probable) IS891/IS1136/IS1341: Transposase IS605 OrfB [Crocospilaera watsonii WH 8501].	51.59
1041025140940	97.92	oslow	1041025333389	0	Null	transposase [Lyngbya sp. PCC 8106].	42.35
1041025306321	98.57	oslow	1041025275290	98.15	roseiflexus.sp.rs1	transposase IS111A/IS1328/IS1533 [Roseiflexus sp. RS-1].	95.45
1041024428162	99.25	oslow	1041024907261	58.96	chloracidobacterium.thermophilum	twin-arginine translocation pathway signal [Anabaena variabilis ATCC 29413].	64.63
1041024880633	98.15	oslow	1041023958020	58.58	chloracidobacterium.thermophilum	twin-arginine translocation pathway signal [Anabaena variabilis ATCC 29413].	64.92
1041025150433	98.62	oslow	1041023079333	56.69	chloroflexus.sp._396-1	uncharacterized conserved protein [Methanothermobacter DSM 1279].	57.21
1101131329366	99.41	oslow	1101131329367	68.28	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	99.58
1101131329466	99.26	oslow	1101131329465	68.28	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	99.58
1101131329409	99.26	oslow	1101131329408	68.63	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	100
1101131329445	99.26	oslow	1101131329444	68.28	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	99.58
1101131329415	99.26	oslow	1101131329414	68.55	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	100
1101131329453	99.12	oslow	1101131329454	69.92	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	100
1101131329594	98.89	oslow	1101131329595	69.69	thermosynechococcus.elongatus.bp-1	unknown function protein [uncultured bacterium].	100
1041025275490	98.77	oslow	1041025156386	68.53	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	66.21
1041025345740	98.74	oslow	1041025345739	63.15	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	70.18
1041024596660	98.56	oslow	1041024807772	57.56	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	65.65
1099474171192	97.93	mslow	1099474159671	64.93	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	72.6
1041023784390	97.92	oslow	1041025141865	63.99	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	67.26
1041024621572	97.87	oslow	1041024643464	65.47	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	70.52
1041024802839	95.78	oslow	1041025271842	63.41	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	65.54
1041025313710	94.86	oslow	1041025274726	54.19	chloracidobacterium.thermophilum	unnamed protein product [Microcystis aeruginosa PCC 7806].	65.27
1041025354817	99.05	oslow	1041024824559	0	Null	unnamed protein product [Microcystis aeruginosa PCC 7806].	50.19
1099474214527	95.06	mslow	1099474235127	0	Null	unnamed protein product [Microcystis aeruginosa PCC 7806].	60.56
1041025150068	93.04	oslow	1041024090044	0	Null	urea carboxylase-associated protein 2 [Cyanotherce sp. PCC 7425].	58.33
1099477832240	98.12	mslow	1099474238461	0	Null	urea carboxylase-associated protein 2 [Cyanotherce sp. PCC 7425].	54.36
1041025345861	98.83	oslow	1041025165792	0	Null	von Willebrand factor type A [Chlorobacter flavus Elin428].	74.63
1041024840414	96.55	oslow	1041024369776	0	Null	von Willebrand factor type A [Chlorobacter flavus Elin428].	68.06
1041025149898	98.83	oslow	1041024089504	0	Null	WD-40 repeat-containing protein [Spirosoma linguale DSM 74].	38.58

APPENDIX C

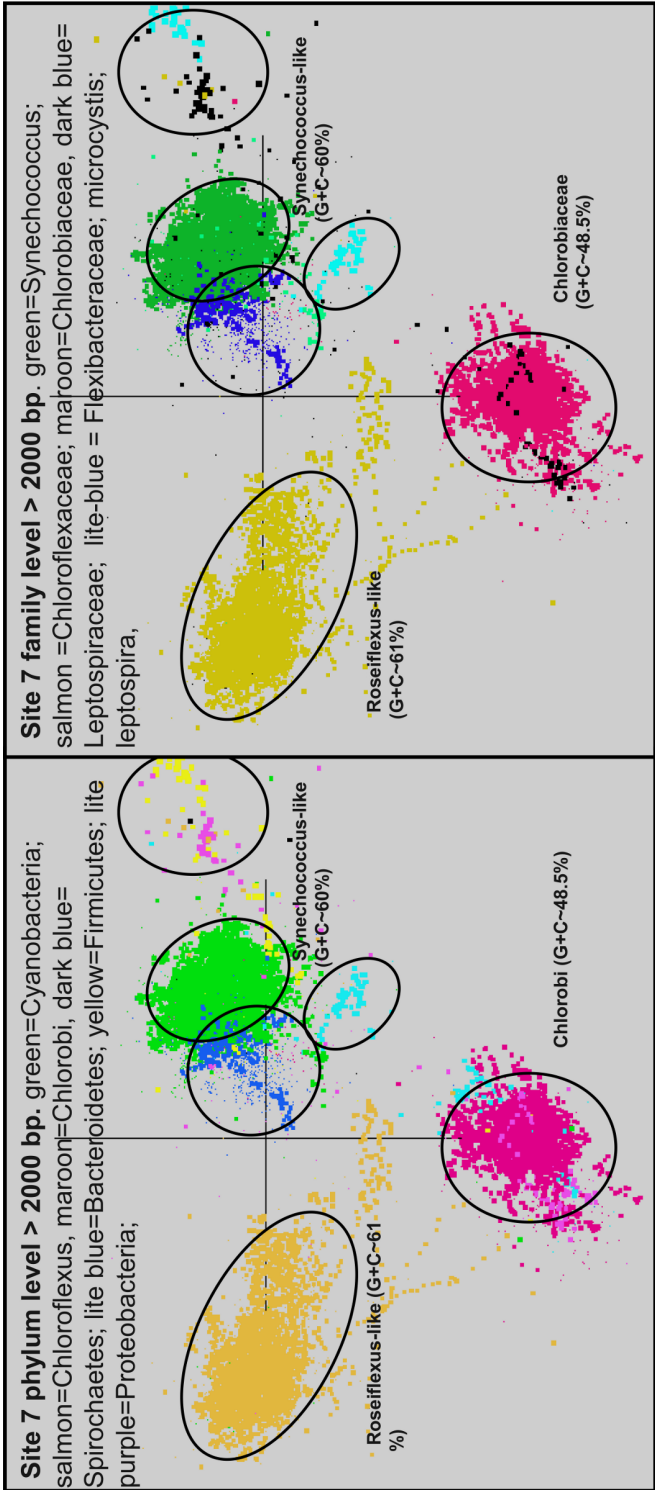
CHAPTER 4 APPENDIX



Supplementary Figure 1 - Rarefaction Curves. OTUs demarcated at the 99% similarity level using the CAP3 assembler and EcoSim.



Supplementary Figure 2 - G+C Composition of Scaffold Clusters. Scaffold clusters greater than 10kbp were demarcated using oligonucleotide frequencies as depicted in Figure 4.4.



Supplementary Figure 3 - Nucleotide word frequency PCA of assembled sequence from Chocolate Pots (CP_7). This community contains predominant phylotypes of *Roseiflexus*-, *Synecococcus*-, Chlorobi- and Spirochaete-like populations as well as minor contributions from the Firmicutes, Proteobacteria and Bacteroidetes.

Supplementary Table 1 - Community structure from 16S sequence libraries. Richness indexes ACE, Chao1 (w/ 95% confidence intervals) and diversity indexes (Fisher's alpha, Shannon-Weaver, and Simpson's Index).

Site	N	99% OTUs	singletons	ACE	Chao1	SD	95U	95D	α	SD	Shannon	Simpson
BLVA_5	305	69	32	120.93	96.55	12.86	134.77	80.54	27.79	2.54	3.24	10.35
BLVA_20	364	130	102	773.87	873.14	318.26	1791.07	462.47	72.34	6.04	3.75	14.1
WC_6	367	69	34	131.33	141.25	36.59	253.32	97.32	25.1	2.14	3.42	19.23
CP_7	380	127	100	402	421.11	94.03	669.1	286.57	66.85	5.44	3.38	9.2
MS_15 Bacteria	314	112	66	214.14	220.9	37.69	322.54	168.32	62.25	5.6	4.1	37.54
MS_15 Archaea	265	35	16	65	77.66	33.23	199.54	46.06	10.8	1.14	2.67	8.94
FG_16	360	137	96	390.33	408.05	87.48	639.37	283.24	80.69	6.82	4.14	32.75

APPENDIX D

CHAPTER 5 APPENDIX

Supplementary Table 1 - Unique ORFs on *Roseiflexus* spp. contigs. These ORFs are located on scaffolds demarcated as *Roseiflexus* spp., but they do not meet the reciprocal blast criterion.

Metagenome ORF ID	Annotation
JCVI.PEP_metagenomic.orf.21522381.1	ABC transporter related — Polyamine-transporting ATPase
JCVI.PEP_metagenomic.orf.21541654.1	amylase- α -16-glucosidase
JCVI.PEP_metagenomic.orf.21218313.1	anti-sigma-factor antagonist
JCVI.PEP_metagenomic.orf.21216237.1	anti-sigma-factor antagonist
JCVI.PEP_metagenomic.orf.21408651.1	CobB/CobQ domain protein glutamine amidotransferase
JCVI.PEP_metagenomic.orf.21106620.1	GCN5-related N-acetyltransferase
JCVI.PEP_metagenomic.orf.21064711.1	histidine kinase
JCVI.PEP_metagenomic.orf.21087361.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21199499.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21078602.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21521011.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21166087.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21480227.1	hypothetical protein
JCVI.PEP_metagenomic.orf.20825976.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21516481.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21153467.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21553710.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21249108.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21040171.1	hypothetical protein
JCVI.PEP_metagenomic.orf.20952696.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21401527.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21166420.1	hypothetical protein
JCVI.PEP_metagenomic.orf.20948596.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21382867.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21241896.1	hypothetical protein
JCVI.PEP_metagenomic.orf.20844639.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21106419.1	hypothetical protein
JCVI.PEP_metagenomic.orf.21135337.1	Iron dependent repressor metal binding and dimerisation domain
JCVI.PEP_metagenomic.orf.21381404.1	LrgA family
JCVI.PEP_metagenomic.orf.21380401.1	LrgB family protein
JCVI.PEP_metagenomic.orf.21577176.1	methyl-accepting chemotaxis sensory transducer
JCVI.PEP_metagenomic.orf.21365353.1	multi-sensor signal transduction histidine kinase
JCVI.PEP_metagenomic.orf.21365353.1	multi-sensor signal transduction histidine kinase
JCVI.PEP_metagenomic.orf.20781948.1	NADH dehydrogenase (quinone) — NADH dehydrogenase (quinone)
JCVI.PEP_metagenomic.orf.21168512.1	protein of unknown function
JCVI.PEP_metagenomic.orf.20994715.1	protein of unknown function
JCVI.PEP_metagenomic.orf.21000294.1	protein of unknown function
JCVI.PEP_metagenomic.orf.21099323.1	putative PAS/PAC sensor protein
JCVI.PEP_metagenomic.orf.20835424.1	putative regulatory protein FmdB family
JCVI.PEP_metagenomic.orf.21296451.1	pyridoxine biosynthesis protein
JCVI.PEP_metagenomic.orf.21577196.1	response regulator receiver protein
JCVI.PEP_metagenomic.orf.20769449.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.20850568.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.20769449.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.20928300.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.21372964.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.20849418.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.20966762.1	transposase IS4 family protein
JCVI.PEP_metagenomic.orf.21297353.1	tRNA-guanine transglycosylases various specificities — Queuine tRNA-ribosyltransferase
JCVI.PEP_metagenomic.orf.20844358.1	type II site-specific deoxyribonuclease — Type II site-specific deoxyribonuclease

Supplementary Table 2 - Unique ORFs on *Chloroflexus* spp. contigs. These ORFs are located on scaffolds demarcated as *Chloroflexus* spp., but they do not meet the reciprocal blast criterion.

Metagenome ORF ID	Annotation
JCVLPEP_metagenomic.orf.21353764.1	ABC-2 type transporter
JCVLPEP_metagenomic.orf.20938692.1	adenine-specific DNA methylase
JCVLPEP_metagenomic.orf.21404274.1	arginyl-tRNA synthetase — Arginine-tRNA ligase
JCVLPEP_metagenomic.orf.21269800.1	ATPase associated with various cellular activities AAA.5
JCVLPEP_metagenomic.orf.21168737.1	ATPase P-type (transporting) HAD superfamily subfamily IC — Calcium-transporting ATPase
JCVLPEP_metagenomic.orf.21140369.1	CRISPR-associated protein Cas1
JCVLPEP_metagenomic.orf.21140789.1	CRISPR-associated protein Cas2
JCVLPEP_metagenomic.orf.21177093.1	divalent cation transporter
JCVLPEP_metagenomic.orf.21515751.1	DNA methylase N-4/N-6 — Site-specific DNA-methyltransferase (adenine-specific)
JCVLPEP_metagenomic.orf.21131376.1	efflux transporter RND family MFP subunit
JCVLPEP_metagenomic.orf.20863335.1	formyl-CoA transferase — Formyl-CoA transferase
JCVLPEP_metagenomic.orf.21224154.1	glycoprotease family
JCVLPEP_metagenomic.orf.21287388.1	Glycosyl hydrolase family 1 — Beta-glucosidase
JCVLPEP_metagenomic.orf.21038486.1	glycosyl transferase family 2
JCVLPEP_metagenomic.orf.20957035.1	glycosyl transferase group 1 — Phosphatidylinositol N-acetylglucosaminyltransferase
JCVLPEP_metagenomic.orf.21153909.1	glyoxalase family protein
JCVLPEP_metagenomic.orf.21160955.1	helicase domain protein
JCVLPEP_metagenomic.orf.21380595.1	histidinol dehydrogenase — Histidinol dehydrogenase
JCVLPEP_metagenomic.orf.21226267.1	histidinol-phosphate aminotransferase — Histidinol-phosphate transaminase
JCVLPEP_metagenomic.orf.21437001.1	hypothetical protein
JCVLPEP_metagenomic.orf.21448658.1	hypothetical protein
JCVLPEP_metagenomic.orf.20969128.1	hypothetical protein
JCVLPEP_metagenomic.orf.21277235.1	hypothetical protein
JCVLPEP_metagenomic.orf.20996118.1	hypothetical protein
JCVLPEP_metagenomic.orf.21136795.1	hypothetical protein
JCVLPEP_metagenomic.orf.21019729.1	hypothetical protein
JCVLPEP_metagenomic.orf.20776959.1	hypothetical protein
JCVLPEP_metagenomic.orf.21421850.1	hypothetical protein
JCVLPEP_metagenomic.orf.20916729.1	hypothetical protein
JCVLPEP_metagenomic.orf.21516653.1	hypothetical protein
JCVLPEP_metagenomic.orf.21180699.1	hypothetical protein
JCVLPEP_metagenomic.orf.21213975.1	hypothetical protein
JCVLPEP_metagenomic.orf.21342601.1	hypothetical protein
JCVLPEP_metagenomic.orf.21275572.1	hypothetical protein
JCVLPEP_metagenomic.orf.21398036.1	hypothetical protein
JCVLPEP_metagenomic.orf.20988831.1	imidazole glycerol phosphate synthase glutamine amidotransferase subunit
JCVLPEP_metagenomic.orf.21359318.1	imidazoleglycerol phosphate synthase cyclase subunit — 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase
JCVLPEP_metagenomic.orf.21331018.1	isoleucyl-tRNA synthetase — Isoleucine-tRNA ligase
JCVLPEP_metagenomic.orf.21166308.1	Lon protease (S16) C-terminal proteolytic domain — Endopeptidase La
JCVLPEP_metagenomic.orf.21338276.1	methyltransferase type 11
JCVLPEP_metagenomic.orf.21178690.1	methyltransferase type 11 — 3-demethylubiquinone-9 3-O-methyltransferase
JCVLPEP_metagenomic.orf.20795289.1	nitroreductase
JCVLPEP_metagenomic.orf.21309111.1	nucleic acid binding OB-fold tRNA/helicase-type
JCVLPEP_metagenomic.orf.21214076.1	pantoate-beta-alanine ligase — Pantoate-beta-alanine ligase
JCVLPEP_metagenomic.orf.21269374.1	PAS domain S-box
JCVLPEP_metagenomic.orf.20880071.1	Peptidase family M20/M25/M40 — Aminoacylase
JCVLPEP_metagenomic.orf.20973066.1	peptidase M48 Ste24p
JCVLPEP_metagenomic.orf.21485390.1	peptidase S9A/B/C families catalytic domain — Acylaminoacyl-peptidase
JCVLPEP_metagenomic.orf.20871950.1	phosphocarrier HPr family — Phosphoenolpyruvate-protein phosphatase
JCVLPEP_metagenomic.orf.20871950.1	phosphocarrier HPr family — Phosphoenolpyruvate-protein phosphatase
JCVLPEP_metagenomic.orf.21272253.1	phosphonate metabolism protein PlnM — Adenine deaminase
JCVLPEP_metagenomic.orf.21272026.1	phosphoribosyl-ATP diphosphatase — Phosphoribosyl-AMP cyclohydrolase
JCVLPEP_metagenomic.orf.21358900.1	phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase — 1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino]imidazole-4-carboxamide isomerase
JCVLPEP_metagenomic.orf.21058423.1	prolipoprotein diacylglycerol transferase
JCVLPEP_metagenomic.orf.21329158.1	protein of unknown function
JCVLPEP_metagenomic.orf.20842723.1	protein RecA — Calcium-transporting ATPase
JCVLPEP_metagenomic.orf.21215921.1	proton-translocating NADH-quinone oxidoreductase chain N — NADH dehydrogenase (quinone)
JCVLPEP_metagenomic.orf.21422025.1	putative oxygen-independent coproporphyrinogen III oxidase — coproporphyrinogen dehydrogenase
JCVLPEP_metagenomic.orf.21518941.1	pyruvate carboxyltransferase — Hydroxymethylglutaryl-CoA lyase
JCVLPEP_metagenomic.orf.21017275.1	Resolvase N terminal domain
JCVLPEP_metagenomic.orf.21218537.1	response regulator receiver and sarp domain protein
JCVLPEP_metagenomic.orf.21357620.1	response regulator receiver sensor signal transduction histidine kinase — histidine kinase
JCVLPEP_metagenomic.orf.20891893.1	RNA methyltransferase TrmH family — tRNA (guanosine-2'-O-)-methyltransferase
JCVLPEP_metagenomic.orf.20941963.1	SMC domain protein
JCVLPEP_metagenomic.orf.20880764.1	tetratricopeptide TPR.2 repeat protein
JCVLPEP_metagenomic.orf.20821182.1	transglutaminase domain protein
JCVLPEP_metagenomic.orf.21025636.1	transposase
JCVLPEP_metagenomic.orf.21025636.1	transposase
JCVLPEP_metagenomic.orf.21492799.1	transposase IS605 OrfB family — DNA (cytosine-5-)-methyltransferase
JCVLPEP_metagenomic.orf.21492799.1	transposase IS605 OrfB family — DNA (cytosine-5-)-methyltransferase
JCVLPEP_metagenomic.orf.20813512.1	transposase IS605 OrfB family — DNA (cytosine-5-)-methyltransferase
JCVLPEP_metagenomic.orf.20944638.1	tRNA-guanine transglycosylases various specificities — Queuine tRNA-ribosyltransferase
JCVLPEP_metagenomic.orf.20992047.1	UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase — UDP-N-acetylmuramoyl-tripeptide-D-alanyl-D-alanine ligase

REFERENCES CITED

- Alber, B., M. Olinger, A. Rieder, D. Kockelkorn, B. Jobst, M. Hügler, G. Fuchs (2006). Malonyl-coenzyme A reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. *Journal of Bacteriology* 188:8551–8559.
- Alber, B. E., G. Fuchs (2002). Propionyl-coenzyme A synthase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO₂ fixation. *The Journal of Biological Chemistry* 277:12137–12143.
- Allewalt, J. P., M. M. Bateson, N. P. Revsbech, K. Slack, D. M. Ward (2006). Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. *Applied and Environmental Microbiology* 72:544–550.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, D. J. Lipman (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.
- Anderson, K. L., T. A. Tayne, D. M. Ward (1987). Formation and fate of fermentation products in hot spring cyanobacterial mats. *Applied and Environmental Microbiology* 53:2343–2352.
- Awramik, S. M. (1992). The oldest records of photosynthesis. *Photosynthesis Research* 33:75–89.
- Bassham, J. A., M. Kirk (1962). The effect of oxygen on the reduction of CO₂ to glycolic acid and other products during photosynthesis by *Chlorella*. *Biochemical and Biophysical Research Communications* 9:376–380.
- Bateson, M. M., D. M. Ward (1988). Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. *Applied and Environmental Microbiology* 54:1738–1743.
- Bauld, J. (1973). *Algal-Bacterial Interactions in Alkaline Hot Spring Effluents*. PhD. dissertation, University of Wisconsin, Madison.
- Bauld, J., T. D. Brock (1973). Ecological studies of *Chloroflexis*, a gliding photosynthetic bacterium. *Archives of Microbiology* 92:267–284.
- Bauld, J., T. D. Brock (1974). Algal excretion and bacterial assimilation in hot spring algal mats. *Journal of Phycology* 10:101–106.
- Becraft, E. D., F. M. Cohan, M. Köhl, S. I. Jensen, D. M. Ward (2011). Fine-scale distribution patterns of *Synechococcus* ecological diversity in microbial mats of Mushroom Spring, Yellowstone National Park. *Applied and Environmental Microbiology* 77:7689–7697.
- Bendtsen, J. D., H. Nielsen, G. von Heijne, S. Brunak (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* 340:783–795.

- Berg, I. A., O. I. Keppen, E. N. Krasilnikova, N. V. Ugolkova, R. N. Ivanovsky (2005). Carbon metabolism of filamentous anoxygenic phototrophic bacteria of the family *Oscillochloridaceae*. *Mikrobiologiya (English translation)* 74:258–264.
- Best, E. A., V. C. Knauf (1993). Organization and nucleotide sequences of the genes encoding the biotin carboxyl carrier protein and biotin carboxylase protein of *Pseudomonas aeruginosa* acetyl coenzyme A carboxylase. *Journal of Bacteriology* 175:6881–6889.
- Bhaya, D., A. R. Grossman, A. Steunou, N. Khuri, F. M. Cohan, N. Hamamura, M. C. Melendrez, M. M. Bateson, D. M. Ward, J. F. Heidelberg (2007). Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *The ISME Journal* 1:703–13.
- Blankenship, R. E. (1992). Origin and early evolution of photosynthesis. *Photosynthesis Research* 33:91–111.
- Boomer, S. M., D. P. Lodge, B. E. Dutton, B. Pierson (2002). Molecular characterization of novel red green nonsulfur bacteria from five distinct hot spring communities in Yellowstone National Park. *Applied and Environmental Microbiology* 68:346–355.
- Boomer, S. M., B. K. Pierson, R. Austinhirst, R. W. Castenholz (2000). Characterization of novel bacteriochlorophyll-*a*-containing red filaments from alkaline hot springs in Yellowstone National Park. *Archives of Microbiology* 174:152–161.
- Brock, T. D. (1973). Lower pH limit for the existence of blue-green algae: evolutionary and ecological implications. *Science* 179:480–483.
- Brock, T. D. (1978). *Thermophilic microorganisms and life at high temperatures*. Springer Verlag, New York.
- Brock, T. D., M. L. Brock (1968). Relationship between environmental temperature and optimum temperature of bacteria along a hot spring thermal gradient. *Journal of Applied Bacteriology* 31:54–58.
- Brock, T. D., H. Freeze (1969). *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *Journal of Bacteriology* 98:289–297.
- Bruce, B. D., R. C. Fuller, R. E. Blankenship (1982). Primary photochemistry in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. *Proceedings of the National Academy of Sciences of the United States of America* 79:6532–6536.
- Bryant, D. A., A. M. G. Costas, J. A. Maresca, A. G. M. Chew, C. G. Klatt, M. M. Bateson, L. J. Tallon, J. Hostetler, W. C. Nelson, J. F. Heidelberg, D. M. Ward (2007). *Candidatus chloracidobacterium thermophilum*: an aerobic phototrophic acidobacterium. *Science* 317:523–6.

- Bryant, D. A., N. Frigaard (2006). Prokaryotic photosynthesis and phototrophy illuminated. *Trends in Microbiology* 14:488–496.
- Bryant, D. A., C. G. Klatt, N. Frigaard, Z. Liu, T. Li, F. Zhao, A. M. Garcia Costas, J. Overmann, D. M. Ward (2012). Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. In: R. L. Burnap, W. Vermaas (eds.), *Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration*, vol. 33. Springer, Dordrecht, The Netherlands., pp. 47–102.
- Buckel, W., B. T. Golding (2006). Radical enzymes in anaerobes. *Annual Review of Microbiology* 60:27–49.
- Buckley, D. H., V. Huangyutitham, S. Hsu, T. A. Nelson (2007). Stable isotope probing with $^{15}\text{N}_2$ reveals novel noncultivated diazotrophs in soil. *Applied and Environmental Microbiology* 73:3196–3204.
- Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T. L. Madden (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421.
- Castenholz, R. W. (1969a). Thermophilic blue-green algae and the thermal environment. *Bacteriological Reviews* 33:476–504.
- Castenholz, R. W. (1969b). The thermophilic cyanophytes of Iceland and the upper temperature limit. *Journal of Phycology* 5:360–368.
- Castenholz, R. W. (1976). The effect of sulfide on the blue-green algae of hot springs. I. New Zealand and Iceland. *Journal of Phycology* 12:54–68.
- Castenholz, R. W. (1977). The effect of sulfide on the blue-green algae of hot springs II. Yellowstone National Park. *Microbial Ecology* 3:79–105.
- Castenholz, R. W. (1978). The biogeography of hot spring algae through enrichment cultures. *Mitt Internat Verein Limnol* 21:296–315.
- Castenholz, R. W. (1988). Culturing of cyanobacteria. In: L. Packer, A. N. Glazer (eds.), *Methods in Enzymology*, vol. 167. Academic Press, San Diego CA, pp. 68–93.
- Castenholz, R. W., B. K. Pierson (1995). Ecology of thermophilic anoxygenic phototrophs. In: R. E. Blankenship, M. T. Madigan, C. E. Bauer (eds.), *Anoxygenic Photosynthetic Bacteria*, vol. 2. Kluwer Academic Publishers, Dordrecht, pp. 87–103.
- Cheng, G., N. Shapir, M. J. Sadowsky, L. P. Wackett (2005). Allophanate hydrolase, not urease, functions in bacterial cyanuric acid metabolism. *Applied and Environmental Microbiology* 71:4437–4445.

- Chew, A. G. M., D. A. Bryant (2007). Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. *Annual Review of Microbiology* 61:113–129.
- Chuakrut, S., H. Arai, M. Ishii, Y. Igarashi (2003). Characterization of a bifunctional archaeal acyl coenzyme A carboxylase. *Journal of Bacteriology* 185:938–947.
- Clesceri, L. S., A. E. Greenburg, A. D. Eaton (eds.) (1998). *Standard Methods for the Examination of Water and Wastewater: Including Bottom Sediments and Sludges*. 20th edn. American Public Health Association.
- Cox, A., E. L. Shock, J. R. Havig (2011). The transition to microbial photosynthesis in hot spring ecosystems. *Chemical Geology* 280:344–351.
- Cronan, J., John E, G. L. Waldrop (2002). Multi-subunit acetyl-CoA carboxylases. *Progress in Lipid Research* 41:407–435.
- Daniel, J., T. Oh, C. Lee, P. E. Kolattukudy (2007). AccD6, a member of the Fas II locus, is a functional carboxyltransferase subunit of the acyl-coenzyme A carboxylase in *Mycobacterium tuberculosis*. *Journal of Bacteriology* 189:911–917.
- Davis, K. E. R., S. J. Joseph, P. H. Janssen (2005). Effects of growth medium, inoculum size, and incubation time on culturability and isolation of soil bacteria. *Applied and Environmental Microbiology* 71:826–834.
- Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox, D. E. Graham, R. Overbeek, M. A. Snead, M. Keller, M. Aujay, R. Huber, R. A. Feldman, J. M. Short, G. J. Olsen, R. V. Swanson (1998). The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* 392:353–358.
- Delcher, A. L., D. Harmon, S. Kasif, O. White, S. L. Salzberg (1999). Improved microbial gene identification with GLIMMER. *Nucleic Acids Research* 27:4636–4641.
- Dempsey, M. P., J. Nietfeldt, J. Ravel, S. Hinrichs, R. Crawford, A. K. Benson (2006). Paired-end sequence mapping detects extensive genomic rearrangement and translocation during divergence of *Francisella tularensis* subsp. *tularensis* and *Francisella tularensis* subsp. *holarctica* populations. *Journal of Bacteriology* 188:5904–5914.
- Denef, V. J., L. H. Kalnejais, R. S. Mueller, P. Wilmes, B. J. Baker, B. C. Thomas, N. C. VerBerkmoes, R. L. Hettich, J. F. Banfield (2010). Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proceedings of the National Academy of Sciences of the United States of America* 107:2383–2390.

- Des Marais, D. J. (1991). Microbial mats, stromatolites and the rise of oxygen in the precambrian atmosphere. *Palaeogeography, Palaeoclimatology, Palaeoecology* 97:93–96.
- Diacovich, L., D. L. Mitchell, H. Pham, G. Gago, M. M. Melgar, C. Khosla, H. Gramajo, S. Tsai (2004). Crystal structure of the beta-subunit of acyl-CoA carboxylase: structure-based engineering of substrate specificity. *Biochemistry* 43:14027–14036.
- Diacovich, L., S. Peir, D. Kurth, E. Rodríguez, F. Podest, C. Khosla, H. Gramajo (2002). Kinetic and structural analysis of a new group of Acyl-CoA carboxylases found in *Streptomyces coelicolor* A3(2). *The Journal of Biological Chemistry* 277:31228–31236.
- Dick, G. J., A. F. Andersson, B. J. Baker, S. L. Simmons, B. C. Thomas, A. P. Yelton, J. F. Banfield (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biology* 10:R85.
- Dillon, J. G., S. Fishbain, S. R. Miller, B. M. Bebout, K. S. Habicht, S. M. Webb, D. A. Stahl (2007). High rates of sulfate reduction in a low-sulfate hot spring microbial mat are driven by a low level of diversity of sulfate-respiring microorganisms. *Applied and Environmental Microbiology* 73:5218–5226.
- Doemel, W. N., T. D. Brock (1974). Bacterial stromatolites: origin of laminations. *Science* 184:1083–1085.
- van Driessche, G., W. Hu, G. Van de Werken, F. Selvaraj, J. D. McManus, R. E. Blankenship, J. J. Van Beeumen (1999). Auracyanin A from the thermophilic green gliding photosynthetic bacterium *Chloroflexus aurantiacus* represents an unusual class of small blue copper proteins. *Protein Science* 8:947–57.
- Eddy, S. R. (1998). Profile hidden markov models. *Bioinformatics (Oxford, England)* 14:755–763.
- Eder, W., R. Huber (2002). New isolates and physiological properties of the Aquificales and description of *Thermocrinis albus* sp. nov. *Extremophiles: Life Under Extreme Conditions* 6:309–318.
- Eisen, M. B., P. T. Spellman, P. O. Brown, D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* 95:14863–14868.
- Ferris, M. J., M. Kühl, A. Wieland, D. M. Ward (2003). Cyanobacterial ecotypes in different optical microenvironments of a 68°C hot spring mat community revealed by 16S-23S rRNA internal transcribed spacer region variation. *Applied and Environmental Microbiology* 69:2893–2898.

- Ferris, M. J., G. Muyzer, D. M. Ward (1996a). Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. *Applied and Environmental Microbiology* 62:340–346.
- Ferris, M. J., A. L. Ruff-Roberts, E. D. Kocczynski, M. M. Bateson, D. M. Ward (1996b). Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. *Applied and Environmental Microbiology* 62:1045–1050.
- Ferris, M. J., K. B. Sheehan, M. Köhl, K. Cooksey, B. Wigglesworth-Cooksey, R. Harvey, J. M. Henson (2005). Algal species and light microenvironment in a low-pH, geothermal microbial mat community. *Applied and Environmental Microbiology* 71:7164–7171.
- Ferris, M. J., D. M. Ward (1997). Seasonal distributions of dominant 16S rRNA-defined populations in a hot spring microbial mat examined by denaturing gradient gel electrophoresis. *Applied and Environmental Microbiology* 63:1375–1381.
- Finn, R. D., J. Tate, J. Mistry, P. C. Coghill, S. J. Sammut, H. Hotz, G. Ceric, K. Forslund, S. R. Eddy, E. L. L. Sonnhammer, A. Bateman (2008). The Pfam protein families database. *Nucleic Acids Research* 36:D281–288.
- Finneran, K. T., C. V. Johnsen, D. R. Lovley (2003). *Rhodoferrax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). *International Journal of Systematic and Evolutionary Microbiology* 53:669–673.
- Fischer, F., W. Zillig, K. O. Stetter, G. Schreiber (1983). Chemolithoautotrophic metabolism of anaerobic extremely thermophilic archaeobacteria. *Nature* 301:511–513.
- Fouke, B. W., G. T. Bonheyo, B. Sanzenbacher, J. Frias-Lopez (2003). Partitioning of bacterial communities between travertine depositional facies at Mammoth Hot Springs, Yellowstone National Park, U.S.A. *Canadian Journal of Earth Sciences* 40:1531–1548.
- Frangeul, L., P. Quillardet, A. Castets, J. Humbert, H. Matthijs, D. Cortez, A. Tolonen, C. Zhang, S. Gribaldo, J. Kehr, Y. Zilliges, N. Ziemert, S. Becker, E. Talla, A. Latifi, A. Billault, A. Lepelletier, E. Dittmann, C. Bouchier, N. Tandeau de Marsac (2008). Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics* 9:274.
- Friedmann, S., B. E. Alber, G. Fuchs (2006a). Properties of succinyl-coenzyme A:D-citramalate coenzyme A transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. *Journal of Bacteriology* 188:6460–6468.

- Friedmann, S., A. Steindorf, B. E. Alber, G. Fuchs (2006b). Properties of succinyl-coenzyme A:L-malate coenzyme a transferase and its role in the autotrophic 3-hydroxypropionate cycle of *Chloroflexus aurantiacus*. *Journal of Bacteriology* 188:2646–2655.
- Frigaard, N., D. A. Bryant (2006). Chlorosomes: Antenna organelles in photosynthetic green bacteria. In: J. M. Shively (ed.), *Complex Intracellular Structures in Prokaryotes*, vol. 2. Springer-Verlag, Berlin/Heidelberg, pp. 79–114.
- Frigaard, N., C. Dahl (2009). Sulfur metabolism in phototrophic sulfur bacteria. *Advances in Microbial Physiology* 54:103–200.
- Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature* 459:193–199.
- Fuhrman, J. A., J. A. Steele (2008). Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquatic Microbial Ecology* 53:69–81.
- Gago, G., D. Kurth, L. Diacovich, S. Tsai, H. Gramajo (2006). Biochemical and structural characterization of an essential acyl coenzyme A carboxylase from *Mycobacterium tuberculosis*. *Journal of Bacteriology* 188:477–486.
- Gao, X., Y. Xin, R. E. Blankenship (2009). Enzymatic activity of the alternative complex III as a menaquinol:auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. *FEBS Letters* 583:3275–3279.
- Garcia Costas, A. M., Z. Liu, L. P. Tomsho, S. C. Schuster, D. M. Ward, D. A. Bryant (2012). Complete genome of *Candidatus Chloracidobacterium thermophilum*, a chlorophyll-based photoheterotroph belonging to the phylum Acidobacteria. *Environmental Microbiology* 14:177–190.
- Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* 14:685–695.
- Gibson, J., N. Pfennig, J. B. Waterbury (1984). *Chloroherpeton thalassium* gen. nov. et spec. nov., a non-filamentous, flexing and gliding green sulfur bacterium. *Archives of Microbiology* 138:96–101.
- Giovannoni, S., N. P. Revsbech, D. M. Ward, R. W. Castenholz (1987). Obligately phototrophic *Chloroflexus*: primary production in anaerobic hot spring microbial mats. *Archives of Microbiology* 147:80–87.
- Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics* 27:857–871.

- Gregor, J., G. Klug (1999). Regulation of bacterial photosynthesis genes by oxygen and light. *FEMS Microbiology Letters* 179:1–9.
- Grimm, F., B. Franz, C. Dahl (2011). Regulation of dissimilatory sulfur oxidation in the purple sulfur bacterium *Allochromatium vinosum*. *Frontiers in Microbiology* 2:51.
- Gupta, R. S., T. Mukhtar, B. Singh (1999). Evolutionary relationships among photosynthetic prokaryotes (*Hellobacterium chlorum*, *Chloroflexus aurantiacus*, cyanobacteria, *Chlorobium tepidum* and proteobacteria): implications regarding the origin of photosynthesis. *Molecular Microbiology* 32:893–906.
- Haft, D. H., B. J. Loftus, D. L. Richardson, F. Yang, J. A. Eisen, I. T. Paulsen, O. White (2001). TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Research* 29:41–43.
- Hallam, S. J., T. J. Mincer, C. Schleper, C. M. Preston, K. Roberts, P. M. Richardson, E. F. DeLong (2006). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biology* 4:e95.
- Hanada, S. (2003). Filamentous anoxygenic phototrophs in hot springs. *Microbes and Environments* 18:51–61.
- Hanada, S., A. Hiraishi, K. Shimada, K. Matsuura (1995). *Chloroflexus aggregans* sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement. *International Journal of Systematic Bacteriology* 45:676–681.
- Hanada, S., S. Takaichi, K. Matsuura, K. Nakamura (2002). *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *International Journal of Systematic and Evolutionary Microbiology* 52:187–193.
- Heidelberg, J. F., W. C. Nelson, T. Schoenfeld, D. Bhaya (2009). Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* 4:e4169.
- Henry, E. A., R. Devereux, J. S. Maki, C. C. Gilmour, C. R. Woese, L. Mandelco, R. Schauder, C. C. Remsen, R. Mitchell (1994). Characterization of a new thermophilic sulfate-reducing bacterium *Thermodesulfovibrio yellowstonii*, gen. nov. and sp. nov.: its phylogenetic relationship to *Thermodesulfovibrio commune* and their origins deep within the bacterial domain. *Archives of Microbiology* 161:62–69.

- Herter, S., A. Busch, G. Fuchs (2002a). L-Malyl-coenzyme A lyase/ β -methylmalyl-coenzyme A lyase from *Chloroflexus aurantiacus*, a bifunctional enzyme involved in autotrophic CO₂ fixation. *Journal of Bacteriology* 184:5999–6006.
- Herter, S., G. Fuchs, A. Bacher, W. Eisenreich (2002b). A bicyclic autotrophic CO₂ fixation pathway in *Chloroflexus aurantiacus*. *The Journal of Biological Chemistry* 277:20277–20283.
- Hesselmann, R. P. X., R. von Rummell, S. M. Resnick, R. Hany, A. J. B. Zehnder (2000). Anaerobic metabolism of bacteria performing enhanced biological phosphate removal. *Water Research* 34:3487–3494.
- Holo, H., R. Sirevåg (1986). Autotrophic growth and CO₂ fixation of *Chloroflexus aurantiacus*. *Archives of Microbiology* 145:173–180.
- Holt, J. G., R. A. Lewin (1968). *Herpetosiphon aurantiacus* gen. et sp. n., a new filamentous gliding organism. *Journal of Bacteriology* 95:2407–2408.
- Huang, X., A. Madan (1999). CAP3: a DNA sequence assembly program. *Genome Research* 9:868–877.
- Huber, T., G. Faulkner, P. Hugenholtz (2004). Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics (Oxford, England)* 20:2317–2319.
- Hugenholtz, P., E. Stackebrandt (2004). Reclassification of *Sphaerobacter thermophilus* from the subclass *Sphaerobacteridae* in the phylum Actinobacteria to the class Thermomicrobia (emended description) in the phylum Chloroflexi (emended description). *International Journal of Systematic and Evolutionary Microbiology* 54:2049–2051.
- Hügler, M., H. Huber, K. O. Stetter, G. Fuchs (2003a). Autotrophic CO₂ fixation pathways in archaea (Crenarchaeota). *Archives of Microbiology* 179:160–173.
- Hügler, M., R. S. Krieger, M. Jahn, G. Fuchs (2003b). Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *European Journal of Biochemistry / FEBS* 270:736–744.
- Hügler, M., C. Menendez, H. Schägger, G. Fuchs (2002). Malonyl-coenzyme A reductase from *Chloroflexus aurantiacus*, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO(2) fixation. *Journal of Bacteriology* 184:2404–2410.
- Hunaiti, A. R., P. E. Kolattukudy (1982). Isolation and characterization of an acyl-coenzyme A carboxylase from an erythromycin-producing *Streptomyces erythreus*. *Archives of Biochemistry and Biophysics* 216:362–371.

- Huson, D. H., A. F. Auch, J. Qi, S. C. Schuster (2007). MEGAN analysis of metagenomic data. *Genome Research* 17:377–386.
- Imhoff, J. F., J. Sling, R. Petri (1998). Phylogenetic relationships among the *Chromatiaceae*, their taxonomic reclassification and description of the new genera *Allochromatium*, *Halochromatium*, *Isochromatium*, *Marichromatium*, *Thiococcus*, *Thiohalocapsa* and *Thermochromatium*. *International Journal of Systematic Bacteriology* 48 Pt 4:1129–1143.
- Inskeep, W. P., G. G. Ackerman, W. P. Taylor, M. Kozubal, S. Korf, R. E. Macur (2005). On the energetics of chemolithotrophy in nonequilibrium systems: case studies of geothermal springs in Yellowstone National Park. *Geobiology* 3:297–317.
- Inskeep, W. P., R. E. Macur, G. Harrison, B. C. Bostick, S. Fendorf (2004). Biomineralization of As(V)-hydrous ferric oxyhydroxide in microbial mats of an acid-sulfate-chloride geothermal spring, Yellowstone National Park. *Geochimica et Cosmochimica Acta* 68:3141–3155.
- Inskeep, W. P., D. B. Rusch, Z. J. Jay, M. J. Herrgard, M. A. Kozubal, T. H. Richardson, R. E. Macur, N. Hamamura, R. d. Jennings, B. W. Fouke, A. Reysenbach, F. Roberto, M. Young, A. Schwartz, E. S. Boyd, J. H. Badger, E. J. Mathur, A. C. Ortmann, M. Bateson, G. Geesey, M. Frazier (2010). Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PloS One* 5:e9773.
- Ivanovsky, R. N., Y. I. Fal, I. A. Berg, N. V. Ugolkova, E. N. Krasilnikova, O. I. Keppen, L. M. Zakharchuc, A. M. Zyakun (1999). Evidence for the presence of the reductive pentose phosphate cycle in a filamentous anoxygenic photosynthetic bacterium, *Oscillochloris trichoides* strain DG-6. *Microbiology* 145:1743–1748.
- Jackson, T. J., R. F. Ramaley, W. G. Meinschein (1973). *Thermomicrobium*, a new genus of extremely thermophilic bacteria. *International Journal of Systematic Bacteriology* 23:28–36.
- Jensen, S. I., A. Steunou, D. Bhaya, M. Khl, A. R. Grossman (2011). In situ dynamics of O₂, pH and cyanobacterial transcripts associated with CCM, photosynthesis and detoxification of ROS. *The ISME Journal* 5:317–328.
- Jiao, Y., D. K. Newman (2007). The *pio* operon is essential for phototrophic Fe(II) oxidation in *Rhodopseudomonas palustris* TIE-1. *Journal of Bacteriology* 189:1765–1773.
- Jrgensen, B. B., D. C. Nelson (1988). Bacterial zonation, photosynthesis, and spectral light distribution in hot spring microbial mats of Iceland. *Microbial Ecology* 16:133–147.

- Kanamori, T., N. Kanou, H. Atomi, T. Imanaka (2004). Enzymatic characterization of a prokaryotic urea carboxylase. *Journal of Bacteriology* 186:2532–2539.
- Kanungo, T., D. Mount, N. Netanyahu, C. Piatko, R. Silverman, A. Wu (2002). An efficient k-means clustering algorithm: analysis and implementation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24:881–892.
- Karlin, S., S. F. Altschul (1990). Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proceedings of the National Academy of Sciences of the United States of America* 87:2264–2268.
- Kettler, G. C., A. C. Martiny, K. Huang, J. Zucker, M. L. Coleman, S. Rodrigue, F. Chen, A. Lapidus, S. Ferriera, J. Johnson, C. Steglich, G. M. Church, P. Richardson, S. W. Chisholm (2007). Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genetics* 3:e231.
- Kiatpapan, P., H. Kobayashi, M. Sakaguchi, H. Ono, M. Yamashita, Y. Kaneko, Y. Murooka (2001). Molecular characterization of lactobacillus plantarum genes for beta-ketoacyl-acyl carrier protein synthase III (fabH) and acetyl coenzyme A carboxylase (accBCDA), which are essential for fatty acid biosynthesis. *Applied and Environmental Microbiology* 67:426–433.
- Kim, E., J. Kim, I. Lee, H. J. Rhee, J. K. Lee (2008). Superoxide generation by chlorophyllide *a* reductase of *Rhodobacter sphaeroides*. *The Journal of Biological Chemistry* 283:3718–3730.
- Kimura, Y., R. Miyake, Y. Tokumasu, M. Sato (2000). Molecular cloning and characterization of two genes for the biotin carboxylase and carboxyltransferase subunits of acetyl coenzyme A carboxylase in *Myxococcus xanthus*. *Journal of Bacteriology* 182:5462–5469.
- Klappenbach, J. A., B. K. Pierson (2004). Phylogenetic and physiological characterization of a filamentous anoxygenic photoautotrophic bacterium 'Candidatus Chlorothrix halophila' gen. nov., sp. nov., recovered from hypersaline microbial mats. *Archives of Microbiology* 181:17–25.
- Klatt, C. G., D. A. Bryant, D. M. Ward (2007). Comparative genomics provides evidence for the 3-hydroxypropionate autotrophic pathway in filamentous anoxygenic phototrophic bacteria and in hot spring microbial mats. *Environmental Microbiology* 9:2067–78.
- Klatt, C. G., J. M. Wood, D. B. Rusch, M. M. Bateson, N. Hamamura, J. F. Heidelberg, A. R. Grossman, D. Bhaya, F. M. Cohan, M. Kühl, D. A. Bryant, D. M. Ward (2011). Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *The ISME Journal* 5:1262–1278.

- Krogh, A., B. Larsson, G. von Heijne, E. L. Sonnhammer (2001). Predicting trans-membrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology* 305:567–580.
- Kumar, S., K. Tamura, M. Nei (2004). MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5:150–163.
- Kunin, V., J. Raes, J. K. Harris, J. R. Spear, J. J. Walker, N. Ivanova, C. von Mering, B. M. Bebout, N. R. Pace, P. Bork, P. Hugenholtz (2008). Millimeter-scale genetic gradients and community-level molecular convergence in a hypersaline microbial mat. *Molecular Systems Biology* 4:198.
- Kunisawa, T. (2010). Evaluation of the phylogenetic position of the sulfate-reducing bacterium *Thermodesulfovibrio yellowstonii* (phylum Nitrospirae) by means of gene order data from completely sequenced genomes. *International Journal of Systematic and Evolutionary Microbiology* 60:1090–1102.
- Langenheder, S., M. T. Bulling, M. Solan, J. I. Prosser (2010). Bacterial biodiversity-ecosystem functioning relations are modified by environmental complexity. *PloS One* 5:e10834.
- Lee, M., M. C. del Rosario, H. H. Harris, R. E. Blankenship, J. M. Guss, H. C. Freeman (2009). The crystal structure of auracyanin A at 1.85 Å resolution: the structures and functions of auracyanins A and B, two almost identical "blue" copper proteins, in the photosynthetic bacterium *Chloroflexus aurantiacus*. *Journal of Biological Inorganic Chemistry* 14:329–345.
- Legendre, L., Pierre Legendre (1998). *Numerical Ecology*. Elsevier, Amsterdam, The Netherlands.
- Li, H., R. Durbin (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25:1754–1760.
- Li, S. J., J. Cronan, J E (1992a). The gene encoding the biotin carboxylase subunit of escherichia coli acetyl-CoA carboxylase. *The Journal of Biological Chemistry* 267:855–863.
- Li, S. J., J. Cronan, J E (1992b). The genes encoding the two carboxyltransferase subunits of escherichia coli acetyl-CoA carboxylase. *The Journal of Biological Chemistry* 267:16841–16847.
- Lin, T., M. M. Melgar, D. Kurth, S. J. Swamidass, J. Purdon, T. Tseng, G. Gago, P. Baldi, H. Gramajo, S. Tsai (2006). Structure-based inhibitor design of AccD5, an essential acyl-CoA carboxylase carboxyltransferase domain of *Mycobacterium*

- tuberculosis. Proceedings of the National Academy of Sciences of the United States of America* 103:3072–3077.
- Liu, Z., C. G. Klatt, M. Ludwig, D. B. Rusch, S. I. Jensen, M. Kühl, D. M. Ward, D. A. Bryant (2011a). *Candidatus* Thermochlorobacter aerophilum: an aerobic chlorophotoheterotrophic member of the phylum Chlorobi. *Submitted*.
- Liu, Z., C. G. Klatt, J. M. Wood, D. B. Rusch, M. Ludwig, N. Wittekindt, L. P. Tomsho, S. C. Schuster, D. M. Ward, D. A. Bryant (2011b). Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *The ISME Journal* 5:1279–1290.
- Lozupone, C. A., M. Hamady, S. T. Kelley, R. Knight (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Applied and Environmental Microbiology* 73:1576–1585.
- Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, K. Schleifer (2004). ARB: a software environment for sequence data. *Nucleic Acids Research* 32:1363–1371.
- Lueders, T., R. Kindler, A. Miltner, M. W. Friedrich, M. Kaestner (2006). Identification of bacterial micropredators distinctively active in a soil microbial food web. *Applied and Environmental Microbiology* 72:5342–5348.
- Lueders, T., M. Manefield, M. W. Friedrich (2004). Enhanced sensitivity of DNA- and rRNA-based stable isotope probing by fractionation and quantitative analysis of isopycnic centrifugation gradients. *Environmental Microbiology* 6:73–78.
- Macur, R. E., C. R. Jackson, L. M. Botero, T. R. McDermott, W. P. Inskeep (2004). Bacterial populations associated with the oxidation and reduction of arsenic in an unsaturated soil. *Environmental Science and Technology* 38:104–111.
- Madigan, M. T. (1984). A novel photosynthetic purple bacterium isolated from a Yellowstone hot spring. *Science* 225:313–315.
- Madigan, M. T., T. D. Brock (1975). Photosynthetic sulfide oxidation by *Chloroflexus aurantiacus*, a filamentous, photosynthetic, gliding bacterium. *J Bacteriol* 122:782–784.
- Madigan, M. T., T. D. Brock (1977). CO₂ fixation in photosynthetically-grown *Chloroflexus aurantiacus*. *FEMS Microbiol Lett* 1:301–304.

- Madigan, M. T., D. O. Jung, E. A. Karr, W. M. Sattley, L. A. Achenbach, M. T. J. van der Meer (2005). Diversity of anoxygenic phototrophs in contrasting extreme environments. In: W. P. Inskeep, T. R. McDermott (eds.), *Geothermal Biology and Geochemistry in Yellowstone National Park*. Montana State University Publications, Bozeman, pp. 203–219.
- Madigan, M. T., S. R. Petersen, T. D. Brock (1974). Nutritional studies on *Chloroflexus*, a filamentous photosynthetic, gliding bacterium. *Archives of Microbiology* 100:97–103.
- Madigan, M. T., R. Takigiku, R. G. Lee, H. Gest, J. M. Hayes (1989). Carbon isotope fractionation by thermophilic phototrophic sulfur bacteria: Evidence for autotrophic growth in natural populations. *Applied and Environmental Microbiology* 55:639–644.
- Majewski, J., P. Zawadzki, P. Pickerill, F. M. Cohan, C. G. Dowson (2000). Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. *Journal of Bacteriology* 182:1016–1023.
- Manefield, M., A. S. Whiteley, N. Ostle, P. Ineson, M. J. Bailey (2002). Technical considerations for RNA-based stable isotope probing: an approach to associating microbial diversity with microbial community function. *Rapid Communications in Mass Spectrometry: RCM* 16:2179–2183.
- Marini, P., S. J. Li, D. Gardiol, J. Cronan, J. E. D. de Mendoza (1995). The genes encoding the biotin carboxyl carrier protein and biotin carboxylase subunits of bacillus subtilis acetyl coenzyme a carboxylase, the first enzyme of fatty acid synthesis. *Journal of Bacteriology* 177:7003–7006.
- McClesky, R. B., J. A. Ball, D. K. Nordstrom, J. M. Holloway, H. E. Taylor (2005). *Water-Chemistry Data for Selected Hot Springs, Geysers, and Streams in Yellowstone National Park, Wyoming, 2001-2002. Open-File Report 2004-1316*. U.S. Geological Survey: Reston, VA.
- McManus, J. D., D. C. Brune, J. Han, J. Sanders-Loehr, T. E. Meyer, M. A. Cusanovich, G. Tollin, R. E. Blankenship (1992). Isolation, characterization, and amino acid sequences of auracyanins, blue copper proteins from the green photosynthetic bacterium *Chloroflexus aurantiacus*. *The Journal of Biological Chemistry* 267:6531–6540.
- van der Meer, M., S. Schouten, S. Hanada, E. Hopmans, J. Sinninghe Damsté, D. Ward (2002). Alkane-1,2-diol-based glycosides and fatty glycosides and wax esters in *Roseiflexus castenholzii* and hot spring microbial mats. *Archives of Microbiology* 178:229–237.

- van der Meer, M. T., S. Schouten, B. E. van Dongen, W. I. Rijpstra, G. Fuchs, J. S. Damsté, J. W. de Leeuw, D. M. Ward (2001). Biosynthetic controls on the ^{13}C contents of organic components in the photoautotrophic bacterium *Chloroflexus aurantiacus*. *The Journal of Biological Chemistry* 276:10971–10976.
- van der Meer, M. T., S. Schouten, J. W. de Leeuw, D. M. Ward (2000). Autotrophy of green non-sulphur bacteria in hot spring microbial mats: biological explanations for isotopically heavy organic carbon in the geological record. *Environmental Microbiology* 2:428–435.
- van der Meer, M. T. J., C. G. Klatt, J. Wood, D. A. Bryant, M. M. Bateson, L. Lammerts, S. Schouten, J. S. Sinninghe Damsté, M. T. Madigan, D. M. Ward (2010). Cultivation and genomic, nutritional, and lipid biomarker characterization of *Roseiflexus* strains closely related to predominant in situ populations inhabiting Yellowstone hot spring microbial mats. *Journal of Bacteriology* 192:3033–3042.
- van der Meer, M. T. J., S. Schouten, M. M. Bateson, U. Nübel, A. Wieland, M. Kühl, J. W. de Leeuw, J. S. Sinninghe Damsté, D. M. Ward (2005). Diel variations in carbon metabolism by green nonsulfur-like bacteria in alkaline siliceous hot spring microbial mats from Yellowstone National Park. *Applied and Environmental Microbiology* 71:3978–3986.
- van der Meer, M. T. J., S. Schouten, J. S. Sinninghe Damsté, J. W. de Leeuw, D. M. Ward (2003). Compound-specific isotopic fractionation patterns suggest different carbon metabolisms among *Chloroflexus*-like bacteria in hot-spring microbial mats. *Applied and Environmental Microbiology* 69:6000–6006.
- van der Meer, M. T. J., S. Schouten, J. S. Sinninghe Damsté, D. M. Ward (2007). Impact of carbon metabolism on ^{13}C signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). *Environmental Microbiology* 9:482–491.
- Melendrez, M. C., R. K. Lange, F. M. Cohan, D. M. Ward (2011). Influence of molecular resolution on sequence-based discovery of ecological diversity among *Synechococcus* populations in an alkaline siliceous hot spring microbial mat. *Applied and Environmental Microbiology* 77:1359–1367.
- Menendez, C., Z. Bauer, H. Huber, N. Gad'on, K. O. Stetter, G. Fuchs (1999). Presence of acetyl coenzyme a (CoA) carboxylase and propionyl-CoA carboxylase in autotrophic Crenarchaeota and indication for operation of a 3-hydroxypropionate cycle in autotrophic carbon fixation. *Journal of Bacteriology* 181:1088–1098.
- Miller, J. R., A. L. Delcher, S. Koren, E. Venter, B. P. Walenz, A. Brownley, J. Johnson, K. Li, C. Mobarry, G. Sutton (2008). Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics (Oxford, England)* 24:2818–2824.

- Miller, S. R. (2003). Evidence for the adaptive evolution of the carbon fixation gene *rbcL* during diversification in temperature tolerance of a clade of hot spring cyanobacteria. *Molecular Ecology* 12:1237–1246.
- Miller, S. R., R. W. Castenholz (2000). Evolution of thermotolerance in hot spring cyanobacteria of the genus *Synechococcus*. *Applied and Environmental Microbiology* 66:4222–4229.
- Miller, S. R., R. W. Castenholz, D. Pedersen (2007). Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Applied and Environmental Microbiology* 73:4751–4759.
- Miller, S. R., M. D. Purugganan, S. E. Curtis (2006). Molecular population genetics and phenotypic diversification of two populations of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Applied and Environmental Microbiology* 72:2793–2800.
- Miller, S. R., A. L. Strong, K. L. Jones, M. C. Ungerer (2009). Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Applied and Environmental Microbiology* 75:4565–4572.
- Nakagawa, T., M. Fukui (2002). Phylogenetic characterization of microbial mats and streamers from a Japanese alkaline hot spring with a thermal gradient. *The Journal of General and Applied Microbiology* 48:211–222.
- Nakamura, Y., T. Kaneko, S. Sato, M. Ikeuchi, H. Katoh, S. Sasamoto, A. Watanabe, M. Iriguchi, K. Kawashima, T. Kimura, Y. Kishida, C. Kiyokawa, M. Kohara, M. Matsumoto, A. Matsuno, N. Nakazaki, S. Shimpo, M. Sugimoto, C. Takeuchi, M. Yamada, S. Tabata (2002). Complete genome structure of the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. *DNA Research* 9:123–130.
- van Niel, C. B., L. A. Thayer (1930). Report on preliminary observations on the microflora in and near the hot springs in yellowstone national park and their importance for the geological formations. *YNP Lib File 7312, Mammoth, WY*.
- Nold, S. C., D. M. Ward (1996). Photosynthate partitioning and fermentation in hot spring microbial mat communities. *Applied and Environmental Microbiology* 62(12):4598–4607.
- Nomata, J., T. Mizoguchi, H. Tamiaki, Y. Fujita (2006). A second nitrogenase-like enzyme for bacteriochlorophyll biosynthesis: reconstitution of chlorophyllide *a* reductase with purified x-protein (BchX) and YZ-protein (BchY-BchZ) from *Rhodobacter capsulatus*. *The Journal of Biological Chemistry* 281:15021–15028.

- Nübel, U., M. M. Bateson, V. Vandieken, A. Wieland, M. Köhl, D. M. Ward (2002). Microscopic examination of distribution and phenotypic properties of phylogenetically diverse *Chloroflexaceae*-related bacteria in hot spring microbial mats. *Applied and Environmental Microbiology* 68:4593–603.
- Olson, J. M. (2006). Photosynthesis in the archaean era. *Photosynthesis Research* 88:109–117.
- Oshima, T., K. Imahori (1974). Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *International Journal of Systematic Bacteriology* 24:102–112.
- Ouchane, S., A. Steunou, M. Picaud, C. Astier (2004). Aerobic and anaerobic Mg-protoporphyrin monomethyl ester cyclases in purple bacteria: a strategy adopted to bypass the repressive oxygen control system. *The Journal of Biological Chemistry* 279:6385–6394.
- Oyaizu, H., B. Debrunner-Vossbrinck, L. Mandelco, J. A. Studier, C. R. Woese (1987). The green non-sulfur bacteria: a deep branching in the eubacterial line of descent. *Systematic and Applied Microbiology* 9:47–53.
- Papke, R. T., N. B. Ramsing, M. M. Bateson, D. M. Ward (2003). Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology* 5:650–659.
- Parenteau, M. N., S. L. Cady (2010). Microbial biosignatures in iron-mineralized phototrophic mats at Chocolate Pots Hot Spring, Yellowstone National Park, United States. *Palaios* 25:97–111.
- Passarge, E., B. Horsthemke, R. A. Farber (1999). Incorrect use of the term synteny. *Nature Genetics* 23:387.
- Pfennig, N. (1974). *Rhodopseudomonas globiformis*, sp. n., a new species of the *Rhodospirillaceae*. *Archives of Microbiology* 100:197–206.
- Pierson, B. K., R. W. Castenholz (1974a). A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Archives of Microbiology* 100:5–24.
- Pierson, B. K., R. W. Castenholz (1974b). Studies of pigments and growth in *Chloroflexus aurantiacus*, a phototrophic filamentous bacterium. *Archives of Microbiology* 100:283–305.
- Pierson, B. K., S. Giovannoni, D. A. Stahl, R. W. Castenholz (1985). *Heliothrix oregonensis*, gen. nov., sp. nov., a phototrophic filamentous gliding bacterium containing bacteriochlorophyll *a*. *Archives of Microbiology* 142:164–167.

- Pierson, B. K., M. N. Parenteau (2000). Phototrophs in high iron microbial mats: microstructure of mats in iron-depositing hot springs. *FEMS Microbiology Ecology* 32:181–196.
- Pierson, B. K., M. N. Parenteau, B. M. Griffin (1999). Phototrophs in high-iron-concentration microbial mats: physiological ecology of phototrophs in an iron-depositing hot spring. *Applied and Environmental Microbiology* 65:5474–5483.
- Pride, D. T., R. J. Meinersmann, T. M. Wassenaar, M. J. Blaser (2003). Evolutionary implications of microbial genome tetranucleotide frequency biases. *Genome Research* 13:145–158.
- Prosser, J. I., B. J. M. Bohannon, Curtis, R. J. Ellis, M. K. Firestone, R. P. Freckleton, J. L. Green, L. E. Green, K. Killham, J. J. Lennon, A. M. Osborn, M. Solan, C. J. van der Gast, J. P. W. Young (2007). The role of ecological theory in microbial ecology. *Nature Reviews Microbiology* 5:384–392.
- R Core Development Team, T. (2011). R: A language and environment for statistical computing - reference index version 2.6.2. <http://www.r-project.org/>.
- Rappé, M. S., S. J. Giovannoni (2003). The uncultured microbial majority. *Annual Review of Microbiology* 57:369–394.
- Raymond, J. (2005). The evolution of biological carbon and nitrogen cycling—a genomic perspective. *Reviews in Mineralogy and Geochemistry* 59:211–231.
- Raymond, J., O. Zhaxybayeva, J. P. Gogarten, R. E. Blankenship (2003). Evolution of photosynthetic prokaryotes: a maximum-likelihood mapping approach. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 358:223–30.
- Raymond, J., O. Zhaxybayeva, J. P. Gogarten, S. Y. Gerdes, R. E. Blankenship (2002). Whole-genome analysis of photosynthetic prokaryotes. *Science* 298:1616–20.
- Revsbech, N. P., D. M. Ward (1984). Microelectrode studies of interstitial water chemistry and photosynthetic activity in a hot spring microbial mat. *Applied and Environmental Microbiology* 48:270–275.
- Reysenbach, A., G. S. Wickham, N. R. Pace (1994). Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Applied and Environmental Microbiology* 60(6):2113–2119.
- Richardson, L. L., R. W. Castenholz (1987). Diel vertical movements of the cyanobacterium *Oscillatoria terebriformis* in a sulfide-rich hot spring microbial mat. *Applied and Environmental Microbiology* 53:2142–2150.

- Roberts, M. S., F. M. Cohan (1993). The effect of DNA sequence divergence on sexual isolation in bacillus. *Genetics* 134:401–408.
- Rocha, E. P. C. (2006). Inference and analysis of the relative stability of bacterial chromosomes. *Molecular Biology and Evolution* 23:513–522.
- Rodríguez, E., C. Banchio, L. Diacovich, M. J. Bibb, H. Gramajo (2001). Role of an essential acyl coenzyme A carboxylase in the primary and secondary metabolism of *Streptomyces coelicolor* A3(2). *Applied and Environmental Microbiology* 67:4166–4176.
- Rodríguez, E., H. Gramajo (1999). Genetic and biochemical characterization of the alpha and beta components of a propionyl-CoA carboxylase complex of *Streptomyces coelicolor* A3(2). *Microbiology* 145:3109–3119.
- Röling, W. F. M., M. Ferrer, P. N. Golyshin (2010). Systems approaches to microbial communities and their functioning. *Current Opinion in Biotechnology* 21:532–538.
- Rowe, J. J., R. O. Fournier, G. W. Morey (1973). *Chemical analysis of thermal waters in Yellowstone National Park, Wyoming, 1960–65*. Geological Survey, Washington, DC.
- Ruan, Q., D. Dutta, M. S. Schwalbach, J. A. Steele, J. A. Fuhrman, F. Sun (2006). Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. *Bioinformatics (Oxford, England)* 22:2532–2538.
- Ruff-Roberts, A. L., J. G. Kuenen, D. M. Ward (1994). Distribution of cultivated and uncultivated cyanobacteria and *Chloroflexus*-like bacteria in hot spring microbial mats. *Applied and Environmental Microbiology* 60:697–704.
- Rusch, D. B., A. L. Halpern, G. Sutton, K. B. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. A. Eisen, J. M. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. E. Venter, K. Li, S. Kravitz, J. F. Heidelberg, T. Utterback, Y. Rogers, L. I. Falcn, V. Souza, G. Bonilla-Rosso, L. E. Eguiarte, D. M. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. R. Ferrari, R. L. Strausberg, K. Nealson, R. Friedman, M. Frazier, J. C. Venter (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biology* 5:e77.
- Sadekar, S., J. Raymond, R. E. Blankenship (2006). Conservation of distantly related membrane proteins: Photosynthetic reaction centers share a common structural core. *Molecular Biology and Evolution* 23:2001–7.

- Sakata, S., J. M. Hayes, A. R. McTaggart, R. A. Evans, K. J. Leckrone, R. K. Togasaki (1997). Carbon isotopic fractionation associated with lipid biosynthesis by a cyanobacterium: relevance for interpretation of biomarker records. *Geochim Cosmochim Acta* 61:5379–5389.
- Saldanha, A. J. (2004). Java treeview—extensible visualization of microarray data. *Bioinformatics (Oxford, England)* 20:3246–3248.
- Samols, D., C. G. Thornton, V. L. Murtif, G. K. Kumar, F. C. Haase, H. G. Wood (1988). Evolutionary conservation among biotin enzymes. *The Journal of Biological Chemistry* 263:6461–6464.
- Sandbeck, K. A., D. M. Ward (1981). Fate of immediate methane precursors in low-sulfate, hot-spring algal-bacterial mats. *Applied and Environmental Microbiology* 41:775–782.
- Say, R. F., G. Fuchs (2010). Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme. *Nature* 464:1077–1081.
- Schaffert, C. S., D. M. Ward, C. G. Klatt, M. Pauley, L. A. Steinke (2011). Identification and distribution of high abundance proteins in an Octopus Spring microbial mat community. *in prep*.
- Schroeder, A., O. Mueller, S. Stocker, R. Salowsky, M. Leiber, M. Gassmann, S. Lightfoot, W. Menzel, M. Granzow, T. Ragg (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* 7:3.
- Sekiguchi, Y., T. Yamada, S. Hanada, A. Ohashi, H. Harada, Y. Kamagata (2003). *Anaerolinea thermophila* gen. nov., sp. nov. and *Caldilinea aerophila* gen. nov., sp. nov., novel filamentous thermophiles that represent a previously uncultured lineage of the domain Bacteria at the subphylum level. *International Journal of Evolutionary Microbiology* 53:1843–1851.
- Shannon, P., A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13:2498–2504.
- Shiea, J., S. C. Brassell, D. M. Ward (1991). Comparative analysis of extractable lipids in hot spring microbial mats and their component photosynthetic bacteria. *Organic Geochemistry* 17:309–319.
- Simmons, S. L., G. DiBartolo, V. J. Denef, D. S. A. Goltsman, M. P. Thelen, J. F. Banfield (2008). Population genomic analysis of strain variation in *Leptospirillum* group II bacteria involved in acid mine drainage formation. *PLoS Biol* 6:e177.

- Sirevåg, R., R. Castenholz (1979). Aspects of carbon metabolism in *Chloroflexus*. *Archives of Microbiology* 120:151–153.
- Skirnisdottir, S., G. O. Hreggvidsson, S. Hjörleifsdottir, V. T. Marteinsson, S. K. Petursdottir, O. Holst, J. K. Kristjansson (2000). Influence of sulfide and temperature on species composition and community structure of hot spring microbial mats. *Applied and Environmental Microbiology* 66:2835–2841.
- Smith, D. R., L. A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert, D. Harrison, L. Hoang, P. Keagle, W. Lumm, B. Pothier, D. Qiu, R. Spadafora, R. Vicaire, Y. Wang, J. Wierzbowski, R. Gibson, N. Jiwani, A. Caruso, D. Bush, J. N. Reeve (1997). Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: functional analysis and comparative genomics. *Journal of Bacteriology* 179:7135–7155.
- Sprague, S. G., L. A. Staehelin, M. J. DiBartolomeis, R. C. Fuller (1981). Isolation and development of chlorosomes in the green bacterium *Chloroflexus aurantiacus*. *Journal of Bacteriology* 147:1021–1031.
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics (Oxford, England)* 22:2688–2690.
- Steinke, L. A., G. Slys, C. G. Klatt, M. S. Lipton, D. A. Bryant, G. Anderson, D. M. Ward (2011). Integration of real-time systems biology in a microbial community. *in prep*.
- Steunou, A., D. Bhaya, M. M. Bateson, M. C. Melendrez, D. M. Ward, E. Brecht, J. W. Peters, M. Köhl, A. R. Grossman (2006). In situ analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proceedings of the National Academy of Sciences of the United States of America* 103:2398–2403.
- Steunou, A., S. I. Jensen, E. Brecht, E. D. Becraft, M. M. Bateson, O. Kilian, D. Bhaya, D. M. Ward, J. W. Peters, A. R. Grossman, M. Köhl (2008). Regulation of *nif* gene expression and the energetics of N_2 fixation over the diel cycle in a hot spring microbial mat. *The ISME Journal* 2:364–78.
- Stolyar, S., S. Van Dien, K. L. Hillesland, N. Pinel, T. J. Lie, J. A. Leigh, D. A. Stahl (2007). Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology* 3:92.
- Strauss, G., W. Eisenreich, A. Bacher, G. Fuchs (1992). ^{13}C -NMR study of autotrophic CO_2 fixation pathways in the sulfur-reducing archaebacterium *Thermoproteus neutrophilus* and in the phototrophic eubacterium *Chloroflexus aurantiacus*. *Eur J Biochem* 214:853–866.

- Strauss, G., G. Fuchs (1993). Enzymes of a novel autotrophic CO₂ fixation pathway in the phototrophic bacterium *Chloroflexus aurantiacus*, the 3-hydroxypropionate cycle. *European Journal of Biochemistry / FEBS* 215:633–643.
- Swingley, W. D., R. E. Blankenship, J. Raymond (2008). Integrating markov clustering and molecular phylogenetics to reconstruct the cyanobacterial species tree from conserved protein families. *Molecular Biology and Evolution* 25:643–654.
- Swingley, W. D., S. Sadekar, S. D. Mastrian, H. J. Matthies, J. Hao, H. Ramos, C. R. Acharya, A. L. Conrad, H. L. Taylor, L. C. Dejesa, M. K. Shah, M. E. O'hallachain, M. T. Lince, R. E. Blankenship, J. T. Beatty, J. W. Touchman (2007). The complete genome sequence of *Roseobacter denitrificans* reveals a mixotrophic rather than photosynthetic metabolism. *Journal of Bacteriology* 189:683–90.
- Taffs, R., J. E. Aston, K. Brileya, Z. Jay, C. G. Klatt, S. McGlynn, N. Mallette, S. Montross, R. Gerlach, W. P. Inskeep, D. M. Ward, R. P. Carlson (2009). In silico approaches to study mass and energy flows in microbial consortia: A syntrophic case study. *BMC Systems Biology* 3:114.
- Tamura, K., J. Dudley, M. Nei, S. Kumar (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* 24:1596–1599.
- Tanenbaum, D., J. Goll, S. Murphy, P. Kumar, N. Zafar, M. Thiagarajan, R. Madupu, T. Davidsen, L. Kagan, S. Kravitz, D. B. Rusch, S. Yooseph (2010). The JCVI standard operating procedure for annotating prokaryotic metagenomic shotgun sequencing data. *Standards in Genomic Sciences* 2:229–237.
- Tang, K., K. Barry, O. Chertkov, E. Dalin, C. Han, L. Hauser, B. Honchak, L. Karch, M. Land, A. Lapidus, F. Larimer, N. Mikhailova, S. Pitluck, B. Pierson, R. Blankenship (2011). Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genomics* 12:334.
- Teeling, H., A. Meyerdierks, M. Bauer, R. Amann, F. O. Glöckner (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environmental Microbiology* 6:938–947.
- Toplin, J. A., T. B. Norris, C. R. Lehr, T. R. McDermott, R. W. Castenholz (2008). Biogeographic and phylogenetic diversity of thermoacidophilic cyanidiales in yellowstone national park, japan, and new zealand. *Applied and Environmental Microbiology* 74:2822–2833.
- Tsukatani, Y., N. Nakayama, K. Matsuura, K. Shimada, S. Hanada, K. Nagashima (2007). Characterization of a blue copper protein auracyanin from the filamentous

- anoxygenic phototroph *Roseiflexus castenholzii*. *Plant and Cell Physiology* 48:S73–S73.
- Tyson, G. W., J. Chapman, P. Hugenholtz, E. E. Allen, R. J. Ram, P. M. Richardson, V. V. Solovyev, E. M. Rubin, D. S. Rokhsar, J. F. Banfield (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428:37–43.
- Ugol'kova, N. V., R. N. Ivanovsky (2000). On the mechanism of autotrophic fixation of CO₂ by *Chloroflexus aurantiacus*. *Microbiology* 69:139–142.
- Vignais, P. M., B. Billoud, J. Meyer (2001). Classification and phylogeny of hydrogenases. *FEMS Microbiology Reviews* 25:455–501.
- Wahlund, T. M., C. R. Woese, R. W. Castenholz, M. T. Madigan (1991). A thermophilic green sulfur bacterium from New Zealand hot springs, *Chlorobium tepidum* sp. nov. *Archives of Microbiology* 156:81–90.
- Walter, M. R., G. Heys (1985). Links between the rise of the metazoa and the decline of stromatolites. *Precambrian Research* 29:149–174.
- Wang, Q., G. M. Garrity, J. M. Tiedje, J. R. Cole (2007). Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology* 73:5261–5267.
- Ward, D. M. (1978). Thermophilic methanogenesis in a hot-spring algal-bacterial mat (71 to 30 degrees C). *Applied and Environmental Microbiology* 35:1019–1026.
- Ward, D. M. (1998). A natural species concept for prokaryotes. *Current Opinion in Microbiology* 1:271–277.
- Ward, D. M., M. M. Bateson, M. J. Ferris, M. Kühl, A. Wieland, A. Koeppel, F. M. Cohan (2006). Cyanobacterial ecotypes in the microbial mat community of Mushroom Spring (Yellowstone National Park, Wyoming) as species-like units linking microbial community composition, structure and function. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences* 361:1997–2008.
- Ward, D. M., J. Bauld, R. W. Castenholz, B. K. Pierson (1992). Modern phototrophic microbial mats: anoxygenic, intermittantly oxygenic/anoxygenic, thermal, eukaryotic, and terrestrial. In: J. W. Schopf, C. Klein (eds.), *The Proterozoic Biosphere: A multidisciplinary study*. Cambridge University Press, Cambridge UK, pp. 309–324.
- Ward, D. M., R. W. Castenholz (2000). Cyanobacteria in geothermal habitats. In: B. A. Whitton, M. Potts (eds.), *Ecology of Cyanobacteria*. Kluwer Academic Publishers, The Netherlands, pp. 37–59.

- Ward, D. M., F. M. Cohan, D. Bhaya, J. F. Heidelberg, M. Kühl, A. Grossman (2008). Genomics, environmental genomics and the issue of microbial species. *Heredity* 100:207–19.
- Ward, D. M., M. J. Ferris, S. C. Nold, M. M. Bateson (1998). A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. *Microbiology and Molecular Biology Reviews* 62:1353–1370.
- Ward, D. M., C. G. Klatt, J. Wood, F. M. Cohan, D. A. Bryant (2012a). Functional genomics in an ecological and evolutionary context: maximizing the value of genomes in systems biology. In: R. L. Burnap, W. Vermaas (eds.), *Functional Genomics and Evolution of Photosynthetic Systems, Advances in Photosynthesis and Respiration*, vol. 33. Springer, Dordrecht, The Netherlands., pp. 1–16.
- Ward, D. M., S. R. Miller, R. W. Castenholz (2012b). Cyanobacteria in geothermal habitats. In: B. A. Whitton (ed.), *Ecology of Cyanobacteria*, 2nd edn. Springer, Dordrecht, The Netherlands., p. in press.
- Ward, D. M., R. T. Papke, U. Nübel, M. C. McKittrick (2002). Natural history of microorganisms inhabiting hot spring microbial mat communities: clues to the origin of microbial diversity and implications for microbiology and macrobiology. In: J. T. Staley, A. Reysenbach (eds.), *Biodiversity of Microbial Life: Foundations of Earth's Biosphere*. John Wiley and Sons, New York., pp. 27–48.
- Ward, D. M., C. M. Santegoeds, S. C. Nold, N. B. Ramsing, M. J. Ferris, M. M. Bateson (1997). Biodiversity within hot spring microbial mat communities: molecular monitoring of enrichment cultures. *Antonie van Leeuwenhoek* 71(1-2):143–150.
- Ward, D. M., J. Shiea, Y. B. Zeng, G. Dobson, S. Brassell, G. Eglinton (1989a). Lipid biochemical markers and the composition of microbial mats. In: Y. Cohen, E. Rosenberg (eds.), *Microbial Mats: Physiological ecology of benthic microbial communities*. American Society of Microbiology, Washington DC, pp. 439–454.
- Ward, D. M., T. A. Tayne, K. L. Anderson, M. M. Bateson (1987). Community structure and interactions among community members in hot spring cyanobacterial mats. *Symposium of the Society for General Microbiology* 41:179–210.
- Ward, D. M., R. Weller, M. M. Bateson (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* 345:63–65.
- Ward, D. M., R. Weller, J. Shiea, R. W. Castenholz, Y. Cohen (1989b). Hot spring microbial mats: Anoxygenic and oxygenic mats of possible evolutionary significance. In: Y. Cohen, E. Rosenberg (eds.), *Microbial Mats: Physiological ecology of benthic microbial communities*. American Society of Microbiology, Washington DC, pp. 3–15.

- Ward, N. L., J. F. Challacombe, P. H. Janssen, B. Henrissat, P. M. Coutinho, M. Wu, G. Xie, D. H. Haft, M. Sait, J. Badger, R. D. Barabote, B. Bradley, T. S. Brettin, L. M. Brinkac, D. Bruce, T. Creasy, S. C. Daugherty, T. M. Davidsen, R. T. DeBoy, J. C. Detter, R. J. Dodson, A. S. Durkin, A. Ganapathy, M. Gwinn-Giglio, C. S. Han, H. Khouri, H. Kiss, S. P. Kothari, R. Madupu, K. E. Nelson, W. C. Nelson, I. Paulsen, K. Penn, Q. Ren, M. J. Rosovitz, J. D. Selengut, S. Shrivastava, S. A. Sullivan, R. Tapia, L. S. Thompson, K. L. Watkins, Q. Yang, C. Yu, N. Zafar, L. Zhou, C. R. Kuske (2009). Three genomes from the phylum Acidobacteria provide insight into the lifestyles of these microorganisms in soils. *Applied and Environmental Microbiology* 75:2046–2056.
- Watanabe, Y., R. G. Feick, J. A. Shiozawa (1995). Cloning and sequencing of the genes encoding the light-harvesting B806-866 polypeptides and initial studies on the transcriptional organization of *puf2B*, *puf2A* and *puf2C* in *Chloroflexus aurantiacus*. *Archives of Microbiology* 163:124–30.
- Weller, R., M. M. Bateson, B. K. Heimbuch, E. D. Kopczynski, D. M. Ward (1992). Uncultivated cyanobacteria, *Chloroflexus*-like and spirochete-like inhabitants of a hot spring microbial mat. *Applied and Environmental Microbiology* 58:3964–3969.
- Wickstrom, C. E., R. W. Castenholz (1973). Thermophilic ostracod: aquatic metazoan with the highest known temperature tolerance. *Science* 181:1063–1064.
- Wickstrom, C. E., R. W. Castenholz (1985). Dynamics of cyanobacterial and ostracod interactions in an Oregon hot spring. *Ecology* 66:1024–1041.
- Wilhelm, L. J., H. J. Tripp, S. A. Givan, D. P. Smith, S. J. Giovannoni (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct* 2:27.
- Wilmes, P., A. F. Andersson, M. G. Lefsrud, M. Wexler, M. Shah, B. Zhang, R. L. Hettich, P. L. Bond, N. C. VerBerkmoes, J. F. Banfield (2008). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *The ISME Journal* 2:853–864.
- Woese, C. R. (1987). Bacterial evolution. *Microbiological Reviews* 51(2):221–271.
- Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J. Shapiro, E. Szeto, N. C. Kyrpides, M. Mussmann, R. Amann, C. Bergin, C. Ruehland, E. M. Rubin, N. Dubilier (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* 443:950–955.
- Wu, M., J. Eisen (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biology* 9:R151.

- Wu, M., Q. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S. A. Sullivan, J. F. Kolonay, D. H. Haft, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb, J. A. Eisen (2005). Life in hot carbon monoxide: the complete genome sequence of *Carboxydotherrmus hydrogenoformans* z-2901. *PLoS Genetics* 1:e65.
- Xiong, J., W. M. Fischer, K. Inoue, M. Nakahara, C. E. Bauer (2000). Molecular evidence for the early evolution of photosynthesis. *Science* 289:1724–1730.
- Xiong, J., K. Inoue, C. E. Bauer (1998). Tracking molecular evolution of photosynthesis by characterization of a major photosynthesis gene cluster from *Heliobacillus mobilis*. *Proceedings of the National Academy of Sciences of the United States of America* 95:14851–14856.
- Xu, J., M. A. Mahowald, R. E. Ley, C. A. Lozupone, M. Hamady, E. C. Martens, B. Henrissat, P. M. Coutinho, P. Minx, P. Latreille, H. Cordum, A. Van Brunt, K. Kim, R. S. Fulton, L. A. Fulton, S. W. Clifton, R. K. Wilson, R. D. Knight, J. I. Gordon (2007). Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biology* 5:e156.
- Yamada, M., H. Zhang, S. Hanada, K. V. P. Nagashima, K. Shimada, K. Matsuura (2005). Structural and spectroscopic properties of a reaction center complex from the chlorosome-lacking filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. *Journal of Bacteriology* 187:1702–1709.
- Yamada, T., H. Imachi, A. Ohashi, H. Harada, S. Hanada, Y. Kamagata, Y. Sekiguchi (2007). *Bellilinea caldifistulae* gen. nov., sp. nov. and *Longilinea arvoryzae* gen. nov., sp. nov., strictly anaerobic, filamentous bacteria of the phylum chloroflexi isolated from methanogenic propionate-degrading consortia. *International Journal of Systematic and Evolutionary Microbiology* 57:2299–306.
- Yamada, T., Y. Sekiguchi, S. Hanada, H. Imachi, A. Ohashi, H. Harada, Y. Kamagata (2006). *Anaerolinea thermolimos* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae* classis nov. and *Caldilineae* classis nov. in the bacterial phylum chloroflexi. *International Journal of Evolutionary Microbiology* 56:1331–1340.
- Yanyushin, M. F., M. C. del Rosario, D. C. Brune, R. E. Blankenship (2005). New class of bacterial membrane oxidoreductases. *Biochemistry* 44:10037–45.
- Youvan, D. C., E. J. Bylina, M. Alberti, H. Begusch, J. E. Hearst (1984). Nucleotide and deduced polypeptide sequences of the photosynthetic reaction-center, B870 antenna, and flanking polypeptides from *R. capsulata*. *Cell* 37:949–957.

- Zarzycki, J., V. Brecht, M. Müller, G. Fuchs (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO₂ fixation cycle in *Chloroflexus aurantiacus*. *Proceedings of the National Academy of Sciences* 106:21317–21322.
- Zarzycki, J., G. Fuchs (2011). Co-Assimilation of organic substrates via the autotrophic 3-hydroxypropionate bi-cycle in *Chloroflexus aurantiacus*. *Applied and Environmental Microbiology* .
- Zeikus, J. G., A. Ben-Bassat, P. W. Hegge (1980). Microbiology of methanogenesis in thermal, volcanic environments. *Journal of Bacteriology* 143:432–440.
- Zeikus, J. G., M. A. Dawson, T. E. Thompson, K. Ingvorsen, E. C. Hatchikian (1983). Microbial ecology of volcanic sulphidogenesis: isolation and characterization of *Thermodesulfobacterium commune* gen. nov. and sp. nov. *J Gen Microbiol* 129:1159–1169.
- Zeikus, J. G., P. W. Hegge, M. A. Anderson (1979). *Thermoanaerobium brockii* gen. nov. and sp. nov., a new chemoorganotrophic, caldoactive, anaerobic bacterium. *Archives of Microbiology* 122:41–48.
- Zeikus, J. G., R. S. Wolfe (1972). *Methanobacterium thermoautotrophicus* sp. n., an anaerobic, autotrophic, extreme thermophile. *Journal of Bacteriology* 109:707–713.
- Zhaxybayeva, O., W. F. Doolittle, R. T. Papke, J. P. Gogarten (2009). Intertwined evolutionary histories of marine *Synechococcus* and *Prochlorococcus marinus*. *Genome Biol Evol* 2009:325–339.
- Zhaxybayeva, O., J. P. Gogarten, R. L. Charlebois, W. F. Doolittle, R. T. Papke (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Research* 16:1099–1108.