MICROBIAL COMMUNITY DYNAMICS AND CARBON FIXATION IN DARK OLIGOTROPHIC VOLCANIC ECOSYSTEMS

By

Richard Eric Davis

A DISSERTATION

Presented to the Division of Environmental & Biomolecular Systems
and the Oregon Health & Science University
School of Medicine
In partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
in
Environmental Science and Engineering

June 2014

Oregon Health & Science University

Institute of Environmental Health

CERTIFICATE OF APPROVAL

This is to certify that the Ph.D. dissertation of

Richard E. Davis

has been approved

Bradley M. Tebo, Ph.D.
Research Advisor

Margo G. Haygood, Ph.D.
Examination Committee Chair

Holly M. Simon, Ph.D.
Examination Committee Member

Craig Moyer, Ph.D.
Western Washington University
External Committee Member

Dedication

For Leslie

Table of Contents

List of Tables

List of Figures

Acknowledgements

Many people have impacted my research in my time at OHSU and helped in my development as a scientist. I would like to begin by thanking by PhD advisor and mentor Brad Tebo who gave me an opportunity to enter his lab. Brad always gave me the freedom to develop my own research project and allowed me to challenge myself to learn new methods to answer questions that I had, but not the ability to answer. I was allowed to travel extensively during my graduate program, with field studies that include two trips to the South Pacific, four trips to Hawaii, and one memorable trip to Antarctica. I was also allowed to attend multiple professional meetings to present my research, including international meetings in Cairns, Australia and Okinawa, Japan. More than any other attribute, I appreciate Brad's patience the most. He never rushed me to get results or made me feel like a failure when research wasn't going my way in the lab. I feel that his patient confidence helped me get through some difficult times in the lab, and aided me in making decisions for my research based on doing things the “right way” rather than rushing results I am not confident in to simply get publications. I hope I am able to show this kind of patience when I advise students of my own.

I also must thank Craig Moyer, whose enthusiasm brought me into the world of microbial ecology. I joined Craig's lab as an undergraduate at Western Washington University where I stayed and earned my Master's degree in his lab. Craig taught me almost everything I know about working in a lab along with how to work in the field, I even met my wife in his lab. We have remained colleagues during my time at OHSU and I am honored to call him a friend.

Margo Haygood always gave me great advise on my research throughout my time at OHSU and gave critical edits and encouragement as the chair of my defense committee. Holly Simon also provided advise and edits on my thesis, both as a member of my defense committee and during my entire graduate program.

Hubert Staudigel has always helped me understand the environments I worked in from a geological perspective. He was instrumental in our research at Antarctica, both conceiving of the research and guiding us to the samples. Greg Dick and Sunit Jain provided critical guidance in metagenomics and gave me remote access to their hi-memory server to learn how to assemble the Loihi metagenome. Katrin Kiesslich interned for a year in our lab and taught me how to extract RNA from difficult samples. My summer REU interns Michael Guzman and Rachel Tullsen worked on samples from Antarctica. Michael provided critical first analyses of the Warren Cave metagenome despite my mind being on my newborn son, and his hard work and dedication was an inspiration to me to always do my best and learn new skills.

Carolyn Sheehan was more than a lab manager during my time in the Tebo lab. She always made sacrifices to make my research happen, including the ultimate sacrifice of driving me around Hawaii for four years. Many trips were made for buckets, rope, and coolers around Hawaii and she almost always did it with a smile. She was one of the fortunate three to experience every day of Femo and I'm not sure what the cruises would have been like without her staying up all night to archive samples and pack boxes for Fedex. I would also like to thank Suzanna Bräuer, Kati Geszvain, Sung-Woo Lee, and Roberto Anitori for helping me in the lab and teaching me new research tricks along with

their friendship during my time in the lab. Nancy Christie always made graduate school problems go away and always gave me great advice when I needed it. I think our entire institute would crumble without her. I would also like to thank Jason Isbell and Lucinda Williams for providing the soundtrack for the last two chapters of this thesis.

Finally, I would like to thank my family for all of their help and support during my time at OHSU. My mother, father, sister, and brothers were always supportive of me during this time. My wife, Leslie supported me and listened to my endless complaints about failed experiments, despite my degree taking way too long. Her sacrifice during this time to allow me to realize my dream is the second greatest gift I could receive. My greatest gift is my son Richie, who despite not allowing me to sleep for the last two years, is the love of my life and my greatest source of pride.

Short read sequencing assays were performed by the OHSU Massively Parallel Sequencing Shared Resource. This work was supported by the Oregon Opportunity Fund, and NSF grants MCB-0348668/0742010 (Molecular and Cellular Biosciences), OCE-1129553 (Ocean Sciences), and OPP-0739731 (Office of Polar Programs).

Abstract

This research characterizes microbial dynamics in dark, oligotrophic volcanic environments. The objective of this thesis is to provide new understanding to community succession at hydrothermal vents, what mechanisms are used for primary productivity in dark, oligotrophic environments, and what are the metabolic strategies of the primary producing populations in these communities. As introduced in Chapter 1, dark oligotrophic volcanic systems are ubiquitous, understudied ecosystems which have a tremendous influence on the biogeochemical cycles on earth. Community succession of primary producing bacteria in these ecosystems is shown in Chapter 2 using a long-term ecological study of microbial mat communities from Loihi Seamount which have been disturbed by an eruptive event. This microbial community succession closely mirrors that of models of hydrothermal fluid evolution following an eruptive event. Chapter 3 explores the great explosion of diversity information from genomic data of autotrophic isolates and presents a review of known carbon fixation pathways along with the phylogenetic diversity of key enzymes used in each pathway. Chapter 4 details the carbon fixation mechanisms and biodiversity of primary producers at multiple microbial mat communities at Loihi Seamount using both PCR-based and metagenome-based analyses. This study also proposes a novel syntrophic model consisting of hydrogen production from anaerobic iron oxidizing bacteria feeding methanogenic archaea and sulfur/iron reducing bacteria. Chapter 5 explores the microbial diversity and metabolic potential of a fumarolic sediment community from Warren Cave, Antarctica using metagenomic data. This chapter shows these communities are supported primary production via the Calvin

Cycle which is fueled by high affinity hydrogenases and/or carbon monoxide dehydrogenase enzymes. Chapter 6 discusses some general considerations about sample collection, processing, and inherent biases of the methods used in this thesis. This is followed by a general summary and future directions for research for each chapter. Taken together, this thesis presents new understanding of mechanisms for community succession, carbon fixation, and metabolic strategies of microbes living in dark, oligotrophic volcanic systems.

Chapter 1: Introduction

Microbial life in dark, nutrient-poor ecosystems is poorly understood with respect to community composition, population dynamics, metabolic mechanisms, and overall contribution to nutrient cycling on Earth. Microbial metabolism and growth in typical oligotrophic systems such as the hadal zone in the deep sea or the deep subsurface is highly dependent on scavenging of rare, recalcitrant organic carbon for growth (Morita, 1997). This lifestyle has adapted many of these organisms to persist in a starvation-survival lifestyle in which metabolic processes occur at a very slow rate and doubling times can be measured in years or even decades (Jörgensen, 2011).

When organic carbon is limiting to microbial communities and a source of reduced redox-active chemicals are available, chemoautotrophic organisms which organisms obtain their energy from inorganic energy sources and use that energy to fix carbon in the form of CO_2 (or other one-carbon compounds) become the dominant members of the communities. In general, such chemoautotrophic-based ecosystems have been less well studied than ecosystems fueled by organic matter. The objectives of this thesis were to provide new understanding to: 1) Community succession at hydrothermal vents; 2) the mechanisms used for primary production in dark, oligotrophic environments; and 3) the metabolic strategies of the primary producing populations in these communities. These objectives are addressed using a combination of PCR-based ecological tools, metagenomic analysis, and bioinformatics on samples from deep-sea hydrothermal vent communities from Loihi Seamount and from fumarolic ice caves on Mt. Erebus, Antarctica.

New interest in deep-sea and subsurface microbial communities came with the

revelation that the combined deep subsurface could harbor one third of all life on Earth (Whitman et al., 1998). This estimation was based on an extrapolation of cell counts from known drill cores at the time. Unfortunately these drill sites were near the continents where subsurface biomass was high due to greater productivity in the surface waters dropping more detritus to the sediment where it is buried. When new cores were drilled in the South Pacific Gyre, cell counts were much lower than they were near the Peru Margin where the original cell counts were made. This new data revealed that the original census for the deep biosphere was overestimated by about an order of magnitude (Kallmeyer et al., 2012). While the biomass estimate decreased dramatically, the dark subsurface still represents a significant portion of the earth's biomass and is an important component of global nutrient cycling (Kallmeyer et al., 2012; Meyers et al., 2014; Roden, 2012).

The deep ocean floor is typically a cold, low biomass ecosystem. When hydrothermal vents were discovered in 1978, scientists were surprised at the high-biomass communities found near these hydrothermal systems (Corliss et al., 1979). Thick microbial mats covered the vents and large populations of animals covered the vent field. It was later discovered that these high biomass systems were being driven by primary production from chemoautotrophic carbon fixation (Jannasch and Mottl, 1985; Jannasch and Wirsen, 1979). The microbial communities were found to be oxidizing sulfide (HS^-) and using the resultant energy to fix carbon. Many of the hydrothermal animals, such as shrimp and segmented worms, were directly feeding on these microbial mats as well as hosting epibiotic bacteria which are grown and used as a food source. Other vent animals, such as Siboglinidae tube worms, were found to harbor large populations of chemoautotophic endosymbiotic bacteria within specialized cells which fix CO_2 and

provide organic carbon to the host (Cavanaugh et al., 1981). These high-biomass sulfur-based ecosystems became the model systems for hydrothermal research, and the majority of research at hydrothermal vents was performed at these systems.

In the late 1980's, a new type of hydrothermal vent was described. This vent system was found off of the coast of the big island of Hawaii and appeared to be covered in thick rust, and did not harbor large communities of animals or thick, white microbial mats (Karl et al., 1988). When the vent fluid chemistry was analyzed, high concentrations of reduced iron were found, along with elevated amounts of dissolved CO_2. When the rusty sediment was examined under a microscope the iron oxides appeared to have a similar structure to biogenic iron oxides that have been found in freshwater systems, where these stalk-like structures are formed by the neutraphilic iron-oxidizing chemoautroph *Gallionella ferruginea* (Karl et al., 1988).

Culture-independent studies of these hydrothermal systems in 1995 revealed they were dominated by Epsilonproteobacteria closely related to sulfur oxidizers, along with lesser amounts of Gammaproteobacteria (Moyer et al., 1995). One phylotype, termed PVB OTU 4 appeared to be a deeply-rooted Gammaproteobacteria, however it did not cluster closely with any known sequence at the time. The study found that there were some sulfur-oxidizing bacteria in the community, however no phylotypes were found that were closely related to any known iron-oxidizing bacteria.

A chemoautotrophic stalk-forming iron-oxidizing bacterium was later isolated from a hydrothermal vent at Loihi Seamount (Emerson and Moyer, 1997). This isolate was an obligate chemoautotroph whose only energy source was reduced Fe(II) which it oxidized to produce the energy needed to fix CO_2 for biomass. This isolate was closely

related to PVB OTU 4 and was later fully described and placed into its own class, the *Zetaproteobacteria*, based on genomic sequence analysis (Emerson et al., 2007). Later cell-independent PCR-based analyses of iron-dominated hydrothermal systems at Loihi and other sites in the Pacific has shown this phylotype to be ubiquitous in these communities and is probably the dominant iron-oxidizing group in brine-dominated deep-sea hydrothermal vents (McAllister et al., 2011).

One aspect of hydrothermal microbial ecology that is poorly understood is community succession over the "lifespan" of the hydrothermal vent. Hydrothermal vents are ephemeral in nature and are usually formed by increased activity of the volcanic system or through new diking events diverting the hydrothermal flow to new areas (Rubin et al., 2012; Figure 1-1). When hydrothermal vents are formed by increased volcanic activity, magmatic sources are closer to the surface of the seafloor and hydrothermal vent chemistry reflects this with a greater magmatic gas input in the form of molecular hydrogen, hydrogen sulfide or sulfite, and dramatically increased ^{3}He concentration (Butterfield et al., 1997; Cheminée et al., 1991). Studies of hydrothermal fluids have shown elevated concentrations of volatile gas components at several actively erupting volcanic systems (Butterfield et al., 2011; Lupton, 1996; Lupton et al., 2006). These systems often have extremely high concentrations of hydrogen, along with sulfur components such as hydrogen sulfide or sulfite. As the depth of the magma decreases and the hydrothermal vent begins to cool, the hydrothermal fluid transitions to brine-based fluid characterized by high concentrations of iron and other chemical species formed through hot water-rock interactions and seawater mixing below the subsurface (Butterfield et al., 1997). This periodic eruptive cycle at active volcanic sites has been

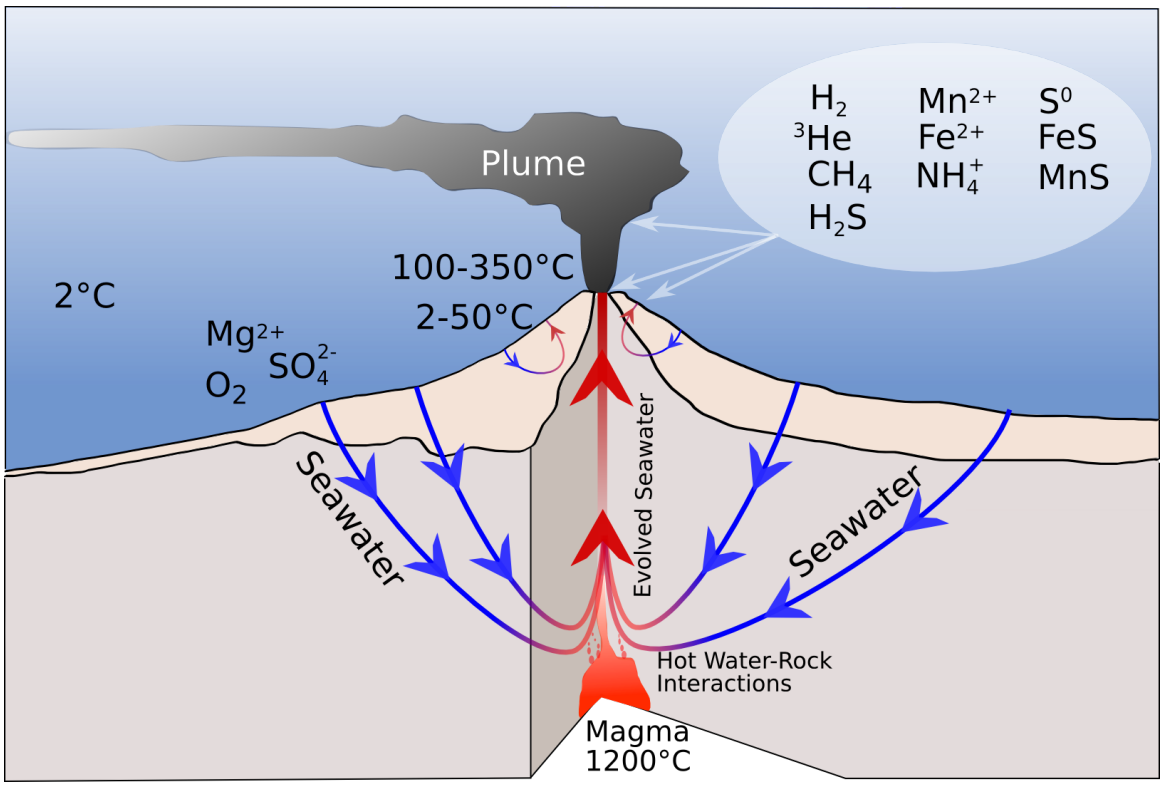

Figure 1-1. Conceptual cross section of a hydrothermal vent. Ocean seawater is heated by magma causing the chemical composition to change through water-rock interactions and inputs of magmatic volatiles. The fluid then rises and escapes the volcano forming hydrothermal vents.

described by Butterfield et al (1997) leading to a general model of volcanic hydrothermal systems consisting of periodic eruptive events increasing the heat flow and volatile gas concentration termed the "vapor phase", followed by a slow transition to cooler vent fluids with chemistry mostly influenced by water-rock interactions and subsurface mixing with seawater, termed the "brine phase" (Figure 1-2). This transition has been observed with vent chemistry data, however the ecological impacts on the hydrothermal vent communities has never been shown.

Chapter two in this thesis is an ecological study of hydrothermal vent communities from Loihi Seamount where the community was sampled before an eruption, shortly after the eruption occurred, and then for 12 years after the eruption. This long-term ecological study is the first to show a hydrothermal system to be disturbed by an eruption and to later form communities that are similar to the pre-eruption communities. This community succession closely parallels the vent fluid composition at the hydrothermal vents over the course of years, which also follows the Butterfield model of hydrothermal fluid evolution after an eruptive event (Butterfield et al., 1997).

Primary production in the oxic photic zone of the earth is dominated by photoautotrophic organisms, almost all of which utilize the Calvin Cycle to fix carbon (Tabita, 1999). Many organism at hydrothermal vents also utilize the Calvin Cycle, especially communities growing in oxic conditions. Organisms using carbon fixation pathways other than the Calvin Cycle at hydrothermal vents have been known for some "rare" species, such as methanogens and hyperthermophilic hydrogen oxidizers, however the discovery that Epsilonproteobacateria utilize the reductive TCA cycle to fix carbon has led to the discovery that many other hydrothermal vent populations utilize carbon

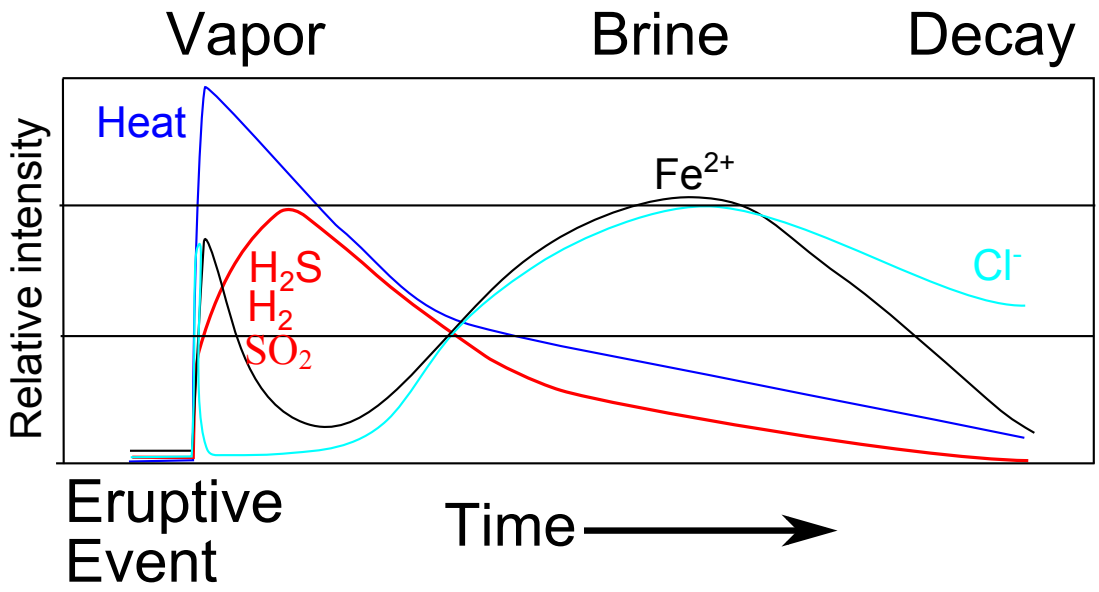

Figure 1-2. The Butterfield model of a response of hydrothermal systems to a volcanic event. As heat (blue line) increases after an event, volatile gas components increase (red line) and the vent is considered in vapor phase. As heat decreases, gas concentration also decreases and reduced iron (black line) and chloride (teal line) increase due to water-rock interactions and subsurface seawater mixing and the vent is considered in the brine phase. As heat continues to decrease, so does the concentration of reduced iron, marking the decay phase of the vent. Figure is modified from Butterfield et al. 1997.

fixation strategies other than the Calvin Cycle (Campbell and Cary, 2004; Nakagawa and Takai, 2008). Each of the carbon fixation pathways have different oxygen and thermal tolerances as well as different energetic "costs" associated with fixing carbon (Hügler and Sievert, 2011). Since carbon fixation is an obligate metabolism in these organisms, these properties will have a dramatic affect on what niche these organisms can exploit in the environment.

Rapid advances in genome sequencing technology has facilitated the sequencing of hundreds of chemoautotrophic isolates in recent years. Many of these isolates were known to be chemoautotrophic, however the metabolic pathway for the isolate was unknown or undiscovered (Bazylinski et al., 2004; Huber et al., 2008; Könneke et al., 2005). Genome sequencing has not only shown what carbon fixation pathways are being used in these isolates, but also has aided in the discovery of new, previously unknown carbon fixation pathways (Berg et al., 2007; Huber et al., 2008). Chapter three in this thesis investigates how these isolates are fixing carbon, and shows the phylogeny of key genes from each of the five known carbon fixation pathways. This chapter shows not only new diversity in known pathways such as the Calvin Cycle, but also shows the phylogeny for key carbon fixation genes which has not been shown before.

Chapter three explores the diversity of carbon fixation genes in isolated cultures with sequenced genomes. While extremely valuable, this information is also highly biased in an ecological context. Culture independent PCR primers for Rubisco, the key gene in the Calvin Cycle, have been designed and used in soil communities (Selesi et al., 2005), but primers designed for the other carbon fixation pathways have either not been designed or the PCR products are extremely short (Campbell et al., 2003). Chapter four

investigates carbon fixation in microbial mat communities at Lohi Seamount using culture-independent techniques. New PCR primers were designed for both the form I and form II Rubisco genes and the beta-subunit of ATP-citrate lyse gene using new sequences as targets to improve both the breadth of targets for each primer pair and producing an amplicon with sufficient length for phylogenetic inference.

Environmental genomics has the potential to revolutionize our knowledge about uncultured organisms (Simon and Daniel, 2011). This method sequences total sheared DNA from a community of uncultured organisms. These sequences can then be assembled into large genomic fragments and clustered into "genomic bins" which usually contains partial genomes from multiple closely-related organisms (Hess et al., 2011). These bins can then be phylogenically characterized using essential housekeeping genes and inferences can be made from the metabolic genes found in each bin.

Chapter four utilizes a metagenomic library to find carbon fixation genes from pathways without using PCR primers and also identify phylotypes from the bin for which the genes are found. This allows us to explore pathways for carbon fixation, and to characterize the genomic phylotypes from which these genes are found. This provides information of not only what carbon fixation pathways are present in the community, but who could be using them and what other metabolic potential each carbon fixing bin is capable of using. Only a single ecologically relevant isolate (Emerson et al., 2007) has ever been isolated and sequenced from a iron-based microbial mat community, with all other information about these communities coming from PCR-based analyses. The genetic information from this metagenome should provide critical insight into the biochemical mechanisms and metabolic potential in these communities.

Deep-sea hydrothermal vents are not the only dark, oligotrophic volcanic ecosystem. Terrestrial volcanic systems may also support large communities of carbon fixing communities through fumarolic venting of gases (Staudigel et al., 2011; Zimbelman et al., 2000). These communities may also be metabolizing reduced metals within the sediments for energy or trace nutrients, such as iron and manganese, causing weathering in the rocks as seen in deep-sea basalts (Staudigel et al., 2008). High altitude thermal environments are difficult to study due to the danger and inaccessibility to samples. The communities are also often contaminated by photosynteticlly-derived organic carbon which is deposited by wind, melting from the surface, or through groundwater flow (Zimbelman et al., 2000). This organic material could mask the autotrophic communities from analysis due to cryopreservation of genetic material or by feeding heterotrophic microbes.

Chapter five investigates a dark, oligotrophic volcanic community from Warren Cave, a fumarolic ice cave near the summit of Mt. Erebus, Antarctica. Mt. Erebus ice caves are formed by worm fumarolic gases melting passageways under permanent ice fields (Giggenbach, 1976). These caves are completely sealed off from each other and positive airflow of gases out of the caves prevents outside organic material from entering the caves (Curtis and Kyle, 2011). There are no animals or plants living near the caves and human impact is limited to seasonal scientists trained in low-impact sampling. These factors together make a study site that is both contained and pristine. A metagenomic library was constructed from Warren Cave sediments sampled from an area near a fumarole vent. This chapter investigates not only the carbon fixation pathways utilized in this community, and the energy and nutrient metabolic potential as well.

This thesis presents the first long-term ecological study of a bacterial community at a hydrothermal vent that has undergone an eruptive event. It shows the transition from a stable brine based community to a community based on the metabolism of volatile gases that comprise a vapor-based eruptive system, and then the transition back to a brine-based community of iron oxidizing bacteria. This thesis also explores the great explosion of diversity information from genomic data of autotrophic isolates and also in uncultured microbes from the environment, including analysis of the first two metagenomic libraries from dark, volcanic systems with sufficient sequencing for assembly of large genomic fragments.

This thesis provides new insights in community succession at hydrothermal vents at Loihi Seamount using a long-term sampling strategy spanning fifteen years, with twelve of those years directly following a significant volcanic eruption (Chapter 2). This thesis also reviews the currently known mechanisms of carbon fixation and diversity of organisms which use each of the pathways (Chapter 3), and then determines what mechanisms are used for carbon fixation at Loihi Seamount hydrothermal vents (Chapter 4) and Warren Cave fumarolic sediments (Chapter 5). Finally, this thesis explores the metabolic potential of primary producing populations in a microbial mat from Loihi Seamount (Chapter 4) and Warren Cave (Chapter 5). These studies all provide novel information about the community dynamics, metabolic potential, and carbon fixation mechanisms in these understudied environments.

Works Cited

Bazylinski, D.A., Dean, A.J., Williams, T.J., Long, L.K., Middleton, S.L., and Dubbels, B.L. (2004). Chemolithoautotrophy in the marine, magnetotactic bacterial strains MV-1 and MV-2. Arch. Microbiol. *182*, 373–387.

Berg, I.A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007). A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. Science *318*, 1782–1786.

Butterfield, D.A., Jonasson, I.R., Massoth, G.J., Feely, R.A., Roe, K.K., Embley, R.E., Holden, J.F., McDuff, R.E., Lilley, M.D., and Delaney, J.R. (1997). Seafloor eruptions and evolution of hydrothermal fluid chemistry. Phil Trans R Soc Lond A *355*, 369–386.

Butterfield, D.A., Nakamura, K., Takano, B., Lilley, M.D., Lupton, J.E., Resing, J.A., and Roe, K.K. (2011). High SO2 flux, sulfur accumulation, and gas fractionation at an erupting submarine volcano. Geology *39*, 803–806.

Campbell, B.J., and Cary, S.C. (2004). Abundance of Reverse Tricarboxylic Acid Cycle Genes in Free-Living Microorganisms at Deep-Sea Hydrothermal Vents. Appl Env. Microbiol *70*, 6282–6289.

Campbell, B.J., Stein, J.L., and Cary, S.C. (2003). Evidence of Chemolithoautotrophy in the Bacterial Community Associated with *Alvinella pompejana*, a Hydrothermal Vent Polychaete. Appl Env. Microbiol *69*, 5070–5078.

Cavanaugh, C.M., Gardiner, S.L., Jones, M.L., Jannasch, H.W., and Waterbury, J.B. (1981). Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila* Jones: possible chemoautotrophic symbionts. Science *213*, 340–342.

Cheminée, J.-L., Stoffers, P., McMurtry, G., Richnow, H., Puteanus, D., and Sedwick, P. (1991). Gas-rich submarine exhalations during the 1989 eruption of Macdonald Seamount. Earth Planet. Sci. Lett. *107*, 318–327.

Corliss, J.B., Dymond, J., Gordon, L.I., Edmond, J.M., von Herzen, R.P., Ballard, R.D., Green, K., Williams, D., Bainbridge, A., Crane, K., et al. (1979). Submarine Thermal Sprirngs on the Galapagos Rift. Science *203*, 1073–1083.

Curtis, A., and Kyle, P. (2011). Geothermal point sources identified in a fumarolic ice cave on Erebus volcano, Antarctica using fiber optic distributed temperature sensing. Geophys. Res. Lett. doi:/10.1029/2011GL048272

Emerson, D., and Moyer, C.L. (1997). Isolation and characterization of novel iron-oxidizing bacteria that grow at circumneutral pH. Appl. Environ. Microbiol. *63*, 4784–4792.

Emerson, D., Rentz, J.A., Lilburn, T.G., Davis, R.E., Aldrich, H., Chan, C., and Moyer, C.L. (2007). A novel lineage of Proteobacteria involved in formation of marine Fe-Oxidizing microbial mat communities. PLoS ONE *2*, e667.

Giggenbach, W. (1976). Geothermal ice caves on Mt Erebus, Ross Island, Antarctica. N. Z. J. Geol. Geophys. *19*, 365–372.

Hess, M., Sczyrba, A., Egan, R., Kim, T.-W., Chokhawala, H., Schroth, G., Luo, S., Clark, D.S., Chen, F., Zhang, T., et al. (2011). Metagenomic Discovery of Biomass-Degrading Genes and Genomes from Cow Rumen. Science *331*, 463 –467.

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I.A., Kockelkorn, D., Eisenreich, W., and Fuchs, G. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic *Archaeum Ignicoccus* hospitalis. Proc. Natl. Acad. Sci. *105*, 7851–7856.

Hügler, M., and Sievert, S.M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Annu. Rev. Mar. Sci. *3*, 261–289.

Jannasch, H.W., and Mottl, M.J. (1985). Geomicrobiology of deep-sea hydrothermal vents. Science *229*, 717–725.

Jannasch, H.W., and Wirsen, C.O. (1979). Chemosynthetic primary production at East Pacific sea floor spreading centers. Bioscience *29*, 592–598.

Jörgensen, B.B. (2011). Deep subseafloor microbial cells on physiological standby. Proc. Natl. Acad. Sci. U. S. A. *108*, 18193–18194.

Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. Proc. Natl. Acad. Sci. *109*, 16213–16216.

Karl, D.M., McMurtry, G.M., Malahoff, A., and Garcia, M.O. (1988). Loihi Seamount, Hawaii: a mid-plate volcano with a distinctive hydrothermal system. Nature *335*, 532–535.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature *437*, 543–546.

Lupton, J.E. (1996). A Far-Field Hydrothermal Plume from Loihi Seamount. Science *272*, 976–979.

Lupton, J., Butterfield, D., Lilley, M., Evans, L., Nakamura, K., Chadwick, W., Resing, J., Embley, R., Olson, E., and Proskurowski, G. (2006). Submarine venting of liquid carbon dioxide on a Mariana Arc volcano. Geochem Geophys Geosyst 7:doi:10.1029/2005GC001152.

McAllister, S.M., Davis, R.E., McBeth, J.M., Tebo, B.M., Emerson, D., and Moyer, C.L. (2011). Biodiversity and Emerging Biogeography of the Neutrophilic Iron-Oxidizing Zetaproteobacteria. Appl Env. Microbiol *77*, 5445–5457.

Meyers, M.E.J., Sylvan, J.B., and Edwards, K.J. (2014). Extracellular enzyme activity and microbial diversity measured on seafloor exposed basalts from Loihi Seamount indicate importance of basalts to global biogeochemical cycling. Appl. Environ. Microbiol. AEM–01038.

Morita, R.Y. (1997). Bacteria in oligotrophic environments. Chapman & Hall. 529p.

Moyer, C.L., Dobbs, F.C., and Karl, D.M. (1995). Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. *61*, 1555–1562.

Nakagawa, S., and Takai, K. (2008). Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. FEMS Microbiol. Ecol. *65*, 1–14.

Roden, E. (2012). Microbial iron-redox cycling in subsurface environments. Biochem. Soc. Trans. *40*, 1249–1256.

Rubin, K.H., Soule, S.A., Chadwick, W.W., Fornari, D.J., Clague, D.A., Embley, R.W., Baker, E.T., Perfit, M.R., Caress, D.W., and Dziak, R.P. (2012). Volcanic eruptions in the deep sea. Oceanography *25*, 142-157.

Selesi, D., Schmid, M., and Hartmann, A. (2005). Diversity of Green-Like and Red-Like Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes (cbbL) in Differently Managed Agricultural Soils. Appl. Environ. Microbiol. *71*, 175–184.

Simon, C., and Daniel, R. (2011). Metagenomic Analyses: Past and Future Trends. Appl Env. Microbiol *77*, 1153–1161.

Staudigel, H., Furnes, H., McLoughlin, N., Banerjee, N.R., Connell, L.B., and Templeton, A. (2008). 3.5 billion years of glass bioalteration: Volcanic rocks as a basis for microbial life? Earth-Sci. Rev. *89*, 156–176.

Staudigel, H., Anitori, R., Davis, R., Connell, L., and Tebo, B. (2011). Dark Oligotrophic Volcanic Ecosystems (DOVEs) in Fumarolic Ice Caves of Mt. Erebus Volcano. AGU Fall Meet. Abstr. *1*, 08.

Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynth. Res. *60*, 1–28.

Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998). Prokaryotes: The Unseen Majority. Proc. Natl. Acad. Sci. *95*, 6578–6583.

Zimbelman, D.R., Rye, R.O., and Landis, G.P. (2000). Fumaroles in ice caves on the summit of Mount Rainier—preliminary stable isotope, gas, and geochemical studies. J. Volcanol. Geotherm. Res. *97*, 457–473.

Chapter 2: Spatial and Temporal Variability of Microbial Communities from Pre- and Post-Eruption Microbial Mats Collected from Loihi Seamount, Hawaii.

Introduction

Deep-sea hydrothermal vent fields are characterized by strong physical and chemical gradients that support biological communities based on chemoautotrophic primary production. Since the discovery of hydrothermal vents in 1977 (Corliss et al., 1979), high-biomass chemoautotrophic microbial communities based on iron, sulfur, hydrogen, and methane have been discovered (Nakagawa and Takai, 2008). While these metabolisms are not mutually exclusive within a community, hydrothermal communities are often characterized by the dominance of either brine-based chemoautotrophic production such as iron oxidation (Davis et al., 2009; Juniper and Fouquet, 1988; Moyer et al., 1995) or by the presence of volatile or gas-based metabolisms based on sulfur, hydrogen, or methane (Campbell et al., 2006; Takai et al., 2004; Toki et al., 2008).

The presence and concentration of reduced chemicals in hydrothermal fluids are influenced by many abiotic factors, but is primarily formed by a complex interaction between volatile magmatic input, hot water-rock interactions, and phase separation (Seyfried et al., 1988; Wheat and Mottl, 2000). Vent fluid chemistry is also shaped by the magmatic depth, the chemical composition and age of the crust, the hydrothermal flow rate and depth of venting, and the degree of subsurface seawater mixing (Butterfield et al., 1997). Biotic factors can also influence the evolution of hydrothermal fluids, especially when fluids are enriched in hydrogen and carbon dioxide allowing methanogenesis (Takai et al., 2004).

The recent discovery of actively erupting submarine volcanoes and long-term monitoring has given insight into the formation and evolution of hydrothermal vent fields

(Embley et al., 2006; Resing et al., 2011). Hydrothermal vents are generally ephemeral in nature, forming during periods of decreased magma depth during eruptive events with activity generally decreasing over time because of reduced heat flow or subsurface conduit changes due to seismic activity and diking events (Delaney et al., 1998). High quality sampling of hydrothermal fluids at actively erupting submarine volcanoes shows high concentrations of volcanic volatiles such as hydrogen gas and carbon dioxide sourced from magmatic degassing along with mixed sulfuric acid and elemental sulfur rather than the more common hydrogen sulfide found in non-erupting sulfur-rich hydrothermal vents (Embley et al., 2006; Lupton et al., 2008; Resing et al., 2007). Long-term monitoring of hydrothermal fluids from the CoAxial vent site on the Juan de Fuca Ridge for several years after an eruptive event showed the decrease of heat and hydrogen sulfide coincided with an increase in reduced iron and chloride concentrations before an eventual decay in all hydrothermal activity after several years (Butterfield et al., 1997).

Loihi Seamount is an active submarine volcano approximately 20 miles off of the southeast shore of the big island of Hawaii. Hydrothermal activity was first observed near the summit in 1987 by scientists using the manned submersible *Alvin* (Karl et al., 1988). Diffuse hydrothermal vents (T_{max}=30°C) covered with iron oxide-encrusted microbial mats were discovered near the summit of the volcano, termed Pisces Peak. Microscopic analysis of these microbial mats showed they were composed of biogenic iron oxides along with some elemental sulfur particles (Karl et al., 1988). Small subunit ribosomal RNA (SSU rRNA) community analysis showed these mats were dominated by *Epsilonproteobacteria* and *Gammaproteobacteria* (Moyer et al., 1995). Loihi experienced an earthquake swarm in 1996 that was greater than any ever observed on a

Hawiian Volcano, and chemical analysis directly after the swarm indicated an eruption occurred (Duennebier et al., 1997). Bathymetric maps made after the swarm showed the former summit had collapsed to form a new pit crater subsequently named Pele's Pit. Hydrothermal vents discovered in Pele's Pit had temperatures in excess of 200°C and were enriched in iron, manganese, sulfur, hydrogen, and methane (Wheat et al., 2000). Diffuse vents were also discovered on the south rift of Loihi with hydrothermal chemistry similar to those found pre-eruption (Malahoff et al., 2006). Since this time, hydrothermal vents in Pele's Pit have continued to have activity, although the heat flux has decreased (T_{max}=60°C) and volatile chemical species are no longer found in high concentrations (Glazer and Rouxel, 2009).

Since the biological communities at hydrothermal vents are intrinsically coupled to the chemical and physical properties of a hydrothermal vent field, the community structure should reflect the hydrothermal vent decay over time. We collected microbial mat samples at Loihi from 1993 to 2009 and preserved them at -80°C. We hypothesize these microbial mat communities will reflect the shift predicted by Butterfield et al. (1997) from chemoautotrophic microbes utilizing brine-associated electron donors such as iron and manganese to metabolisms using volatile chemicals such as hydrogen sulfide and hydrogen gas. As the vents have returned to near pre-eruptive chemical and physical conditions, the microbial mat communities should begin to resemble the communities found in pre-eruption samples. In work described in this chapter, I utilized clone library, TRFLP molecular fingerprinting, and quatitative PCR analysis on representative microbial mat samples collected pre- and post-eruption to monitor the community succession after this eruption.

Materials and Methods

Sample collection

Microbial mat samples were obtained using either the manned submersible *Pisces V* (1993-2004) or the unmanned ROV *Jason II* (2006-2009). Microbial mat material was sampled using either a suction sampler or a resealable PVC scoop. All samples were immediately frozen and maintained at -80°C until DNA extraction was performed on shore.

DNA extraction

Total genomic DNA (gDNA) was extracted from each sample in duplicate using the FastDNA Spin Kit for Soil following the manufacturer's protocol (Qbiogene, Irvine, CA). Extracted gDNA from each sample was pooled, cleaned, and concentrated using Montáge PCR centrifugal filter devices (Millipore, Bedford, MA). The gDNAs were then quantified using a Nanodrop ND-1000 spectrophotometer and were diluted to 10 ng DNA/mL using filter sterilized 10 mM Tris 0.1 mM EDTA (pH 8.0).

T-RFLP preparation

Three replicate PCRs were performed, each using 50 ng of total gDNA and domain specific primers 68F (TNANACATGCAAGTCGRRCG) and 1492R (RGYTACCTTGTTACGACTT), where R is purine analog K, Y is pyrimidine analog P, and N is an equal mixture of both analogs at a single position (Glen Research, Sterling,

VA.). PCR conditions were as previously described (Emerson et al. 2002). The forward primers were labeled with 6-FAM (6-carboxyfluorescein) on the 5' end. The PCR products were visually assayed for size by 1% agarose gel electrophoresis against a 1-kb ladder DNA size standard. Only reaction mixtures yielding the expected DNA fragment size with corresponding no amplification of negative controls were used. The remaining fluorescently labeled PCR products were desalted using Montáge PCR centrifugal filter device (Millipore). The PCR products (15 mL) were then partitioned into eight aliquots and separately digested overnight with 5 U of *Hae* III, *Hha* I, *Alu* I, *Mbo* I, *Msp* I, *Rsa* I, *Hinf* I, and *BstU* I (New England Biolabs, Berverly, MA) in a total volume of 30 mL at 37°C, with the exception of the *BstU* I reaction which was incubated at 60°C. The restriction fragments were then desalted using Sephadex G-75 (Amersham Biosciences, Uppsala, Sweden) and dehydrated. The fragments were resuspended in 15 ml formamide and 0.33 ml Genescan ROX-500 internal size standard (Applied Biosystems, Foster City, CA.), denatured by heating for 5 min at 95°C, and separated by capillary electrophoresis using an ABI 3130XL genetic analyzer with a 50 cm capillary array using POP6 polymer (Applied Biosystems). Each T-RFLP digestion was separated and visualized at least twice to ensure reproducibility of the analysis.

T-RFLP analysis

The fluorescently labeled 5' terminal-restriction fragments (T-RFs) were sized against the Genescan 500 LIZ internal size standard using Genemapper v3.7 (Applied Biosystems). Only fragments between 50 and 500 nucleotides were included in the analysis. Electropherograms were then imported into the program BioNumerics (Applied

Maths, Sint-Martens-Latem, Belgum). Community fingerprints were compared in BioNumerics using average Pearson product moment correlation and unweighted pair group method with arithmetic mean (UPGMA; Applied Maths) cluster analysis of all eight restriction digests using the relative fluorescent proportions of each electropherogram. Peak detection was limited to peaks between 50 and 500 base pairs in size and with height at least 3% of the maximum value of the fingerprint.

Clone library construction

Five replicate PCRs were performed on gDNA samples PV-247 B18, PV-340 Sc1, and PV-342 B56 using identical PCR conditions and the same SSU rRNA primers that were used for T-RFLP without 5' fluorochrome labeling on the forward primer. The PCR amplicons were visually assayed for size and purity by gel electrophoresis as described above. The remaining sample was gel purified using the GeneJET extraction kit (Thermo-Fisher) and cloned into the pJET1.2 cloning vector using the CloneJet PCR cloning kit (Thermo-Fisher). Ligation reactions were transformed into chemically competent E. coli cells (Active Motif, Carlsbad, CA). All putative clones were streaked for isolation and the insert was assayed for size using PCR with the primers pJetF and pJetR and running the products against a 1 kb size standard by 1% agarose electrophoresis. Clones with inserts were then end-sequenced using the pJetF and pJetR primers. Vector sequence was identified and removed and the sequences were trimmed to only include 400 base pairs from the end of the 68F primer. The sequences were then aligned to each other using the automated sequence alignment tool MUSCLE and a nucleotide similarity matrix was calculated using DNADIST from the PHYLIP package. Operational taxonomic units

(OTUs) were then defined as groups of at least two clones with a minimum of 97% similarity grouped using the furthest neighbor algorithm calculated with the program DOTUR (Schloss et al. 2005) with the similarity matrix from each clone library. Both strands of a representative clone from each OTU was sequenced and contiguous sequences were assembled using the Geneious software program. The OTUs were checked for chimeras using the Bellerophon server. Putative chimeras were further analyzed with the program Pintail (Ashelford et al. 2005) by comparing the putative chimera to any potential parent sequence.

Phylogenetic analysis

The fully sequenced OTUs were then aligned to the Silva database using the online SINA aligner. The sequences were imported into the Silva reference reference database (release 108) using the ARB phylogenetic software environment. The sequences were then inserted into the reference tree using the parsimony quickadd option to determine the approximate phylogenetic relatedness of the OTUs. Representative nearest neighbors were then exported along with other relevant sequences, such as closely related type strains to aid in phylogenetic analysis. The sequences were then opened in the sequence editor BioEdit and the alignment was manually checked for misaligned bases and a mask was constructed to remove positions in the alignment that were ambiguously aligned. A maximum likelihood tree was then calculated using RAxML v7.2.6 using the general time reversible model of nucleotide substitution (Tavaré, 1986). The trees were calculated 20 consecutive times using randomized starting trees to find the tree with the greatest likelihood to be the best tree. The treefiles were then visualized within the

software Mega and annotated using the vector drawing program Inkscape.

Quantitative PCR primer design

All full length sequences belonging to the Zetaproteobacteria and Epsilonproteobacteria from Loihi were identified in the Silva 108 database using the ARB sequence program (Ludwig, et al. 2004). Quantitative PCR (QPCR) primers specific for the Zetaproteobacteria- Zeta_537F (GAAAGGWGCRAGCGYTGTT) and Zeta_671R (TGCTACACACGGAATTCCGC) and the Sulfurimonas- Sulf_731F (TAGGAATACCNAAAGCGAAG) and Sulf_853R (TTAATSYGTTARSTGCATCAC) and Nitratiruptor- Nitro_729F (CAGGAATACCSATKGCGA) and Nitro_834R (ACGATGRATGCTAGTCGTT) groups of the Epsilonproteobaceria were designed using the Probe find function in ARB. General bacteria primers Bact533F (GTGYCAGCMGCCGCGGTAA) and Bact684R (TCTACGSATTTYACYSCTAC) were chosen to quantify the total bacterial SSU rRNA copies in each sample.

Quantitative PCR optimization

The primers were optimized for efficiency and specificity using a StepOnePlus real-time PCR system (Life Technologies) and DyNAmo Flash SYBR green PCR master mix (ThermoFisher Scientific, Waltham, MA). Each reaction contained 1X master mix, 0.5 μM of each primer, 1X ROX reference dye, and either 1ng of gDNA or a serial dilution of plasmid DNA. Serial dilutions of plasmids containing the SSU rRNA gene containing the target of each respective primer set were used to optimize the PCR efficiency and gDNA samples from different microbial mats from Loihi Seamount were

used to optimize the specificity of the primers by raising the annealing temperatures (54˚C to 64˚C) until only a single product was seen on a melting curve which was confirmed by running the completed reactions on a 3% agarose gel.

Quantitative PCR data analysis

Copy numbers of total bacteria and Zetaproteobacteria, Sulfurimonas, and Nitratiruptor were quantified in 1 ng of gDNA using absolute quantification against plasmid that was linearized by digestion with the restriction enzyme *Xho* I (New England Biolabs). Relative population proportions were calculated by dividing the copy number of each group by the copy number found using the general bacteria primer set. All reactions were performed in triplicate and checked for specificity with a melting curve.

Results

Site descriptions

Pre-eruption microbial mats

Thick iron oxide-encrusted microbial mats were first observed on Loihi in 1987 at Pele's Vents near the summit of the volcano. The mats were draped around hydrothermal vents which were emitting fluids that were highly enriched with CO_2 (300 mM) and Fe(II) (1.01 mM) and ranged from 15-30°C (Karl et al., 1988). These mats were later shown to be dominated by *Epsilonproteobacteria* and *Gammaproteobacteria*, and the first zetaproteobacterial sequence was also found in these pre-eruption communities sampled between 1987 and 1991 (Moyer et al., 1995).

The vents were visited in 1993 and appeared to be largely unchanged. Focused flow near the summit of Loihi at Pele's Vents were covered with red, flocculate microbial mats with some white streamer mats near the vent orifices. Diffuse vent fields were located downslope of Pele's Vents with more crusty microbial mats covering a large area (10-15 m^2). The venting around these mats was through 4-8 cm iron oxide chimneys and the fluid temperature was only 2-10°C above ambient seawater.

1-3 year post-eruption microbial mats

Loihi was visited ~2 months after it erupted with the manned submersible Pisces V (Malahoff et al., 2006). The newly-formed pit crater, named Pele's Pit, was deemed too dangerous to enter with a manned submersible at the time, however 77°C vents were discovered at the Lohiuau Vent Field near the entrance of the pit (Figure 2-1). These mats consisted of thick, white streamers near the vent orifices, and iron oxide-encrusted mat material further away from the focused flow. Pohaku, Kapos, and Naha diffuse vent fields were also discovered in 1996 on the southern flank of the volcano (Wheat et al., 2000). The vent fluids at these sites were between 10-30°C and were devoid of reduced sulfur species. The microbial mats at these vents were iron oxide-encrusted and often 10-20 cm thick with copious amounts of polysaccharide material coating the outsides of the mat.

Pele's pit was first explored in 1997 and hydrothermal vents were found at Ikaika Vents venting fluids that were in excess of 200°C covered with white microbial mat material. The pit was further explored in 1998 and the Jet Vents were discovered venting fluids at 165°C. Vents at Lohiau cooled to 63°C, and the south rift vents remained at a somewhat constant temperature of 10-30°C.

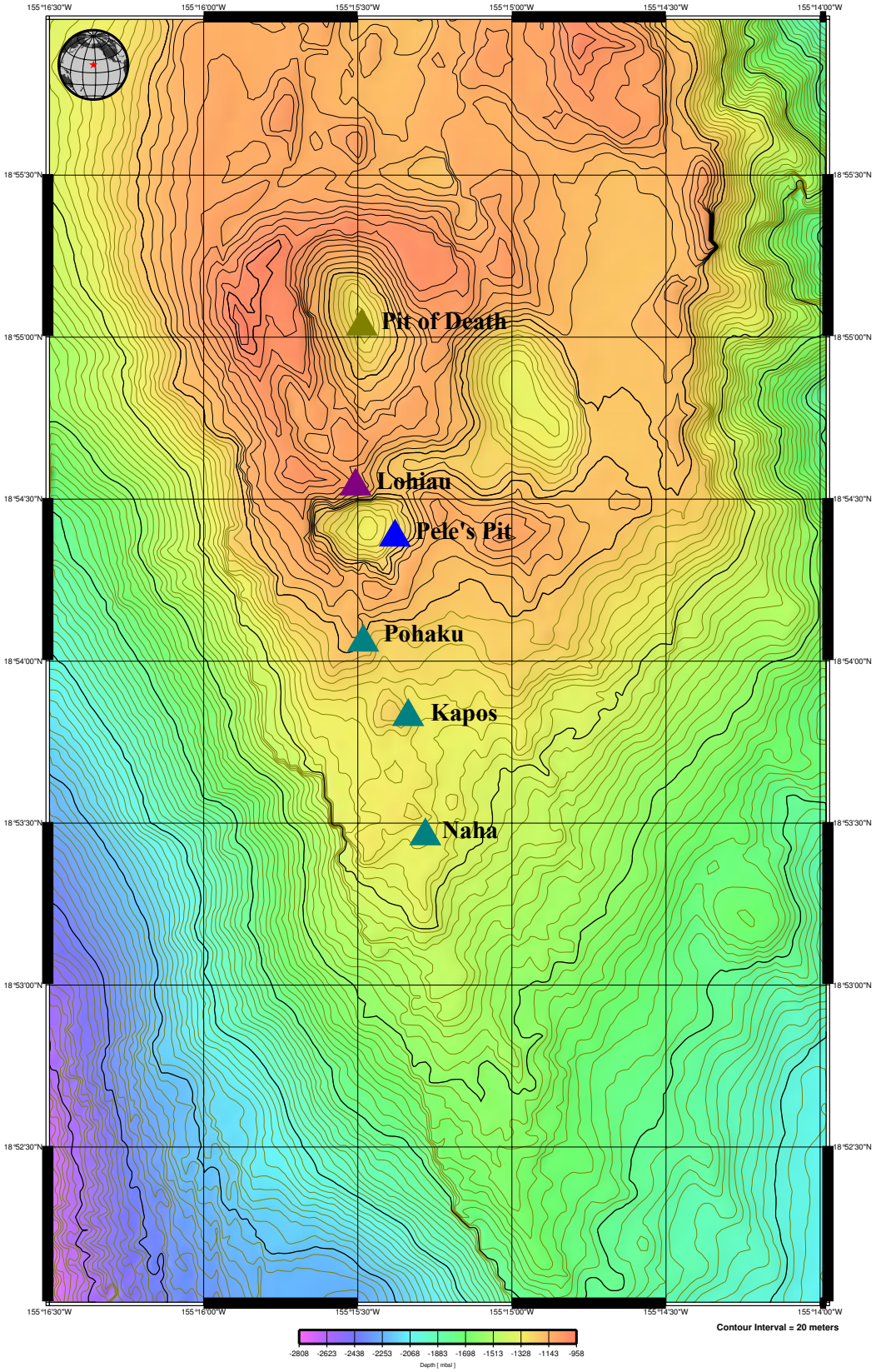

Figure 2-1. Bathymetric map showing the known active hydrothermal vent fields at Loihi Seamount. Original map was drawn at Earthref.org and is courtesy of Hubert Staudigel and Anthony Koppers.

7-13 year post-eruption microbial mats

We next visited Loihi in 2004, and annually from 2006-2009. Lohiau Vent Field was still active, however the venting had cooled to 21-28°C in 2004 and further cooled to only 1 degree Celsius above ambient in 2009. The Jet Vents had cooled to 61°C and were not active in 2006. Active venting was discovered in the Hiolo Ridge area of the pit, with thick iron-encrusted mats in the Lower Hiolo Vent Field (Markers 34 and 38), while the upper Hiolo Ridge vents (Markers 39 and 36) often had long white streamers growing in the direct vicinity of the vent along with thick iron oxide-encrusted mats. Temperatures in the Hiolo Ridge region were generally 40-63°C and did not change appreciably through 2009.

Naha Vents had cooled to 10°C in 2004 and was only 2-4°C above ambient in 2009. Microbial mats were still thick and covered with polysacchardes in 2004, however they became mineralized by 2007 indicating they were no longer actively growing. The microbial mats were coated with manganese oxides in 2009 suggesting iron(II) was no longer present in the hydrothermal fluids. Pohaku Vents were still active when visited in 2008 and 2009. The vents were 27°C and the vent field was covered in thick, luxurious iron oxide encrusted mat. Kapos Vents was no longer active in 2006.

'Ula Nui Vent Field was discovered in 2006 near the southern base of Loihi at 5000 meters below sea level. This vent field contained thick (up to 1 meter) iron-encrusted microbial mats covered by a more mineralized mat consisting of laminated iron oxide and manganese-oxide layers. The temperature anomaly of these mats was less than 1°C, however almost all of the iron oxides were biologically formed and were of

hydrothermal origin.

T-RFLP Results

Cluster analysis of T-RFLP community fingerprints indicated two distinctly divergent community clusters, which we named Loihi Group 1 (LG1) and Loihi Group 2 (LG2) (Figure 2-2). LG1 is the larger of the two clusters (n=55) and primarily contained iron oxide-encrusted microbial mats that did not have any white streamer material near the vent. This clade contained all but one pre-eruption samples, most of the 7-13 year post-eruption samples, and all of the South Rift samples. In general, microbial mats collected from the same vent in different years clustered together in LG1, and mats with similar temperatures also clustered together. LG2 contained three distinct subclusters within the group. Two of the subclusters contained communities from 1-3 years post-eruption Lohiau and Pele's Pit vents and a single pre-eruption sample that was enriched with white streamer material. The third subcluster contains communities from Upper Hiolo Ridge that contained white streamer material in addition to iron oxide-encrusted mats.

Clone Library Analysis

Nine clone libraries from microbial mats were analyzed that were representative of communities clustering within LG1 and LG2 (Table 2-1). The samples were chosen for clone library construction based on the T-RFLP clustering, the year sampled, and the site's interest for other studies. One pre-eruption and five post-eruption LG1 communities and three post-eruption LG2 communities were analyzed for community composition and

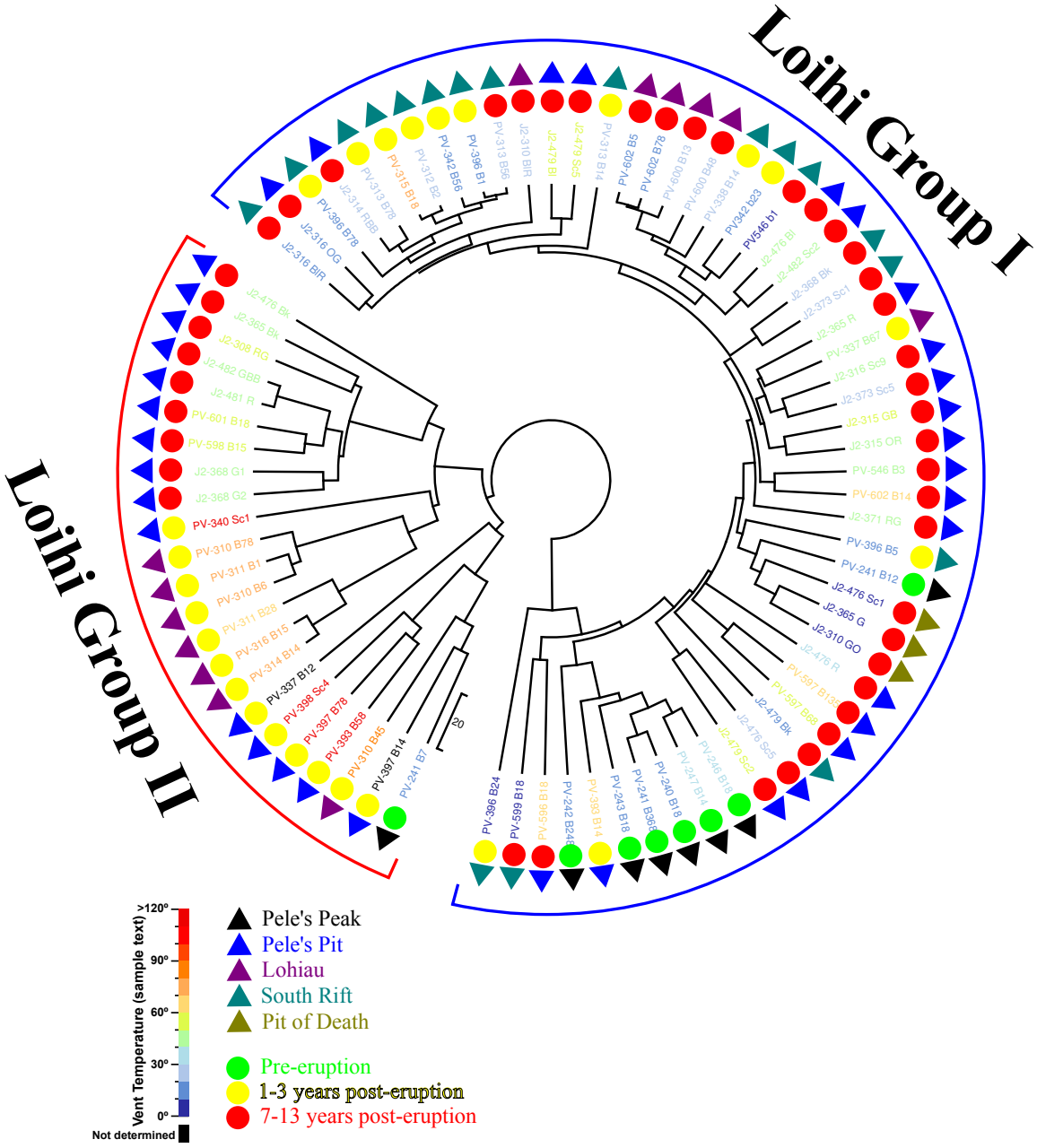

Figure 2-2. UPGMA/Pearson product-moment cluster analysis of T-RFLP bacterial community fingerprints from distinct microbial mat communities from Loihi Seamount, HI. Text color on sample names indicte the vent fluid temperature when the sample was obtained. Colored circles indicate the time the sample was taken, with green circles being from 1993, yellow circles from 1996-1998, and red circles are from 2004-2009. Triangles indicate the general location of the vents on Loihi Volcano.

Table 2-1. Table showing the temperature, percent Zetaproteobacteria, and percent Epsilonprotiobacteria found in SSU rRNA clone libraries from Loihi Seamount.

Year	Loihi Group	Sample	Location	Temp (°C)	%Zeta	%Epsilon	Source
1993	LG I	PV-246 B18	Boulder Patch	31	17	11	This Study
1996	LG II	PV-310 B6	Lohiau (M2)	77	0	59	This Study
1997	LG I	PV-342 B56	Pohaku	11	57	0	This Study
2004	LG II	PV-601 B18	Upper Hiolo (M39)	57	12	24	McAllister et al., 2011
2004	LG I	PV-602 B14	Lower Hiolo (M34)	63	42	0	McAllister et al., 2011
2006	LG I	J2-244Blu	'Ula Nui	2	8	0	Edwards et al., 2011
2007	LG II	J2-308 RG	Upper Hiolo (M39)	53	5	5	McAllister et al., 2011
2007	LG I	J2-310 BR	Lohiau (M55)	22	9	0	McAllister et al., 2011
2008	LG I	J2-373 Sc1	Pohaku	27	59	0	McAllister et al., 2011

phylogenentic diversity.

LG1 clone libraries

The pre-eruption sample (PV-246 B18) was from the Boulder Patch Vent Field located on Pele's Peak. The sample was dominated by *Deltaproteobacteria* (29%) and deeply-rooted *Gammaproteobacteria* (27%). The community also had populations of *Zetaproteobacteria* (17%) and *Epsilonproteobacteria* (11%). The *Zetaproteobacteria* were most closely related to other other Loihi phylotypes from post-eruption communities from Pohaku Vent Field (Figure 2-3). The *Epsilonproteobacteria* from Pele's Peak cluster in the *Sulfurimonas* Group most closely to phylotypes found at Pele's Peak in 1991 (Figure 2-4). These phylotypes are closely related to *Sulfurimonas autotrophica*, an aerobic obligate chemoautotrophic sulfur oxidizer (Inagaki et al., 2003).

Clone libraries from Pohaku Vent Field on the South Rift area of Loihi were dominated by *Zetaproteobacteria*. The *Zetaproteobacteria* comprised 57% of the sequences cloned from a 1997 sample, and 59% from a 2008 sample. The zetabacterial phylotypes from Pohaku were diverse and not dominated by a single phylotype (Figure 2-3). A clone library from Lower Hiolo Ridge in Pele's Pit from 2004 was also dominated by Zetaproteobacteria with 43% of the sequences clustering within the class. The community also contained lesser populations of *Alphaproteobacteria* and *Bacteroidetes*. A clone library from a ferromanganous microbial mat from the 5000 mbsl hydrothermal vent field 'Ula Nui collected in 2006 was dominated by phylotypes related to the Planctomycete (27%), Deltaproteobacteria (24%), Gammaproteobacteria (20%), and the Zetaproteobacteria (13%) (Edwards et al., 2011). The bacterial community was also much

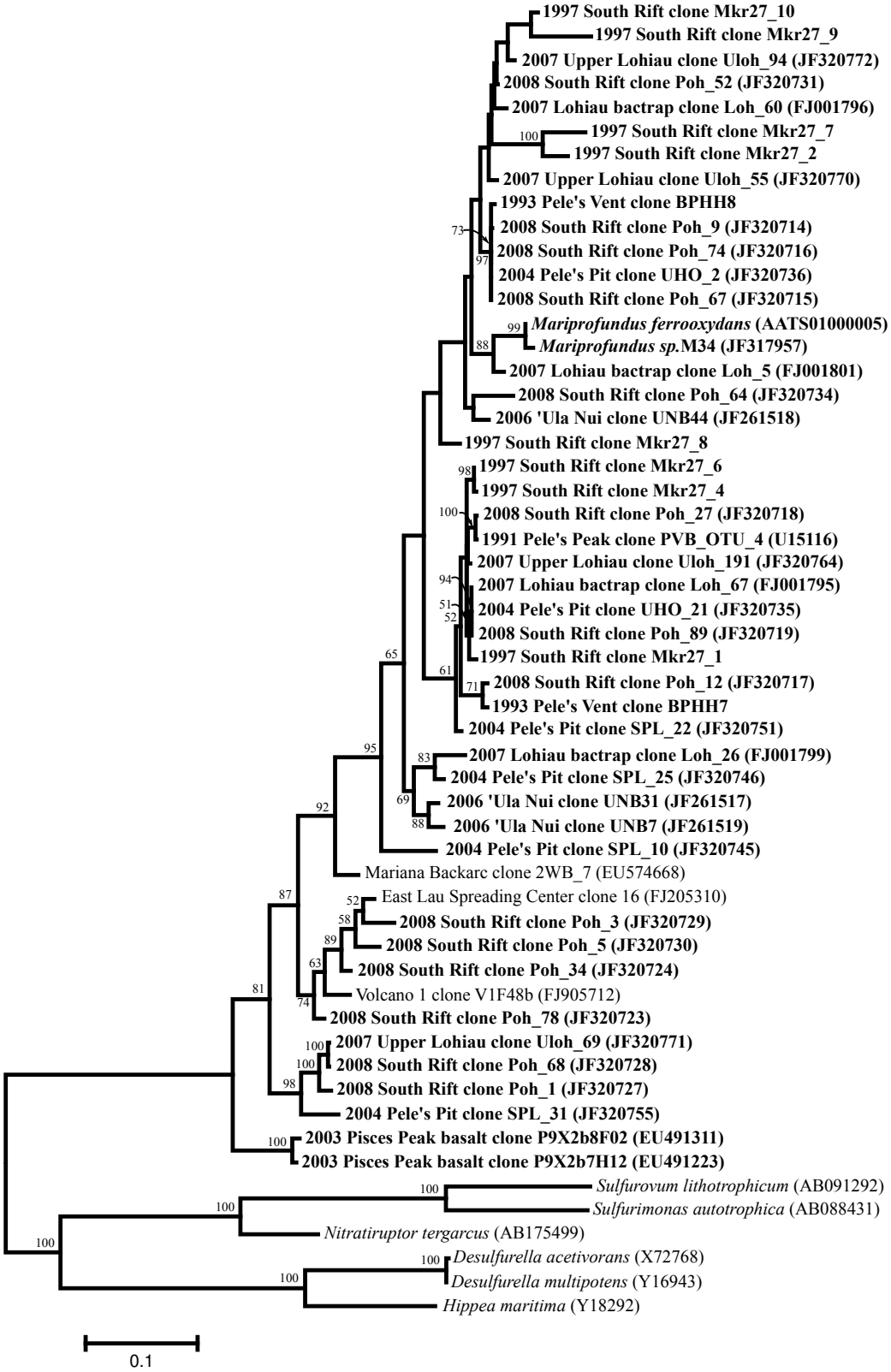

Figure 2-3. Maximum likelihood phylogenetic tree of the Zetaproteobacteria. Sequences from Loihi Seamount are in bold.

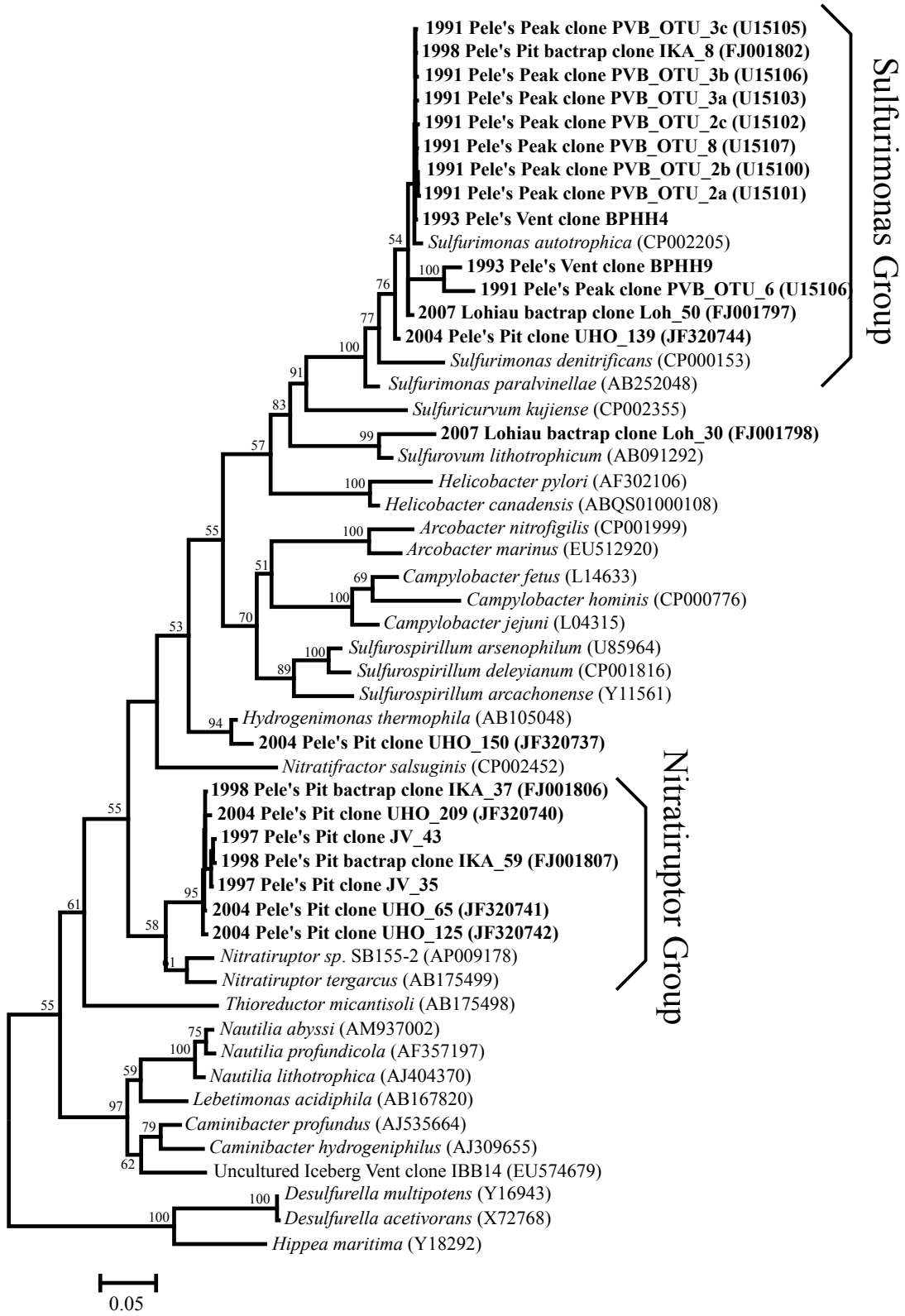

Figure 2-4. Maximum likelihood phylogenetic tree of the Epsilonproteobacteria. Sequences from Loihi Seamount are in bold.

more diverse than mat communities from Pele's Pit hydrothermal vent systems. A library from Upper Lohiau Vent Field from 2007 was also very diverse and was dominated by Deltaproteobacteria (16%), Alphaproteobacteria, Actinobacteria, and Zetaproteobacteria which were 9% of the bacterial community. Of these clone libraries, Epsilonproteobacteria were only found in the Upper Lohiau Vent microbial community where they composed only 2% of the bacterial community.

LGII clone libraries

A clone library from a Lohiau Vent microbial mat sampled shortly after the eruption in 1996 was dominated by phylotypes most closely related to to the *Nitratiruptor tergarcus* (59%), a hydrogen-oxidizing *Epsilonproteobacterium* (Figure 2-4) and also by phylotypes clustering within the *Thiomicrospira* genus (26%), which are sulfur-oxidizing bacteria belonging to the Gammaproteobacteria class.

Two clone libraries were performed on microbial mats from Marker 39 vents at the Upper Hiolo Ridge Vent Field in Pele's Pit. In 2004, the microbial mat was almost equally dominated by phylotypes clustering within the *Nitrospira* (35%) and the *Epsilonproteobacteria* (33%) classes. A clone library constructed from a community sampled in 2007 was dominated by equal numbers of *Nitrospira* and *Alphaproteobacteria* (21%); no *Epsilonproteobacteria* were found in this library.

Quantitative PCR Results

Quantitative PCR primers were designed for the *Sulfuromonas* and *Nitratiruptor* groups of *Epsilonproteobacteria* based on their abundance in pre- and post-eruption clone

libraries and well characterized metabolisms. A third quantitative PCR primer set was designed to amplify the dominant members of the *Zetaproteobacteria* that were found in the pre- and post-eruption microbial mat communities. These three populations were quantified in representative communities from pre-eruption samples from Pele's Peak and post-eruption samples from Lohiau, Pele's Pit, and vent fields on the South Rift of Loihi.

In five pre-eruption samples, the *Zetaproteobacteria* averaged 17% of the community (Figure 2-5). The *Sulfuromonas* group averaged 5% of the community with the greatest proportion (10%) occurring at warmer (25°C) hydrothermal vents where white mat material was observed inside the vent orifice. The *Nitrotiruptor* group averaged only 1% of the pre-eruption community (Figure 2-5).

The microbial mats at Lohiau Vent Field were dominated by the *Nitrotiruptor* group (52%) directly after the eruption in 1996 with smaller proportions of *Sulfurominas* (14%) and *Zetaproteobacteria* (6%). The *Nitrotiruptor* and *Sulfuromonas* groups were a much smaller portion of the community in 1997 with each group only composing 3% of the community, and the *Zetaproteobacteria* (40%) became a much larger portion of the community. The community were dominated by the *Zetaproteobacteria* in 2004 where they were 58% of the community while the *Nitrotiruptor* and *Sulfuromonas* groups were only 1% of the community. Microbial mats from 2006, 2007, and 2009 had only 0-1% *Nitrotiruptor* and *Sulfuromonas* groups and decreasing proportions of *Zetaproteobacteria* (25% to 6%) in the community.

The microbial mats in Pele's Pit were not sampled until 1997, however the mat that was sampled at the Forbidden Vent Field was dominated by the *Nitrotiruptor* group which composed of 85% of the population. The *Sulfuromonas* group was 6% and the

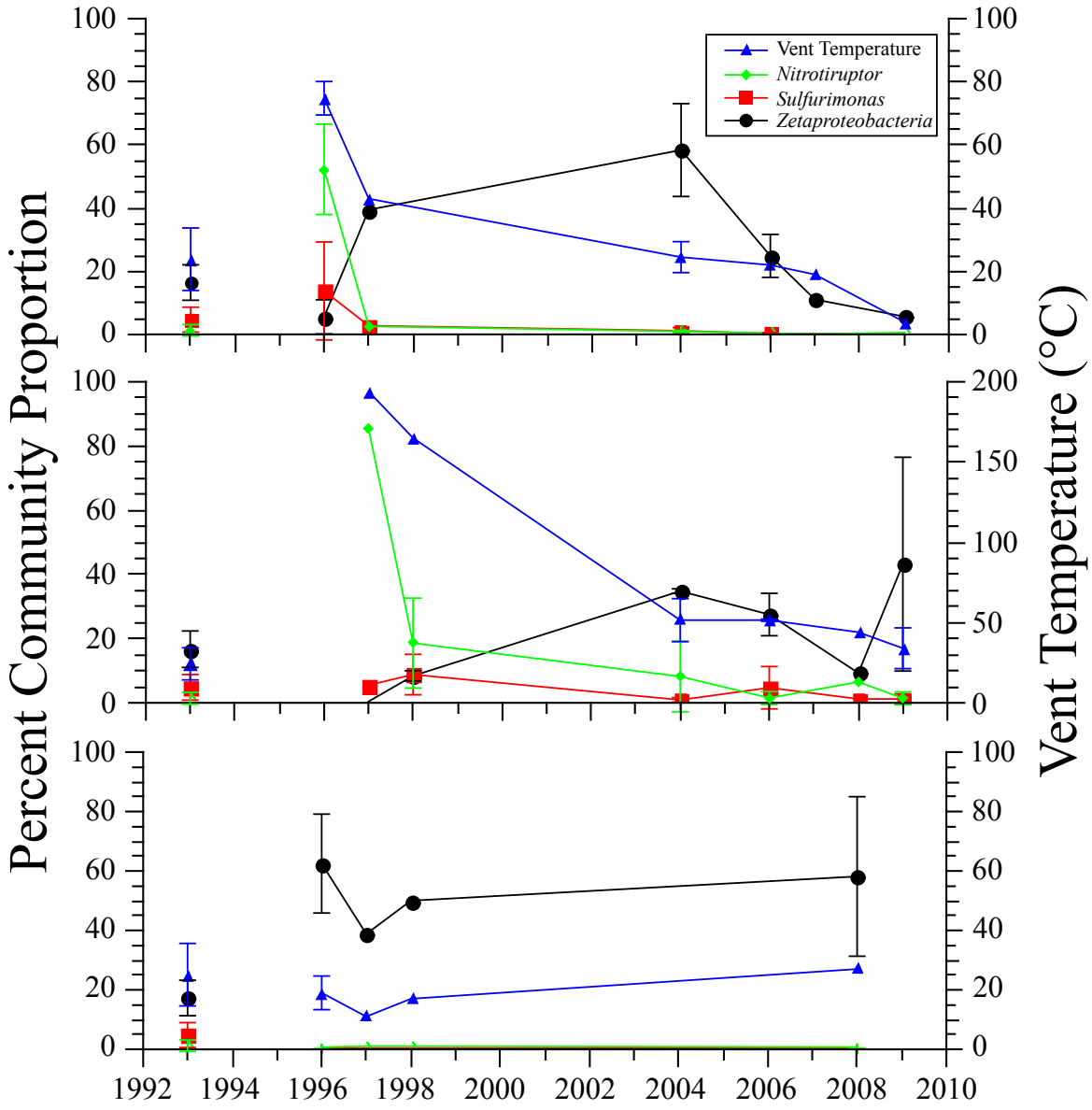

Figure 2-5. Graph showing QPCR and temperature data for Lohiau Vents (top), Pele's Pit (middle), and the South Rift (bottom). Data points at 1993 are from Pele's Peak Vents.

Zetaproteobacteria were less than 1% of the community. The proportion of *Nitrotiruptor* decreased in 1998 to 19% of the community, while the *Sulfuromonas* and *Zetabacteria* both increased becoming 9% and 8% of the community, respectively. From 2004 to 2009 the *Zetaproteobacteria* population was greater than the *Epsilonproteobacterial* populations in Pele's Pit communities. The *Zetaproteobacteria* was between 9% and 35% during these years, while neither of the *Epsilonproteobacterial* populations were over 8% of the community during this time.

Microbial mats on the South Rift of Loihi were dominated by *Zetaproteobacteria* immediately after the eruption where they were 62% of the community. The *Zetaproteobacteria* remained a dominant member of the bacterial community in 1997 and 1998, and still consisted of 58% of the bacterial community in 2008. The *Epsilonproteobacteria* were never over 1% of the bacterial community on the South Rift of Loihi.

Discussion

The bacterial community at Loihi Seamount changed radically after the eruptive event in 1996. Bacterial community fingerprinting shows that there have been two vastly different microbial mat communities at Loihi which we have named Loihi Group I and Loihi Group II (Figure 2-2). Clone library analysis of representative communities from both groups shows that Loihi Group I communities are primarily dominated by members of the *Zetaproteobacteria*, and contain little to no *Epsilonproteobacteria* (Table 2-1). Loihi Group II bacterial communities are dominated by *Epsilonproteobacteria* and *Nitrospira* groups with little to no *Zetaproteobacteria*.

The hydrothermal vents on Pele's peak prior to the 1996 eruption contained extremely high concentrations of dissolved Fe(II) and the microbial mats were encrusted with biogenic iron oxides that are similar to those formed by the chemoautotrophic iron-oxidizer *Mariprofundus ferrooxydans*, suggesting active iron oxidation was occurring in these mats (Emerson and Moyer, 2002). White material seen in the orifices of focused hydrothermal vents on the summit of Pele's Peak appeared similar to sulfur-cycling microbial mats from other vent fields and the presence of *Epsilonproteobacteria* closely related to *Sulfuromonas autotrophica* in the microbial mats strongly suggests some oxidation of reduced sulfur compounds or reduction of reduced of oxidized sulfur in these mats.

Microbial mat communities growing at Lohiau changed radically in the 2 years following the eruption. The temperatures of vent fluids were at least 77°C after the eruption indicating a prolonged increase in heat flux in the hydrothermal system (Wheat et al., 2000). The bacterial communities growing at these vents clustered within Loihi Group II (Figure 2-2), and were dominated by *Epsilonproteobacteria* related to *Nitrotiruptor tergarcus* which is a thermophilic hydrogen-oxidizing bacterium (Nakagawa et al., 2005). The vents cooled to 43°C by 1997 and the *Nitrotiruptor* population became a much smaller part of the bacterial community, and the *Zetaproteobacteria* became a much larger part of the community (Figure 2-5).

From 8-13 years following the eruption, the microbial mat communities from Lohiau cluster within Loihi Group I. The *Zetaproteobacteria* were up to 69% of the community in 2004, and then steadily declined to only being 6% in 2009. This decline also mirrored the heat flux of the vent which was 25°C in 2004 and near ambient in 2009

with very little visible fluid flow at the vent. The decreasing amount of *Zetaproteobacteria* at Lohiau may be explained by the decrease in fluid flow and therefore the concentration of dissolved iron around the vent site.

Microbial mats in Pele's Pit were also dominated by phylotypes closely related to *Nitrotiruptor tergarcus* after the eruption. The proportions of *Nitrotiruptor* phylotypes decreased from 85% to 19% from 1997 to 1998, while the *Zetaproteobacteria* increased from being not measurable in 1997 to 8% in 1998. The heat flux in Pele's Pit stabilized between 2004 and 2009 unlike the vents at Lohiau, and the mats maintain high proportions of *Zetaproteobacteria* (9% to 43%) during this time. Microbial mat communities growing at the Marker 39 Vent Field at North Hiolo Ridge in Pele's Pit still cluster within Loihi Group II, however their communities have transitioned from being dominated by *Epsilonproteobacteria* in 2004 to being dominated by phylotypes clustering within the *Nitrospira* class of bacteria. These phylotypes cluster closely with uncultured phylotypes from other hot iron-rich hydrothermal systems rather than with cultured nitrite-oxidizing *Nitrospira*.

Hydrothermal vent fluids on the South Rift of Loihi were never measured warmer than 27°C even shortly after the eruption in 1996. These vents were always devoid of reduced sulfur compounds in the vent fluids and highly enriched in dissolved iron (Wheat et al., 2000). Microbial mats growing near these vents always clustered within Loihi Group I and were dominated by Zetaproteobacteria.

The community succession observed at Loihi Seamount closely mirrors the observations and conceptual models of hydrothermal fluid evolution following a deep sea eruptive event (Butterfield et al., 1997). Hydrothermal vent fluids are dominated by

volatile gasses shortly after an eruption due to phase separation and direct input from magmatic gases. With time, the heat flux of the hydrothermal system decreases and phase separation ceases and volatile inputs from the magma decrease or cease to occur. The hydrothermal fluids become dominated by brine-dominated fluids that are characterized by high concentrations of chloride and iron ions and low concentration of volatile molecules such as hydrogen sulfide and hydrogen gas. The hydrothermal fluids at Loihi followed this model during and after the eruptive event. The iron/manganese ratio in hydrothermal fluids at Lohiau was 4 in 1996 and the ratio in Pele's Pit was 0.8 in 1997 (Wheat et al., 2000) which implies large inputs of sulfide reacting with the reduced iron forming sulfide minerals deep in the subsurface. The iron/manganese ratio at the South Rift was 58 directly after the eruption which is consistent with the ratio found in Hawaiian basalts suggesting no iron sulfide precipitation at these vents (Huang et al., 2007).

The succession of microbial communities at Lohiau and Pele's Pit closely mirrored the hydrothermal fluid evolution after the eruption (Figure 2-5). The samples from communities growing at the vents shortly after the eruption were dominated by phylotypes most closely related to hydrogen-oxidizing thermophiles. As the vents cooled and volatile molecules such as hydrogen gas and reduced sulfur compounds decreased in concentration in the vent fluids, iron-oxidizing bacteria became the dominant members of the bacterial communities. As the vents have cooled at Lohiau, the proportions of *Zetaproteobacteria* has decreased, presumably because of lower iron(II) concentrations. Pele's Pit hydrothermal vents have not cooled to extent that Lohiau has, and the microbial mats remain dominated by *Zetaproteobacteria*. The hydrothermal vents on the South Rift

did not follow the model of hydrothermal fluid evolution. They did not contain measurable amounts of volatile compounds in the hydrothermal fluids and they were always dominated by *Zetaproteobacteria*.

The period of time that a hydrothermal vent system transitions from a vapor-dominated system to a brine-dominated system to finally decaying can be highly variable, with some vents transitioning from a vapor-dominated system to decay in days or weeks, and others may be persistent for decades (Rubin et al., 2012). The vents at Lohiau were vapor-dominated when they were sampled in 1996 only a few months after the eruption, but were beginning to be brine dominated by 1997. The vents were dominated by iron-oxidizing bacteria by 2004 indicating the vents were fully brine-dominated by this time, and the vents decayed slowly over the next five years until they were nearly extinct in 2009 (Figure 2-5). Pele's Pit was still vapor dominated in 1997 and the community still had a significant proportion of *Nitrotiruptor* phylotypes in 1998 suggesting there was still a large vapor component to the hydrothermal fluids. The vents in Pele's Pit were dominated by iron-oxidizers in 2004, and the vents did not undergo decay through 2009, and are still dominated by Zetaproteobacteria. The hydrothermal vents on the South Rift were brine dominated when they were discovered in 1996 and it is unknown if they ever contained a large vapor component at these vents. The South Rift vents at Pohaku were still active in 2009 and dominated by Zetaproteobacteria.

This study shows that succession of hydrothermal vent microbial communities at erupting is intrinsically tied to the evolution of the vent fluids in the time after the eruptive event. Loihi Seamount had three distinct hydrothermal venting zones after the eruption in 1996, and each has shown different hydrothermal activity after the eruption.

Two of the venting regions had microbial communities that were dominated by hydrogen-oxidizing bacteria from 1-3 years after the eruption, and then they transitioned to being dominated by iron-oxidizing bacteria by 9 years after the eruption. Lohiau Vent Field has decayed to a point were there was negligible venting fluids in 2009 that was near ambient in temperature. Pele's Pit hydrothermal vents stabilized after 2004 and were still venting warm waters (10-63°C) at multiple vent sites. The hydrothermal vents on the South Rift were not vapor dominated when sampled in 1996, and may have never had a strong magmatic gaseous component.

The hydrothermal vent communities on Loihi Seamount showed spatial and temporal variability across the active volcano that was directly dependent on the hydrothermal fluid evolution following the eruptive event of 1995. The variability of the communities reinforces the necessity for long-term, high resolution ecological studies of complex eruptive systems to better understand the physical and chemical factors that influence microbial succession at these dynamic systems.

Works Cited

Butterfield, D.A., Jonasson, I.R., Massoth, G.J., Feely, R.A., Roe, K.K., Embley, R.E., Holden, J.F., McDuff, R.E., Lilley, M.D., and Delaney, J.R. (1997). Seafloor eruptions and evolution of hydrothermal fluid chemistry. Phil Trans R Soc Lond A *355*, 369–386.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006). The versatile [epsi]-proteobacteria: key players in sulphidic habitats. Nat Rev Micro *4*, 458–468.

Corliss, J.B., Dymond, J., Gordon, L.I., Edmond, J.M., von Herzen, R.P., Ballard, R.D., Green, K., Williams, D., Bainbridge, A., Crane, K., et al. (1979). Submarine Thermal Springs on the Galapagos Rift. Science *203*, 1073–1083.

Davis, R.E., Stakes, D.S., Wheat, C.G., and Moyer, C.L. (2009). Bacterial Variability within an Iron-Silica-Manganese-rich Hydrothermal Mound Located Off-axis at the Cleft Segment, Juan de Fuca Ridge. Geomicrobiol. J. *26*, 570–580.

Delaney, J.R., Kelley, D.S., Lilley, M.D., Butterfield, D.A., Baross, J.A., Wilcock, W.S. D, Embley, R.W., and Summit, M. (1998). The Quantum Event of Oceanic Crustal Accretion: Impacts of Diking at Mid-Ocean Ridges. Science *281*, 222–230.

Duennebier, F., Becker, N., Caplan-Auerbach, J., Clague, D., Cowen, J., Cremer, M., Garcia, M., Goff, F., Malahoff, A., McMurtry, G., et al. (1997). Researchers rapidly respond to submarine activity at Loihi volcano, Hawaii. Eos Trans AGU *78*, 229–233.

Edwards, K.J., Glazer, B.T., Rouxel, O.J., Bach, W., Emerson, D., Davis, R.E., Toner, B.M., Chan, C.S., Tebo, B.M., Staudigel, H., et al. (2011). Ultra-diffuse hydrothermal venting supports Fe-oxidizing bacteria and massive umber deposition at 5000m off Hawaii. ISME J. *5*, 1748-1758.

Embley, R.W., Chadwick, W.W., Baker, E.T., Butterfield, D.A., Resing, J.A., de Ronde, C.E.J., Tunnicliffe, V., Lupton, J.E., Juniper, S.K., Rubin, K.H., et al. (2006). Long-term eruptive activity at a submarine arc volcano. Nature *441*, 494–497.

Emerson, D., and Moyer, C.L. (2002). Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition. Appl Env. Microbiol *68*, 3085–3093.

Glazer, B.T., and Rouxel, O.J. (2009). Redox speciation and distribution within diverse iron-dominated microbial habitats at Loihi Seamount. Geomicrobiol. J. *26*, 606–622.

Huang, S., Humayun, M., and Frey, F.A. (2007). Iron/manganese ratio and manganese content in shield lavas from Ko'olau Volcano, Hawai'i. Geochim. Cosmochim. Acta *71*, 4557–4569.

Inagaki, F., Takai, K., Kobayashi, H., Nealson, K.H., and Horikoshi, K. (2003). *Sulfurimonas autotrophica* gen. nov., sp. nov., a novel sulfur-oxidizing epsilon-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. Int J Syst Evol Microbiol *53*, 1801–1805.

Juniper, S.K., and Fouquet, Y. (1988). Filamentous iron-silica deposits from modern and ancient hydrothermal sites. Can Miner. *26*, 859–869.

Karl, D.M., McMurtry, G.M., Malahoff, A., and Garcia, M.O. (1988). Loihi Seamount, Hawaii: a mid-plate volcano with a distinctive hydrothermal system. Nature *335*, 532–535.

Lupton, J.E., Lilley, M.D., Butterfield, D.A., Evans, L., Embley, R.W., Massoth, G.J., Christenson, B., Nakamura, K., and Schmidt, M. (2008). Venting of a Separate CO2-Rich Gas Phase from Submarine Arc Volcanoes - Examples from the Mariana and Tonga-Kermadec Arcs. J Geophys Res.113, B08S12, doi:10.1029/2007JB005467.

Malahoff, A., Kolotyrkina, I.Y., Midson, B.P., and Massoth, G.J. (2006). A decade of exploring a submarine intraplate volcano: Hydrothermal manganese and iron at Lo'ihi volcano, Hawai'i. Geochem Geophys Geosyst *7*, Q06002.

Moyer, C.L., Dobbs, F.C., and Karl, D.M. (1995). Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. *61*, 1555–1562.

Nakagawa, S., and Takai, K. (2008). Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. FEMS Microbiol. Ecol. *65*, 1–14.

Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K., and Sako, Y. (2005). *Nitratiruptor tergarcus* gen. nov., sp. nov. and *Nitratifractor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the ε-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. Int. J. Syst. Evol. Microbiol. *55*, 925–933.

Resing, J.A., Lebon, G., Baker, E.T., Lupton, J.E., Embley, R.W., Massoth, G.J., Chadwick, W.W., and de Ronde, C.E.J. (2007). Venting of Acid-Sulfate Fluids in a High-Sulfidation Setting at NW Rota-1 Submarine Volcano on the Mariana Arc. Econ. Geol. *102*, 1047–1061.

Resing, J.A., Rubin, K.H., Embley, R.W., Lupton, J.E., Baker, E.T., Dziak, R.P., Baumberger, T., Lilley, M.D., Huber, J.A., Shank, T.M., et al. (2011). Active submarine eruption of boninite in the northeastern Lau Basin. Nat. Geosci *4*, 799–806.

Rubin, K.H., Soule, S.A., Chadwick, W.W., Fornari, D.J., Clague, D.A., Embley, R.W., Baker, E.T., Perfit, M.R., Caress, D.W., and Dziak, R.P. (2012). Volcanic eruptions in the deep sea. Oceanography *25*, 142-157.

Seyfried, W.E., Berndt, M.E., and Seewald, J.S. (1988). Hydrothermal alteration processes at mid-ocean ridges; constraints from diabase alteration experiments, hot spring fluids and composition of the oceanic crust. Can Miner. *26*, 787–804.

Takai, K., Gamo, T., Tsunogai, U., Nakayama, N., Hirayama, H., Nealson, K.H., and Horikoshi, K. (2004). Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. Extremophiles *8*, 269–282.

Toki, T., Tsunogai, U., Ishibashi, J., Utsumi, M., and Gamo, T. (2008). Methane enrichment in low-temperature hydrothermal fluids from the Suiyo Seamount in the Izu-Bonin Arc of the western Pacific. Ocean. J. Geophys. Res. *113*, B08S13.

Wheat, C.G., and Mottl, M.J. (2000). Composition of pore and spring waters from Baby Bare: global implications of geochemical fluxes from a ridge flank hydrothermal system. Geochim Cosmochim Acta *64*, 629–642.

Wheat, C.G., Jannasch, H.W., Plant, J.N., Moyer, C.L., Sansone, F.J., and McMurtry, G.M. (2000). Continuous sampling of hydrothermal fluids from Loihi Seamount after the 1996 event. J. Geophys. Res. *105*, PP. 19,353–19,367.

Chapter 3: Identification and Phylogenetic Analysis of Key Genes in the Six Known Carbon Fixation Pathways.

Introduction

Autotrophic organisms are able to derive all of their organic carbon from inorganic sources. This anabolic process is very energy intensive and autotrophic organisms are further classified by the energy source harvested to generate the reducing power needed to convert oxidized carbon (CO_2) to organic molecules. Organisms which derive energy from photosynthesis and fix CO_2 are characterized as photoautotrophs, while organisms which derive energy from the oxidation of reduced chemicals are called chemotrophs. Organisms which derive their energy solely from reduced inorganic chemicals and fix CO_2 are called chemolithoautotrophs.

There are currently six known carbon fixation pathways; the Calvin Cycle, the reductive TCA (rTCA) cycle, the Wood-Ljungdahl (reductive acetyl-CoA) pathway, the 3-hydroxypropionate (3HP) cycle, 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) Cycle, and the Dicarboxylate/4-Hydroxybutyrate (DC/4HB) Cycle. The Calvin Cycle is by far the most historically studied carbon fixation pathway (Tabita, 1999) and is assumed to be responsible for the majority of carbon fixation on earth. The ecological significance of the other five pathways has only begun to be realized in the past ten years as a significant number of genomes and metagenomes have been sequenced, as well as directed PCR-based studies in diverse environments (Hügler and Sievert, 2011).

The genomic era has not only confirmed the carbon fixation pathways used in many autotrophic microbes, it has directly aided in the identification of new pathways that may be ecologically significant (Berg et al., 2007). Ecological studies of carbon fixation pathways depend on identifying a "key" enzyme in the pathway whose presence

is used as an indicator of the existence of the entire pathway. All carbon fixation pathways have intermediate steps which feed into other biosynthetic or catabolic pathways. The key enzyme must catalyze a step in the carbon fixation pathway which is unique to the pathway and is not used in any other known biochemical pathways within the organism or the phylogeny of the autotrophic organisms is easily distinguished from the heterotrophic organisms that utilize a homologous enzyme. The gene for the key enzyme would also ideally contain highly conserved regions to allow for universal primer design and be >900 base pairs long to allow for phylogenetic inference.

This chapter will give a short review of the historical and ecological relevance of each carbon fixation pathway. I will then discuss the key enzymes for each pathway and evaluate their potential to be used for ecological studies. I will then present a phylogenetic analysis of one key gene for each autotrophic pathway using all known homologs for the gene that have been archived by the integrated microbial genome database at the Joint Genome Institute as of January 2014 (Markowitz, 2006).

Materials & Methods

Database construction

Homologous genes for each key gene were identified and downloaded using the Integrated Microbial Genomes system (IMG) website. Full length sequences were then aligned using the multiple alignment tool BioEdit (Hall, 1999) by translating the sequence and aligning the amino acid sequences before toggling the translation back to the original nucleotides. Once the sequences were aligned, they were imported into the

ARB software environment along with pertinent metadata about each sequence making a master alignment.

Phylogenetic tree construction

Full length sequences were translated and exported from ARB. The alignment was masked to include only unambiguously aligned positions and positions containing insertions or deletions in less than 50% of the sequences. The alignment was then exported in PHYLIP format and the phylogenetic tree was calculated using RaxML v8.0 (Stamatakis, 2014) using the Whelan and Goldman model of amino acid substitution (Whelan and Goldman, 2001). Each tree was calculated 10 times with randomized starting trees and the tree with the best likelihood value was used. Bootstrap values were calculated using 100 bootstrap trees calculated using the same parameters as the phylogenetic tree, only without recalculation to find the best tree. The treefiles were visualized using the software Mega and annotated using the vector drawing program Inkscape (http://inkscape.org).

Calvin Cycle

The Calvin Cycle, also called the Calvin-Benson-Bassham (CBB) Cycle and the reductive pentose phosphate cycle, was the first carbon fixation pathway fully elucidated (Bassham et al., 1956). The Calvin Cycle is used by many chemoautotrophic and photoautotrophic bacteria, including the *Proteobacteria*, *Cyanobacteria*, *Chloroflexi*, *Verrucomicrobia*, *Firmicutes*, *Thermus*, *Actinobacteria*, and all photosynthetic eukaryotes. The pathway was elucidated in the laboratory of Melvin Calvin using then-

revolutionary $^{14}CO_2$ techniques using algal *Scenedesmus* cells (Seaborg and Benson, 2008). The Calvin Cycle shares many common intermediates and enzymes with the pentose phosphate pathway, but it converts the pathway into a cycle with the addition of a carbon fixation step.

The cycle begins with the fixation of three carbon dioxide molecules with three ribulose-1,5-bisphosphate molecules to form six 3-phosphoglycerate molecules (Figure 3-1). These 3-phosphoglycerate molecules are eventually transformed to six glyceraldehyde 3-phosphate molecules, one of which leaves the cycle to be incorporated into biomass. Ribulose-1,5-bisphosphate is then regenerated through a multi-step process involving sugar-phosphate molecules found in the pentose phosphate pathway.

Each complete cycle yields a single glyceraldehyde 3-phosphate molecule and costs 9 ATP and 6 NADPH molecules making it the most costly carbon fixation pathway in terms of energetic costs. Despite the large energetic cost to fix carbon using this cycle, the Calvin Cycle is the most used carbon fixation pathway on Earth. The success of this cycle is probably due to the lack of oxygen sensitivity compared to other carbon fixation pathways which allows oxygenic photoautotrophes to use this cycle as well as chemoautotrophs that live in oxic conditions and use molecular oxygen as a terminal electron acceptor.

Key enzymes in the Calvin Cycle

The two key enzymes in the Calvin Cycle are Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco; Figure 3-1 enzyme 1) which catalyzes the formation of 3-phosphoglycerate molecules from ribulose-1,5-bisphosphate and phosphoribulokinase

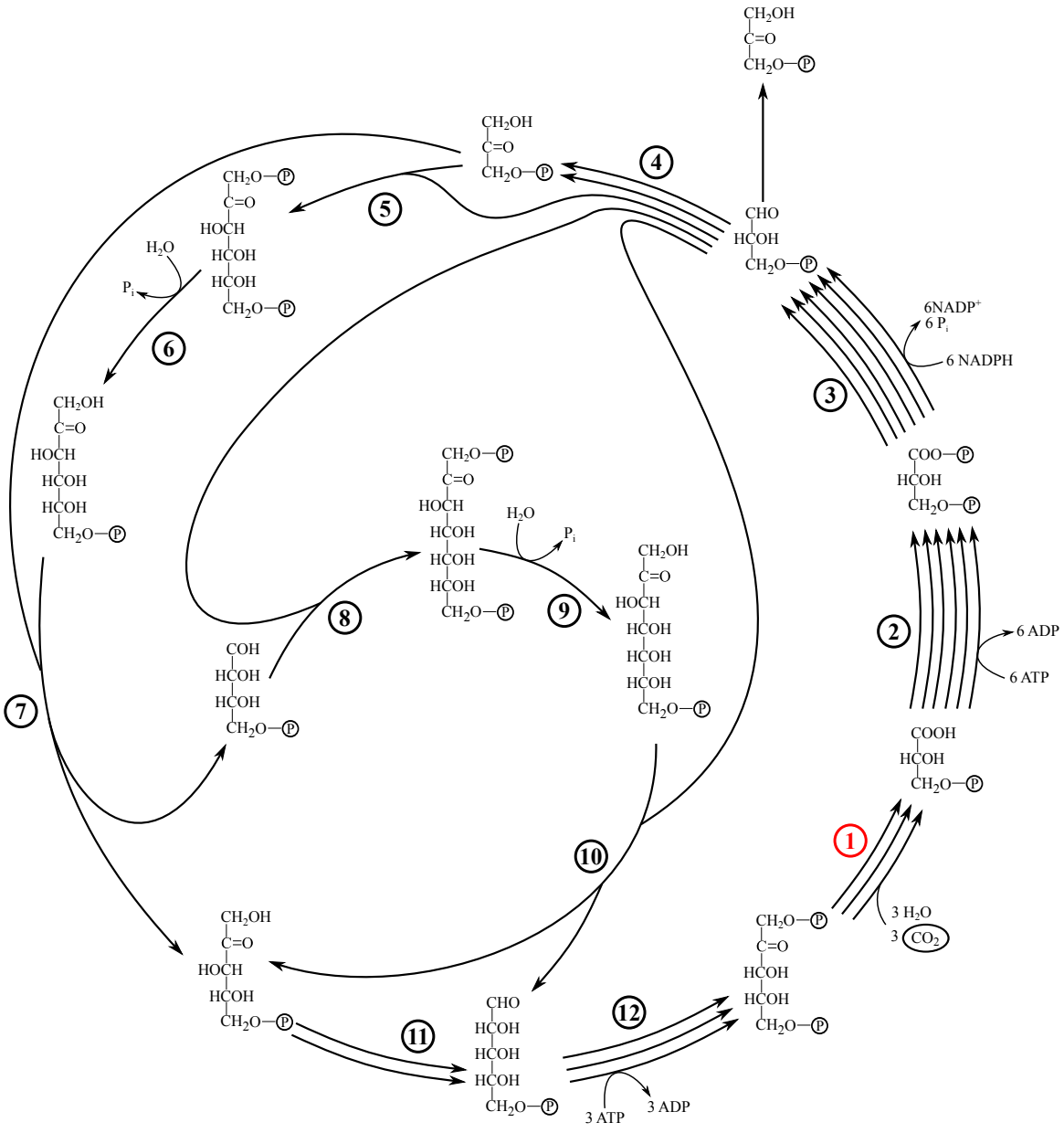

Figure 3-1. Schematic pathway showing the Calvin Cycle. Enzymes are depicted in the figure as circled numbers: 1. ribulose-1,5-bisphosphate carboxylase/oxygenase 2. 3-phosphoglycerate kinase 3. glyceraldehyde-3-phosphate dehydrogenase 4. triose-phosphate isomerase 5. fructose-bisphosphate aldolase 6. furctose-bisphosphate phosphatase 7. transketolase 8. sedoheptulose-bisphosphate aldolase 9. sedoheptulose-bisphosphate phosphatase 10. ribose-phosphate isomerase 11. ribulose-phosphate epimerase 12. phosphoribulokinase. Cycle is based on Bassham et al, (1956) and Berg (2011).

(Figure 3-1 enzyme 12) which forms ribulose 1,5-bisphosphate from ribulose 5-phosphate and ATP. When compared to phosphoribulokinase, Rubisco is much more abundant in the cell, it is often considered the most abundant protein on earth (Ellis, 1979), which allows it to be isolated and enzymatically characterized much easier than phosphoribulokinase (Tabita, 1999). The Rubisco large subunit is ~1400-1500 nucleotides in length and is highly conserved allowing more robust phylogenetic analysis than the shorter (800-900bp) and less conserved phosphoribulokinase gene. For these reasons, Rubisco is by far the more studied enzyme and is almost always used as the sole key enzyme of the Calvin Cycle, however both enzymes must be present to differentiate between a functional Calvin Cycle and a remnant or partial cycle.

Rubisco phylogeny has been characterized historically as having four different 'forms' or 'types' (Tabita, 1999). Form I Rubisco is by far the most common of the forms and is found in all aerobic autotrophes that utilize the Calvin Cycle. The Form I Rubisco consists of a catalytic large subunit that contains a Mg^{2+} metal center and a small subunit which increases the rate of catalysis of the large subunit, but does not directly contribute to substrate specificity (Andersson and Backlund, 2008). The subunits form a quaternary structure consisting of L_8S_8 subunits. The phylogeny of the Form I Rubisco was phylogenetically characterized by identifying four monophyletic groups, designated A-D (Tabita, 1999). Group IA consisted of aerobic chemoautotrphic bacteria, IB was marine cyanobacteria, IC was primarily 'Purple Bacteria', which are phototrophic Alphaproteobacteria and Betaproteobacteria. Group ID was comprised of the non-green algae. Groups IA and IB are often called the 'Green Type' Rubiscos, and IC and ID are called the 'Red Type' Rubisco which is based on the green algae in Group IB and the red

algae in Group ID. A second Form of Rubisco, called 'Form II' or 'Type II' was isolated from Rhodospirillum rubrum and found to consist of only a large subunit that forms a dimer with itself (Gibson and Tabita, 1977). Many organisms with a Form II Rubisco also have a Form I Rubisco encoded in their genomes, with the Form II being expressed under anaerobic conditions and the Form I Rubisco being expressed in oxic environments. A Form III Rubisco was identified in various members of thermophilic Archaea (Bult et al., 1996), however this form was found to be used in AMP recycling, rather than autotrophic carbon fixation (Sato et al., 2007). Form IV Rubisco was identified in several bacteria strains that are incapable of autotrophic growth. This form was later found to be involved in the methionine salvage pathway, and not used in carbon fixation (Grundy and Henkin, 1998). Recently, a fifth form of Rubisco has been identified in *Methanococcoides burtonii* and in several metagenomes. While this form of Rubisco has not been characterized, it is theorized that this 'Hybrid Form II/III' Rubisco has a function analogous to the Form III Archaeal Rubisco, and is not used in a functional Calvin Cycle (Wrighton et al., 2012).

Rubisco is primarily characterized by the inefficiency in which the enzyme fixes carbon. Rubisco is a slow catalyst, with catalytic turnover values (k_{cat}) of $4.3s^{-1} \pm 2.9$ in bacterial isolates under ideal conditions (Hartman and Harpel, 1994). The inefficiency of the turnover rate is further amplified by the inherent oxygenase activity of the enzyme. In the presence of oxygen, Rubisco can cleave ribulose-1,5 bisphosphate and form 2-phosphoglycolate and glyceraldehyde 3-phosphate molecules. Glyceraldehyde 3-phosphate can be quickly metabolized by re-entering the Calvin Cycle, however 2-phosphoglycolate is a potent inhibitor of triose-phosphate isomerase which causes the

Calvin Cycle to cease to function and is toxic to the cell if it is not quickly degraded in the cell (Tabita, 1999). 2-phosphoglycolate is degraded in either the glycerate pathway or the glycolate cycle, with both processes resulting in a loss of CO_2 from the cell. An obligate autotrophic cell utilizing the Calvin Cycle can therefore only survive in conditions where the carboxylase activity is greater than the oxygenase activity of the Rubisco enzyme found in the organism, which can be highly variable between different organisms (Hartman and Harpel, 1994).

Early in the study of Rubisco activity, radically different CO_2/O_2 substrate specificity (Ω values) was found in different bacterial species. The affinity of Rubisco for CO_2 is defined as:

$$\Omega = V_C K_{O2} / V_O K_{CO2}$$

With V_C and V_O representing the V_{max} of carboxylation and oxygenation, respectively (Chen and Spreitzer, 1992). Organisms with high O_2 substrate specificity (low Ω) will favor the oxygenase activity of Rubisco when found in an oxic environment and cannot survive. Organisms with the highest Rubisco CO_2 substrate specificity are within the Type ID Rubisco clade. Rubisco in clade IC and IB generally have lower CO_2 specificity than Type ID Rubisco, and Type IA Rubisco has the lowest CO_2 specificity of the Form I Rubiscos. The Form II Rubisco has CO_2 specificity that is an order of magnitude lower than the Type ID Rubisco which limits the effectiveness of Form II Rubisco carboxylase activity to microaerophilic or anaerobic environments (Chen and Spreitzer, 1992).

The Michaelis-Menten constant of CO_2 (K_{CO2}) for Rubisco is also highly variable,

with values as low as 5μM in some Group ID algae, and as high as 250μM in some form II Rubisco enzymes (Tabita, 1999). The Michaelis-Menten constant is a measurement of the affinity of CO_2 for Rubisco, with high values (low affinity) indicating that a high concentration of CO_2 is necessary for effective carboxylase activity, since K_{CO2} is the concentration in which the enzyme functions at one-half of the maximum reaction rate (V_{max}) the enzymes would function at extremely low rates at CO_2 concentrations below their ideal K_{CO2} concentration.

Rubisco phylogeny

Form I Rubisco

Form I Rubisco phylogeny of 162 sequenced bacterial isolates shows a similar topology to the phologenetic tree presented by Tabita in 1999 which utilized the 18 sequences that were known at that time (Figure 3-2). The Form IA clade contains aerobic chemoautotrophic bacteria from the *Alpha- Beta- Gamma-* and *Zetaproteobacteria* classes. Members of this group include the Nitrogen-oxidizing *Nitrosomonas* and *Nitrobacter* genera, the sulfur-oxidizing *Beggiatoa*, *Halothiobacillus*, *Thioalkalivibrio*, and *Thiomicrospira* genera, the iron-oxidizing *Acidithiobacillus*, *Sideroxydans*, and *Mariprofundus* genera, and the hydrogen-oxidizing *Rhodobacter* and *Cupriavidus* genera. Form IB is a paraphyletic group that consists of photosynthetic cyanobacteria, with a large clade consisting of *Prochlorococcus* and *Synechococcus* genera and another clade consisting of a diverse group of cyanobacteria that includes the *Nostoc*, *Anabaena*, *Cyanothece*, and *Arhrospira* genera. Form IC clade includes the anaerobic phototrophic

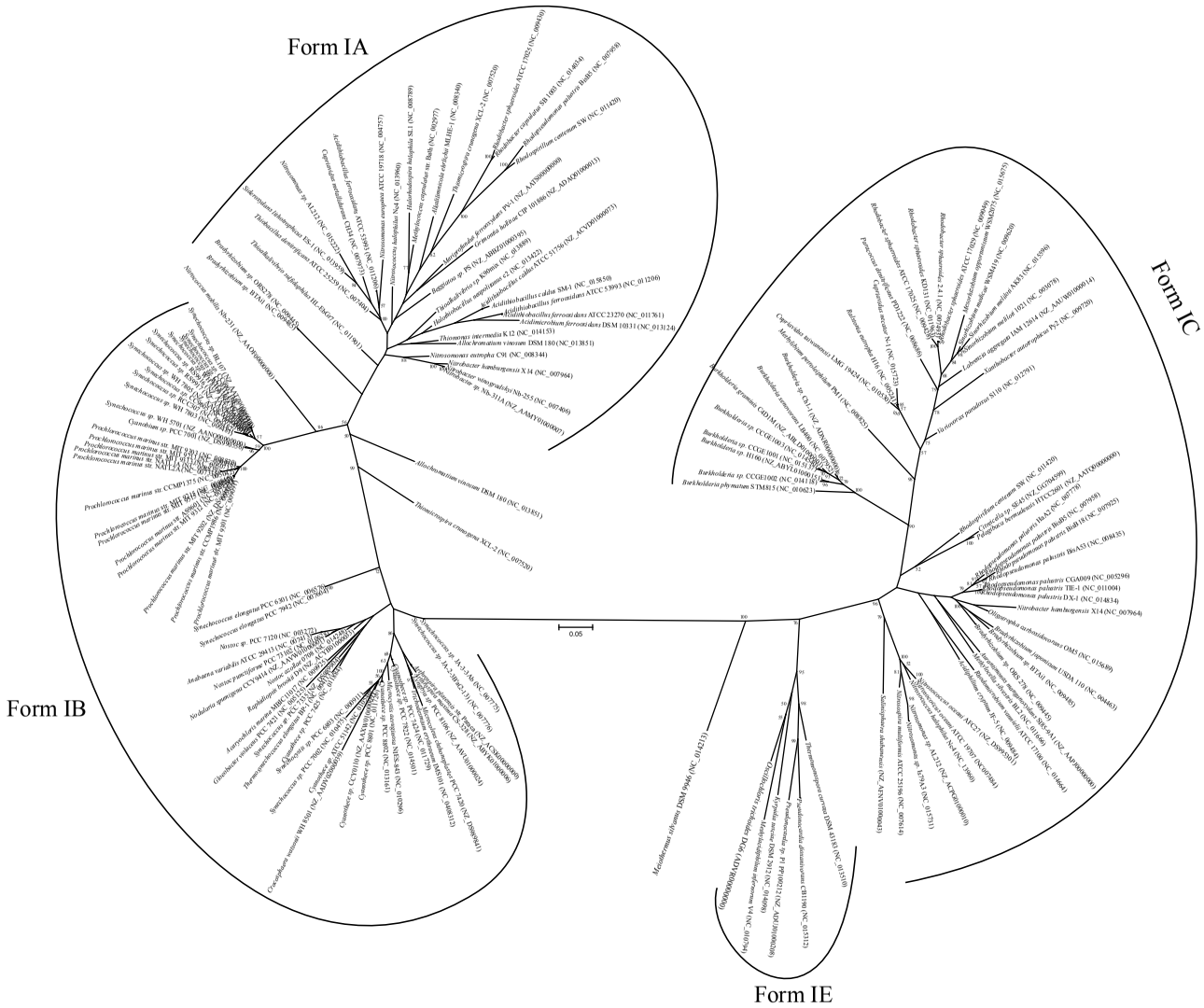

Figure 3-2. Unrooted maximum likelihood phylogenetic tree of the large subunit of the form I ribulose-1,5-bisphosphate carboxylase/oxygenase gene (cbbL).

'purple bacteria' belonging to the *Rhodobacter* genus along with hydrogen oxidizing (Knallgas) *Ralstonia*, *Xanthobacter*, and *Cupriavidus* gerera. The clade also includes the nitrogen fixing *Sinorhizobium* and *Mesorhizobium* genera whose autotrophic metabolism is not yet understood, however the genomes contain carbon monoxide dehydrogenase genes, suggesting aerobic carbon monoxide oxidation is a potential autotrophic metabolism in these isolates. The autotrophic metabolism of the Burkholderia genera that cluster in the Form IC clade is also unknown, although aerobic carbon monoxide oxidation and other C1 metabolisms have been proposed (Anandham et al., 2009). Another clade contains the *Rhodospirillum*, *Rhodomicrobium*, and *Rhodopeudomonas* 'Purple Bacteria' which are facultative anaerobic phototrophes. *Bradyrhizobium japanicum* can grow as a chemoautotroph using thiosulfate and carbon monoxide as energy sources (Lorite et al., 2000; Masuda et al., 2010). *Oligotropha carboxidovorans* is cabable of supporting chemoautotrophic growth with the oxidation of carbon monoxide (Santiago et al., 1999). *Acidiphilium cryptum* is a hydrogen-oxidizer that can reduce Iron(III) oxides (Kusel et al., 1999). The ammoniua oxidizing bacteria *Nitrosococcus*, *Nitrosomonas*, Nitrosospira, and *Salinisphaera* genera also cluster within this clade. Clade 1D was not included in this analysis because it only consisted of algal sequences, which are not included in the tree.

A fifth clade, called Form IE here, has emerged from sequences from newly isolated chemoautotrophs. This clade includes the Actinobacteria *Thermomonospora* and *Pseudonocardia* isolates with no characterized autotrophic metabolisms. The thermophillic hydrogen-oxidizing *Kyrpidia tusciae* is the only known Firmicute capable of autotrophic growth using the Calvin Cycle (Bonjour and Aragno, 1984). The

acidiphilic thermophile *Methylacidiphilum infernorum* is an obligate methanotrophic Verrucomicrobium isolated from geothermal soil (Hou et al., 2008). *Oscillochloris trichoides* is an anearobic phototroph belonging to the *Chloroflexi* class and is capable of photoautotrophic growth in the presence of sulfide or hydrogen gas (Ivanovsky et al., 1999). The organisms in this clade represent novel Rubisco diversity with representatives having diverse or unknown autotrophic metabolisms not belonging to the *Cyanobacteria* or *Proteobacteria* phyla.

Form II Rubisco

The Form II Rubisco phylogenetic tree contains 33 sequences from described species (Figure 3-3). The *Alphaproteobacteria*, consisting primarily of the Purple Bacteria, are represented by 14 sequences that cluster together. These bacteria fix carbon via anoxic phototrophy and often contain both a Form I and a Form II Rubisco in their genomes (Lu et al., 2010). These organisms primarily utilize the faster Form II Rubisco under high CO_2 conditions, and revert to using the Form I enzyme when CO_2 is limiting. The *Rhodopseudomonas spp.* have also been shown to utilize the Form II Rubisco as an electron sink during photoheterotrophic growth (Joshi et al., 2009). The sulfur oxidizing magnetotactic bacterium *Magnetospirillum magneticum* AMB-1 and the sulfur reducing hydrogen oxidizing deltaproteobacterium *Desulfovibrio aespoeensis* Aspo-2 also cluster with the Purple bacteria.

The stalk-forming iron oxidizing betaproteobacterium *Gallionella capsiferriformans* and the zetaproteobacterium *Mariprofundus ferooxydans* cluster together, however their Rubiscos are highly divergent to other organisms in the tree,

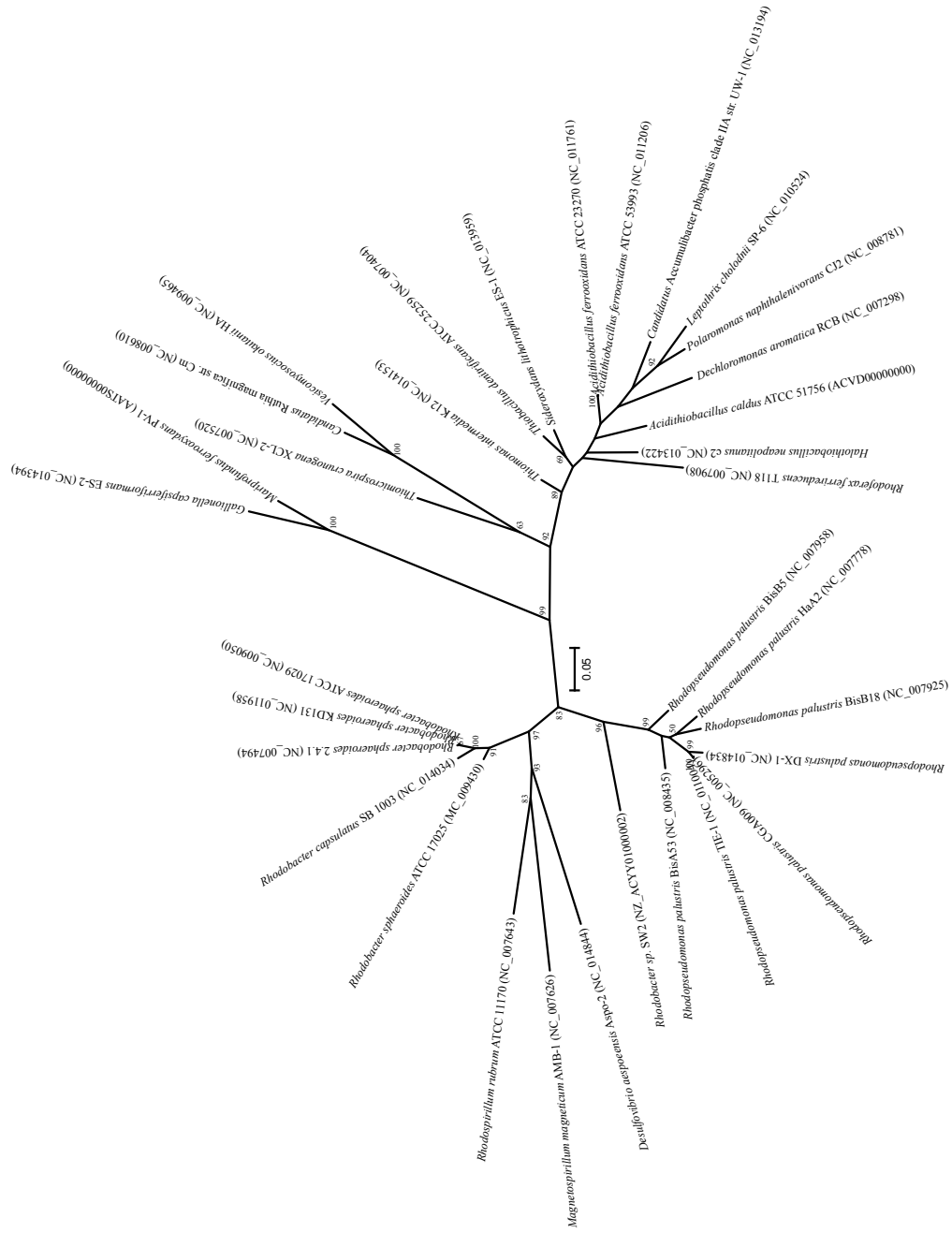

Figure 3-3. Unrooted maximum likelihood phylogenetic tree of the form II ribulose-1,5-bisphosphate carboxylase/oxygenase gene (cbbM).

including the other iron oxidizing betaproteobacteria *Sideroxydans lithotrophicus* and the sheath-forming *Leptothrix cholodnii* SP-6. The close clustering of these divergent Rubisco sequences from organisms that are known to have similar physiology and inhabit similar niches suggests lateral gene transfer may have occurred between ancestors of these two organisms (Emerson et al., 2013). *Sideroxydans lithotrophicus* clusters closely with the facultative anaerobic sulfur oxidizing bacterium *Thiobacullus denitrificans*.

The thioautotrophic intracellular bivalve symbionts *Vesicomyosocius okutanii* and *Candidatus* Ruthia magnifica cluster together along with the distantly related sulfur oxidizing bacterium *Thiomicrospira crunogena*. Other sulfur oxidizing bacteria such as the aerobic carboxysome-forming bacteria *Thiomonas intermedia* and *Halothiobacillus neapolitanus* do not closely cluster with any known organisms.

The Reductive Citric Acid Cycle

The reductive citric acid cycle, also called the reverse TCA cycle or the Arnon-Buchanan Cycle, was first proposed in 1966 for the anaerobic 'green sulfur' bacterium *Chlorobium limicola* (then called *C. thiosulfatophilium*) (Evans et al., 1966). The discovery of a semi-functional Rubisco in Chlorobium limicola cell extracts raised concerns that the proposed pathway was utilized for amino acid synthesis rather than as an autotrophic carbon fixation cycle (Buchanan and Sirevåg, 1976; Tabita et al., 1974). This controversy was finally settled in 1980 with careful use of ^{14}C intermediates, and the final unknown enzymatic step in the cycle was elucidated with the purification and characterization of ATP citrate lyase in 1982 (Fuchs et al., 1980). The genome of *C. limicola* has since been sequenced and it contains a copy of a Form IV Rubisco, which

may have been the protein that was purified by Tabita et al. (1974).

The cycle shares many of the intermediates and enzymes used in the TCA Cycle, only the reductive cycle runs in reverse with the eventual formation of acetyl-CoA and the consumption of energy and CO_2 (Figure 3-4). Three enzymes must be replaced in order for the cycle to be run in reverse. 1) The oxidation of succinate to fumarate by succinate dehydrogenase is considered to be an irreversible reaction due to the coupling of the $FADH_2$ bound to the enzyme being passed directly to the electron transport chain and is replaced by fumarate reductase in the reductive citric acid cycle (Evans et al., 1966). 2) The NAD^+ dependent decarboxylation of 2-oxoglutarate to form Succinyl-CoA is catalyzed by the enzyme 2-oxoglutarate synthase. 3) Citrate cleavage is performed either by ATP citrate lyase or via a two step cleavage utilizing citryl-CoA synthetase to form citryl-CoA which is then converted by citryl-CoA synthetase to oxaloacetate and acetyl-CoA (Antranikian et al., 1982; Aoshima, 2007; Robidart et al., 2008).

Each revolution of the reductive citric acid cycle costs three ATP, oxidizes one $FADH_2$ and one NADPH, oxidizes one ferrodoxin, and yields one acetyl-CoA molecule that can be converted to pyruvate with the cost of a second reduced ferrodoxin. The reductive citric acid cycle is generally restricted to microaerophilic or anaerobic bacteria due to oxygen sensitivity of 2-oxoglutarate synthase. The cycle seems to be less sensitive to high temperatures than the Calvin Cycle, with many hyperthermophilic aquificales utilizing the reductive citric acid cycle (Hügler et al., 2007).

Key enzymes in the reductive citric acid cycle

Of the three key enzymes found in the reductive citric acid cycle, ATP citrate

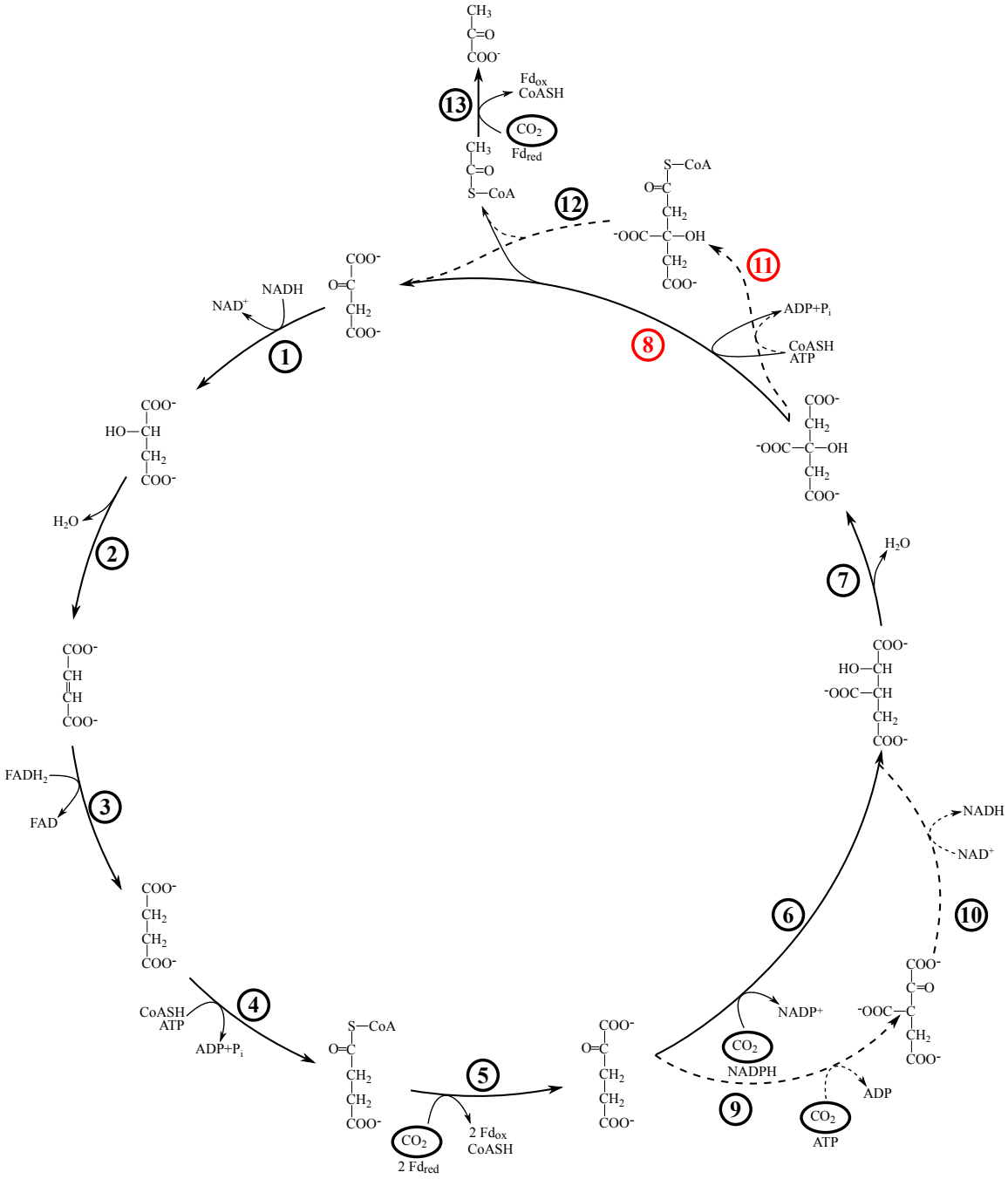

Figure 3-4. Schematic pathway showing the reductive TCA cycle. Enzymes are depicted in the figure as circled numbers: 1. malate dehydrogenase 2. fumarate hydratase 3. fumarate reductase 4. succinyl-CoA synthase 5. 2-oxogluarate synthase 6. isocitrate dehydrogenase 7. aconitase 8. ATP citrate lyase 9. 2-oxoglutarate carboxylase 10. oxalosuccinate reductase 11. citryl CoA synthase 12. citryl CoA lyase 13. pyruvate synthase. Cycle is based on Evans et al., (1966) and Hügler et al, (2007)

lyase (Figure 3-4 enzyme 8) is the gene that is most often used as an indicator of the presence of the cycle. The discovery of ATP citrate lyase in *Chlorobium limicola* solved the longstanding question of how citrate was cleaved in the reductive citric acid cycle (Antranikian et al., 1982; Ivanovsky et al., 1980). ATP citrate lyase catalyzes the formation of oxaloacetate and acetyl CoA from the cleavage of citrate with the cost of an ATP hydrolysis. *Eukaryotic* ATP citrate lyase is utilized in acetyl-CoA production for the biosynthesis of fats in all animals and some plants, fungi, and protists. The enzyme also requires a divalent metal as a cofactor, with Mg^{+2} having the greatest activity followed by Mn^{+2}, and some bacterial ATP citrate lyase enzymes can also utilize Cu^{+2} (Ivanovsky et al., 1980). ATP citrate lyase was found in the *Chlorobium*, *Epsilonproteobacteria* and the *Aquificales*, however an alternative citrate cleavage pathway was found in *Hydrogenobacter thermophilus* TK-S that utilizes a two-step process to regenerate oxaloacetate (Hügler et al., 2007). Citrate is first converted to citryl-CoA by citryl-CoA synthetase (Figure 3-4 enzyme 11) which is then cleaved to form oxaloacetate and acetyl-CoA by citryl-CoA lyase (Figure 3-4 enzyme 12). A 'Form II' ATP citrate lyase has been proposed from the genome of the alphaproteobacterium *Magnetococcus marinus*, however citrate cleavage has not been experimentally shown using this enzyme (Schübbe et al., 2009).

ATP citrate lyase phylogeny

Phylogenetic analysis of the ATP citrate lyase small subunit gene (*aclB*) revealed three distinct clusters (Figure 3-5). One cluster contains all the anaerobic phototrophic *Chlorobium* sequences and the aerobic nitrite-oxidizing nitrospirum *Candidatus*

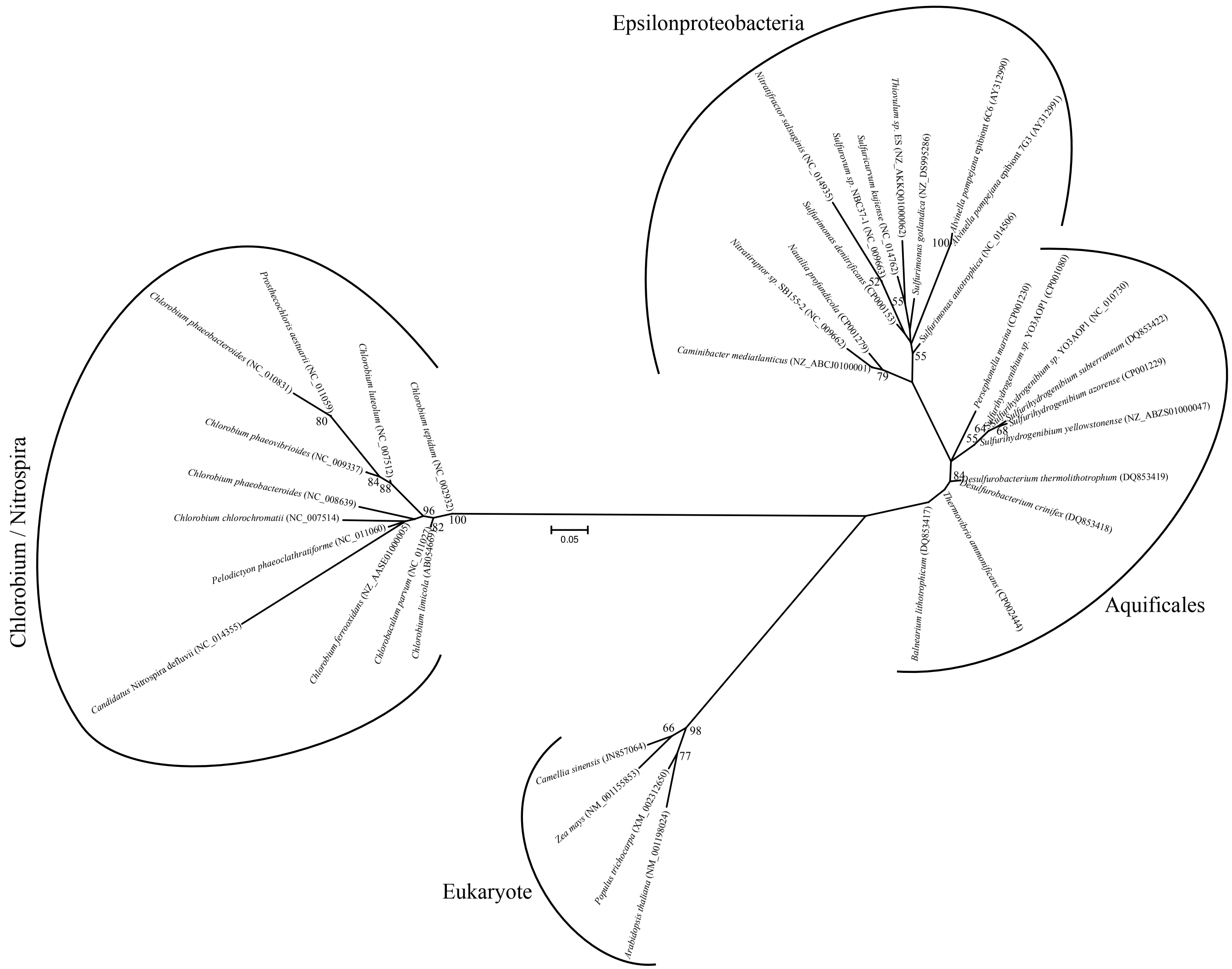

Figure 3-5. Unrooted maximum likelihood phylogenetic tree of the large subunit of the ATP Citrate Lyase gene (aclB).

Nitrospira defluvii. The genome of *Candidatus Nitrospira defluvii* was reconstructed from a mixed culture of activated sewage sludge and represents the only sequenced genome from this phylum. It is currently unknown what carbon fixation pathway other chemoautotrophic *Nitrospira* utilize. The remaining sequences in this group belong to the *Chlorobi* phylum, also known as the green sulfur bacteria. These anaerobic phototrophs couple photosynthesis with the oxidation of sulfur, hydrogen, or iron as electron sources. A second clade contains both the *Aquifacales* and the *Epsilonproteobaceria* sequences. The chemoautotrophic *Epsilonproteobacteria* can oxidize reduced sulfur or hydrogen compounds and are often either microaerophilic, such as the *Sulfuromonas* genus (Inagaki et al., 2003), or anaerobic such as the *Nautilia* genus (Miroshnichenko et al., 2002). These bacteria are often found in marine environments where reduced sulfur compounds come into contact with a potential oxidant, such as oxygen or nitrate. High biomass communities of *Epsilonproteobacteria* have been found in the suboxic zone of the Black and Baltic Seas (Lin et al., 2006) and as thick microbial mats at deep-sea hydrothermal vents (Huber et al., 2003). The *Aquificales* containing ATP citrate lyase are diverse thermophilic or hyperthermophilic organisms capable of utilizing various reduced molecules, such as hydrogen, sulfur, and iron, and are often microaerophilic or anaerobic nitrogen or sulfur reducers. The Eukaryotic ATP citrate lyase sequences branch near the Aquificales sequences and form a monophyletic cluster isolated from the bacterial aclB genes.

Phylogenetic analysis of the citryl-CoA synthetase alpha subunit (*ccsA*) indicates three distinct clusters (Figure 3-6). The *Aquificales* cluster includes members of the *Hydrogenobacter*, *Thermocrinis*, *Hydrogenobaculum*, *Hydrogenivirga*, and *Aquifex*

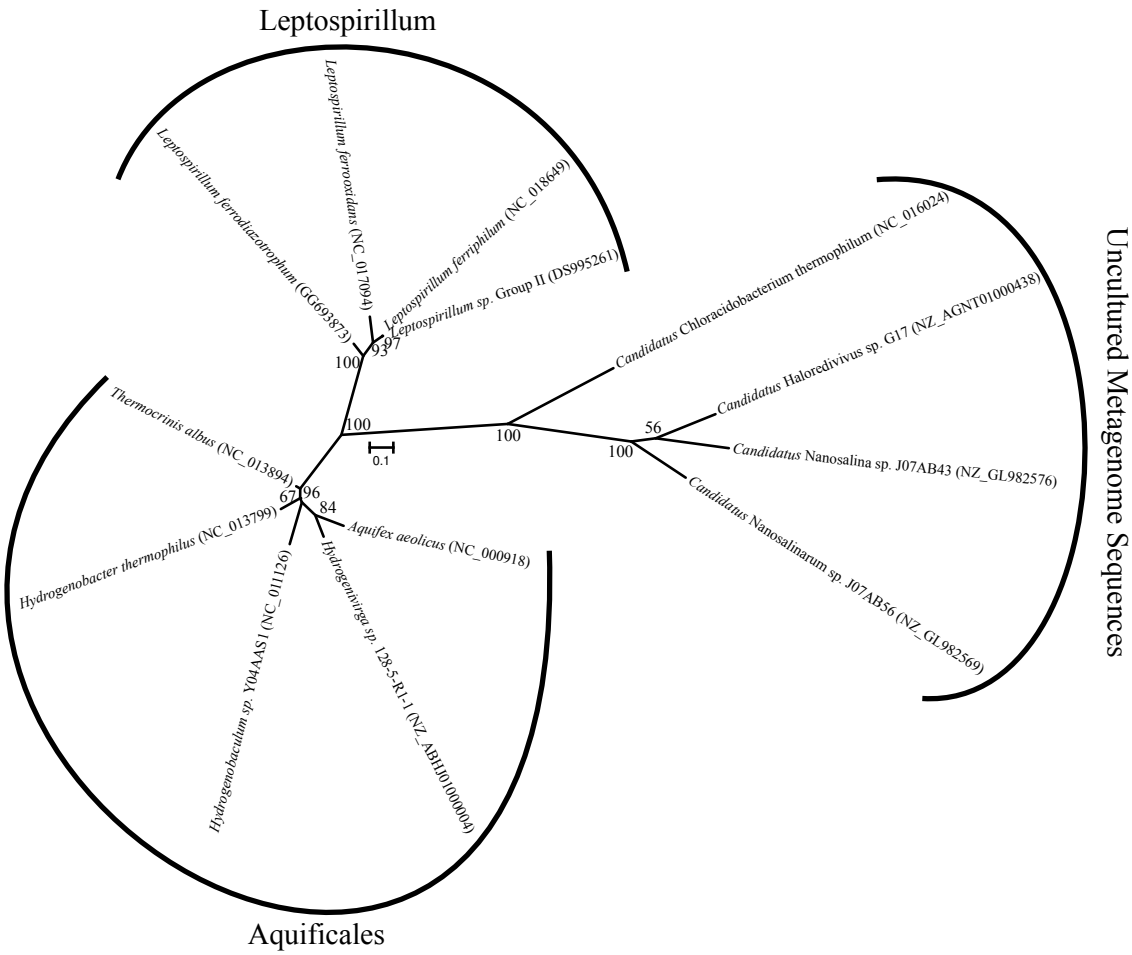

Figure 3-6. Unrooted maximum likelihood phylogenetic tree of the large subunit of the citryl-CoA synthetase alpha subunit (*ccsA*).

genera. This diverse group of *Aquificales* are all capable of growth by coupling hydrogen oxidation with oxygen reduction. A second clade includes the acidophilic iron-oxidizing *Nitrospira* bacteria belonging to the *Leptospirillum* genus. These isolates were originally thought to utilize the Calvin Cycle (Tyson et al., 2005), however the Rubisco found in previously sequenced Leptospirillum metagenomes is a Form IV Rubisco that is not used in the Calvin Cycle to fix inorganic carbon. The third cluster consists of metagenomic sequences of uncultured or enriched microbes. *Candidatus* Chloracidobacterium thermophilum is a photoheterotrophic bacterium belonging to the Acidobacteria phylum that was enriched from a photosynthetic microbial mat growing in Yellowstone National Park (Bryant et al., 2007). Three sequences are from uncultured members of the candidate class 'Nanohaloarchaea', which is a newly identified group within the Euryarchaeota phylum that is a dominant member of hypersaline waters. Analysis of draft genomes from this group of Archaea suggest they are aerobic heterotrophic archaea, and it is unknown if they are capable of autotrophic metabolism (Narasingarao et al., 2012).

The 'form II' ATP citrate lyase gene (Figure 3-7) was identified as a possible enzyme for citrate cleavage in the magnetotactic chemoautotroph *Magnetococcus marinus* (Schübbe et al., 2009). This bacterium has been shown to utilize the reductive TCA cycle, however the molecular mechanism of citrate cleavage was not deduced until the genome was sequenced. The 'form II' ATP citrate lyase has low sequence similarity to known, functional ATP citrate lyase genes. The enzymes have not been purified or expressed in a laboratory so citrate cleavage by this enzyme has not been experimentally shown. In addition to *Magnetococcus marinus*, the 'form II' ATP citrate lyase gene was found in symbionts of the *Vestimeniferan* tubeworms *Riftia pachyptila* and *Tevnia*

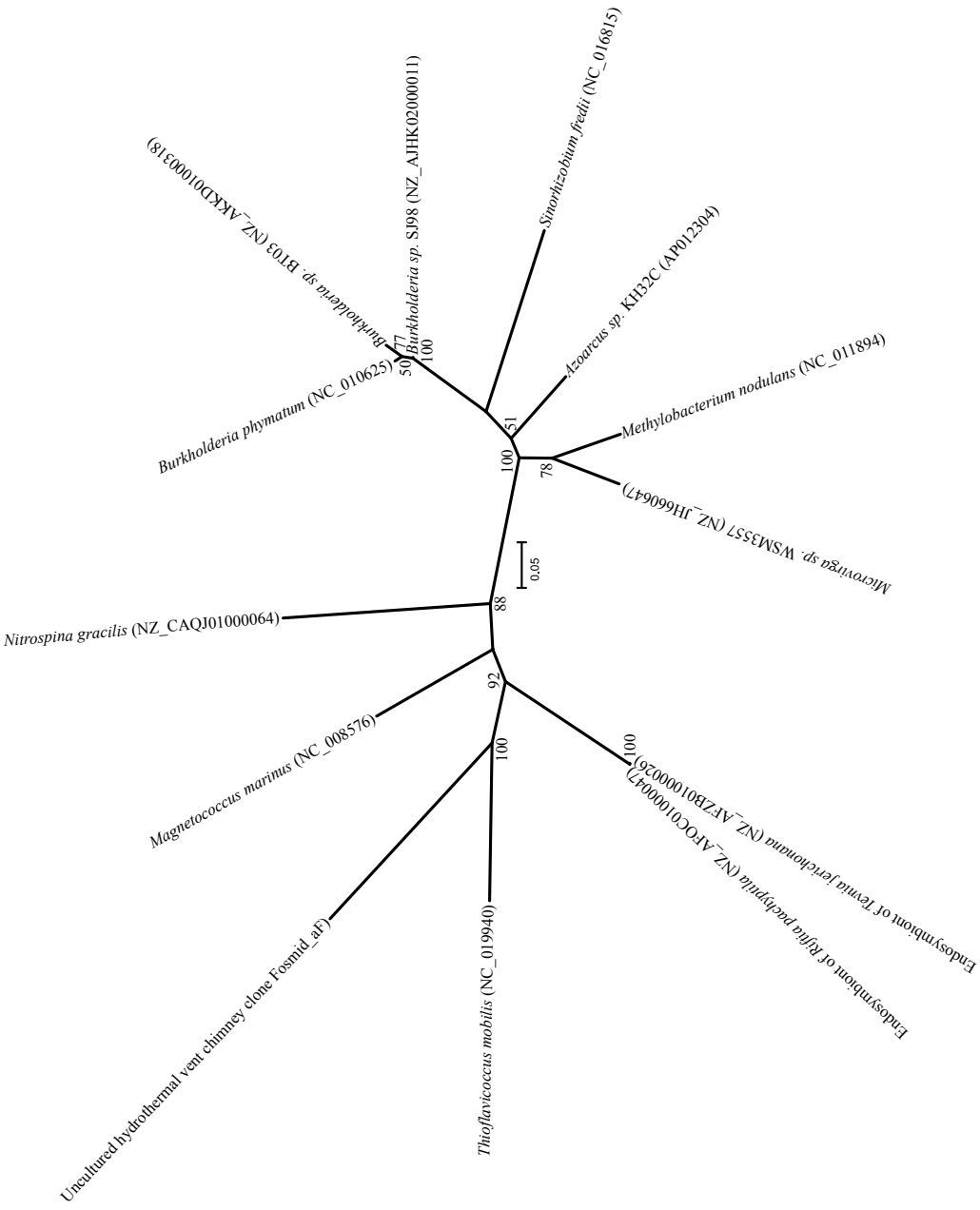

Figure 3-7. Unrooted maximum likelihood phylogenetic tree of the large subunit of the form II ATP Citrate Lyase gene (*aclB*).

jerichonana suggesting that these symbionts fix carbon using both the Calvin Cycle and the Reductive Citric Acid Cycle (Robidart et al., 2008). *Thioflavicoccus mobilis*, an anaerobic photosynthetic 'purple bacterium', also contains a copy of the 'form II' ATP Citrate Lyase and also contains a complete Calvin Cycle including a form II Rubisco gene. *Nitrospina gracilis* is a chemoautotrophic nitrite-oxidizing bacterium that has also been shown to utilize the Reductive Citric Acid Cycle (Lucker et al., 2013). This bacterium does not contain genes associated with the Calvin Cycle nor does it contain another form of ATP citrate lyase. Other 'Form II' sequences belong to rhizosphere-associated *Alphaproteobacteria* which do not have a defined autotrophic metabolism, however carbon monoxide, hydrogen, and C1 oxidation have been implicated in their genomes.

Reductive Acetyl-CoA Pathway

The reductive Acetyl-CoA pathway is unique among the carbon fixation pathways in that it is used by both *Bacteria* and *Archaea* and is an obligately anaerobic process. It is also the only carbon fixation pathway that is not a cyclic reaction, but is instead a linear pathway. The pathway was first elucidated by Lars Ljungdahl and Harland Wood using the bacterium *Moorella thermoaceticum* (formerly *Clostridium thermoaceticum*) which generates acetate from the reduction of carbon dioxide (Ljungdahl et al., 1965). The pathway was soon discovered in the sulfate reducing Bacteria as well as the methanogenic Archaea. Recently, the pathway has been found in the anaerobic ammonia-oxidizing “annomox” Planctomycetes which are ubiquitous anaerobic soil and sediment bacteria (Jetten et al., 2009).

The pathway operates by reducing CO_2 via two different pathways, which are then coupled together to form acetyl-CoA which is used to build biomass (Figure 3-8). The "methyl branch" of CO_2 reduction is variable between the *Bacteria* and the *Archaea*, however the end result is an activated methyl group attached to a cofactor. The acetogenic pathway involves the reduction of CO_2 to form free formate which is transferred to a carrier tetrahydrofolate where it is reduced to a methyl group. The archaeal pathway utilizes a methanofuran cofactor to fix CO_2 in the methyl branch rather than using formate, and the formate group is then passed to a tetrahydrobiopterin where it is also reduced to an activated methyl group, however the archaeal process utilizes different enzymes than the acetogenic pathway. The "carbonyl branch" of the pathway reduces CO_2 to a trapped carbonyl (CO) utilizing the nickel-containing enzyme Carbon Monoxide dehydrogenase/acetyl-CoA synthase which is conserved in all organisms using the reductive acetyl-CoA pathway (Ragsdale and Wood, 1991). This large, oxygen sensitive enzyme then catalyzes the formation of acetyl-CoA by combining the methyl group formed in the methyl branch with the CO formed in the carbonyl branch along with coenzyme-A which can be further utilized to fix a third CO_2 molecule to form pyruvate.

The energetic requirement of the pathway is different for the Bacteria and the Archaea due to different methyl branch pathways. The acetogenic bacterial pathway requires one NADPH and one NADH along with one ATP molecule. The methanogenic archaeal pathway does not require ATP or NAD-derived electron shuttles, but instead utilizes the oxidation of a reduced ferredoxin and the splitting of two H_2 molecules. The reduction of CO_2 to CO also requires the oxidation of a reduced ferredoxin group in all pathways. The pathway is also able to incorporate other C1 compounds, such as formate

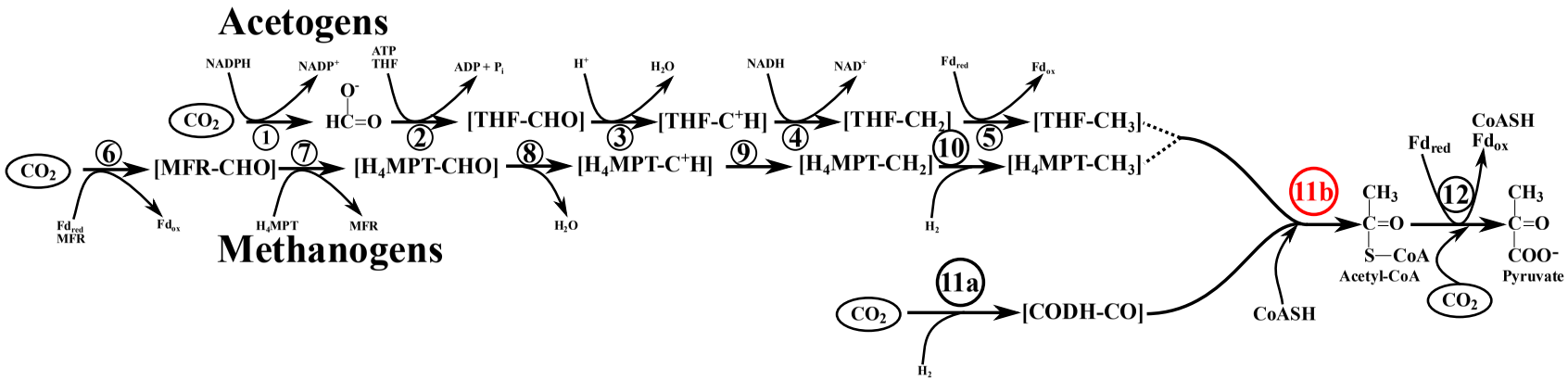

Figure 3-8. Schematic pathway showing the reductive acetyl-CoA pathway. Enzymes are depicted in the figure as circled numbers:1. Formate dehydrogenase 2. formyl-THF synthetase 3. formyl-MFR:tetrahydromethanopterin formyltransferase 4. methylene-THF dehyrogenase 5. methylene-THF reductase 6. formyl-MFR dehydrogenase 7. formyl-MFR:tetrahydromethanopterin formyltransferase 8. methenyl-tetrahydromethanopterin cyclohydrolase 9. methylene-tetrahydromethanopterin dehydrogenase 10. methylene-tetrahydromethanopterin reductase 11a. Carbon monoxide dehydrogenase 11b. Acetyl-CoA synthase 12. pyruvate synthase. Pathway is based on Ragsdale and Kumar (1996) and Berg (2011).

and methanol through intermediate pathways entering the reductive acetyl-CoA pathway (Hügler and Sievert, 2011). The pathway can also be run in reverse to oxidize organic molecules as a way to remove excess reducing potential in the cell.

Key enzyme of the reductive Acetyl-CoA pathway

The key enzyme in the reductive acetyl-CoA pathway is the carbon monoxide dehydrogenase / acetyl-CoA synthase enzyme (Figure 3-8 enzyme 11a and b) which is conserved in all organisms utilizing the pathway. The tertiary structure of the enzyme consists of a pair of dimers, with the pair of β-subunits catalyzing the reduction of CO_2 and the pair of α-subunits forming acetyl-CoA (Ragsdale and Kumar, 1996). Carbon monoxide oxidation/reduction occurs at a Ni-[3Fe-4S] active site on the β-subunit with additional Fe-S active sites passing the electrons to a ferrodoxin. The CO is trapped in a "gas tunnel" where it travels to the active site of the α-subunit where it is used to synthesize acetyl-CoA via a Ni-[4Fe-4S] active site (Amara et al., 2009). The α-subunit is used as a phylogenetic marker for this pathway due to its conservation among all organisms using the reductive acetyl-CoA pathway.

Acetyl-CoA synthase phylogeny

The α-subunit of the acetyl-CoA synthase gene (Figure 3-8 enzyme 11b) formed four large monophyletic clusters consisting of two *Firmicutes* clusters, one *Deltaproteobacteria* cluster, and one methanogenic cluster (Figure 3-9). The *Firmicute* cluster I are all acetogens which generally form acetate through the reduction of CO_2 coupled to the oxidation of hydrogen. The *Firmicute* cluster 2 contains the thermophilic

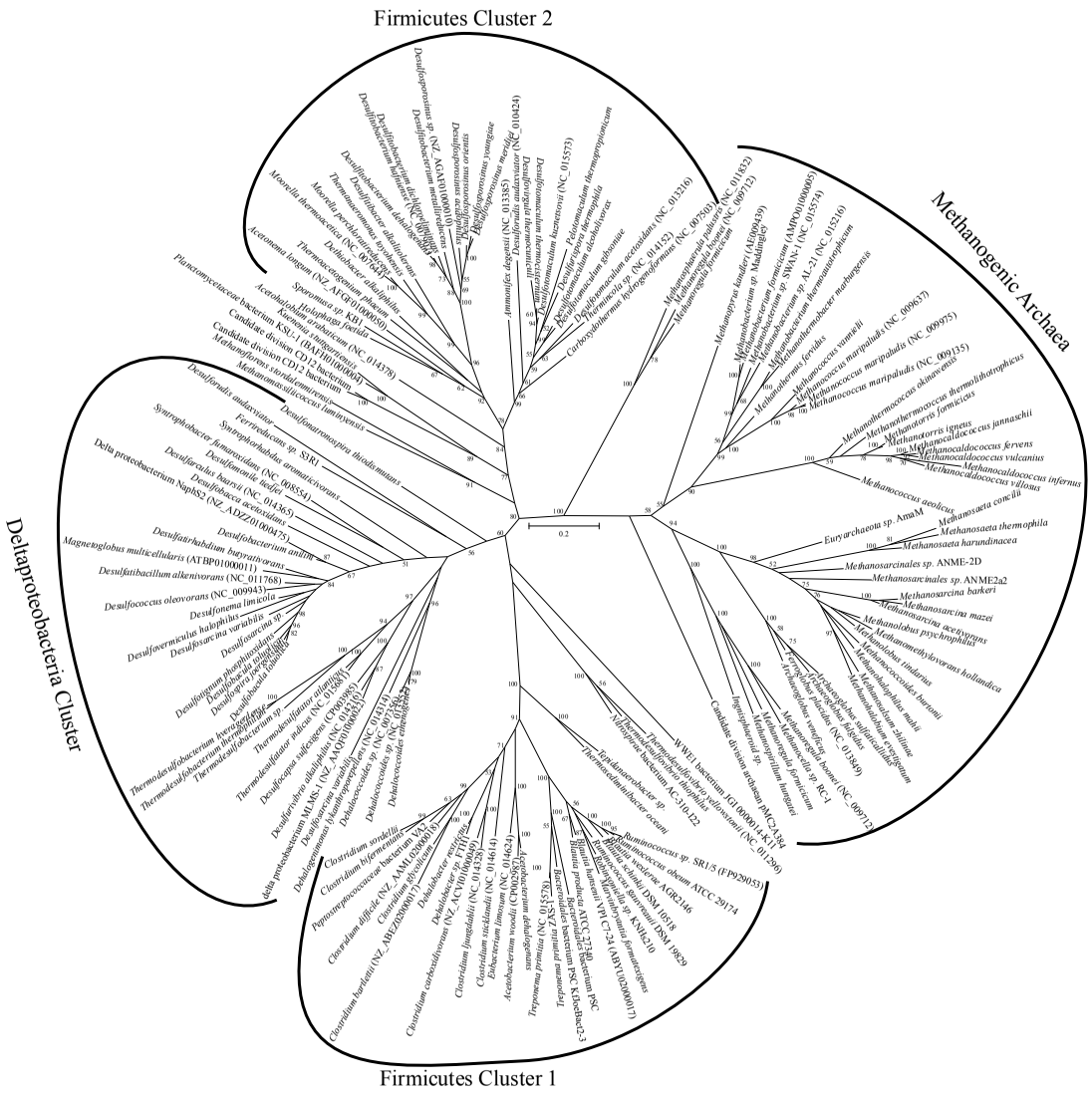

Figure 3-9. Unrooted maximum likelihood phylogenetic tree of the α-subunit of the carbon monoxide dehydrogenase / acetyl-CoA synthase gene (ACS).

acetogens and the "sulfate-reducing" *Firmicutes* which can generally oxidize hydrogen and reduce numerous compounds, including sulfate. The *Deltaproteobacteria* form a large monophyletic clade consisting primarily of sulfate-reducing hydrogen oxidizing isolates. Finally, the methanogenic *Archaea* form a large, monophyletic cluster which contains all of the methanogenic archaea with the exception of the "Rice Cluster II" methanogen *Methanoflorens stodalenmirensis* and the unique methanogen *Methanomassiliicoccus luminyensis (Gorlas et al., 2012)*, both of which do not form a strong cluster with any group. The cluster does not contain non-methanogenic *Bacteria* or *Archaea*. Other sequences that do not cluster within these groups are the anammox *Planctomycete* and sequences from candidate division "*Aerophobetes*" or CD12 that appear to be anaerobic hydrogen oxidizing bacteria (Rinke et al., 2013).

Taken together, the phylogeny of this gene supports the hypothesis of the reductive acetyl-CoA pathway being the most ancient carbon fixation pathway known. The pathway is less complex than the carbon fixation cycles and is the only carbon fixation pathway that is used by both *Bacteria* and *Archaea*, which suggests that the pathway predates the evolutionary events that gave rise to the three domains, and the lack of apparent horizontal gene transfer of the Acetyl-CoA Synthase gene between the Archaea and the Bacteria supports this theory. The key genes in the pathway also utilize metal centers such as nickel that are thought to be ancient in origin and predate the oxygen-tolerant enzymes that are utilized in the aerobic carbon fixation cycles (Righter, 2003).

The 3-Hydroxypropionate Bi-Cycle

The three hydroxypropionate cycle was discovered in 1989 by Helge Holo in the anaerobic phototroph *Chloroflexus aurantiacus* as a cycle which formed glyoxylate from fixing two CO_2 molecules (Holo, 1989). The pathway was later refined by the lab of Georg Fuchs who made the discovery that the cycle was actually a bi-cycle where the glyoxylate feeds into a second cycle where it is converted to pyruvate and acetyl-CoA (Zarzycki et al., 2009). This second cycle is necessary because glyoxylate is not a central metabolite in the *Chloroflexi*.

The cycle requires five ATP and four NADPH per pyruvate formed, making the pathway highly energetically costly compared to the other carbon fixation pathways utilized by anaerobes (Figure 3-10). The pathway appears to be intact in many photoheterotrophic *Chloroflexi*, however *Chloroflexus aurantiacus* is the only isolate able to grow chemoautotrophically (Zarzycki and Fuchs, 2011). The other Chloroflexi are only able to grow mixotrophically requiring some organic carbon for growth. One theory for this observation is that the photosynthetic Chloroflexi were once capable of autotrophic growth, but they have adapted to rely on photoheterotrophy due to the increased growth yield of using this metabolism (Hügler and Sievert, 2011). The lack of autotrophic Chloroflexi using this pathway may also be a result of culturing bias, as many of the pure chloroflexi cultures were originally isolated using media that selects for photoheterotrophs.

Key genes of the 3-Hydroxypropionate bi-cycle

The key genes for the 3-Hydroxypropionate bi-cycle include malonyl-CoA

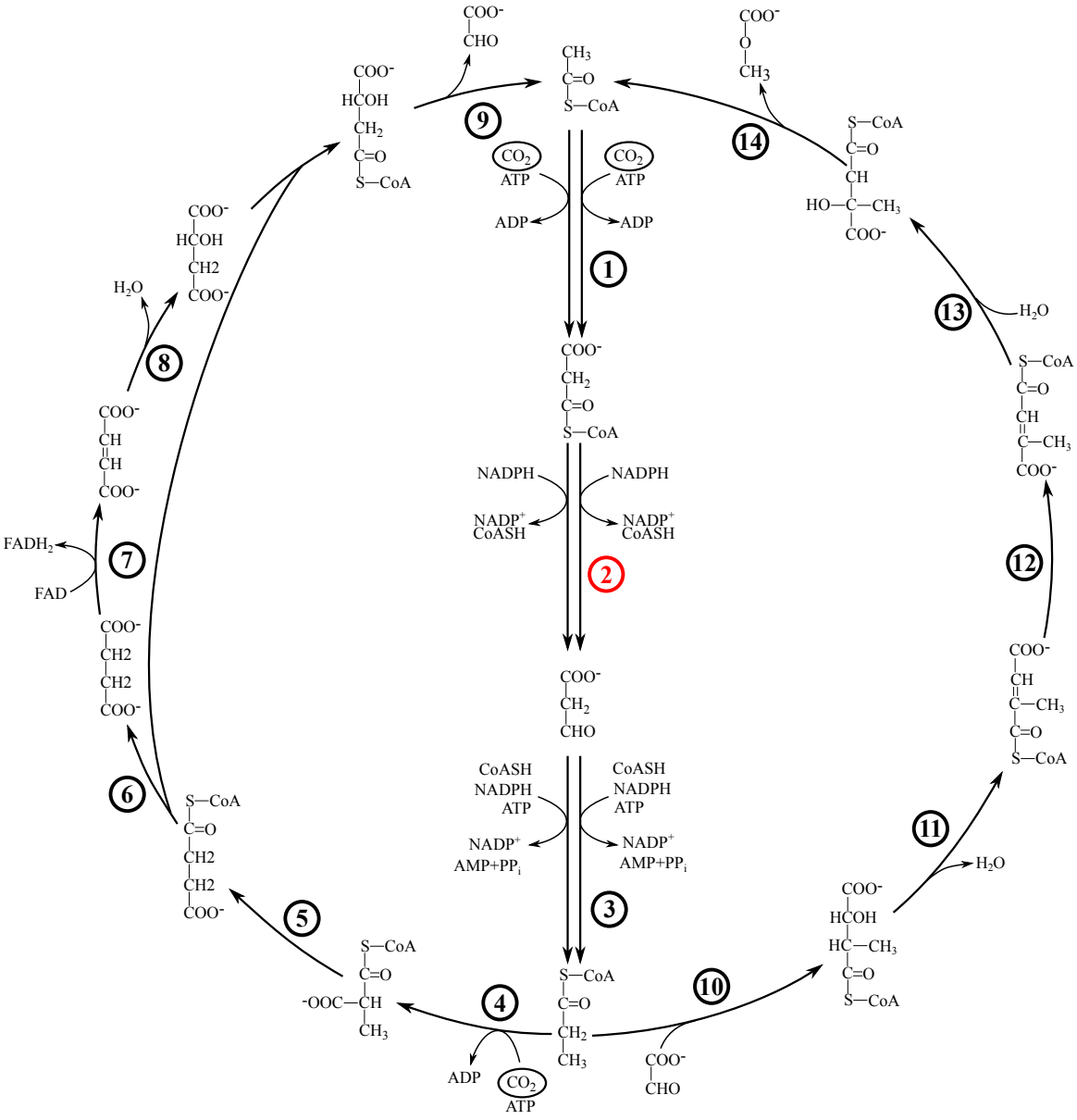

Figure 3-10. Schematic pathway showing the 3-hydroxypropionate bi-cycle. Enzymes are depicted in the figure as circled numbers: 1. acetyl-CoA carboxylase 2. malonyl-CoA reductase 3. propionyl-CoA synthase 4. propionyl-CoA carboxylase 5. methylmalonyl-CaA epimerase 6. syccinyl-CoA:malate CoA transferase 7. succinate dehydrogenase 8. fumarate hydratase 9. MMC lyase 10. MMC lyase 11. mesaconyl-C1-CoA 12. mesaconyl-CoA C1-C4 CoA transferase 13. mesaconyl-C4-CoA hydratase 14. MMC lyase. Cycle is based on Zarzycki et al., (2009).

reductase (Figure 3-10 enzyme 2), propionyl-CoA synthase (Figure 3-10 enzyme 3), and MMC lyase (Figure 3-10 enzyme 10 and 14). Malonyl-CoA reductase has been proposed to be the key gene for the bi-cycle due to its specificity for the bi-cycle (Zarzycki et al., 2009). The enzyme is bifunctional in that it is both an aldehyde dehydrogenase and an alcohol dehydrogenase. In the 3- hydroxypropionate bi-cycle, malonyl-CoA reductase converts malonyl-CoA to 3-hydroxypropionate using 2 NADPH and malonate semialdehyde as an intermediate. The bifunctional properties of the enzyme is unique, as other pathways which convert malonyl-CoA to 3-hydroxypropionate use two enzymes, a malonyl-CoA reductase and a malonate semialdehyde reductase.

Malonyl-CoA reductase phylogeny

Malonyl-CoA reductase (Figure 3-10 enzyme 2) was only found in 20 genomes, with 7 of them in the Chloroflexi phylum (Figure 3-11). *Chloroflexus auranticus* clustered with *Chloroflexus aggregans* and two other *Chloroflexi*. The chemoautotrophic chloroflexi *Oscillochloris trichoides* is capable of carbon fixation, however it uses the Calvin Cycle. Two *Roseiflexus* chloroflexi isolates also contain all genes necessary for the 3-hydroxypropionate bi-cycle, however they have never been successfully grown on inorganic carbon alone.

One striking feature of the malonyl-CoA reductase tree is that all of the described isolates containing the gene are capable of anoxygenic photosynthesis or are heavily pigmented. *Candidatus* Chloracidobacterium thermophilum is a recently isolated phototrophic acidobacterium that grows photoheterotrophically as does the *Porphyrobacter* and *Blastomonas* groups of *Alphaproteobacteria*. It is currently unknown

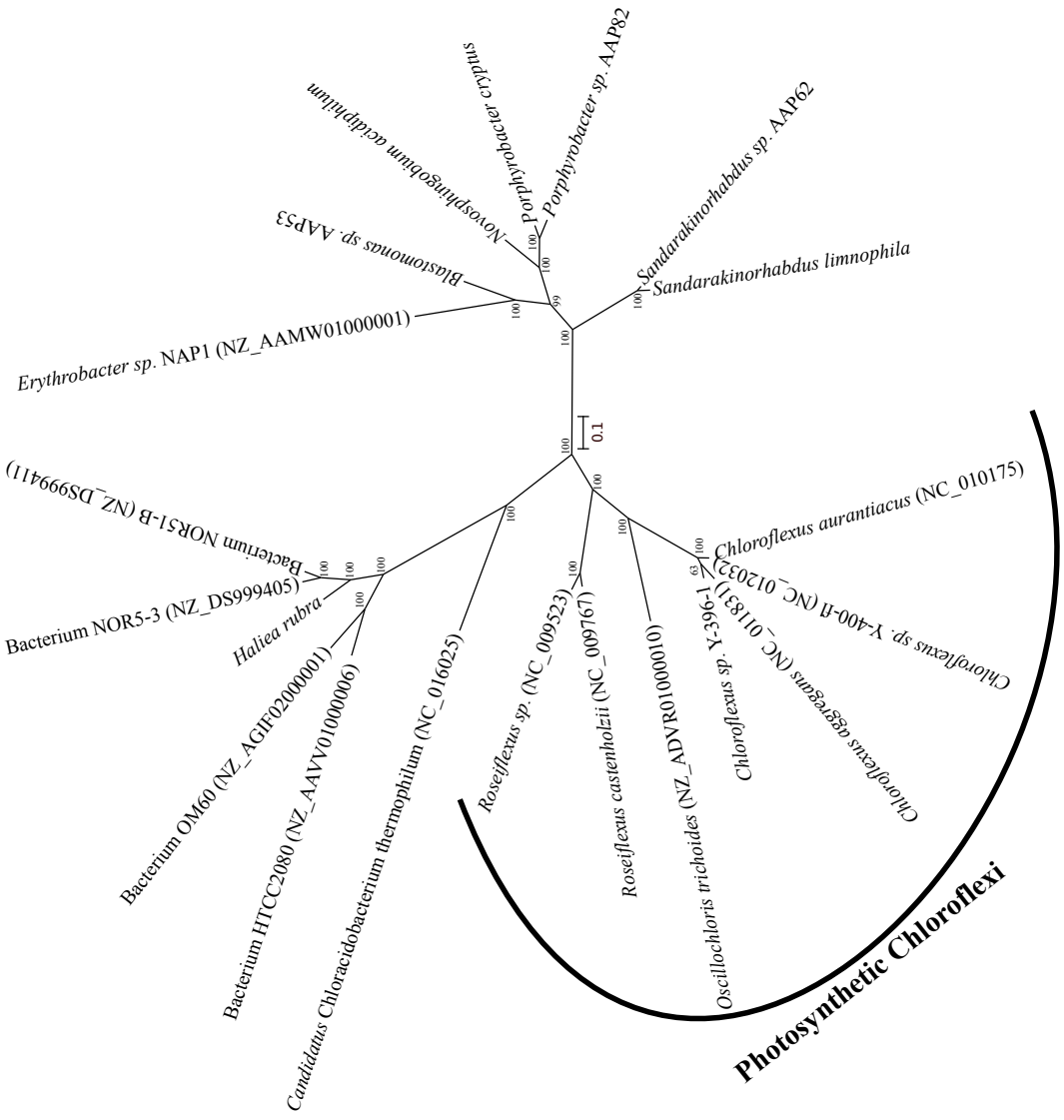

Figure 3-11. Unrooted maximum likelihood phylogenetic tree of the malonyl-CoA reductase gene.

if these isolates grow autotrophically using the 3-Hydroxypropionate bi-cycle or if the cycle is used as a method of mixotrophic growth with some carbon intermediates generated through heterotrophic processes.

3-Hydroxypropionate/4-Hydroxybutyrate Cycle

The 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle was discovered by Ivan Berg in 2007 using the microaerophilic Crenarchaea groups Sulfolobus, Acidianus, and Metallosphaera (Berg et al., 2007). The cycle was found to use many of the intermediates used in the 3-hydroxypropionate bi-cycle, but forms succinate semialdehyde from succinyl-CoA which is then ultimately converted to two molecules of acetyl-CoA, one of which leaves the cycle to be used to build biomass.

The cycle uses four ATP and five NADPH for each acetyl-CoA formed. In addition to the crenarchaotes that use the cycle, recent work has shown that the chemoautotrophic *Thaumarchaeota* also use this cycle to fix carbon suggesting that these abundant organisms have adapted the cycle to the oxic conditions these archaea inhabit (La Cono et al., 2010; Hallam et al., 2006; Figure 3-12). Key genes from this pathway are shared with the dicarboxylate/4-hydroxybutyrate cycle and will be discussed after the next section.

Dicarboxylate/4-Hydroxybutyrate Cycle

The dicarboxylate/4-hydroxybutyrate cycle is the latest carbon fixation pathway that has been discovered, being elucidated by Georg Fuchs in 2008 using the obligate anaerobe *Ignicoccus hospitalis (Huber et al., 2008)*. This cycle appears to be a hybrid

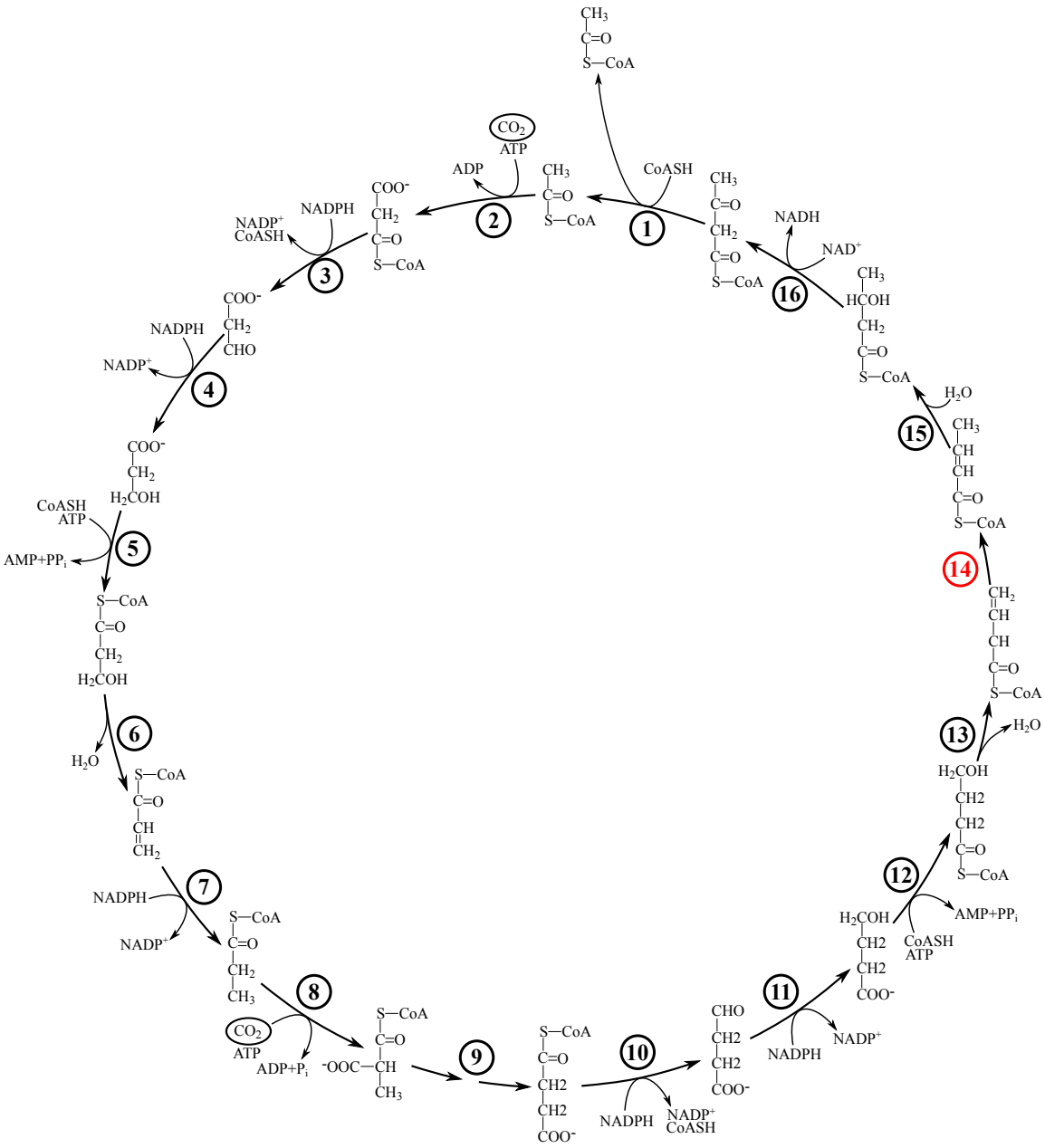

Figure 3-12. Schematic pathway showing the 3-hydroxypropionate/4-hydroxybutyrate cycle. Enzymes are depicted in the figure as circled numbers: 1. acetoacetyl-CoA C-acetyltransferase 2. acetyl-CoA carboxylase 3. malonyl-CoA reductase 4. malonic semialdehyde reductase 5. 3-hydroxypropionate-CoA ligase 6. 3-hydroxypropionyl-CoA dehydratase 7. acryloyl-CoA reductase 8. acetyl-CoA/propionyl-CoA carboxylase 9. methylmalonyl-CoA epimerase and mutase 10. succinyl-CoA reductase 11. succinic semialdehyde reductase 12. 4-hydroxybutyrate-CoA ligase 13. 4-hydroxybutyryl-CoA syntase 14. 4-hydroxybutyryl-CoA dehydratase 15. crotonyl-CoA hydratase 16. 3-hydroxybutyryl-CoA dehydrogenase. Pathway is based on Berg (2011).

cycle which includes many of the intermediates of the 3HP/4HB cycle along with the reductive TCA cycle (Figure 3-13). The cycle converts acetyl-CoA to pyruvate, and eventually succinyl-CoA utilizing many of the same enzymes used in the reductive TCA cycle. The succinyl-CoA is then converted to succinic semialdehyde where it uses the same pathway as the 3HP/4HB cycle to generate two molecules of acetyl-CoA.

The cycle requires three ATP, one NADH, one FADH2, and two NADPH for each acetyl-CoA generated. The cycle is currently thought to be used by the obligately anaerobic Crenarchaeota groups *Desulfurococcales* and *Thermoproteales (Berg et al., 2010)*. Many of these organisms were once thought to use the reductive TCA cycle to fix carbon, however it has now been shown that only Bacteria utilize the reductive TCA cycle.

Key genes of the 3-Hydroxypropionate/4-Hydroxybutyrate and Dicarboxylate/4-Hydroxybutyrate cycles

There is not a unique key enzyme for either the 3HP/4HB or the Dicarboxylate/4-Hydroxybutyrate cycles, however there is one key enzyme found in both cycles that can be used as an indicator for these autotrophic cycles. The key gene for the 3HP/4HB and Dicarboxylate/4-Hydroxybutyrate cycles has been proposed to be 4-hydroxybutyryl-CoA dehydratase (Figure 3-12 enzyme 14; Figure 3-13 enzyme 12) which catalyzes the conversion of 4-hydroxybutyryl-CoA to crotonyl-CoA in both cycles (Huber et al., 2008). The enzyme has a [4Fe-4S] active site and is also utilized by anaerobic Bacteria which form vinylacetyl-CoA as part of a fermentation pathway. The gene has also been found in marine bacteria where crotonyl-CoA is formed as an intermediate in

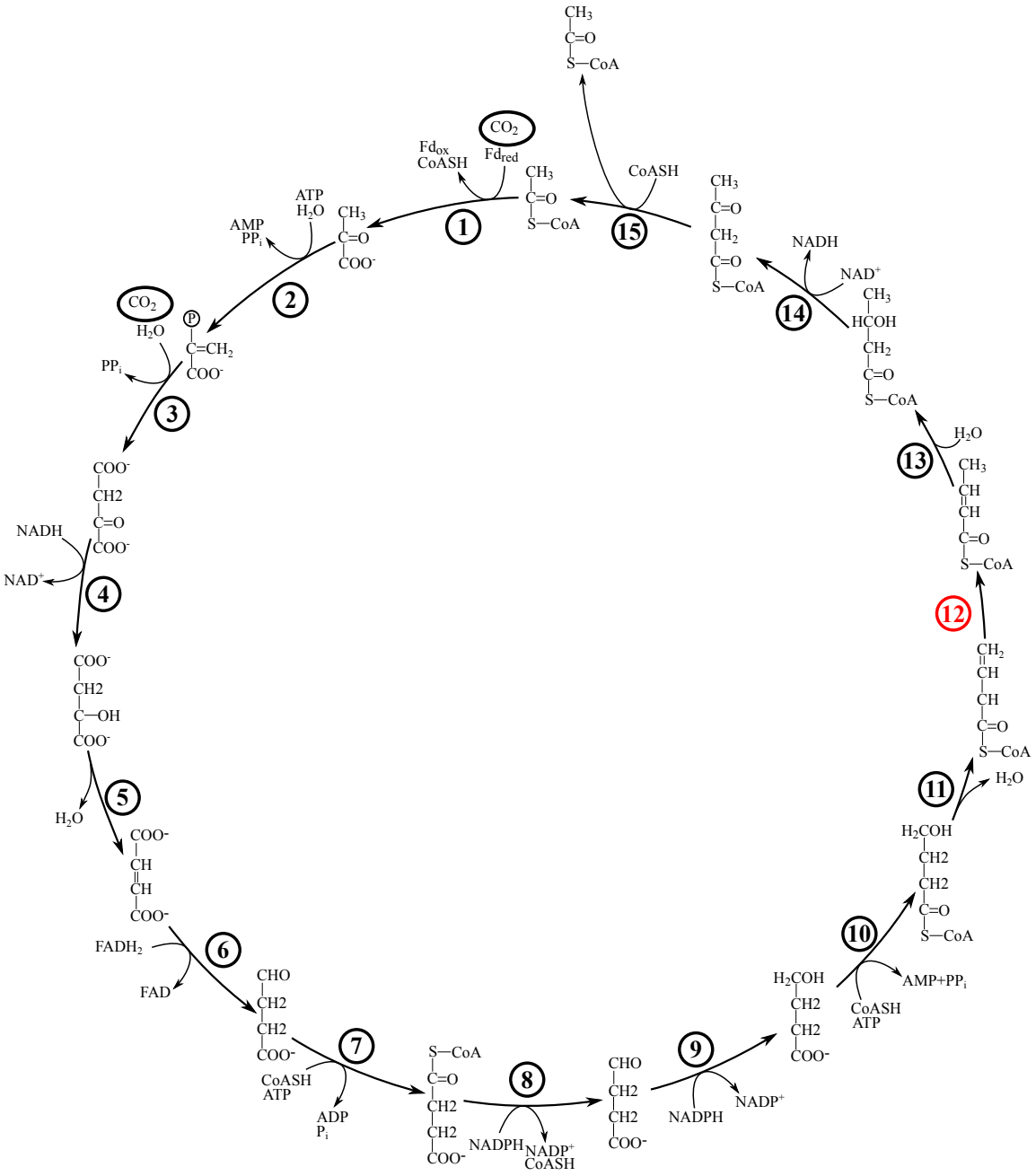

Figure 3-13. Schematic pathway showing the dicarboxylate/4-hydroxybutyrate cycle. Enzymes are depicted in the figure as circled numbers: 1. acetoacetyl-CoA C-acetyltransferase 2. pyruvate synthase 3. PEP carboxylase 4. malate dehydrogenase 5. fumarate hydratase 6. fumarate reductase 7. succinyl-CoA synthetase 8. succinyl-CoA reductase 9. succinic semialdehyde reductase 10. 4-hydroxybutyrate-CoA ligase 11. 4-hydroxybutyryl-CoA syntase 12. 4-hydroxybutyryl-CoA dehydratase 13. crotonyl-CoA hydratase 14. 3-hydroxybutyryl-CoA dehydrogenase. Cycle is based on Huber et al., (2008)

dimethylsulfonopropionate (DMSP) metabolism (Schneider et al., 2012).

4-hydroxybutyryl-CoA Dehydratase Phylogeny

The 4-hydroxybutyryl-CoA dehydratase gene was found in four phylogenetically distinct groups, the thermophilic Archaea, the *Thaumarchaeota*, a large cluster of anaerobic *Firmicutes*, and two small clusters of marine proteobacteria (Figure 3-14). The thermophilic archaea formed two distinct clusters, the *Sulfolobus* which use the 3HP/4HB cycle and another cluster that includes the *Thermoproteales* and *Desulfurococcales* which use the dicarboxylate/4-Hydroxybutyrate cycle. *Archaeoglobus fulgidus* also contained a 4-hydroxybutyryl-CoA dehydratase gene, however this archaean has been shown to fix carbon using the reductive acetyl-CoA pathway (Klenk et al., 1997). This gene is probably used in a similar pathway as the anaerobic *Firmicutes* which use it in the aminobutyrate fermentation pathway. The gene is also found in some marine proteobacteria, however it is unknown what function this gene has in these organisms. The 4-hydroxybutyryl-CoA dehydratase gene is also found in the chemoautotrophic *Thaumarchaeota* isolates and single-cell genomes. This abundant group may represent a large portion of the fixed organic carbon in the deep ocean, however it is not known what proportion of the *Thaumarchaeota* are autotrophic, and if the autotrophic isolates are obligate chemoautotrophs (Berg et al., 2010).

Concluding Remarks

Recent advances in DNA sequencing technology has resulted in hundreds of autotrophic bacteria being sequence in recent years. This, along with the discovery of

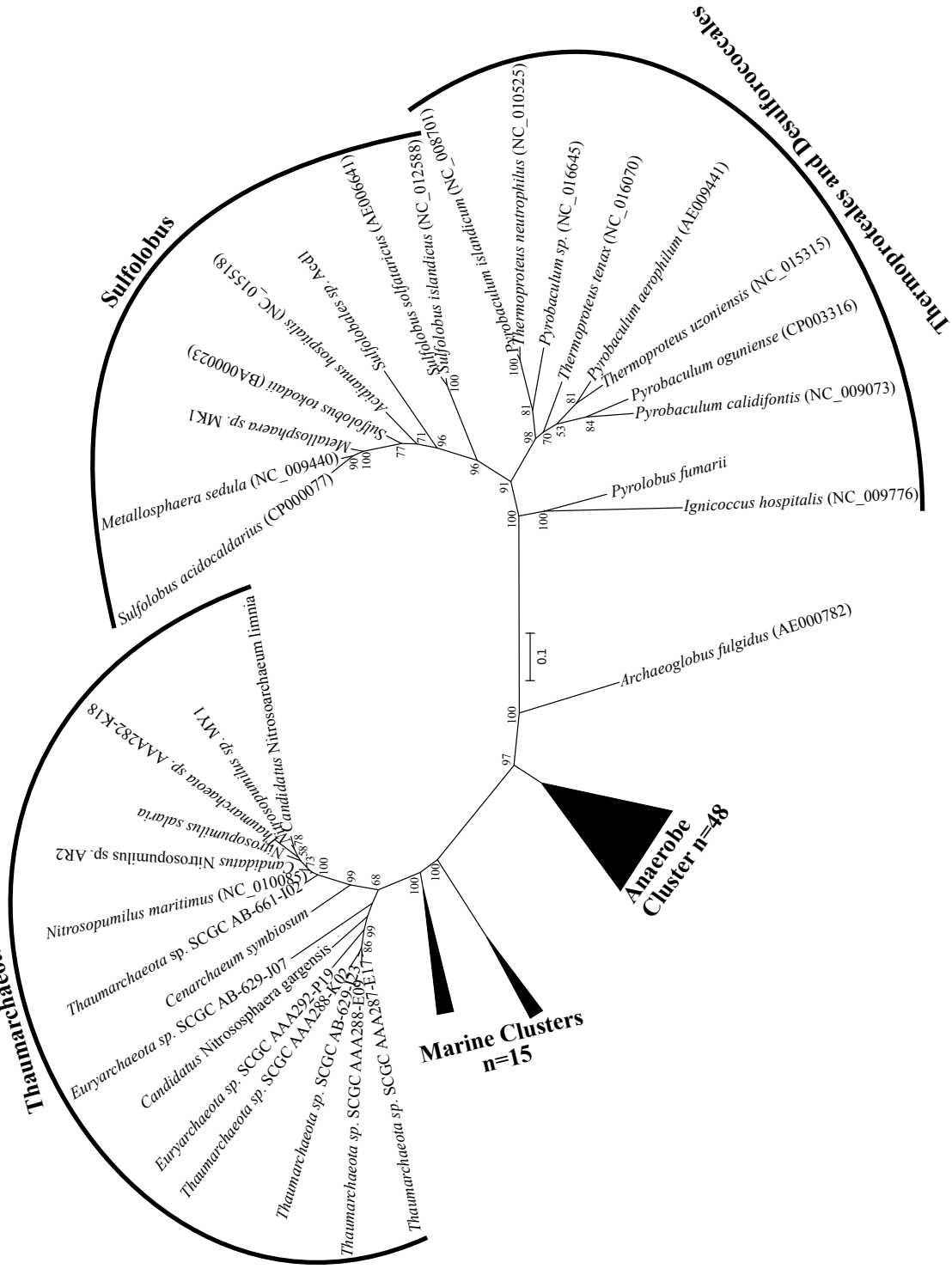

Figure 3-14. Unrooted maximum likelihood phylogenetic tree of the 4-hydroxybutyryl-CoA dehydratase gene.

new carbon fixation pathways over the last decade has resulted in a great leap in our understanding of the diversity of organisms capable of fixing inorganic carbon, as well as a greater understanding of the mechanisms of carbon fixation. With a few exceptions, the sequences used in this analysis were sourced from bacteria that have been isolated and maintained in a pure culture. This presents a bias toward bacteria that we are simply able to cultivate, and may not represent the true diversity of microorganisms in the natural environment.

While current culturing techniques are attempting to target new isolates that are environmentally relevant, culture-free methods of genomics may lead to new understanding of carbon fixation in the natural environment. Environmental genomics (metagenomics) and single-cell sequencing have the potential to provide unbiased mechanistic and phylogenetic information about the metabolic potential of uncultured, but dominant members of chemoautotrophic populations. This information should lead to new discoveries of previously unknown autotrophic organisms along with potentially discovering entirely new autotrophic pathways.

Works Cited

Amara, P., Cavazza, C., Fontecilla-Camps, J.C., Nicolet, Y., and Volbeda, A. (2009). Structure-function relationships of anaerobic gas-processing metalloenzymes. Nature *460*, 814-822.

Anandham, R., Gandhi, P.I., Kwon, S.W., Sa, T.M., Kim, Y.K., and Jee, H.J. (2009). Mixotrophic metabolism in *Burkholderia kururiensis* subsp. *thiooxydans* subsp. nov., a facultative chemolithoautotrophic thiosulfate oxidizing bacterium isolated from rhizosphere soil and proposal for classification of the type strain of *Burkholderia kururiensis* as *Burkholderia kururiensis* subsp. *kururiensis* subsp. nov. Arch. Microbiol. *191*, 885–894.

Andersson, I., and Backlund, A. (2008). Structure and function of Rubisco. Plant Physiol. Biochem. *46*, 275–291.

Antranikian, G., Herzberg, C., and Gottschalk, G. (1982). Characterization of ATP citrate lyase from *Chlorobium limicola*. J. Bacteriol. *152*, 1284–1287.

Aoshima, M. (2007). Novel enzyme reactions related to the tricarboxylic acid cycle: phylogenetic/functional implications and biotechnological applications. Appl. Microbiol. Biotechnol. *75*, 249–255.

Bassham, J.A., Barker, S.A., Calvin, M., and Quarck, U.C. (1956). Intermediates in the photosynthetic cycle. Biochim. Biophys. Acta *21*, 376–377.

Berg, I.A., Kockelkorn, D., Buckel, W., and Fuchs, G. (2007). A 3-Hydroxypropionate/4-Hydroxybutyrate Autotrophic Carbon Dioxide Assimilation Pathway in Archaea. Science *318*, 1782–1786.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., Alber, B.E., and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. Nat Rev Micro *8*, 447–460.

Bonjour, F., and Aragno, M. (1984). Bacillus tusciae, a new species of thermoacidophilic, facultatively chemolithoautotrophic hydrogen oxidizing sporeformer from a geothermal area. Arch. Microbiol. *139*, 397–401.

Bryant, D.A., Costas, A.M.G., Maresca, J.A., Chew, A.G.M., Klatt, C.G., Bateson, M.M., Tallon, L.J., Hostetler, J., Nelson, W.C., Heidelberg, J.F., et al. (2007). Candidatus *Chloracidobacterium thermophilum*: An Aerobic Phototrophic *Acidobacterium*. Science *317*, 523–526.

Buchanan, B.B., and Sirevåg, R. (1976). Ribulose 1,5-diphosphate carboxylase and *Chlorobium thiosulfatophilum*. Arch. Microbiol. *109*, 15–19.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., FitzGerald, L.M., Clayton, R.A., Gocayne, J.D., et al. (1996). Complete Genome Sequence of the Methanogenic Archaeon, *Methanococcus jannaschii*. Science *273*, 1058–1073.

Chen, Z., and Spreitzer, R.J. (1992). How various factors influence the CO2/O2 specificity of ribulose-1,5-bisphosphate carboxylase/oxygenase. Photosynth. Res. *31*, 157–164.

La Cono, V., Smedile, F., Ferrer, M., Golyshin, P.N., Giuliano, L., and Yakimov, M.M. (2010). Genomic signatures of fifth autotrophic carbon assimilation pathway in bathypelagic *Crenarchaeota*. Microb. Biotechnol. *3*, 595–606.

Ellis, R.J. (1979). The most abundant protein in the world. Trends Biochem. Sci. *4*, 241–244.

Emerson, D., Field, E.K., and Woyke, T. (2013). Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. Front. Evol. Genomic Microbiol. *4*, 254.

Evans, M.C., Buchanan, B.B., and Arnon, D.I. (1966). A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. Proc. Natl. Acad. Sci. *55*, 928–934.

Fuchs, G., Stupperich, E., and Eden, G. (1980). Autotrophic CO2 fixation in *Chlorobium limicola*. Evidence for the operation of a reductive tricarboxylic acid cycle in growing cells. Arch. Microbiol. *128*, 64–71.

Gibson, J.L., and Tabita, F.R. (1977). Isolation and preliminary characterization of two forms of ribulose 1,5-bisphosphate carboxylase from *Rhodopseudomonas capsulata*. J. Bacteriol. *132*, 818–823.

Gorlas, A., Robert, C., Gimenez, G., Drancourt, M., and Raoult, D. (2012). Complete Genome Sequence of *Methanomassiliicoccus luminyensis*, the Largest Genome of a Human-Associated Archaea Species. J. Bacteriol. *194*, 4745–4745.

Grundy, F.J., and Henkin, T.M. (1998). The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in Gram-positive bacteria. Mol. Microbiol. *30*, 737–749.

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser *41*, 95–98.

Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and DeLong, E.F. (2006). Pathways of Carbon Assimilation and Ammonia Oxidation Suggested by Environmental Genomic Analyses of Marine *Crenarchaeota*. PLoS Biol *4*, e95.

Hartman, F.C., and Harpel, M.R. (1994). Structure, Function, Regulation, and Assembly of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. Annu. Rev. Biochem. *63*, 197–232.

Holo, H. (1989). *Chloroflexus aurantiacus* secretes 3-hydroxypropionate, a possible intermediate in the assimilation of CO2 and acetate. Arch. Microbiol. *151*, 252–256.

Hou, S., Makarova, K.S., Saw, J.H., Senin, P., Ly, B.V., Zhou, Z., Ren, Y., Wang, J., Galperin, M.Y., Omelchenko, M.V., et al. (2008). Complete genome sequence of the extremely acidophilic methanotroph isolate V4, *Methylacidiphilum infernorum*, a representative of the bacterial phylum *Verrucomicrobia*. Biol. Direct *3*, 26.

Huber, H., Gallenberger, M., Jahn, U., Eylert, E., Berg, I.A., Kockelkorn, D., Eisenreich, W., and Fuchs, G. (2008). A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum *Ignicoccus hospitalis*. Proc. Natl. Acad. Sci. *105*, 7851–7856.

Huber, J.A., Butterfield, D.A., and Baross, J.A. (2003). Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption. FEMS Microbiol Ecol *43*, 393–409.

Hügler, M., and Sievert, S.M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Annu. Rev. Mar. Sci. *3*, 261–289.

Hügler, M., Huber, H., Molyneaux, S.J., Vetriani, C., and Sievert, S.M. (2007). Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum *Aquificae*: evidence for two ways of citrate cleavage. Environ. Microbiol. *9*, 81–92.

Inagaki, F., Takai, K., Kobayashi, H., Nealson, K.H., and Horikoshi, K. (2003). Sulfurimonas autotrophica gen. nov., sp. nov., a novel sulfur-oxidizing epsilon-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. Int J Syst Evol Microbiol *53*, 1801–1805.

Ivanovsky, R.N., Sintsov, N.V., and Kondratieva, E.N. (1980). ATP-linked citrate lyase activity in the green sulfur bacterium *Chlorobium limicola* forma *thiosulfatophilum*. Arch. Microbiol. *128*, 239–241.

Ivanovsky, R.N., Fal, Y.I., Berg, I.A., Ugolkova, N.V., Krasilnikova, E.N., Keppen, O.I., Zakharchuc, L.M., and Zyakun, A.M. (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.

Jetten, M.S.M., Niftrik, L. van, Strous, M., Kartal, B., Keltjens, J.T., and Op den Camp, H.J.M. (2009). Biochemistry and molecular biology of anammox bacteria. Crit. Rev. Biochem. Mol. Biol. *44*, 65–84.

Joshi, G.S., Romagnoli, S., VerBerkmoes, N.C., Hettich, R.L., Pelletier, D., and Tabita, F.R. (2009). Differential Accumulation of Form I RubisCO in Rhodopseudomonas palustris CGA010 under Photoheterotrophic Growth Conditions with Reduced Carbon Sources. J. Bacteriol. *191*, 4243–4250.

Klenk, H.-P., Clayton, R.A., Tomb, J.-F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., et al. (1997). The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. Nature *390*, 364–370.

Kusel, K., Dorsch, T., Acker, G., and Stackebrandt, E. (1999). Microbial Reduction of Fe(III) in Acidic Sediments: Isolation of *Acidiphilium cryptum* JF-5 Capable of Coupling the Reduction of Fe(III) to the Oxidation of Glucose. Appl Env. Microbiol *65*, 3633–3640.

Lin, X., Wakeham, S.G., Putnam, I.F., Astor, Y.M., Scranton, M.I., Chistoserdov, A.Y., and Taylor, G.T. (2006). Comparison of Vertical Distributions of Prokaryotic Assemblages in the Anoxic Cariaco Basin and Black Sea by Use of Fluorescence *In Situ* Hybridization. Appl. Environ. Microbiol. *72*, 2679–2690.

Ljungdahl, L., Irion, E., and Wood, H.G. (1965). Total Synthesis of Acetate from CO2. I. Co-Methylcobyric Acid and Co-(Methyl)-5-methoxybenzimidazolylcobamide as Intermediates with *Clostridium thermoaceticum.* Biochemistry (Mosc.) *4*, 2771–2780.

Lorite, M.J., Tachil, J., Sanjuán, J., Meyer, O., and Bedmar, E.J. (2000). Carbon Monoxide Dehydrogenase Activity in *Bradyrhizobium japonicum*. Appl. Environ. Microbiol. *66*, 1871–1876.

Lu, Y.-K., Marden, J., Han, M., Swingley, W., Mastrian, S., Chowdhury, S., Hao, J., Helmy, T., Kim, S., Kurdoglu, A., et al. (2010). Metabolic flexibility revealed in the genome of the cyst-forming alpha-1 proteobacterium *Rhodospirillum centenum*. BMC Genomics *11*, 325.

Lucker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013). The Genome of *Nitrospina gracilis* Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. Front. Microbiol. 4:27. doi: 10.3389/fmicb.2013.0002

Markowitz, V.M. (2006). The integrated microbial genomes (IMG) system. Nucleic Acids Res. *34*, D344–D348.

Masuda, S., Eda, S., Sugawara, C., Mitsui, H., and Minamisawa, K. (2010). The cbbL gene is required for thiosulfate-dependent autotrophic growth of *Bradyrhizobium japonicum*. Microbes Environ. JSME *25*, 220–223.

Miroshnichenko, M.L., Kostrikina, N.A., L'Haridon, S., Jeanthon, C., Hippe, H., Stackebrandt, E., and Bonch-Osmolovskaya, E.A. (2002). Nautilia lithotrophica gen. nov., sp. nov., a thermophilic sulfur-reducing epsilon-proteobacterium isolated from a deep-sea hydrothermal vent. Int J Syst Evol Microbiol *52*, 1299–1304.

Narasingarao, P., Podell, S., Ugalde, J.A., Brochier-Armanet, C., Emerson, J.B., Brocks, J.J., Heidelberg, K.B., Banfield, J.F., and Allen, E.E. (2012). De novo metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. ISME J. *6*, 81–93.

Ragsdale, S.W., and Kumar, M. (1996). Nickel-Containing Carbon Monoxide Dehydrogenase/Acetyl-CoA Synthase. Chem Rev *96*, 2515–2540.

Ragsdale, S.W., and Wood, H.G. (1991). Enzymology of the Acetyl-CoA Pathway of CO_2 Fixation. Crit. Rev. Biochem. Mol. Biol. *26*, 261–300.

Righter, K. (2003). Metal-Silicate Partitioning of Siderophile Elements and Core Formation in the Early Earth. Annu. Rev. Earth Planet. Sci. *31*, 135–174.

Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.-F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature. *499*. 431-437.

Robidart, J.C., Bench, S.R., Feldman, R.A., Novoradovsky, A., Podell, S.B., Gaasterland, T., Allen, E.E., and Felbeck, H. (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. Environ. Microbiol. *10*, 727–737.

Santiago, B., Schübel, U., Egelseer, C., and Meyer, O. (1999). Sequence analysis, characterization and CO-specific transcription of the cox gene cluster on the megaplasmid pHCG3 of *Oligotropha carboxidovorans*. Gene *236*, 115–124.

Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal Type III RuBisCOs Function in a Pathway for AMP Metabolism. Science *315*, 1003–1006.

Schneider, K., Asao, M., Carter, M.S., and Alber, B.E. (2012). *Rhodobacter sphaeroides* Uses a Reductive Route via Propionyl Coenzyme A To Assimilate 3-Hydroxypropionate. J. Bacteriol. *194*, 225–232.

Schübbe, S., Williams, T.J., Xie, G., Kiss, H.E., Brettin, T.S., Martinez, D., Ross, C.A., Schüler, D., Cox, B.L., Nealson, K.H., et al. (2009). Complete Genome Sequence of the Chemolithoautotrophic Marine Magnetotactic Coccus Strain MC-1. Appl. Environ. Microbiol. *75*, 4835–4852.

Seaborg, G.T., and Benson, A.A. (2008). Melvin Calvin. 8 April 1911 — 8 January 1997. Biogr. Mem. Fellows R. Soc. *54*, 59–70.

Stamatakis, A. (2014). RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics btu033.

Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynth. Res. *60*, 1–28.

Tabita, F.R., McFadden, B.A., and Pfennig, N. (1974). d-ribulose-1,5-bisphosphate carboxylase in Chlorobium thiosulfatophilum Tassajara. Biochim. Biophys. Acta BBA - Enzymol. *341*, 187–194.

Tyson, G.W., Lo, I., Baker, B.J., Allen, E.E., Hugenholtz, P., and Banfield, J.F. (2005). Genome-Directed Isolation of the Key Nitrogen Fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an Acidophilic Microbial Community. Appl. Environ. Microbiol. *71*, 6319–6324.

Whelan, S., and Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Biol. Evol. *18*, 691 –699.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., et al. (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. Science *337*, 1661–1665.

Zarzycki, J., and Fuchs, G. (2011). Coassimilation of Organic Substrates via the Autotrophic 3-Hydroxypropionate Bi-Cycle in *Chloroflexus aurantiacus*. Appl. Environ. Microbiol. *77*, 6181–6188.

Zarzycki, J., Brecht, V., Müller, M., and Fuchs, G. (2009). Identifying the missing steps of the autotrophic 3-hydroxypropionate CO2 fixation cycle in *Chloroflexus* aurantiacus. Proc. Natl. Acad. Sci. *106*, 21317–21322.

Chapter 4: Carbon Fixation in Iron-Cycling Microbial Mats From Loihi Seamount, HI.

Introduction

The deep sea (>1000 mbsl) is an extremely oligotrophic environment characterized by organic carbon starvation resulting in extremely low biomass in both planktonic and sediment microbial communities (Edwards et al., 2005; Jannasch and Taylor, 1984). In contrast, deep-sea hydrothermal vents support high-biomass microbial and animal communities that are sustained by primary production derived from chemoautotrophic carbon fixation (Jannasch and Mottl, 1985; Nakagawa and Takai, 2008). These microbes harvest energy from the reduced chemicals in the hydrothermal fluid and reduce CO_2 into biological carbon. There are six known carbon fixation pathways, all of which may be utilized by different microbes at hydrothermal vents (Berg, 2011, see chapter 2).

The majority of carbon fixation studies at hydrothermal vents have been focused on sites dominated by sulfur cycling and hydrogen oxidizing microbes (Campbell et al., 2003; Crépeau et al., 2011) which found key enzymes from both the Calvin and the rTCA Cycles. Using ATP citrate lyase (aclB) as an indicator for the rTCA cycle, the above studies found sequences closely related to the *Epsilonproteobacteria* and the *Aquificales* groups of bacteria. The *Epsilonproteobacteria* dominate most hydrothermal vent microbial mats where sulfur or hydrogen are thought to be the primary electron donors (Campbell et al., 2006). All cultured *Epsilonproteobacteria* and *Aquificale* chemoautotrophic bacteria utilize the rTCA Cycle for carbon fixation (Nakagawa and Takai, 2008). Other known chemoautotrophic sulfur-oxidizing bacteria isolated at hydrothermal vents, including the *Gammaproteobacteria* such as *Thiomicrospira* and

Beggiatoa that utilize the Calvin Cycle. Uncultured sulfur oxidizing *Gammaproteobacterial* symbionts and epibionts of hydrothermal vent macrofauna also utilize the Calvin cycle, with the exception of the symbiont within the *Vestimentiferan* tubeworm *Riftia pachyptila* which contains and utilizes both the Calvin and rTCA Cycles (Robidart et al., 2008). All methanogenic archaea isolated from hydrothermal vents utilize the Wood-Ljungdahl Pathway. *Deltaproteobacteria* which couple hydrogen oxidation with sulfate reduction have been found to utilize both the rTCA Cycle and the Wood-Ljungdahl Pathway, while all chemoautotrophic thermophilic and hyperthermophilic sulfate reducing archaea utilize the Wood-Ljungdahl Pathway (Hügler and Sievert, 2011). Uncultured archaea within the *Thaumarchaeota* have also been found at hydrothermal vent communities. Although a *Thaumarchaeota* has not been cultured from a hydrothermal vent, cultured representatives from other environments are obligate chemoautotrophic ammonium oxidizers utilizing the 3-hydroxypropionate/4-hydroxybutyrate cycle (Berg et al., 2010).

Carbon fixation at circumneutral iron-dominated hydrothermal vents has been studied far less than vents with high sulfur and hydrogen concentrations. Many iron oxide-encrusted microbial mats have been found to be dominated by the *Zetaproteobacteria* class of bacteria (Davis and Moyer, 2008; Emerson et al., 2007). The only confirmed chemoautotrophic iron-oxidizing bacterium from a hydrothermal vent is *Mariprofundus ferrooxydans*, which belongs to the *Zetaproteobacteria* clade (Emerson et al., 2007). *M. ferrooxydans* utilizes the Calvin Cycle for carbon fixation, and contains a copy of both form I and form II large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco) enzymes (Singer et al., 2011). There has only been one

study of carbon fixation in an iron-dominated microbial mat to date at the Fryer Hydrothermal Field, Mariana Backarc (Elsaied and Naganuma, 2001). That study utilized primers for the cbbL and cbbM genes, however the primers used in the study were designed utilizing only five total sequences, and would not amplify the majority of Rubisco genes from the environment, including those from *M. ferrooxydans*.

In this study, I examined the carbon fixation pathways which may be utilized in iron cycling microbial mats from Loihi Seamount, Hawaii. These microbial mats have been shown to be dominated by Zetaproteobacteria, along with lesser populations of *Epsilonproteobacteria*, *Gammaproteobacteria*, *Thaumarchaeota*, and uncultued *Euryarchaeota* using culture-independent PCR-based analysis. Taken together, these groups contain chemoautotrophs that potentially utilize four pathways of carbon fixation. I utilized PCR-based analysis of these communities using newly designed primers for Rubisco and aclB genes in multiple mats with either genomic DNA (gDNA) or complementary DNA (cDNA) as a template. I also utilized a metagenome from Loihi Seamount to find key genes for all six carbon fixation pathways. This is the first study to perform a comprehensive survey of carbon fixation pathways at a hydrothermal vent system utilizing

Materials and Methods

Sample collection

Samples were collected from vents at Markers 34, 36, and 39 using the suction sampler from the ROV Jason II. This sampler collects bulk material over an area of approximately 1.5 m^2 around the vent orifice. The suction sampler produces a

homogenized sample nearly encompassing the entire microbial mat. The mat collected at Pohaku was sampled with a PVC scoop sampler that was designed to sample a smaller area of mat than the suction sampler, while further minimizing any cross-contamination from other samples or surface seawater. The sampler is constructed with 3" PVC pipe and is washed with 70% ethanol before being rinsed and filled with filter-sterilized deionized water and sealed with a ball valve before deployment. The valve is opened immediately prior to sampling and closed directly after the sample collected and remains sealed until the sample is recovered on the ship. All samples were allowed to settle at 4°C for approximately 2 hours before the overlying seawater was decanted and the mat material was frozen at -80°C until DNA was extracted in the lab.

DNA extraction

Total genomic DNA (gDNA) was extracted from each sample in duplicate using the FastDNA Spin Kit for Soil following the manufacturer's protocol (Qbiogene, Irvine, CA). Extracted gDNA from each sample was pooled, cleaned, and concentrated using Montáge PCR centrifugal filter devices (Millipore, Bedford, MA). The gDNAs were then quantified using a Nanodrop ND-1000 spectrophotometer and were diluted to 10 ng DNA/ml using filter sterilized 10 mM Tris 0.1 mM EDTA (pH 8.0).

RNA extraction

Total RNA extraction was performed using a modified Trizol protocol with all procedures being performed at 4°C with RNAse-free reagents and plastic. Approximatedly 0.5 g of mat material (wet weight) was added to a 2.0 ml screw top tube

along with 0.01 g polyvinylpyrrolidone (Sigma-Aldridge), 0.2 g Chelex (Biorad), and 0.2 g sterilized 0.1 mm “zirconia”/silica beads (Biospec Products). One milliliter of cold Trizol was added to the tube and was immediately homogenized in a FastPrep instrument (MP Biomedicals) for 60 seconds at 5.5 m/sec. The tubes were then incubated for 5 minutes with gentle mixing and centrifuged for 10 minutes. The supernatant was then transferred to two 2.0 ml tubes, and the mat material was extracted with bead beating a second time using another 1.0 ml Trizol. Tubes with the Trizol supernatant were then diluted with an additional 1.5 ml Trizol and 0.3 ml chloroform was added to each tube and incubated for 3 minutes with gentle mixing. The tubes were then centrifuged for 15 minutes and the top aqueous phase was extracted and pooled from each tube. The extract was extracted again by adding 0.3 ml acid phenol-chloroform (Life Technologies: pH 4.2) to the pooled extract and incubated for 5 minutes with gentle shaking. The aqueous phase was then transferred to a new tube and the acid phenol-chlorororm extraction was repeated. The extract was then purified by adding 0.3 ml chloroform and 0.1 volume 3M sodium acetate to the tube, incubating the tube for 3 minutes with gentle shaking, and centrifuging the tube for 15 minutes. The aqueous phase was then removed and the chloroform/sodium acetate extraction was repeated. Total RNA was then precipitated overnight at 20°C using 0.7 volume isopropanol, 0.1 volume 3M sodium acetate, and 3 µl glycogen (Life Technologies). The tubes were then centrifuged for 30 minutes to pellet the RNA and washed with 70% ethanol. The RNA pellet was then air dried and dissolved in 30µl DEPC-treated water. The RNA was then treated with TURBO DNase (Life Technologies) to remove any remaining DNA contamination. Complementary DNA (cDNA) was synthesized using the SuperScript III first-strand synthesis system (Life

Technologies) using 300 ng of RNA and random hexamers as the annealing mix.

Primer design and optimization

All Form I and Form II Rubisco and Form I aclB (ATP citrate lyase) DNA sequences were exported from the Integrated Microbial Genomes system (IMG) website using the homology search function on well-described Rubisco and aclB sequences representative of the phylogenetic diversity of the gene. Sequences from environmental samples and isolates not found in IMG were downloaded from the NCBI nr database using blastn. The sequences were temporarily translated and aligned using the multiple sequence alignment editor BioEdit. The sequences were then reverse-translated back to the original nucleic acid sequence and imported into the ARB software environment. Primers were designed with the goal to maximize the length of the amplicon while minimizing the degeneracy needed to make the primers as inclusive to each group as possible. One primer set was designed for the aclB gene which amplifies a ~1kb fragment of the gene termed 178F- CCNGAYATGYTNTTTGGWAA and 1314R-CCNWNYTCRTARTTWGGNCC. Three primers were needed to amplify the majority of Rubisco sequences. One set of primers was designed for the Form II Rubisco 484F-GGNACNATCATCAARCCNAA and 1146R-TTNTCRAAGAARCCNGGNA. Two primers were designed for the form I Rubisco, TIGF- CARCCNTTCMWRCGBTGG and TIGR-GTNCCDCCDCCRAAYTG and another pair TIRF-AAGGAYGACGAGAACATCAAYT and TIRR- AAYCGSRTNGCSCTSGA.

PCR reactions were optimized using genomic DNA from *Mariprofundus ferrooxydans* for the Form II and TIGF and TIGR primers, and *Aurantimonas*

manganoxydans gDNA was used to optimize the TIRF and TIRR primers. The aclB primers were optimized with a genomic DNA sample from NW Rota-1 volcano that is nearly a pure community of Epsilonproteobacteria that utilize the rTCA cycle to fix carbon, along with a cloned environmental aclB fragment from the same sample.

PCR amplification and clone library construction, and sequencing of Rubisco and aclB genes.

PCR was performed in triplicate 20 µl reactions in a mix containing 400 µM dNTPs, 0.5 µM each primer, 20 ng of template DNA or cDNA, 10 µg molecular biology grade BSA and 1 µL Phire polymerase. PCR cycling was performed with a 98°C hot start for 30 s, followed by 30 cycles of denaturation at 98°C for 10 s, 60°C for 20 s, and 72°C for 30 s, and the cycle was completed with a final extension at 72°C for 60 s. Positive PCR amplifications were excised from a 1.5% agarose gel and purified with a GeneJET gel extraction kit (Thermo Fisher) following the manufacturer's protocol. The PCR products were ligated into the pJET1.2/blunt cloning vector using the CloneJET PCR cloning kit (Thermo Fisher) at a 5:1 insert to vector ratio. The plasmids were transformed into competent E. coli cells (Active Motif) and spread onto LB-Amp plates. The resultant clone library was produced by randomly picking clones and transferring them onto a grid plate. Forty eight clones were sequenced from each clone library using the primers pJetF and pJetR which anneal to the plasmid 30-50 bp upstream of the insert.

Overlapping contiguous sequences (contigs) were assembled and trimmed of plasmid sequence using the program Geneious. Sequences from each library were translated and aligned using the multiple sequence alignment program Muscle and

operational taxonomic units were determined using the program DOTUR (Schloss and Handelsman, 2005) with a 10% amino acid sequence similarity as a cutoff.

Metagenome sequencing

Two micrograms of gDNA was sheared using a Covaris S2 sonicator and 400bp fragments were size selected for library construction. The library was constructed using the TruSeq DNA sample prep kit (Illumina) following the manufactures protocol. Sequencing was performed using an Illumina HiSeq 2000 sequencer. Two lanes of paired-end sequences were generated for a total of 632 million 101 bp sequence reads.

Metagenome quality control and assembly

Sequences were quality controlled and trimmed using the program Sickle. The trimmed sequences were then digitally normalized using the program Khmer (Crusoe et al., 2014) utilizing a two-pass strategy (Brown et al., 2012). The sequences were first normalized by median to reduce sequences with high-abundance kmers using a kmer size of 20bp and a maximum coverage overlap value of 20. The sequences were then filtered to remove low-abundance kmers and Illumina sequencing artifacts using the Khmer script filter-below-abund with a minimum overlap of 2 kmers. The remaining sequences after preprocessing were used for sequence assembly.

The normalized reads were assembled using the de Bruijn graph-based assembler Velvet. Four independent assemblies were constructed with kmer sizes of 41, 51, 61, and 71. The contigs were assembled into "supercontigs" using the program Geneious with a minimum overlap of 25bp. Contigs longer than 1000bp were exported and used in further

analysis. Read coverage was calculated by mapping the pre-normalized de-replicated reads to the assembled contigs using the program Bowtie 2 (Langmead and Salzberg, 2012). The resulting Sam file was converted to a sorted Bam file from which mean coverage for each contig was calculated using a pileup file generated using the program Samtools (Li et al., 2009).

Metagenome binning and sequence identification

Contigs were grouped to form phylotype clusters by constructing an emergent self organizing map (ESOM) calculated from the tetranucleotide frequencies of each contig with mean coverage greater than 30X (Dick et al., 2009). Tetranucleotide frequencies were calculated as using a minimum contig size of 1 kb and sequences larger than 50 kb were broken into 50 kb fragments. The frequencies were loaded into the ESOM program (Ultsch and Mörchen, 2005) and normalized using RobustZT. The ESOM was calculated using the K-batch training alogrithm with a map size of 1210 rows X 262 columns. The map was visualized with a UMatrix background using a hot gradient clipped to 25% with 16 colors. The clusters were manually drawn on a tiled map. Split contigs that mapped to different clusters were generally solved using consensus-rules or via blast homology.

Sequences from each cluster were separated and genes were predicted with the program Prodigal version v2.60 using the metagenomic procedure for gene identification. Sequences for diagnostic key enzyme genes for each carbon fixation pathway were found using the program Hmmer. Well characterized sequences for each key enzyme was downloaded from NCBI and homologous sequences were downloaded from the JGR IMG website. Each gene was aligned using Muscle and a HMM profile was calculated

using the hmmbuild function of Hmmer (Eddy, 2011). All assembled contigs were searched for homologous genes for each HMM profile using the hmmsearch function of HMMER. Only full length sequences with an hmmsearch inclusion threshold of 1e-50 were used for phylogenetic analysis. Clusters containing carbon fixation genes were characterized by querying the translated genes against the NCBI-NR database using blastp with a minimum evalue cutoff of 1e-20 and maximum descriptions set at 10. The blastp results for each bin were separately imported into MEGAN4 (Huson et al., 2011) and KEGG pathways (Kanehisa and Goto, 2000) were calculated and visualized using the default settings.

Phylogenetic analysis

Sequences from each key pathway were queried against the NCBI nr database using blastp, and sequences with high sequence similarity were downloaded along with sequences with well characterized or phylogenetically informative sequences. Sequences were aligned using MUSCLE and the alignments were visualized and corrected using the multiple sequence alignment editor Bioedit (Hall, 1999). The alignments were masked to include only amino acid positions that were unambiguously aligned. Phylogenetic trees were calculated using RAxML (Stamatakis, 2006) using the WAG (Whelan and Goldman, 2001) model of amino acid substitution and using the best tree from 10 alternative runs. Bootstrap values were calculated from a consensus tree of 100 bootstrap replicates. The trees were visualized using Mega4 (Tamura et al., 2007) and annotated using the vector drawing program Inkscape (http://www.inkscape.org).

Results

Site descriptions

Samples were collected from Loihi Seamount during the 2007 and 2008 field seasons. Samples were chosen for analysis that were representative of the most common geochemical and spatial habitats currently found on the Loihi hydrothermal system. Hydrothermal vents in the Hilo region of Pele's Pit have high concentrations of reduced iron as well as low concentrations of sulfur compounds, generally seen as HS^- (Table 4-1). There are multiple low temperature diffuse vents in the vent field area, with the vents near Marker 39 and Marker 36 the most studied. Marker 39 vent field is a large rock outcrop with multiple hydrothermal vents emitting fluid in several cracks. Microbial mats growing near these vents are often encrusted in iron oxides which is typical of hydrothermal vents at Loihi, however they also contain white stringy mats similar in appearance to mats dominated by Epsilonproteobacteria at sulfur-dominated hydrothermal vents. Marker 36 microbial mats are primarily iron-encrusted and do not have the long white streamers that are seen at Marker 39, however the hydrothermal vents do contain small amounts of HS^- in addition to high concentrations of iron (Table 4-1). The spillway vent field of Pele's Pit also contains multiple diffuse hydrothermal vents, however the vent fluids are generally devoid of measurable HS^- (Table 4-1). Marker 34 vent is found at the base of a large rock structure and is covered in thick (3-7 cm) iron oxide-encrusted microbial mat material. Pohaku Vent Field is outside Pele's Pit caldera and is deeper on the south rift. Vent fluids from Pohaku are very high in iron concentration, have no measurable sulfide, and have a lower temperature than Marker 34

Table 4-1. Temperature and chemistry data from the four hydrothermal vent fields used in this study. Data is from (Glazer and Rouxel, 2009)

Sample	Site	Year	Temp (ºC)	Fe (μM)	Mn (μM)	Fe/Mn	HS^- (μM)
308RG	Hiolo Marker 39	2007	51	578	25	24	16
308OB	Spillway Marker 34	2007	50	355	12.3	29	nd
365B	Hiolo Marker 36	2008	44	538	25	22	6
373Sc1	Pohaku Marker 57	2008	21	773	19	41	nd

nd = not detected

Vent (Table 4-1). The mats at Pohaku are the thickest iron mats observed on Loihi and appear to grow more rapidly than at other vent sites in Pele's Pit (McAllister et al., 2011).

PCR amplification

All samples had strong amplification using the Rubisco Form II primers, while weak amplification was seen with both Rubisco Form I primer sets when DNA was used as a template. When cDNA was used as the template, strong amplification was again seen with the Rubisco Form II primer set, however no amplification was seen with either of the Form I Rubisco primers. Using aclB primers, only DNA samples from Marker 39 and Marker 36 amplified, there was no amplification with cDNA or Marker 34 DNA. All sequenced plasmids with inserts contained the gene of interest indicating the primers had high specificity for the targeted genes.

Metagenome construction and analysis

A metagenome of a scoop sample from Marker 57 was constructed to both look for additional carbon fixation pathways where universal PCR primers could not be designed and also to identify the metabolic potential of phylotypes containing carbon fixation genes. Marker 57 from Pohaku vent field was chosen as the sample because the lack of reduced sulfur found at these vents coupled with high reduced iron concentration should provide information about primary production in iron-dominated hydrothermal communities. The community has been shown to have low bacterial diversity and was dominated by Zetaproteobacteria (McAllister et al., 2011) which should allow for high sequence coverage in a group of bacteria which are not well characterized.

After assembly of the metagenome, there were 112,979 contigs greater than 1000 bp in length with a total metagenome size of 375 Mbp. Contigs with coverage greater than 30X were clustered using a self-organizing map of tetranucleotide frequencies. Seventy six clusters were formed on the map with each cluster containing contigs from either a single genome or from closely related organisms (Table 4-2).

Key genes in every known carbon fixation pathway were queried against the metagenome using the hmmsearch function in Hmmer. Putative homologs of *cbbL*, *cbbM*, *aclB*, 4-hydroxybutyryl-CoA dehydratase, and acetyl-CoA synthase (ACS) were found in the metagenome, with *cbbM* (Form II Rubisco) and ACS having the greatest gene coverage, and therefore the most gene copies within the community (Table 4-2). Sequences for *cbbL*, *aclB*, and 4-hydroxybutyryl-CoA dehydratase are present, but have much lower gene coverage than cbbM and ACS (Table 4-2). In addition to these genes, several copies of the "Form III" archaeal Rubisco sequences were found in the metagenome that were most closely related to methanogenic Archaea. This form of Rubisco has been shown to have very active carboxylase activity, but is utilized in the recycling of AMP and does not appear to be utilized in autotrophic carbon fixation (Sato et al., 2007). An additional "hybrid form II / form III" Rubisco was also found from a bin closely related to the nanoarchaea, however this form of Rubisco has also been theorized to not be involved in carbon fixation (Wrighton et al., 2012).

Rubisco sequence phylogeny

The majority of Form II Rubisco sequences from the metagenome formed two distinct clusters (Figure 4-1). One group of sequences clustered near *Mariprofundus*

Table 4-2. ESOM cluster, Class-level phylogeny of each cluster, Key carbon fixation gene, Sequence coverage for each carbon fixation gene, and Bin size in nucleotides.

Cluster	Phylogeny	Carbon Fixation Gene	Gene Coverage	Bin Size (Mbp)
32	Zetaproteobacteria	cbbM (FII)	755	3.19
20	Zetaproteobacteria	cbbM (FII)	718	6.24
27	Zetaproteobacteria	cbbM (FII)	340	6.76
6	Zetaproteobacteria	cbbM (FII)	298	3.41
22	Zetaproteobacteria	cbbM (FII)	292	3.32
19	Zetaproteobacteria	cbbM (FII)	290	5.44
6	Zetaproteobacteria	cbbM (FII)	239	3.41
62	Zetaproteobacteria	cbbM (FII)	130	3.12
20	Zetaproteobacteria	cbbM (FII)	119	6.24
27	Zetaproteobacteria	cbbM (FII)	93	6.76
76	Gammaproteobacteria	cbbM (FII)	72	4.09
64	Alphaproteobacteria	cbbM (FII)	66	3.20
31	Zetaproteobacteria	cbbM (FII)	62	1.52
7	Alphaproteobacteria	cbbM (FII)	56	2.98
9	Gammaproteobacteria	cbbM (FII)	52	3.77
19	Zetaproteobacteria	cbbM (FII)	51	5.44
19	Zetaproteobacteria	cbbM (FII)	45	5.44
NA	ND	cbbM (FII)	24	NA
NA	ND	cbbM (FII)	13	NA
77	Chloroflexi	cbbL (FI)	50	4.64
42	Deltaproteobacteria	aclB (FI)	201	5.87
18	Deltaproteobacteria	aclB (FII)	233	4.47
NA	ND	aclB (FII)	26	NA
NA	ND	aclB (FII)	18	NA
75	Thaumarchaeota	hpaB	37	1.92
75	Thaumarchaeota	hpaB	68	1.92
37	Methanomicrobia	ACS	539	0.51
46	Methanomicrobia	ACS	492	4.70
55	Methanomicrobia	ACS	169	2.13
NA	ND	ACS	20	NA
NA	ND	ACS	18	NA

NA = Not clustered; ND = Not determined.

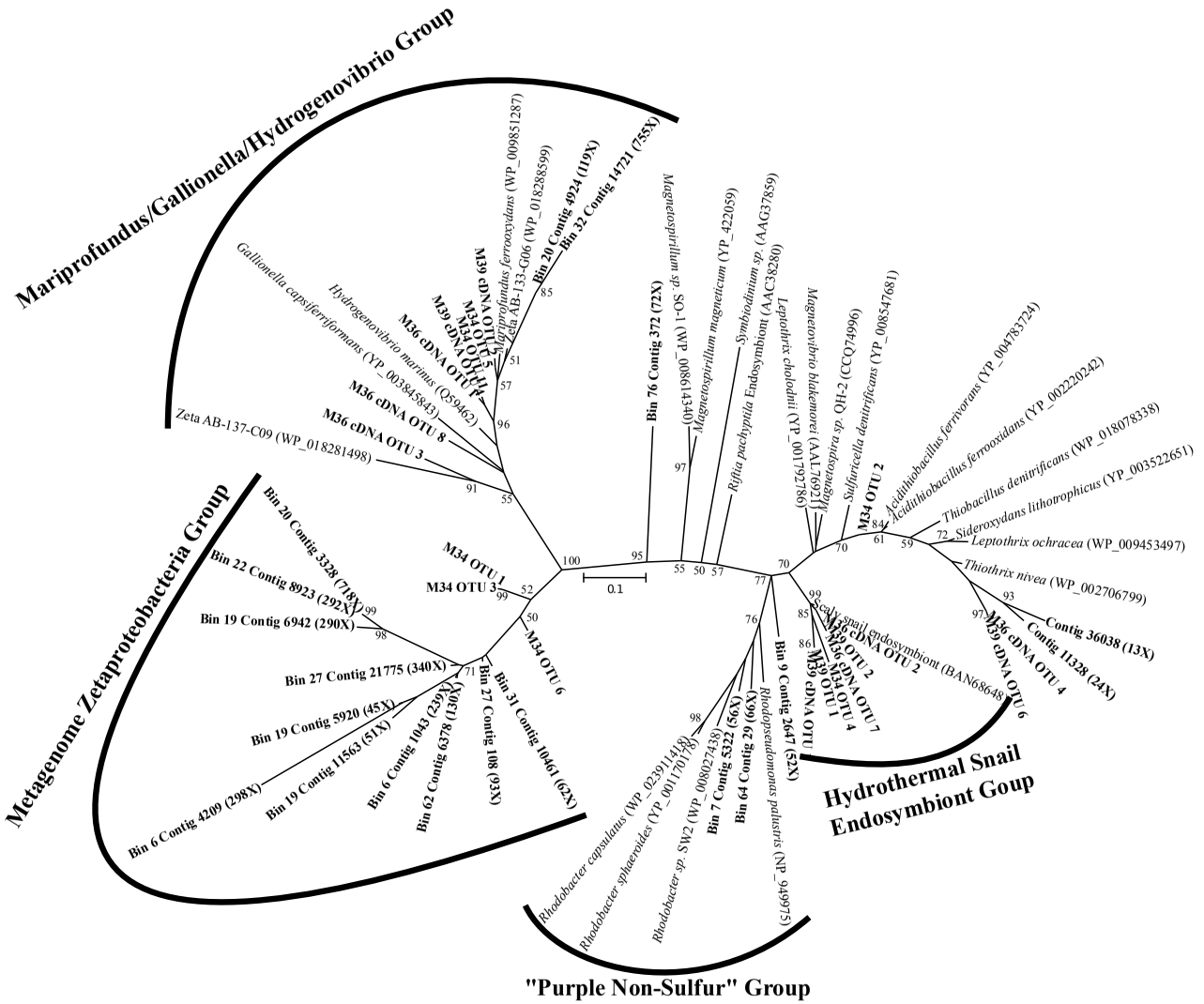

Figure 4-1. Unrooted maximum likelihood phylogenetic tree of form II Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (Rubisco) gene.

ferrooxydans and the other formed a large monophyletic cluster that does not include any described species or uncultured PCR clones from other studies. All of these metagenome sequences with coverage greater than 30X belonged to the *Zetaproteobacteria* clade. Multiple PCR clones from Loihi also clustered within these two groups, including the most abundant Marker 36 cDNA clones. Only three metagenome sequences clustered outside of these two groups. Contig 372 clustered near the magnetotactic *Magnetosprillum* group. This contig grouped in bin 76 which are deeply-rooted *Gammaproteobacteria*. Two other contigs cluster with the alphaproteobacterial "Purple Bacteria" which are anaerobic phototrophs (Niel, 1932). The bins containing these two contigs are also associated with the *Alphaproteobacteria* and do not contain genes associated with photosynthesis. Multiple PCR clone sequences clustered with the endosymbiont of a scaly-foot snail that was collected from the Central Indian Ridge (Warén et al., 2003). This endosymbiont was shown to oxidize sulfur and possibly hydrogen while fixing organic carbon which is passed to the host (Nakagawa et al., 2014). Other low-abundance sequences clustered near the sulfur oxidizer *Thiothrix nivea* and the iron oxidizing betaproteobacterium *Leptospirillum ochracea* and *Sideroxydans lithotrophicus*.

Only a single sequence from the metagenome grouped with the form I Rubisco (Figure 4-2). The sequence does not group with any other described species or PCR clone from previous studies. The sequence branches between the Form IB and Form IC clades and clusters with clone sequences from Marker 36 and Marker 39. The sequence is within a contig that is 516 kbp in length and was clustered into bin 77. Sequences from this bin had high similarity with sequenced members of the *Chloroflexi* phylum which has been

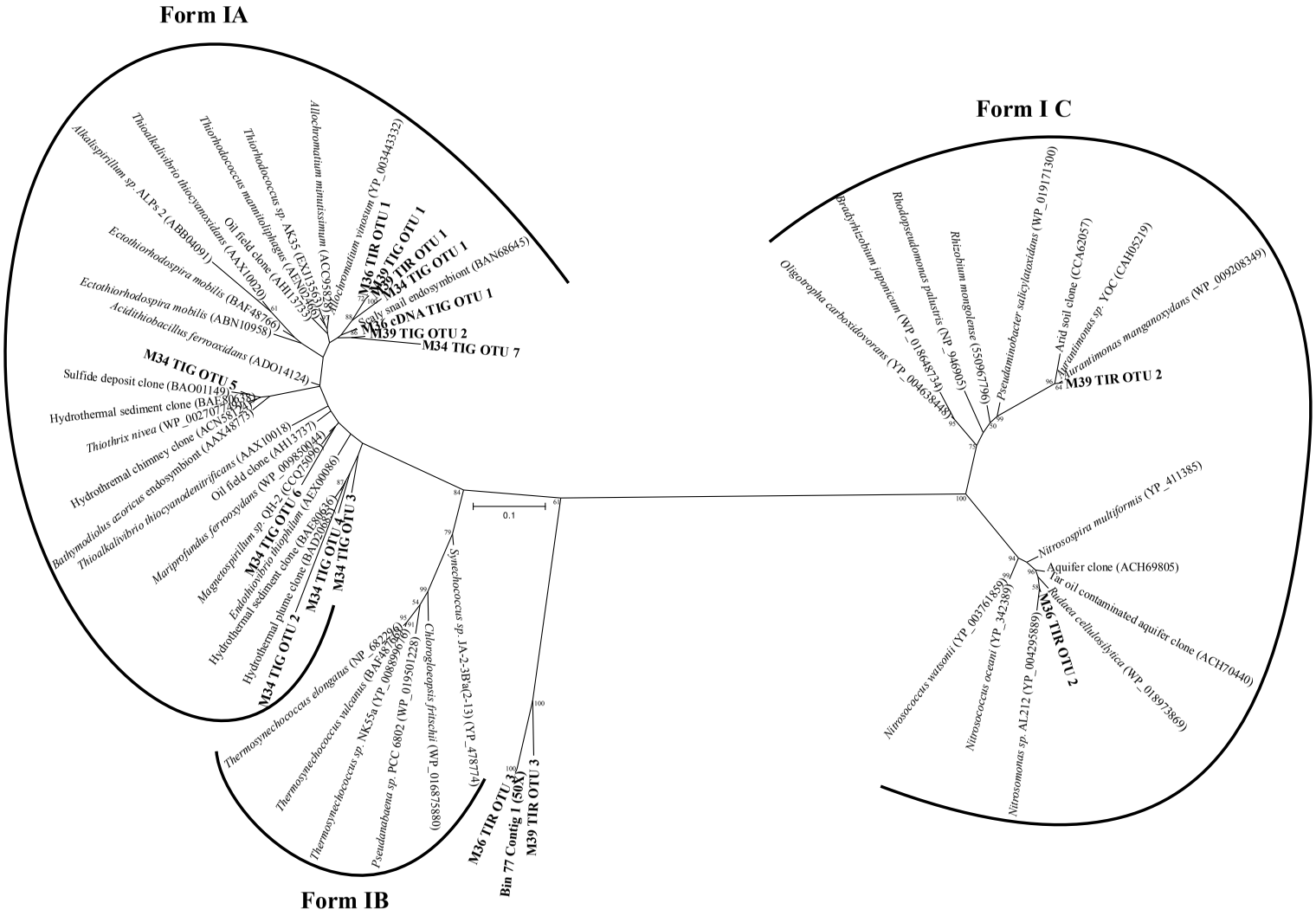

Figure 4-2. Unrooted maximum likelihood phylogenetic tree of form I Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (Rubisco) gene.

found in other clone libraries from microbial mats at Loihi. The most abundant clone sequences found at Loihi were the form I copy of Rubisco from the scaly-foot snail. Other lower abundance sequences cluster within two uncultured groups in the form IA Rubisco clade with other hydrothermal PCR clones. Two sequences clustered within the form I "red type" Rubisco with one clone from Marker 36 clustering with the ammonia oxidizing *Proteobacteria*, and a clone from Marker 39 which clusters with the *Aurantimonas* genus, which are generally heterotrophs but are also known to be able to oxidize manganese (Anderson et al., 2009; Caspi et al., 1996).

4-Hydroxybutyryl-CoA dehydratase phylogeny

Two 4-hydroxybutyryl-CoA dehydratase sequences were found in the metagenome, and both clustered with sequences from the *Thaumarchaeota* (Figure 4-3). Contig 723 clustered with the cultured *Nitrosopumilus* group which are aerobic ammonia-oxidizing Archaea (Könneke et al., 2005). Contig 32857 clustered with sequences from uncultured organisms, including a deep-sea metagenome sample from 4000m deep off the coast of Hawaii and a single-cell genome from a *Thaumarchaeota*. Both sequences clustered in Bin 75 which is a mix of at least two *Thaumarchaeota* genomes.

ATP citrate lyase phylogeny

A single type I *aclB* sequence was found in the metagenome (Figure 4-4). This sequence did not have high similarity with any known sequence and weakly clustered within the *Aquificales aclB* sequences. The sequence clustered within Bin 42 which is most closely related to the *Desulfurella* genus in the *Deltaproteobacteria*. A phylotype

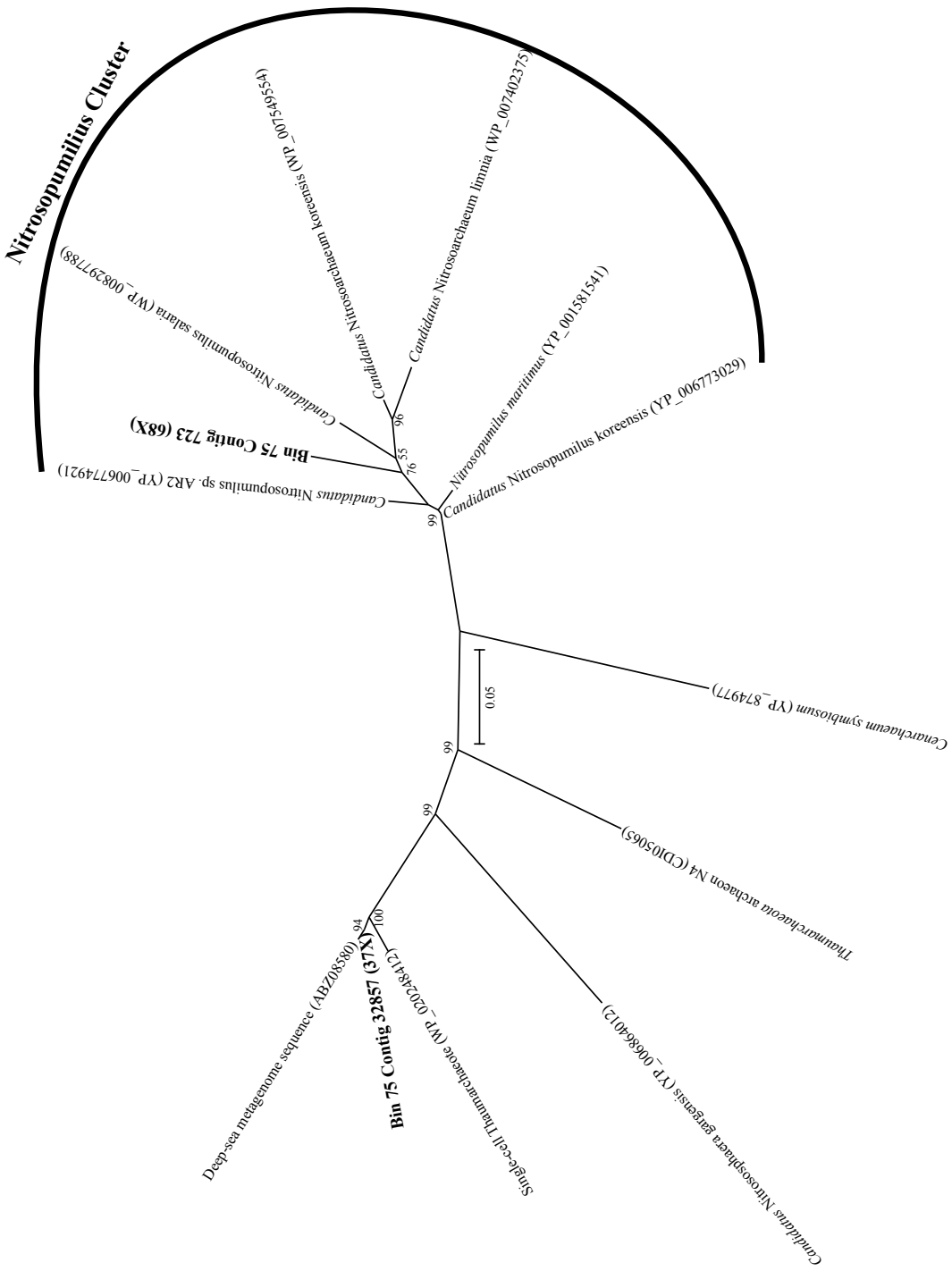

Figure 4-3. Unrooted maximum likelihood phylogenetic tree of the 4-hydroxybutyryl-CoA dehydratase gene (*hpaB*).

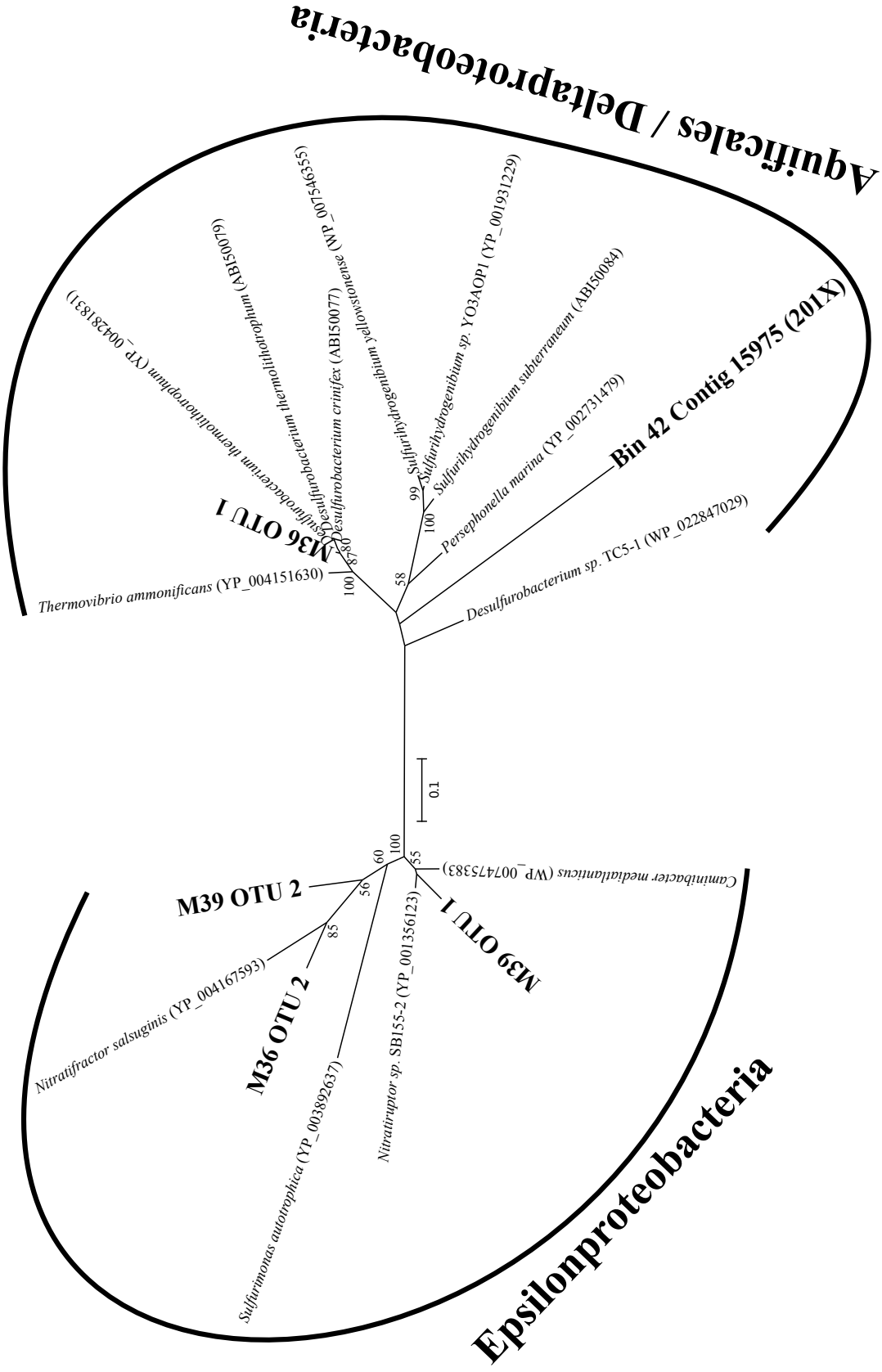

Figure 4-4. Unrooted maximum likelihood phylogenetic tree of the beta subunit of the ATP citrate lyase gene.

from Marker 36 clustered with the sulfate-reducing hydrogen-oxidizing *Desulfurobacterium* bacteria. Three other phylotypes clustered with the epsilonproteobacterial sequences, with Marker 36 and Marker 39 OTUs clustering with *Nitratifractor* and Marker 39 OTU 1 clustering with *Nitratiruptor* groups, both of which are obligate hydrogen oxidizing nitrate reducers (Nakagawa et al., 2005).

Three type II aclB sequences were also found in the metagenome (Figure 4-5). Contig 15723 clusters near the chemoautotrophic nitrite-oxidizing *Nitrospina* isolates. This contig clusters in Bin 18 which is a *Deltaproteobacteria* bin and contains multiple [Ni-Fe]-hydrogenase and dissimilatory sulfite reductase genes. Two other metagenome sequences clustered near Contig 15723, however these sequences had low coverage and were not clustered into bins.

Anaerobic acetyl-CoA synthase phylogeny

Five Acetyl-CoA synthase genes were found in the metagenome, with three of the genes clustering with the methanogenic Archaea having high coverage (Figure 4-6). All three contigs cluster together forming a monophyletic clade with an Eel River methane seep sediment cosmid sequence being the closest neighbor. All three metagenomic sequences clustered into distinct bins with each of them clustering to the *Methanomicrobia* class of methanogens. These obligate anaerobes produce methane using a variety of C1 and C2 compounds, but not with hydrogen oxidation like other archaeal methanogens. Two other low-coverage metagenome sequences cluster with either the sulfate-reducing Deltaproteobaceria or weakly with the acetogenic *Firmicutes*, both of which are obligate anaerobic hydrogen oxidizers.

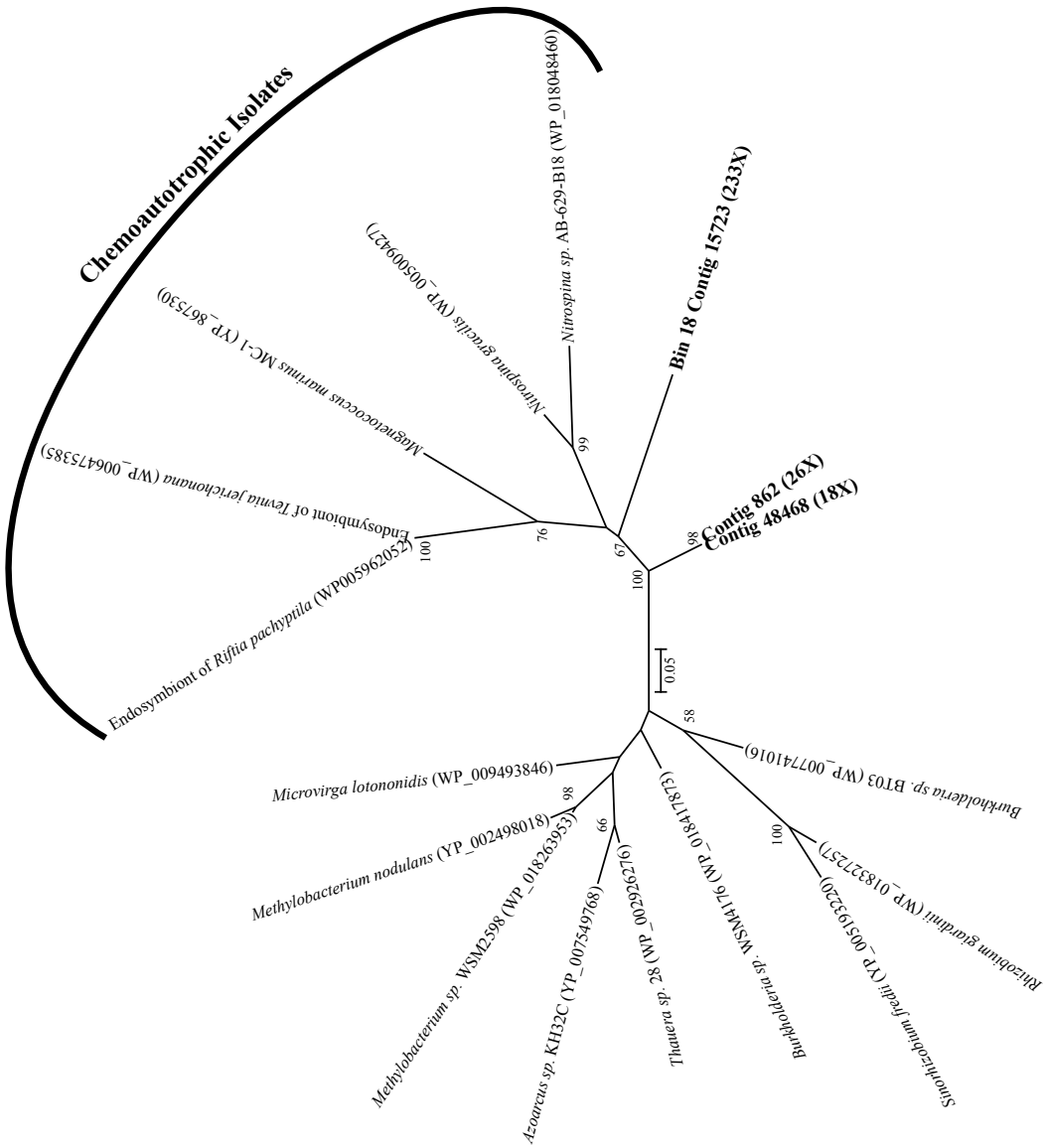

Figure 4-5. Unrooted maximum likelihood phylogenetic tree of the beta subunit of the form II ATP citrate lyase gene.

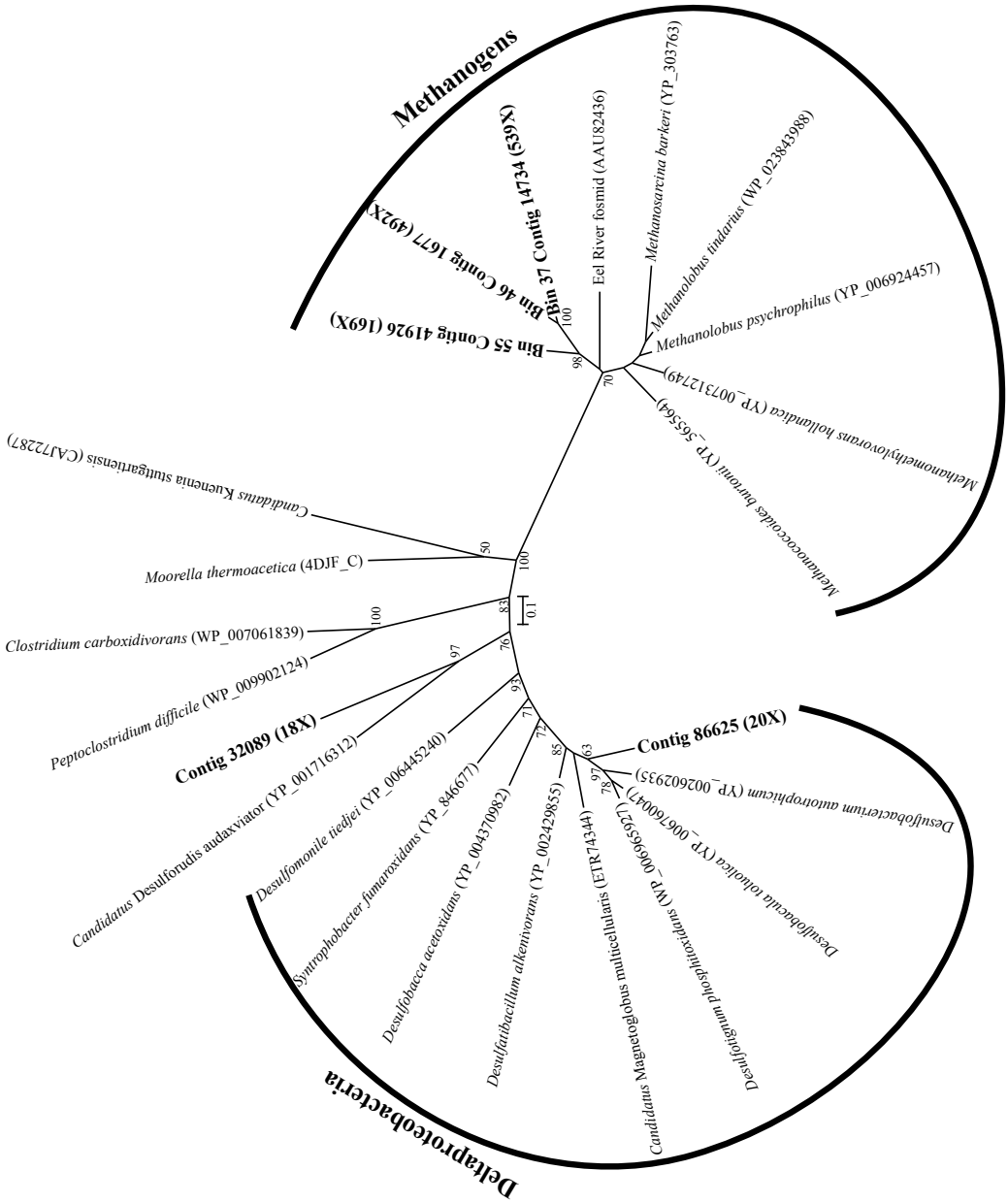

Figure 4-6. Unrooted maximum likelihood phylogenetic tree of the anaerobic acetyl-CoA synthase gene (ACS).

Discussion

The sites utilized in this study give a good representation of the metabolic potentials thought to be currently used at Loihi. The Maker 34 and Marker 56 vents are iron-dominated vents with very low or no measurable sulfide in the vent fluids (Table 4-1). The carbon fixing populations in these communities should be primarily dominated by brine-based metabolisms with iron oxidation being the most likely metabolism due to vent chemistry, the presence of a microbial mat comprised of biogenic iron oxides, and previous phylogenetic analysis (Butterfield et al., 1997; Emerson and Moyer, 2002; McAllister et al., 2011). Markers 36 and 39 are also iron dominated systems, however there is measurable amounts of sulfide in these vent fluids. The morphological changes of the microbial mats is particularly striking at Marker 39 where large white streamers are often seen near the vent orifices which is very similar to sulfur-oxidizing mat streamers found at sulfur dominated vents such as those found at Axial Seamount. These vents should contain populations of both sulfur and iron-oxidizing bacteria along with the possibility of other vapor-dominated metabolisms such as hydrogen and methane oxidation.

Metagenomic and cDNA results strongly suggest that the Rubisco being utilized for carbon fixation by autotrophs in the Loihi microbial mats is the Form II type which has a much faster rate of reaction, but less specificity toward the carboxylase reaction and is therefore used in anaerobic or extremely microaerophilic conditions to minimize the carbon-wasting oxygenase reaction (Tabita, 1999). Previous analysis showed the Pohaku vent field microbial mat community to be dominated by *Zetaproteobacteria* (McAllister

et al., 2011). The type-strain of the *Zetaproteobacteria* is the obligate chemoautotrophic iron-oidizing bacterium *Mariprofundus ferrooxydans* (Singer et al., 2011). All sequenced isolates from the *Mariprofundus* group fix carbon using the Calvin Cycle, however these isolates have two copies of Rubisco, one form I and one form II (Singer et al., 2011). This may help explain the observation of why only a small minority of *Zetaproteobacteria* that have been found by culture-independent methods cluster with the *Mariprofundus* isolates (McAllister et al., 2011). It is possible that the gradient tube method of isolation of neutrophilic iron oxidizing bacteria is selecting for isolates with higher oxygen tolerance than the dominant *Zetaproteobacteria* in these volcanic microbial mats (Emerson and Floyd, 2005). Further evidence for this theory is provided by copies of dissimilatory nitrate reductase genes in most of the *Zetaproteobacteria* bins, but missing in *Mariprofundus ferrooxydans*, suggesting that anaerobic iron oxidation coupled with nitrate reduction is the energy producing metabolism in these uncultured bacteria.

The Rubisco form II sequences from suspected Zetaproteobacteria form two distinct phylogenetic clusters. One cluster groups closely with *Mariprofundus* and other neutrophilic iron oxidizers, which includes two metagenome sequences with high coverage along with the dominate OTU phylotype from the Marker 36 cDNA library. The other cluster consists of zetaproteobacterial sequences from the metagenome along with two phylotypes from Marker 34. This cluster forms a large monophyletic group that is phylogenically unique among these samples. Other metagenomic form II Rubisco sequences include a sequence from a deeply-rooted *Gammaproteobacteria* that did not cluster with any other sequence and *Alphaproteobacteria* that cluster with the 'Purple Bacteria'. These phylotypes are often found in Loihi microbial mats and may also oxidize

iron (Fleming et al., 2013; McAllister et al., 2011). Other phylotypes from all of the clone libraries clustered with the scaly-foot snail endosymbiont which is a sulfur-oxidizing *Gammaproteobacteria*, suggesting these phylotypes have a similar metabolism.

Although there have been culture-independent studies of carbon fixation of communities from hydrothermal communities using primers for the form II Rubisco, the primers that have been used would not have amplified the Rubisco genes from the *Zetaproteobacteria* due the small number of sequences that were available to be used when the primers were designed (Elsaied and Naganuma, 2001). A recent study has designed new PCR primers for the form II Rubisco that is more universal, however the sequences used in their analysis has not been deposited in a publicly available database and could not be used in this analysis (Kato et al., 2012). The primers used in this study amplified genes across the known diversity of the form II Rubisco gene and had good specificity for the gene.

Only a single form I Rubisco phylotype clustered with the form I Rubisco found in *Mariprofundus ferrooxydans*, further supporting the theory that the uncultured *Zetaproteobacteria* are more sensitive to oxygen for growth than the *Mariprofundus* isolates. The majority of form I sequences from the PCR-based analyses grouped with the scaly-foot snail endosymbiont, which also has copies of both forms of Rubisco (Nakagawa et al., 2014). The only form I Rubisco sequence from the metagenome was from a *Chloroflexi* with an unknown metabolism. Phylotypes from Marker 36 and Marker 39 also clustered with this sequence and *Chloroflexi* phylotypes have been found in other culture-independent studies of Loihi Seamount microbial mats, suggesting this group is common in these communities, however their energetic pathways are currently unknown.

Only one other *Chloroflexi* has been shown to utilize the Calvin Cycle, however the sequence clusters in the "Red Type" form I Rubisco sequences and does not appear to be related to this phylotype (Keppen et al., 1994). Other sequences cluster in the "Red Type" group of Form I Rubisco, with one sequence clustering with the Aurantimonas group of Alphaproteobacteria which are typically methylotrophic bacteria and also includes the bacterium *Aurantimonas manganoxydans* which has been theorized to grow through mixotrophy using manganese oxidation, however this has never been experimentally proven (Dick et al., 2008). Another sequence clusters with the ammonia-oxidizing bacteria which, along with the presence of Thaumarchaeota in the metagenome, suggests that ammonia oxidation may be an important autotrophic pathway in these microbial mats.

Two *4*-Hydroxybutyryl-CoA Dehydratase genes were found in the genome, and both sequences clustered in Bin 75 despite having distinctly different phylogeny. Only a single ammonia monooxygenase operon was found in the bin which had high nucleotide sequence similarity (98% amoA gene) with the autotrophic ammonia-oxidizing archaean *Nitrosopumilus maritimus*. There are no housekeeping genes, such as ribosomal proteins, that are present in more than a single copy. This suggests that the gene found on contig 32857 is either a second copy of the 4-hydroxybutyryl-CoA dehydratase gene or it is a very small part of the community, since no housekeeping genes were found in the bin that were not closely related to the *Nitrosopumilus* group. Finally, the presence of carbon fixation genes from both ammonia oxidizing Bacteria and Archaea suggests that these populations are an actively growing part of the microbial mat community, rather than simple entrainment in the mat from deep-sea planktonic *Archaea*.

Form I ATP citrate lyase genes were also found in the metagenome and in the DNA-based PCR analysis, but not when using cDNA as a template. This indicates that these genes were not being actively transcribed when the cDNA samples were taken. One sequence from the metagenome grouped with the Aquificales bacteria, but with a long branch due to the low sequence similarity between the cultured *aclB* genes and the sequence from the metagenome. Other *Deltaproteobacteria* have been shown to use the reductive TCA cycle, however they have since been shown to use the form II *aclB* gene (Lucker et al., 2013) or have not been sequenced as is the case with the hydrogen oxidizing bacterium *Desulfobacter hydrogenophilus* (Londry et al., 2004). The PCR-based phylotype from Marker 36 which clustered with the *Desulfurobacterium* group is interesting because neither a *Desulfitobacterium* or an *Aquificales* have been found in iron oxide microbial mats. Whether or not this sequence is from a *Desulfurobacterium* or is the product of lateral gene transfer from another bacterium is not known. The other PCR-based form I *aclB* phylotypes cluster with *aclB* genes from the *Epsilonproteobacteria*, which are common sulfur oxidizing bacteria often found at hydrothermal vents, including at Loihi Seamount (Moyer et al., 1995).

A form II version of aclB has recently been found through bioinformatic inferences of chemoautotrophic organisms (Hügler and Sievert, 2011; Markert et al., 2007). The gene has not been confirmed to be involved in the mechanism for citrate cleavage experimentally, however, so analyses about carbon fixation using this gene as a proxy must be made with this in mind. The bin with the form II aclB appears to be a sulfate reducing hydrogen oxidizing deltproteobacterium. This sequence clusters near the chemoautotrophic organisms with a form II aclB and the bin contains the gene for 2-

oxoglutarate synthase which is the second key enzyme for the reductive TCA cycle and no other genes for citrate cleavage.

Anaerobic Acetyl-CoA synthase sequences from methanogens represent a potentially significant source of fixed carbon in Loihi microbial mats that was previously unknown. The only published clone library from Loihi which used primers specific for Archaea was from a pre-eruption Pele's Peak sample which found the community consists of Thaumarchaeota along with Marine Group II Euryarchaeota which are related to the Thermoplasmatales (Moyer et al., 1998). Ribosomal RNA genes from methanogenic Archaea comprised over 20% of the total ribosomal RNA sequences in the metagenome. Bin 37 is an incomplete genome at only 0.51 Mbp and does not contain a full suite of metabolic pathways, however Bins 46 and 55 appear to be nearly complete genomes and contain multiple copies of hydrogenase genes and a complete pathway to generate methane from the oxidation of hydrogen coupled to the reduction of carbon dioxide. No genes were found for the oxidation of methanol or acetate which are a common metabolisms of the *Methanosarcina* in addition to hydrogen oxidation (Blaut and Gottschalk, 1984).

The autotrophic populations in the microbial mat sampled from Pohaku appears to be dominated by microaerophilic iron- and hydrogen-oxidizing Bacteria and Archaea. The majority of the Rubisco sequences found in the metagenome are from zetaproteobacterial bins, which are currently thought to be solely autotrophic iron-oxidizers based on pure and enrichment cultures along with correlation of the biogeography of the class. The lack of form I Rubisco sequences along with the presence of nitrate reducing genes suggests that iron oxidation in these vents is occurring under

anaerobic conditions. *In situ* microprobe analysis of a Marker 36 microbial mat showed no measurable oxygen at a depth of only 1 mm below the surface of the mat, further supporting the anaerobic nature of the microbial mat community (Glazer and Rouxel, 2009).

Anaerobic oxidation of iron may be the source of hydrogen utilized by the methanogenic *Archaea* and the *Deltaproteobacteria* which utilize the reductive TCA cycle to fix carbon. The production of hydrogen ions through the anaerobic oxidation of iron can produce molecular hydrogen through direct reduction of hydronium ions under slightly acidic conditions (Reardon, 1995). Hydrogen can also be produced through the reduction of hydrogen ions by hydrogenase enzymes to decrease reducing equivalents inside cells. This production of H_2 can be directly feeding the methanogenic *Archaea* and the sulfate- or iron-reducing *Deltaproteobacteria* which inhabit the mat (Figure 4-7). The hydrogen could also be a component of vent fluid at Marker 57, however diffuse hydrothermal vents with low sulfide concentrations rarely have a high molecular hydrogen concentration due to reduced magmatic gas input (Butterfield et al., 1997). Unfortunately there has never been an accurate measurement of hydrogen at a post-eruption hydrothermal vent at Loihi.

This study is the first comprehensive study of carbon fixation at a brine dominated hydrothermal vent system. The communities are dominated by chemoautotrophic populations of Zetaproteobacteria and methanogenic Archaea along with populations of hydrogen-oxidizing Deltaproteobacteria that utilize the reductive TCA cycle. Smaller populations of ammonia-oxidizing Archaea and Bacteria also inhabit the mats, along with PCR-derived phylotypes related to sulfur-oxidizing

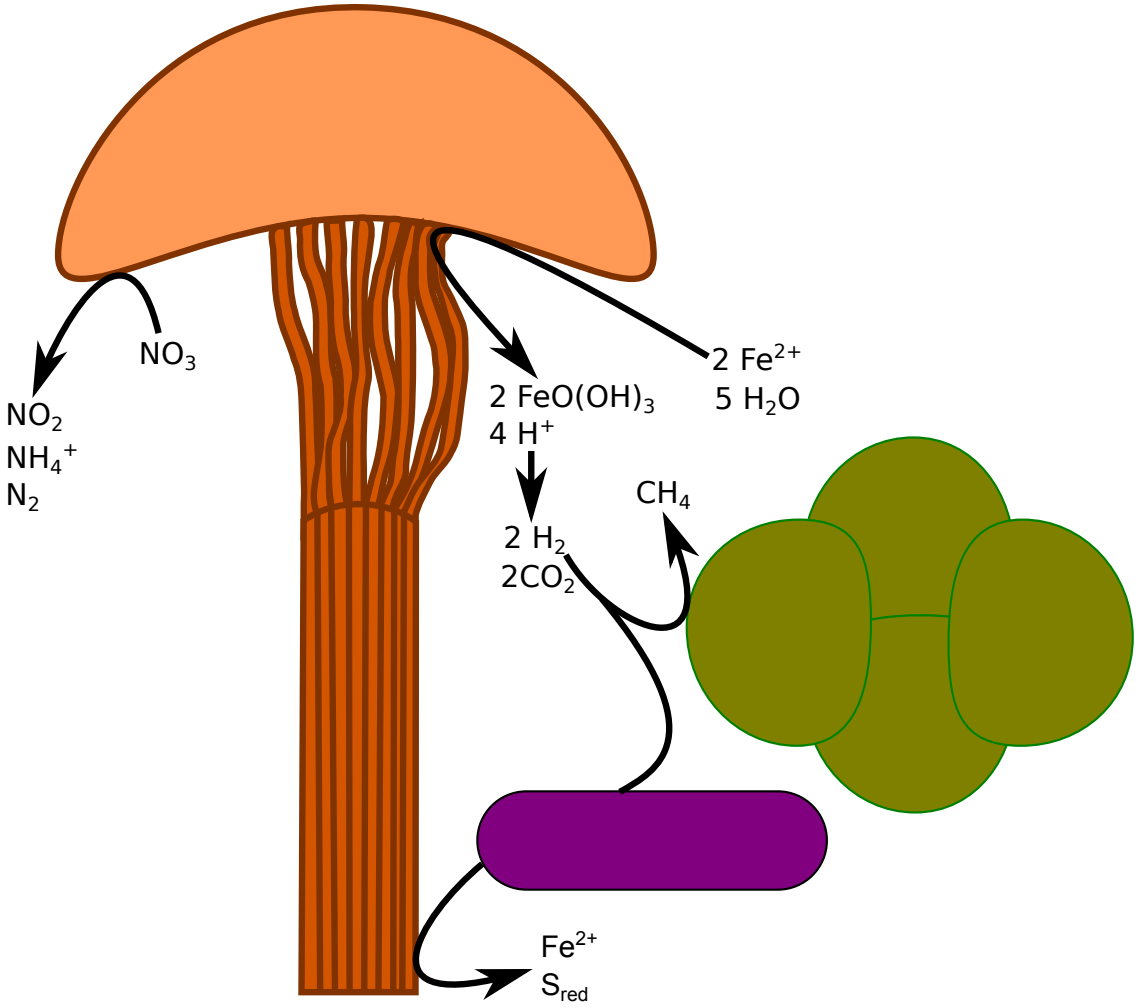

Figure 4-7. Conceptual model of a syntrophic relationship between anaerobic iron oxidizing bacteria (orange cell) with methanogenic archaea (green cells) and sulfur/iron reducing bacteria (purple cell).

Gammaproteobacteria and *Epsilonproteobacteria*. This study shows that future studies of the microbial mats at Loihi Seamount need to include an analysis of the Archaeal populations within the community as they may provide a significant contribution to the primary production at these carbon-starved hydrothermal systems.

Works Cited

Anderson, C.R., Dick, G.J., Chu, M.-L., Cho, J.-C., Davis, R.E., Bräuer, S.L., and Tebo, B.M. (2009). *Aurantimonas manganoxydans*, sp. nov. and *Aurantimonas litoralis*, sp. nov.: Mn(II) Oxidizing Representatives of a Globally Distributed Clade of alpha-Proteobacteria from the Order Rhizobiales. Geomicrobiol. J. *26*, 189.

Berg, I.A. (2011). Ecological Aspects of the Distribution of Different Autotrophic CO_2 Fixation Pathways. Appl Env. Microbiol *77*, 1925–1936.

Berg, I.A., Kockelkorn, D., Ramos-Vera, W.H., Say, R.F., Zarzycki, J., Hügler, M., Alber, B.E., and Fuchs, G. (2010). Autotrophic carbon fixation in archaea. Nat Rev Micro *8*, 447–460.

Blaut, M., and Gottschalk, G. (1984). Coupling of ATP synthesis and methane formation from methanol and molecular hydrogen in *Methanosarcina barkeri*. Eur. J. Biochem. *141*, 217–222.

Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., and Brom, T.H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. ArXiv Prepr. ArXiv12034802.

Butterfield, D.A., Jonasson, I.R., Massoth, G.J., Feely, R.A., Roe, K.K., Embley, R.E., Holden, J.F., McDuff, R.E., Lilley, M.D., and Delaney, J.R. (1997). Seafloor eruptions and evolution of hydrothermal fluid chemistry. Phil Trans R Soc Lond A *355*, 369–386.

Campbell, B.J., Stein, J.L., and Cary, S.C. (2003). Evidence of Chemolithoautotrophy in the Bacterial Community Associated with *Alvinella pompejana*, a Hydrothermal Vent Polychaete. Appl Env. Microbiol *69*, 5070–5078.

Campbell, B.J., Engel, A.S., Porter, M.L., and Takai, K. (2006). The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Micro *4*, 458–468.

Caspi, R., Haygood, M.G., and Tebo, B.M. (1996). Unusual ribulose-1,5-bisphosphate carboxylase/oxygenase genes from a marine manganese-oxidizing bacterium. Microbiology *142*, 2549–2559.

Crépeau, V., Cambon Bonavita, M., Lesongeur, F., Randrianalivelo, H., Sarradin, P., Sarrazin, J., and Godfroy, A. (2011). Diversity and function in microbial mats from the Lucky Strike hydrothermal vent field. FEMS Microbiol. Ecol. *76*, 524–540.

Crusoe, M.R., Edvenson, G., Fish, J., Howe, A., McDonald, E., Nahum, J., Nanlohy, K., Pell, J., Simpson, J., Scott, C., et al. (2014). The khmer software package: enabling efficient sequence analysis. doi: 10.6084/m9.figshare.979190

Davis, R.E., and Moyer, C.L. (2008). Extreme spatial and temporal variability of microbial mat communities along the Mariana Island Arc/Backarc system. J Geophys Res *113*, B08S15.

Dick, G.J., Podell, S., Johnson, H.A., Rivera-Espinoza, Y., Bernier-Latmani, R., McCarthy, J.K., Torpey, J.W., Clement, B.G., Gaasterland, T., and Tebo, B.M. (2008). Genomic Insights into Mn(II) Oxidation by the Marine Alphaproteobacterium *Aurantimonas sp*. Strain SI85-9A1. Appl. Environ. Microbiol. *74*, 2646–2658.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009). Community-wide analysis of microbial genome sequence signatures. Genome Biol. *10*, R85.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput Biol *7*, e1002195.

Edwards, K.J., Bach, W., and McCollom, T.M. (2005). Geomicrobiology in oceanography: microbe-mineral interactions at and below the seafloor. TRENDS Microbiol *13*, 449–456.

Elsaied, H., and Naganuma, T. (2001). Phylogenetic Diversity of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large-Subunit Genes from Deep-Sea Microorganisms. Appl. Environ. Microbiol. *67*, 1751–1765.

Emerson, D., and Floyd, M. (2005). Enrichment and isolation of iron-oxidizing bacteria at neutral pH. Methods Enzymol. *397*, 112–123.

Emerson, D., and Moyer, C.L. (2002). Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition. Appl Env. Microbiol *68*, 3085–3093.

Emerson, D., Rentz, J.A., Lilburn, T.G., Davis, R.E., Aldrich, H., Chan, C., and Moyer, C.L. (2007). A novel lineage of Proteobacteria involved in formation of marine Fe-Oxidizing microbial mat communities. PLoS ONE *2*, e667.

Fleming, E.J., Davis, R.E., McAllister, S.M., Chan, C.S., Moyer, C.L., Tebo, B.M., and Emerson, D. (2013). Hidden in plain sight: discovery of sheath-forming, iron-oxidizing Zetaproteobacteria at Loihi Seamount, Hawaii, USA. FEMS Microbiol. Ecol. *85*, 116–127.

Glazer, B.T., and Rouxel, O.J. (2009). Redox speciation and distribution within diverse iron-dominated microbial habitats at Loihi Seamount. Geomicrobiol. J. *26*, 606–622.

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser *41*, 95–98.

Hügler, M., and Sievert, S.M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Annu. Rev. Mar. Sci. *3*, 261–289.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res. *21*, 1552–1560.

Jannasch, H.W., and Mottl, M.J. (1985). Geomicrobiology of deep-sea hydrothermal vents. Science *229*, 717–725.

Jannasch, H.W., and Taylor, C.D. (1984). Deep-sea microbiology. Annu. Rev. Microbiol. *38*, 487–487.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27 –30.

Kato, S., Nakawake, M., Ohkuma, M., and Yamagishi, A. (2012). Distribution and phylogenetic diversity of cbbM genes encoding RubisCO form II in a deep-sea hydrothermal field revealed by newly designed PCR primers. Extremophiles. *16*, 277-283.

Keppen, O.I., Baulina, O.I., and Kondratieva, E.N. (1994). *Oscillochloris trichoides* neotype strain DG-6. Photosynth. Res. *41*, 29–33.

Könneke, M., Bernhard, A.E., de la Torre, J.R., Walker, C.B., Waterbury, J.B., and Stahl, D.A. (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature *437*, 543–546.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Londry, K.L., Jahnke, L.L., and Marais, D.J.D. (2004). STable 4-Carbon Isotope Ratios of Lipid Biomarkers of Sulfate-Reducing Bacteria. Appl. Environ. Microbiol. *70*, 745–751.

Lucker, S., Nowka, B., Rattei, T., Spieck, E., and Daims, H. (2013). The Genome of Nitrospina gracilis Illuminates the Metabolism and Evolution of the Major Marine Nitrite Oxidizer. Front. Microbiol. doi: 10.3389/fmicb.2013.00027.

Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S.M., Hügler, M., Albrecht, D., Robidart, J., Bench, S., Feldman, R.A., et al. (2007). Physiological Proteomics of the Uncultured Endosymbiont of *Riftia pachyptila*. Science *315*, 247–250.

McAllister, S.M., Davis, R.E., McBeth, J.M., Tebo, B.M., Emerson, D., and Moyer, C.L. (2011). Biodiversity and Emerging Biogeography of the Neutrophilic Iron-Oxidizing *Zetaproteobacteria*. Appl Env. Microbiol *77*, 5445–5457.

Moyer, C.L., Dobbs, F.C., and Karl, D.M. (1995). Phylogenetic diversity of the bacterial community from a microbial mat at an active, hydrothermal vent system, Loihi Seamount, Hawaii. Appl. Environ. Microbiol. *61*, 1555–1562.

Moyer, C.L., Tiedje, J.M., Dobbs, F.C., and Karl, D.M. (1998). Diversity of deep-sea hydrothermal vent Archaea from Loihi Seamount, Hawaii. Deep-Sea Res II *45*, 303–317.

Nakagawa, S., and Takai, K. (2008). Deep-sea vent chemoautotrophs: diversity, biochemistry and ecological significance. FEMS Microbiol. Ecol. *65*, 1–14.

Nakagawa, S., Takai, K., Inagaki, F., Horikoshi, K., and Sako, Y. (2005). *Nitratiruptor tergarcus* gen. nov., sp. nov. and *Nitratifractor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of the ε-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough. Int. J. Syst. Evol. Microbiol. *55*, 925–933.

Nakagawa, S., Shimamura, S., Takaki, Y., Suzuki, Y., Murakami, S., Watanabe, T., Fujiyoshi, S., Mino, S., Sawabe, T., Maeda, T., et al. (2014). Allying with armored snails: the complete genome of gammaproteobacterial endosymbiont. ISME J. *8*, 40–51.

Niel, C.B. van (1932). On the morphology and physiology of the purple and green sulphur bacteria. Arch. Für Mikrobiol. *3*, 1–112.

Reardon, E.J. (1995). Anaerobic Corrosion of Granular Iron: Measurement and Interpretation of Hydrogen Evolution Rates. Environ. Sci. Technol. *29*, 2936–2945.

Robidart, J.C., Bench, S.R., Feldman, R.A., Novoradovsky, A., Podell, S.B., Gaasterland, T., Allen, E.E., and Felbeck, H. (2008). Metabolic versatility of the *Riftia pachyptila* endosymbiont revealed through metagenomics. Environ. Microbiol. *10*, 727–737.

Sato, T., Atomi, H., and Imanaka, T. (2007). Archaeal Type III RuBisCOs Function in a Pathway for AMP Metabolism. Science *315*, 1003–1006.

Schloss, P.D., and Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl Env. Microbiol *71*, 1501–1506.

Singer, E., Emerson, D., Webb, E.A., Barco, R.A., Kuenen, J.G., Nelson, W.C., Chan, C.S., Comolli, L.R., Ferriera, S., Johnson, J., et al. (2011). *Mariprofundus ferrooxydans* PV-1 the First Genome of a Marine Fe(II) Oxidizing *Zetaproteobacterium*. PLoS ONE *6*, e25386.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688 –2690.

Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynth. Res. *60*, 1–28.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Ultsch, A., and Mörchen, F. (2005). ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany, No. 46.

Warén, A., Bengtson, S., Goffredi, S.K., and Dover, C.L.V. (2003). A Hot-Vent Gastropod with Iron Sulfide Dermal Sclerites. Science *302*, 1007–1007.

Whelan, S., and Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Biol. Evol. *18*, 691 –699.

Wrighton, K.C., Thomas, B.C., Sharon, I., Miller, C.S., Castelle, C.J., VerBerkmoes, N.C., Wilkins, M.J., Hettich, R.L., Lipton, M.S., Williams, K.H., et al. (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. Science *337*, 1661–1665.

Chapter 5: A Metagenomic Analysis of the Pristine Oligotrophic Microbial Community of Warren Cave, Antarctica

Introduction

It is widely speculated that the deep subsurface of the earth supports a vast microbial biosphere that could contain significant amounts of the total earth's biomass (Fry et al., 2008; Kallmeyer et al., 2012). Most of these ecosystems are deep within the earth's crust in highly oligotrophic conditions devoid of biomass generated by photosynthetic processes. These communities have been theorized to be extremely slow growing and supported by lithotrophic processes (Nealson et al., 2005; Schippers et al., 2005). In environments without volcanic inputs, hydrogen has been theorized to be the primary electron donor for energy production (Morita, 1999). Hydrogen can be generated from anaerobic water-rock interactions (Stevens and McKinley, 1995), metamorphic transformations such as serpentinization (McCollom and Bach, 2009), and radiolysis of water (Lin et al., 2005) at concentrations generally thought to be able to support life. Ecosystems deep enough to have inputs from the mantle or with close proximity to volcanic centers may thrive on volatile magmatic inputs such as hydrogen, methane, carbon monoxide, and ammonium (Takai et al., 2004). Because of the highly oligotrophic nature of these semi-closed ecosystems, the majority of carbon must be fixed from inorganic sources which would consume a large portion of the energy budget of these organisms.

Initial microbiological studies of the deep subsurface has focused on deep crustal regions in drill holes or terrestrial-based caves and mines (Kallmeyer et al., 2012; Macalady et al., 2007; Takai et al., 2001). Extremely low indigenous biomass in these communities magnify the potential for contamination in these samples. Drill hole samples

are often contaminated by drilling fluid that is pumped into the borehole as the drilling is taking place (Smith et al., 2000). Cave and mine samples often have high amounts of human contamination from mining employees and recreational spelunkers. Some studies have attempted to minimize the impact of contamination in microbial studies by identifying and subtracting putative human borne microbes from the analysis (Onstott et al., 2003).

This study focuses on Warren Cave (Figure 5-1), a fumarolic ice cave on Mt. Erebus, as a model system for studying microbial communities in oxic oligotrophic ecosystems with volcanic inputs. Warren Cave was formed by warm fumarolic gases melting passageways beneath permanent ice cover (Giggenbach, 1976). Positive gas pressure within the cave maintains a warmer ($Temp_{ave}$= 0-4°C) humid condition relative to the cold ($Temp_{ave}$= -30°C), dry environment outside the caves (Curtis and Kyle, 2011). There are no known multicellular organisms living inside the cave, and the harsh environment on the surface of Mt. Erebus along with the permanently frozen ground between ice caves practically eliminates the surface-bound life as a source of photoautotrophically produced organics that could filter down through melt waters into the caves. Access to the caves is restricted to scientists who are trained to limit contact and exposure inside the caves and minimize cross-cave contamination of soils and ice.

A previous microbial study of Mt. Erebus fumarolic caves (Tebo et al., 2014) utilized PCR amplified clone libraries of the small subunit ribosomal gene (SSU rRNA) and the Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (Rubisco) showed that Warren Cave had very low diversity and was dominated by *Chloroflexi*, *Acidobacteria*, and *Alphaproteobacteria* that are not similar to any described isolates.

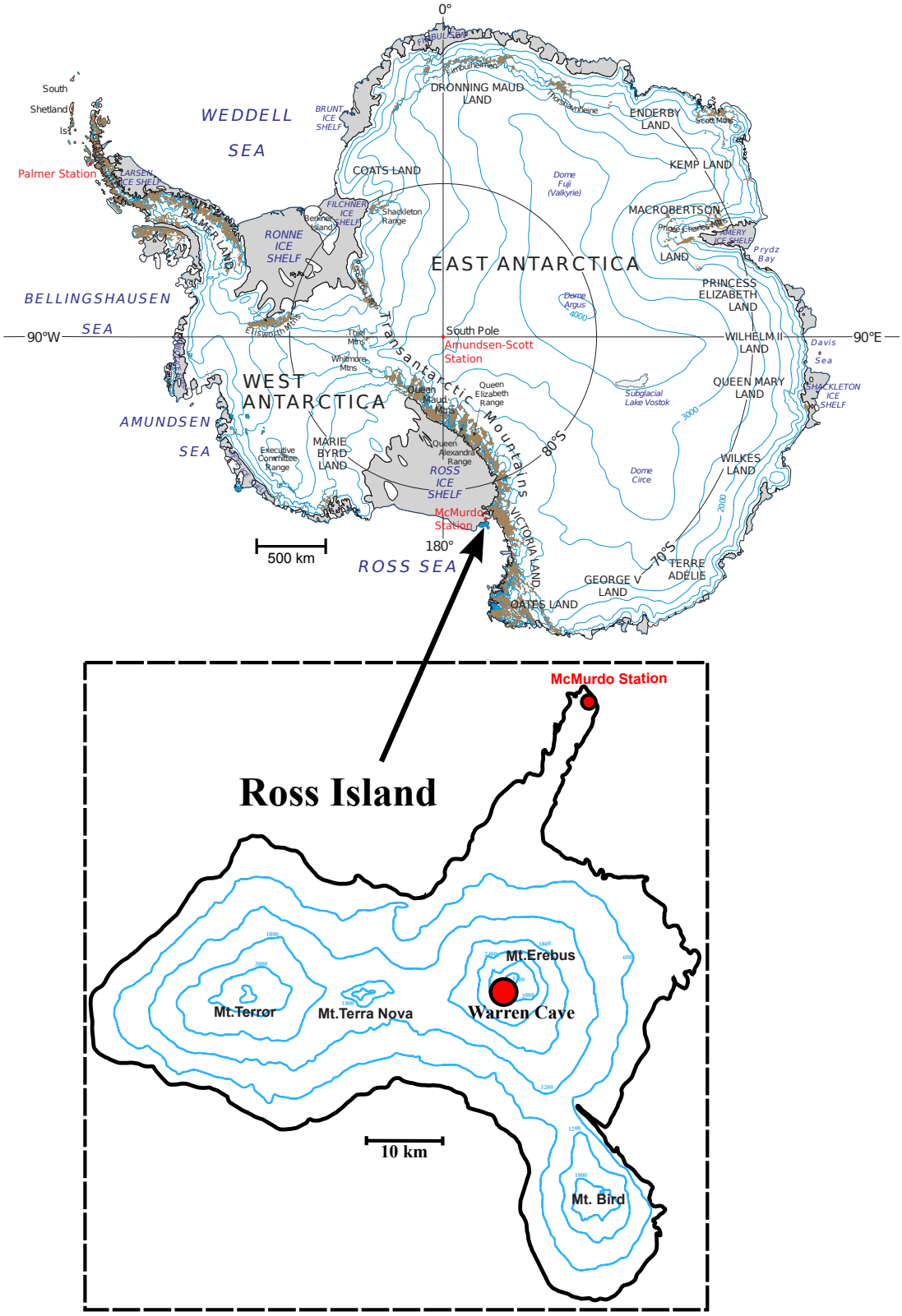

Figure 5-1. Map of Antarctica. Insert is Ross island with the location of Warren Cave. Map is modified from the Lansat Image Mosaic of Antarctica, available: http://lima.usgs.gov.

Phylogenetic analysis of the Rubisco genes showed that the community contained genes that clustered within the form I 'Red Type' Rubisco, many of which did not cluster with described isolates. These analyses show that the community at Warren Cave are capable of carbon fixation, but do not show which organisms are fixing the carbon and what chemoautotrophic metabolisms are utilized to generate the energy needed for carbon fixation.

To determine which organisms are fixing carbon and what mechanisms are used to generate biological energy, we chose a metagenomic approach to study these cave communities. The low diversity of Warren Cave should allow de novo assembly of large genomic fragments using next generation sequencing. This will allow analysis of the mechanisms of chemoautotrophic growth in oxic oligotrophic environments with volcanic inputs.

Material and Methods

Sample collection

Fumarole sediments were collected in November 2010 near a fumarole vent emitting gas at 18.5°C. The sediments consisted of porous glassy material with an average size of approximately 1 mm diameter. The sediments were collected aseptically using sterile 50 mm conical tubes and immediately sealed. The samples were kept cold until they were transported to the McMurdo laboratory where they were frozen at -80°C and shipped to the US and stored until DNA extraction.

DNA extraction

Total genomic DNA (gDNA) was extracted from eight grams of soil using 10 replicate extractions with the FastDNA Spin Kit for Soil following the manufacturer's protocol (Qbiogene, Irnive, CA). Extracted gDNA was pooled, cleaned, and concentrated using Montáge PCR centrifugal filter devices (Millipore, Bedford, MA) and eluted in 10 mM Tris 0.1 mM EDTA (pH 8.0). The gDNA was then quantified using a Nanodrop ND-1000 (Thermo Scientific, Wilmington, DE) spectrophotometer and frozen at -20°C until library construction.

Metagenomic library construction and sequencing

A metagenomic sequencing library was constructed from 50 ng of gDNA using the Nextera DNA sample preparation kit (Illumina, Inc., San Diego, CA). The libraries were quantified using QPCR and sequenced with two lanes of single-end 101 bp runs on a HiSeq 2000 sequencer resulting in 444,243,598 sequence reads.

Pre-assembly taxonomic analysis

Small subunit ribosomal RNA sequences (SSU rRNA; 16S rRNA) were extracted from the raw sequences using the program Hmmer (Eddy, 2011) using a profile hidden Markov model (profile HMM) which was calculated using all SSU rRNA sequences from the Silva SSU Ref NR database (Pruesse et al., 2007) (release 111; 286,858 sequences) using the hmmsearch function. Putative SSU rRNA sequences were exported and classified using the SINA standalone aligner (Pruesse et al., 2012) with the Silva SSU Ref NR database using the Greengenes taxonomy with a 70% sequence similarity cutoff.

Sequence pre-processing

Sequence ends were trimmed to remove transposon sequence contamination from the Nextera library kit using the FASTA/Q clipper program from the FAST-X Toolkit version 0.0.13 (Available: http://hannonlab.cshl.edu/fastx_toolkit) using the search string CTGTCTCTTATA as the adapter sequence. The sequences were digitally normalized using the program Khmer (Crusoe et al., 2014) utilizing a two-pass strategy (Brown et al., 2012). The sequences were first normalized by median to reduce sequences with high-abundance kmers using a kmer size of 20bp and a maximum coverage overlap value of 20. The sequences were then filtered to remove low-abundance kmers and Illumina sequencing artifacts using the Khmer script filter-below-abund with a minimum overlap of 2 kmers. The remaining sequences after preprocessing were used for sequence assembly.

Sequence assembly

Contiguous sequences (contigs) were assembled using the de Bruijn graph-based assembler Velvet (Zerbino and Birney, 2008). Two independent assemblies were constructed, one with a kmer size of 41 and a second with a kmer of 51. The contigs were then assembled into supercontigs using the program Geneious (Geneious, Auckland, NZ) with a minimum overlap of 25bp. Contigs longer than 1000bp were then exported and used in further analysis. Read coverage was calculated by mapping the pre-normalized de-replicated reads to the assembled contigs using the program Bowtie 2 (Langmead and Salzberg, 2012). The resulting Sam file was converted to a sorted Bam file using

BamTools (Barnett et al., 2011) from which mean coverage for each contig was calculated using a pileup file.

Sequence clustering and binning

Contigs were grouped to form phylotype clusters by constructing an emergent self organizing map (ESOM) calculated from the tetranucleotide frequencies of each contig (Dick et al., 2009). Tetranucleotide frequencies were calculated as using a minimum contig size of 2.5 kb and contigs larger than 10 kb were split into 5 kb fragments. The frequencies were loaded into Databionic ESOM tools (Ultsch and Mörchen, 2005) and normalized using RobustZT. The ESOM was calculated using the K-batch training alogrithm with a map size of 238 rows X 1074 columns. The map was visualized with a UMatrix background using a hot gradient clipped to 25% with 16 colors. The clusters were manually drawn on a tiled map. Split contigs that mapped to different clusters were generally solved using consensus-rules or via blast homology.

Contigs from each ESOM clusters were then plotted by coverage to determine if more than one genomic bin could be extracted from a cluster. The sequences were sorted by mean read coverage and visualized on a scatter plot using the data mining software Orange (available:http://orange.biolab.si/). Clusters showing multiple groupings of contig coverage values were then sub-clustered by converting the coverage values to Euclidean distances and utilizing hierarchical clustering with Ward's linkage with a cutoff value of approximately 1.0. Contigs were extracted from the clusters and were now called genome bins and numbered sequentially by decreasing mean read coverage of the contigs in each bin.

Functional analysis

Genes were predicted using the program Prodigal version v2.60 (Hyatt et al., 2010) using the metagenomic procedure for gene identification. The translated genes were queried against the NCBI-NR database using blastp with a minimum evalue cutoff of 1e-20 and maximum descriptions set at 10. The blastp results for each bin were separately imported into MEGAN4 (Huson et al., 2011) and KEGG pathways (Kanehisa and Goto, 2000) were calculated and visualized using the default settings.

Phylogenetic analysis

Sequences for phylogenetic analysis were found using HMMER. Sequences for elongation factor G, *recA*, aerobic carbon monoxide dehydrogenase (CODH), and ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco) were downloaded from the Integrated Microbial Genomes (Markowitz, 2006) website and the sequences were aligned using the program MUSCLE (Edgar, 2004) and the resulting alignments were used to build a profile HMM file for each gene. All assembled contigs were then searched for homologous genes for each HMM profile using the hmmsearch function of HMMER. Full length sequences with an hmmsearch inclusion threshold of 1e-50 were extracted from contigs and similar sequences were downloaded from the NCBI nr database. Sequences were aligned using MUSCLE and the alignments were visualized and corrected using the multiple sequence alignment editor Bioedit (Hall, 1999). The alignments were then masked to include only amino acid positions that were unambiguously aligned. Phylogenetic trees was calculated using RAxML (Stamatakis,

2006) using the WAG (Whelan and Goldman, 2001) model of amino acid substitution and using the best tree from 10 alternative runs. Bootstrap values were calculated from a consensus tree of 100 boostrap replicates. The trees were visualized using Mega4 (Tamura et al., 2007) and annotated using the vector drawing program Inkscape (http://www.inkscape.org).

Genomes that were classified as belonging to the Chloroflexi class were aligned using the BLAST Ring Image Generator (BRIG) (Alikhan et al., 2011) using *Ktedonobacter racemifer* as the reference genome.

Results

Metagenomic library construction and assembly

Warren Cave soils exhibit low biomass, so it was necessary to pool, clean, and concentrate multiple DNA extractions to obtain genomic DNA that was concentrated and pure enough for metagenome construction. The Illumina Nextera kit was used to construct the library due to the low quantity of DNA needed for the transposon / PCR based kit (50 ng) vs. the more traditional physical shearing method of library construction (2 μg). This method produced many reads that were 50-75 bp in length and needed to be trimmed of contaminating adapters from our 101 bp reads. The library was digitally normalized to produce a higher quality assembly while reducing the memory requirement required by de Bruijn graph-based assemblers. High abundance sequences which would not contribute additional sequence information and low abundance sequences which would not assemble into contigs were removed from the analysis (Brown et al., 2012). The normalized sequence reads were then assembled into contigs using two different

kmer values (41 and 51) and then merging the result. Utilizing radically different kmer lengths during construction helps minimize assembly biases found in unevenly distributed sequences (Chikhi and Medvedev, 2014).

Sequences were binned using tetranucleotide frequencies coupled with ESOM sorting. The sequences sorted into 20 distinct clusters (Figure 5-2). The sequences from each cluster was extracted and initial functional and taxonomic screenings showed that most of the bins consisted of multiple genomes of closely related bacteria (data not shown). The clusters were then subdivided to form “genome bins' using the mean coverage of each contig as an indicator of the organismal abundance within a mixed cluster (Figure 5-3). The bins were numbered sequentially using the mean coverage from each bin (Table 5-1). Bins with mean contig coverage below 30X were found to contain an unaccepTable 5-amount (>20%) of mis-binned or chimeric contigs and were removed from further genome analysis, but were included in phylogenetic analyses of single genes.

Microbial diversity of the Warren Cave metagenome

The microbial diversity of the Warren Cave metagenome was assessed both pre- and post-assembly. Small subunit ribosomal RNA genes were used to assess the pre-assembly diversity. Reads containing putative SSU rRNA sequences were identified and 25,710 of these reads were classified to the phylum or class level. The reads were dominated by the *Chloroflexi* phylum (47%) with lesser amounts of *Actinobacteria* (15%), *Acidobacteria* (14%), and *Alphaproteobacteria* (13%) (Figure 5-4). *Eukaryotes* and *Archaea* represented less than 0.05% of the sequences.

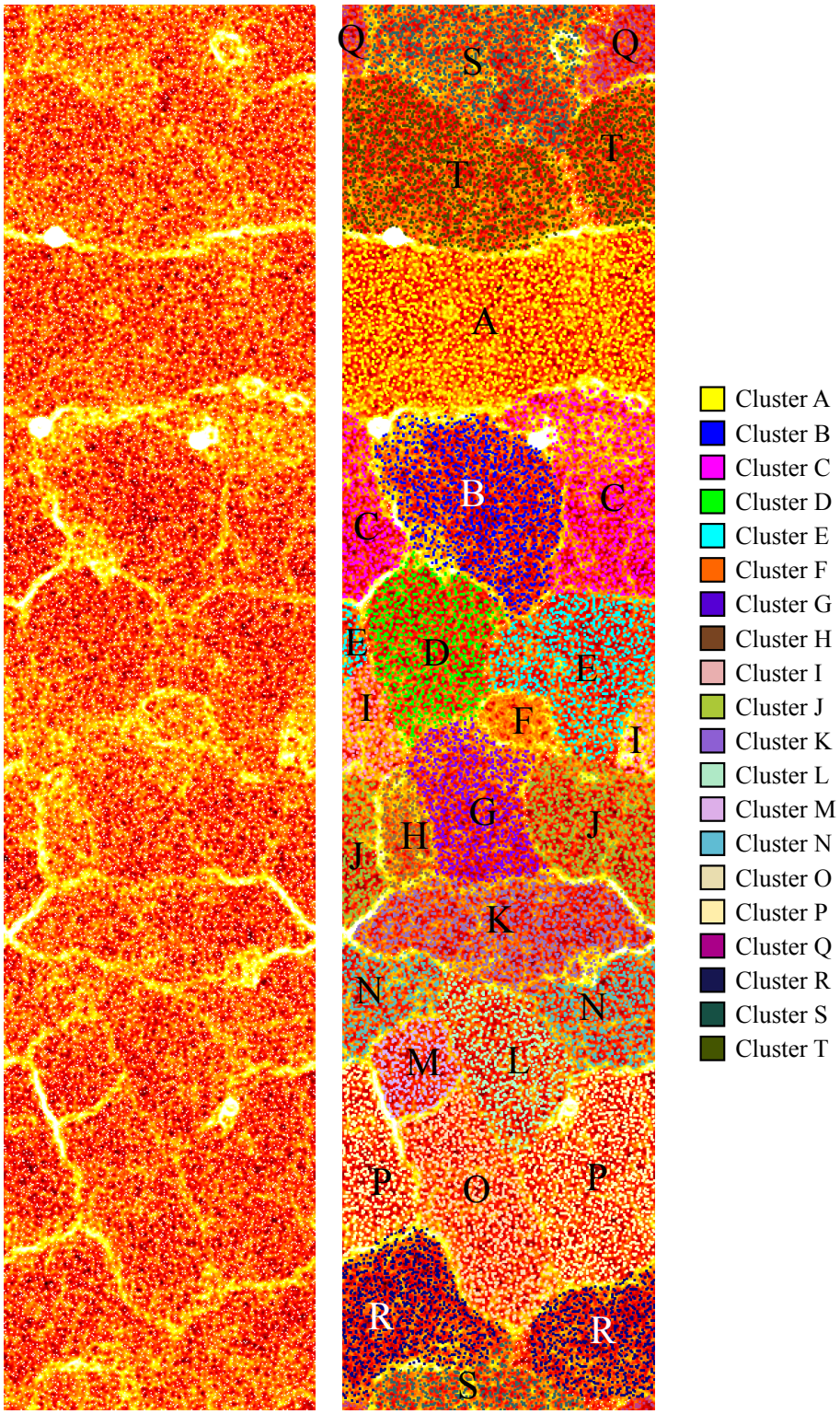

Figure 5-2. Edgeless emergent self-organized map of contigs using tetranucleotide frequencies. Map on left is the finished map with contigs represented by white dots, map on right shows the 20 clusters formed by the map with contigs within a group colored for easier visualization.

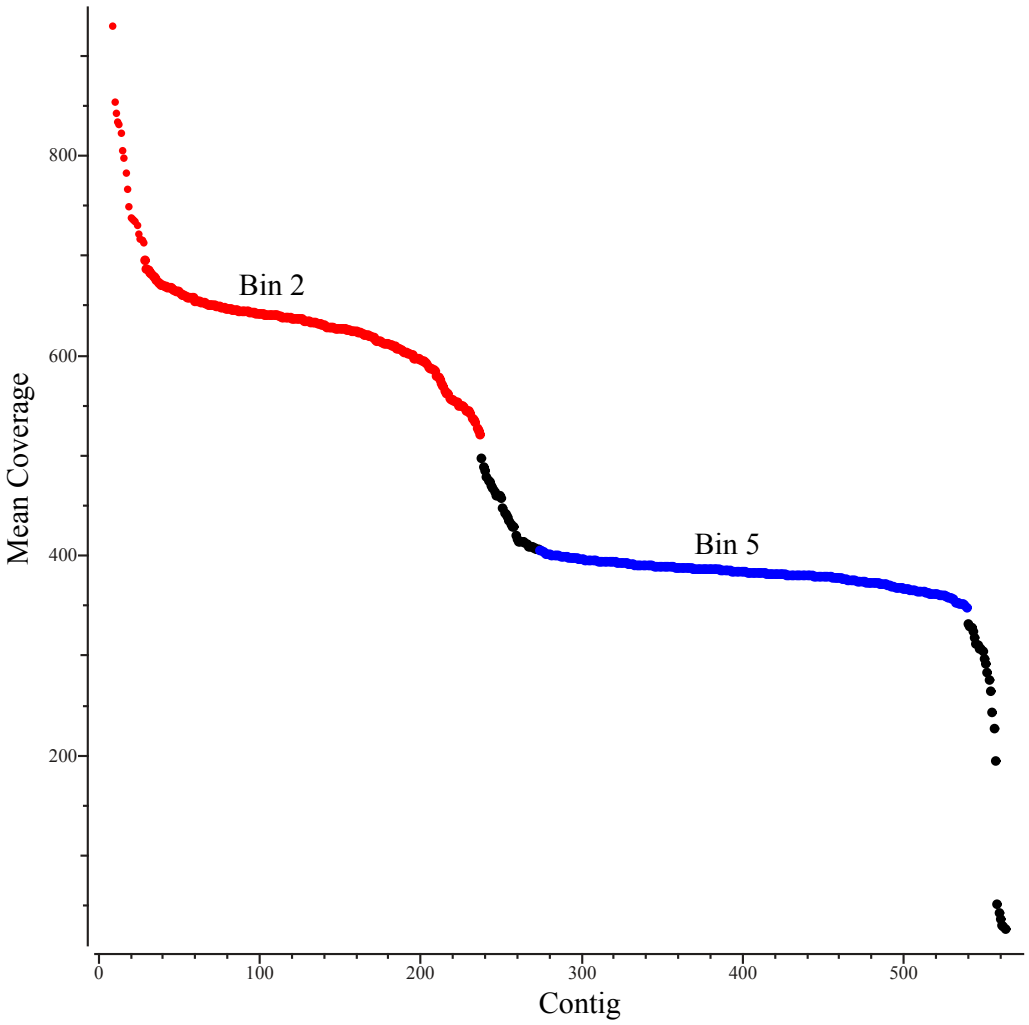

Figure 5-3. Coverage plot of contigs found in ESOM cluster A. Contigs in red were grouped into Bin 2, blue in Bin 5, and black were not placed in a bin.

Table 5-1. Genomic and phylogenetic characterization of genome bins from Warren Cave.

Genome Bin	ESOM Cluster	Phylogenetic Group	Mean Coverage	GC Content (%)	Bin Size (Mbp)	Number of Contigs	N50 (bp)	Gene Count	Estimated Completeness (%)
1	T	Acidobacteria	807 ± 72	56	3.84	135	61094	3368	87
2	A	Chloroflexi	624 ± 55	52	6.57	228	48484	5979	110
3	P	Chloroflexi	568 ± 41	69	2.50	83	63619	2539	87
4	C	Chloroflexi	406 ± 24	68	2.81	185	22635	2756	103
5	A	Chloroflexi	381 ± 12	50	5.43	264	31707	4961	32
6	P	Chloroflexi	362 ± 47	73	8.01	361	35974	8209	194
7	S	Alphaproteobacteria	310 ± 20	61	0.79	99	9496	973	0
8	R	Alphaproteobacteria	274 ± 32	65	3.35	108	44987	3372	106
9	E	Actinobacteria	221±14	68	4.68	123	78255	4901	94
10	I	Actinobacteria	183 ± 13	64	0.34	41	10256	417	10
11	C	Chloroflexi	181 ± 6	67	2.35	123	31374	2462	84
12	O	Actinobacteria	178 ± 26	68	3.50	115	59122	3445	48
13	S	Alphaproteobacteria	139 ± 14	60	2.35	382	6937	2573	71
14	B	Planctomycete	74 ± 3	65	6.93	387	26585	5451	97
15	J	Chloroflexi	66 ± 6	63	2.81	103	50170	2772	100
16	G	Betaproteobacteria	59 ± 3	63	3.45	143	38109	3280	74
17	M	Chloroflexi	61 ± 8	71	2.33	91	40698	2384	94
18	D	Actinobacteria	52 ± 3	67	4.15	177	42128	4278	97
19	R	Alphaproteobacteria	43 ± 4	65	4.92	530	12582	4897	61
20	T	Acidobacteria	41 ± 4	56	3.56	124	48082	3054	90
21	L	Chloroflexi	35 ± 3	68	4.14	385	14191	3890	84

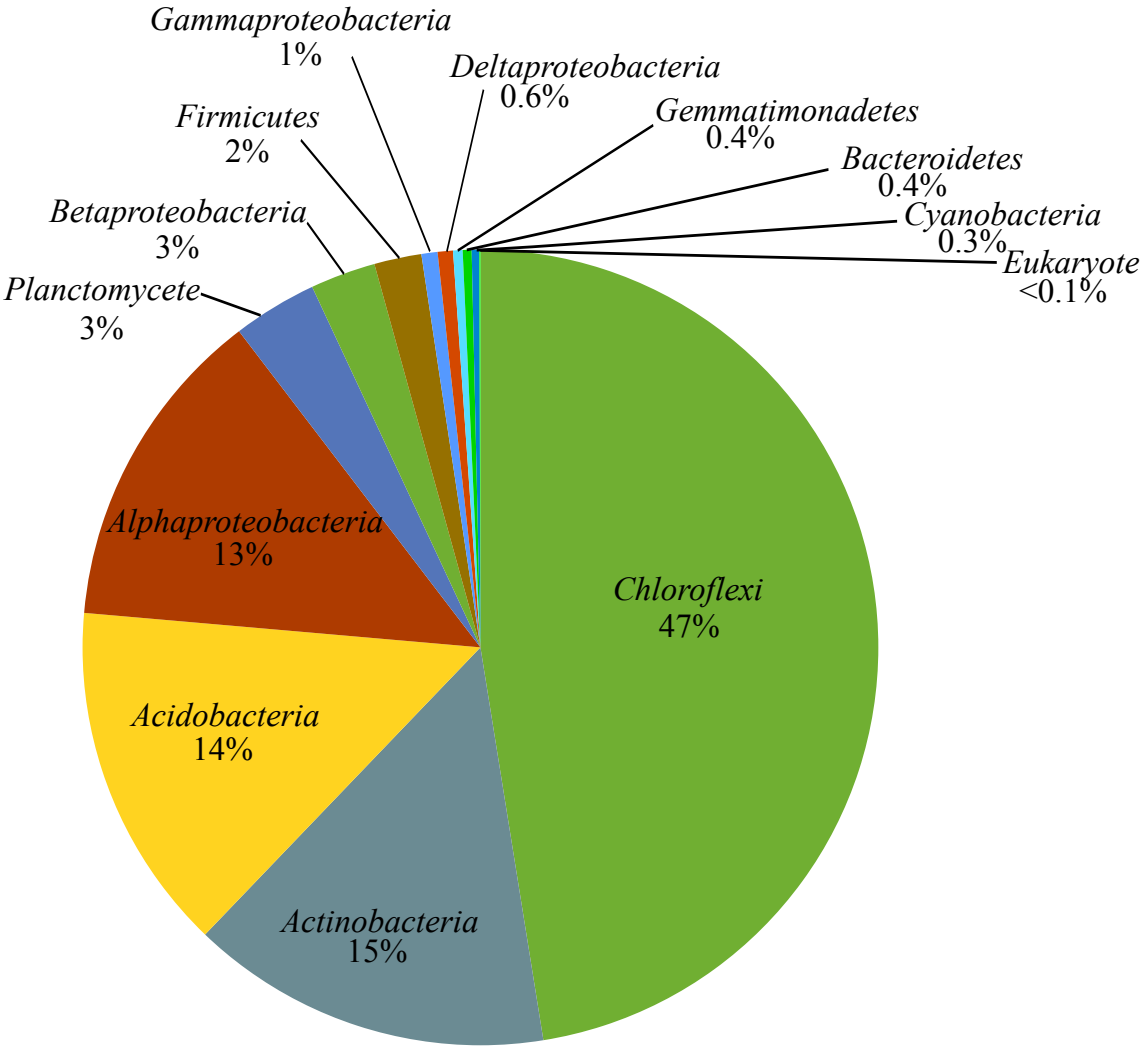

Figure 5-4. Relative proportions of phylotypes derived from taxonomic classification of unassembled SSU rRNA genes.

Ribosomal RNA genes did not assemble well in our analysis, which is consistent with other metagenomic assemblies utilizing short reads. This is possibly due to the highly variable mutational clock speeds of the gene preventing assembly of complete genes or because of multiple copies of the rrn operon breaking the assembly. Post-assembly diversity was assessed using the *recA* (Figure 5-5) and elongation factor G (EF-G) (Figure 5-6) genes which are highly conserved housekeeping genes with well-known phylogenetic properties (Cammarano et al., 1992; Lloyd and Sharp, 1993).

The genome bin with the highest coverage belonged to the *Acidobacteria* phylum, which are commonly found in culture-independent phylogenetic studies of microbial soil communities (Jones et al., 2009, Table 5-1), however Bin 1 and Bin 20 did not cluster closely with any sequenced Acidobacteria species. The next five most abundant bins all grouped within the *Chloroflexi* phylum. Bins two and five were closely related to *Ktedonobacter racemifer*, while Bins three, four, and six formed a distinct cluster within the Chloroflexi along with Bins 11 and 15. Bins seven and eight belong to the *Alphaproteobacteria* class and are most related to the *Hyphomicrobiaceae* family of "purple" bacteria, while Bin 13 clusters within the methylotrophic *Alphaproteobacteria*. Bins 9, 12, 15, and 18 clustered with the *Actinobacteria* phylum. Bin 9 clustered with *Acidimicrobium ferroxidans* which is a moderately thermophilic iron-oxidizing bacteria that was isolated from a geothermal source in Iceland, while Bin 18 clustered within the Pseudonocardiaceae family which are commonly found in soils. Bin 14 clusters within the *Planctomycetes* phylum clustering with an uncultured cosmid clone from a deep gold mine. Bin 16 clustered within the *Betaproteobacteria* class and clustered with *Leeia oryzae*, which is a aerobic bacterium isolated from rice paddy soil (Lim et al., 2007).

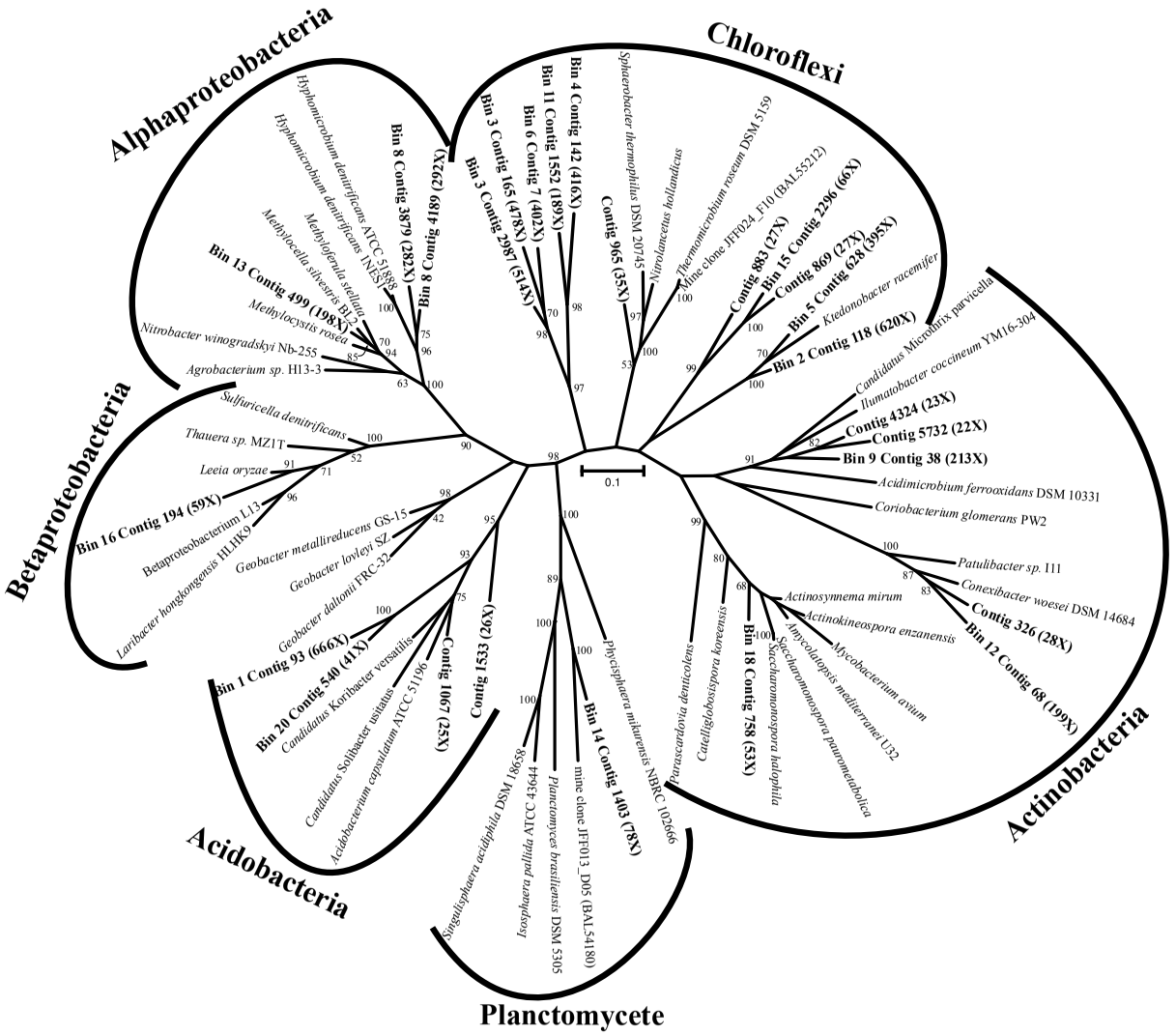

Figure 5-5. Unrooted maximum likelihood phylogenetic tree of recA genes. Sequences in bold are metagenomic sequences from this study with bin numbers if binned. Numbers in parentheses are deduplicated mean read coverage of the contigs which the sequences are found.

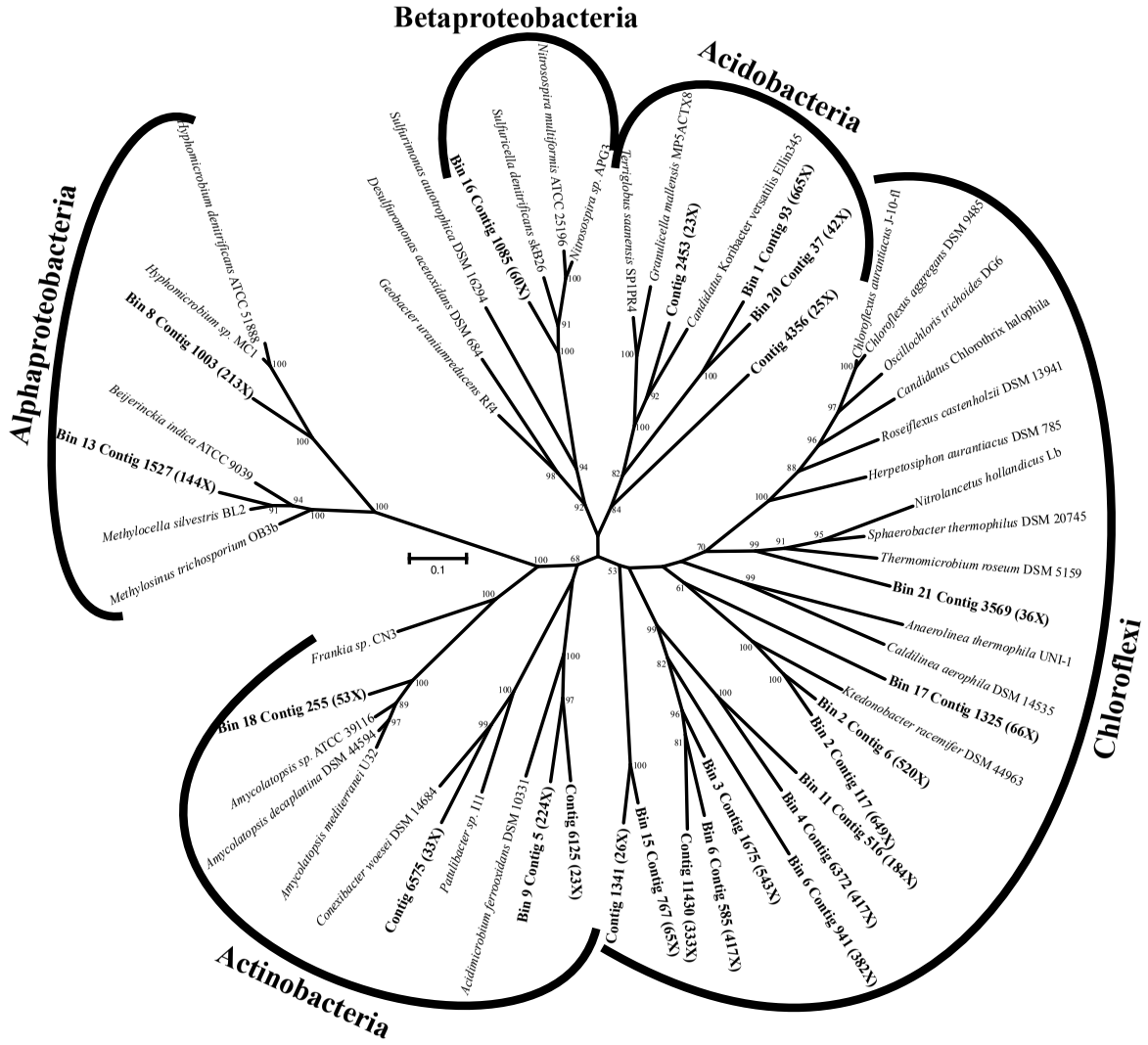

Figure 5-6. Unrooted maximum likelihood phylogenetic tree of Elongation factor G genes. Sequences in bold are metagenomic sequences from this study with bin numbers if binned. Numbers in parentheses are deduplicated mean read coverage of the contigs which the sequences are found.

Carbon fixation genes

The only carbon fixation pathway found in the Warren Cave metagenome was the Calvin Cycle. Key enzymes from the reductive TCA (rTCA) cycle, the Wood-Ljungdahl (reductive acetyl-CoA) pathway, the 3-hydroxypropionate cycle, 3-hydroxypropionate/4-hydroxybutyrate cycle, and the Dicarboxylate/4-Hydroxybutyrate cycle were not found using Hmmer or Blast homology searches. All of the Rubisco sequences grouped in the 'Form ID' and 'Form IE' clades (Figure 5-7).

Rubisco genes clustering within the 'Form 1D' cluster include Bin 1 and several low-coverage sequences which clusters with the Alphaproteobacteria purple bacteria, Bin 16 which clusters with the deeply rooted Gammaproteobacteria and nitrogen-fixing mesorhizobium, and Bin 19 which clusters with uncultured clones from a hydrothermal plume.

Rubisco genes clustering within the 'Form 1E' cluster were Bin 2 which clusters with the gram+ bacteria *Sulfobacillus acidophilus*, an iron and sulfur oxidizer, and *Kypidia tusciae* which oxidizes hydrogen. A sequence from Bin 18 clusters with *Pseudonocardia sp.* P1 which was isolated from a leafcutter ant. Sequences from Bins 3,4,6,9, and 12 clustered together in a cluster that does not include a described isolate. The only other sequences falling within this bin were uncultured soil clones from the arid regions of Northwest China and from agricultural topsoil.

Nitrogen assimilation pathways

Assimilatory nitrate and nitrite reduction genes were found in many of the

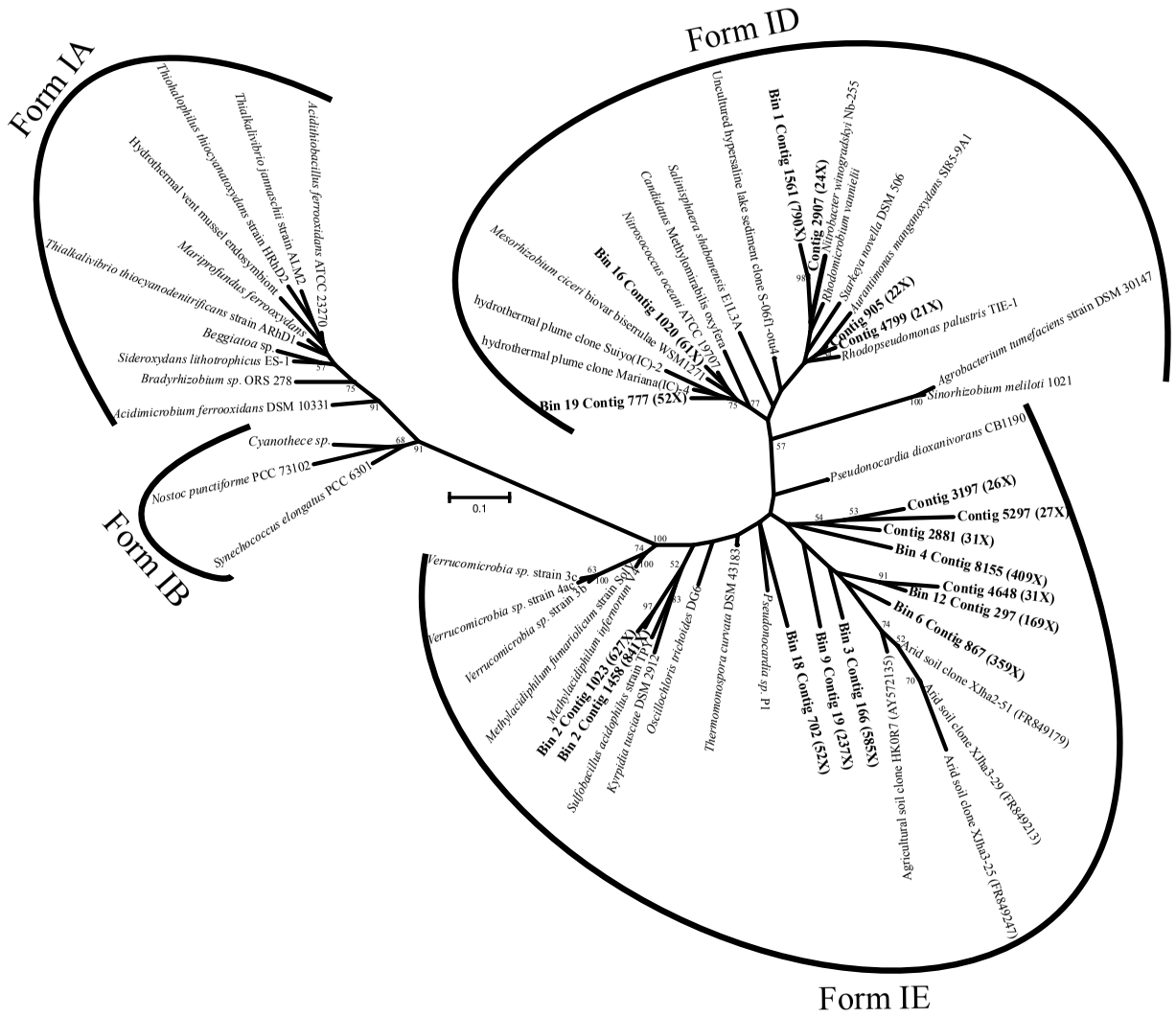

Figure 5-7. Unrooted maximum likelihood phylogenetic tree of Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (Rubisco) genes. Sequences in bold are metagenomic sequences from this study with bin numbers if binned. Numbers in parentheses are deduplicated mean read coverage of the contigs.

autotrophic bins. No genes were found for dissimilatory nitrate reduction, ammonia oxidation, or nitrogen fixation were found in the metagenome.

Lithotrophic pathways

Three known lithotrophic pathways were identified in the Warren Cave metagenome that could potentially fuel autotrophic carbon fixation. Functional genes or phylogenetic groups associated with known metal oxidation (Belchik et al., 2012; Geszvain et al., 2013) were not found. Sulfur oxidation genes (sox system) were found only in low abundance Alphaproteobacterial bins 15 and 19. The only other lithotrophic, energy yielding metabolisms that were found are carbon monoxide oxidation and molecular hydrogen oxidation.

The aerobic carbon monoxide dehydrogenase (CODH) gene was found in many of the most abundant bins (Figure 5-8). Sequences belonging to Bins 5 and 2 clustered closely with sequences from *Ktedonobacter racemifer* along with uncultured CODH clones from historic volcanic lava flows (Dunfield and King, 2004; Weber and King, 2010). Sequences from Bins 4 and 9 also clustered within the form I CODH group, with the Bin 9 sequence clustering with the Deltaproteobacterium *Haliangium ochracem* and Bin 4 clustering with uncultured sequences from Qinghai-Tibetan lakes and the Kilauea Caldera Rim. Other low-abundance sequences clustered within the Actinobactrerial and Proteobacterial Form I clades. Sequences from Bins two, five, and nine have sequences that cluster within the Form II as well as the Form I. Other putative CODH sequences clustered within the Form II clade include sequence from Bins 6, 11, 12, 15, 16, 19, and 20.

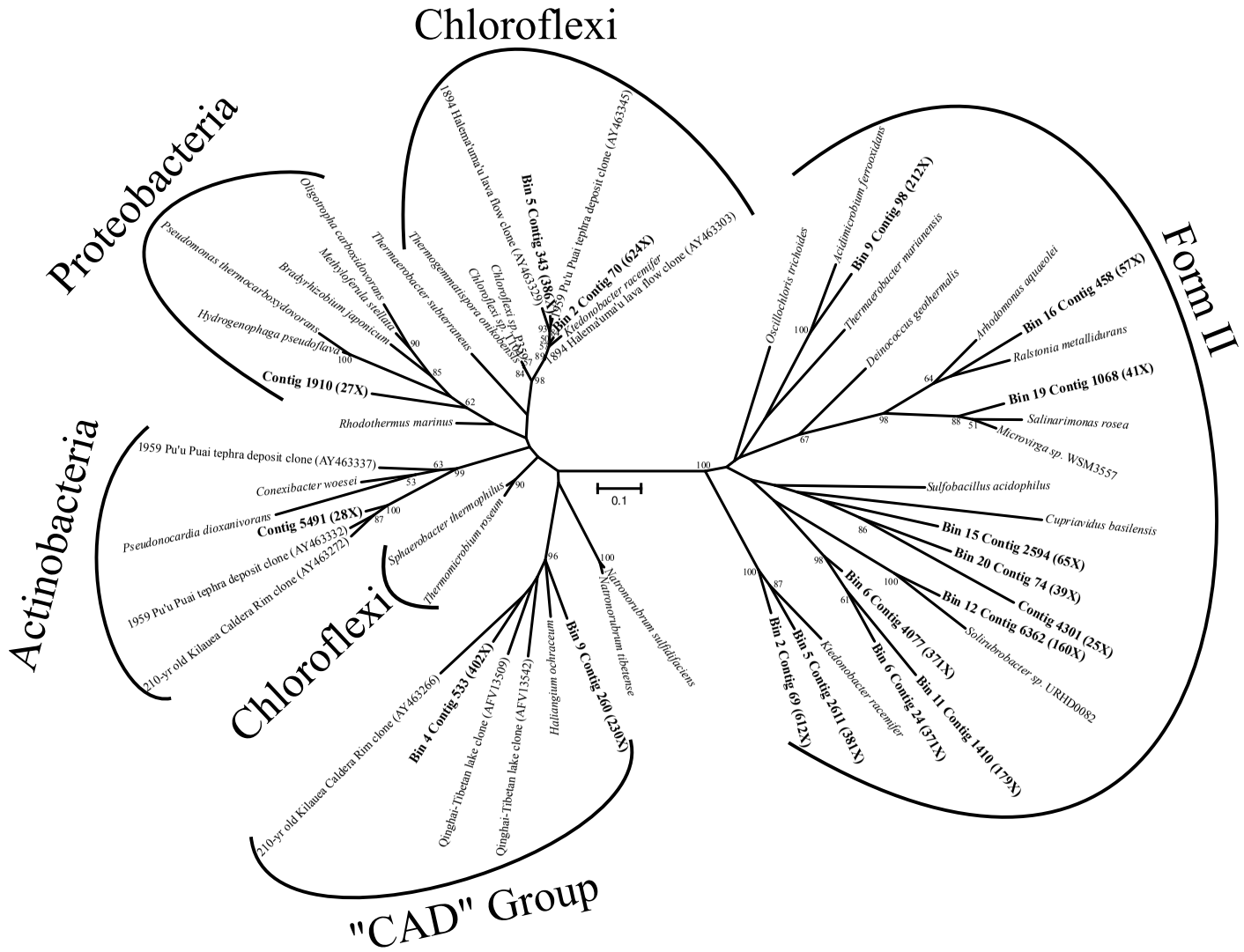

Figure 5-8. Unrooted maximum likelihood phylogenetic tree of aerobic carbon monoxide dehydrogenase (CODH) genes. Sequences in bold are metagenomic sequences from this study with bin numbers if binned. Numbers in parentheses are deduplicated mean read coverage of the contigs

Multiple sequences clustering within the 'high-affinity' [NiFe]-hydrogenases were also found in the metagenome (Figure 5-9). These hydrogenases have been shown to oxidize molecular hydrogen gas at atmospheric concentrations (Constant et al., 2010). Sequences from seven of the ten most abundant bins clustered within this group, including sequences from all six most abundant groups. The sequence from Bin 1 clustered with *Acidobacteria* isolates suggesting that this metabolic process is common in this phylum. Seqeunces from bins that are from the *Chloroflexi* phylum cluster either with *Ktedonobacter racemifer*, as is the case with Bins 2 and 5, or weakly with *Methylocystis parvus* which is a methanotrophic Alphaproteobacteria (Hou et al., 1979). The *Actinomycete* bins clustered with other Actinomycete sequences with Bins 9 and 18 clustring with *Pseudonocardiacae* and *Corynebacterineae* actinomycetes, while two sequences from Bin 12 cluster with *Conexibacter woesei* and uncultured soil clones.

Discussion

The metagenome assembly and the unassembled SSU rRNA genes had good overall agreement with diversity and relative abundances of phylotypes found in PCR clone libraries generated from Warren Cave (Tebo et al., 2014). Clustering the contigs using tetranucleotide frequencies utilizing self-emerging maps successfully divided the contigs into groups that contained highly related phylotypes. Attempts to utilize Blast homologies coupled with NCBI taxonomy to initially cluster the sequences failed even when using a phylum-level cutoff, probably due to the large evolutionary distance between genes on the contigs and sequenced isolates. After annotating the genes found in each cluster, it became apparent that multiple, closely related genomes were in many of

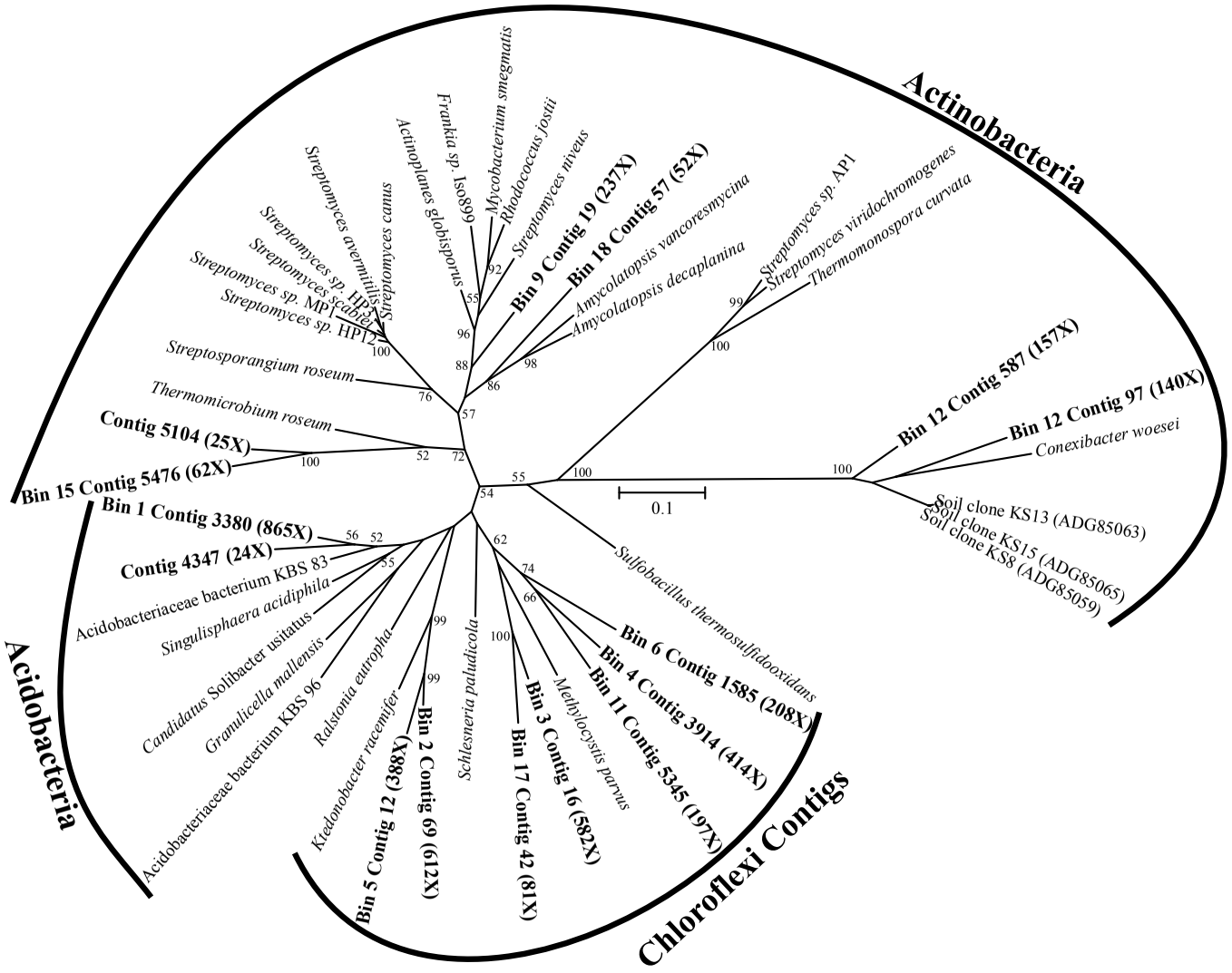

Figure 5-9. Unrooted maximum likelihood phylogenetic tree of 'high-affinity' [NiFe]-hydrogenases genes. Sequences in bold are metagenomic sequences from this study with bin numbers if binned. Numbers in parentheses are deduplicated mean read coverage of the contigs which the sequences are found.

the clusters. Because of the low diversity and low evenness in the community, these populations were then separated by the mean coverage of each contig. Analysis of single-copy housekeeping genes indicate that this method worked in all clusters except cluster 16 which had at least two genomes with almost identical coverage.

The *Chloroflexi* dominate the cave metagenome in both the unassembled reads and the bin coverage, and also is in agreement with a SSU rRNA clone library generated from the same sample collection (Tebo et al., 2014). Sequences with high similarity to these SSU rRNA *Chloroflexi* clones have been found in high altitude soils, mine tailings and other oligotrophic soil environments suggesting similar organisms are common in soil communities. Genomic analysis of housekeeping genes shows that these organisms are not similar to cultured Chloroflexi whose genomes have been sequenced, and a genomic alignment of these cave *Chloroflexi* with the soil bacterium *Ktedonobacter racemifer* shows that metagenomic Bins 2, 5, 4, and 6 are more similar to this isolate than to the other closely related sequenced *Thermomicrobia* chloroflexi hot spring isolates *Thermomicrobium roseum* and *Sphaerobacter thermophilus* (Figure 5-10). The *Ktedonobacter racemifer* genome is unique as it has the second largest genome (13.7 Mbp) of any prokaryote yet sequenced and contains a large number of redundant sequences associated with transposon insertions (Chang et al., 2011). Despite high sequence similarity to *Ktedonobacter racemifer* housekeeping and functional genes, the Warren Cave *Chloroflexi* Bins 2 and 5 have bin sizes of 6.57 Mbp (110% conmplete) and 5.43 Mbp (32% complete) respectively indicating that these genomes do not contain the same sequence redundancy or diversity that *Ktedonobacter racemifer* genome does. The remaining *Chloroflexi* bins appear to be from an uncultured clade within the *Chloroflexi*,

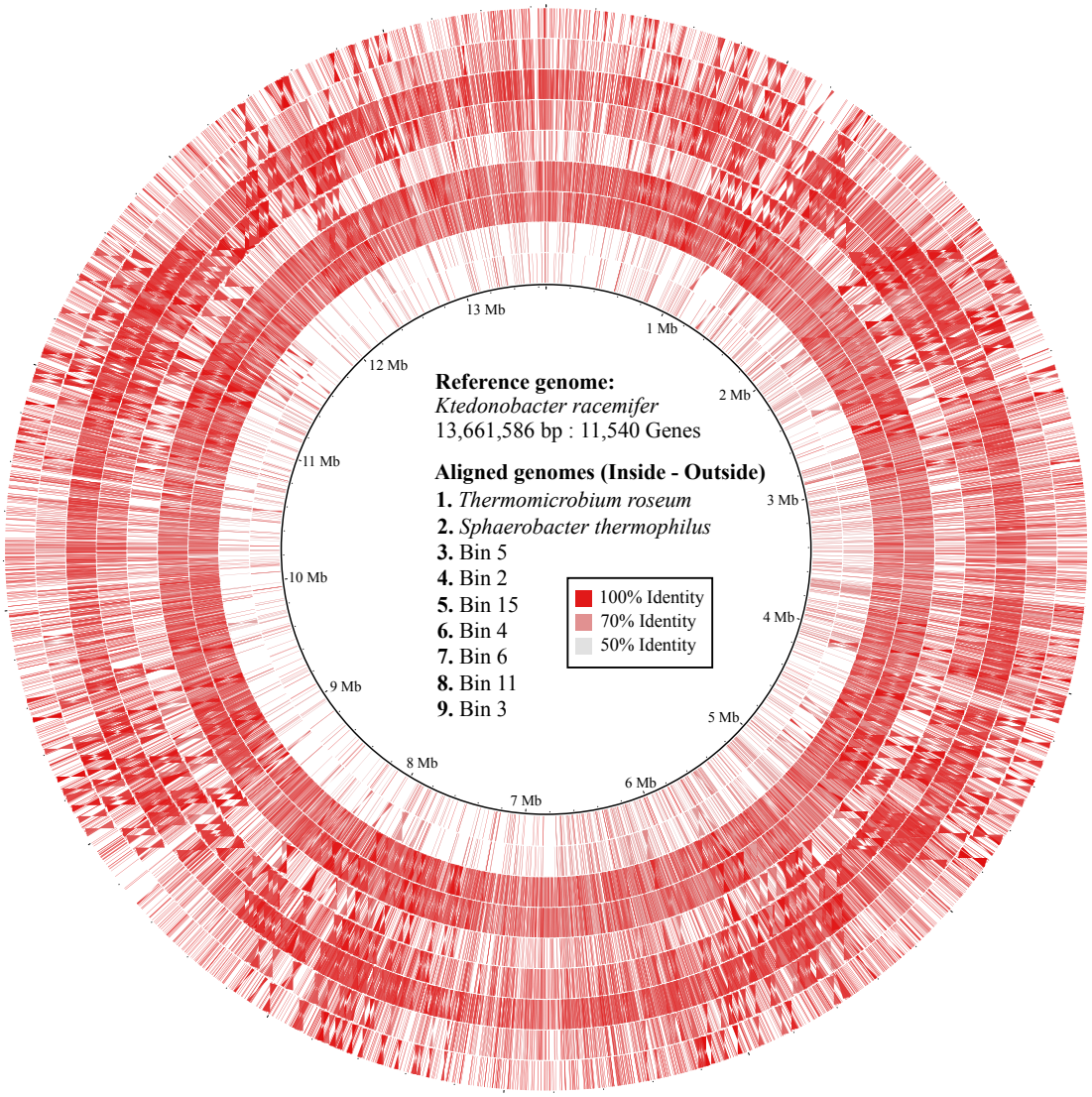

Figure 5-10. Genome alignment of bins associated with the *Chloroflexi* phylum using *Ktedonobacter racemifer* as the reference genome. Dark red portions of the genomes represent regions of high sequence similarity with the K. racemifer genome, while white areas represent areas showing very low or no sequence similarity.

as neither the housekeeping or functional genes cluster closely with sequenced organisms. These bins were generally smaller than *Ktedonobacter* bins and had a GC content greater than 63% which is consistent for some other *Chloroflexi*, such as *Thermomicrobium roseum* (64%) (Wu et al., 2009) and *Sphaerobacter thermophilus* (68%) (Pati et al., 2010), but not for *Ktedonobacter racemifer* (54%).

The Acidobacteria bins, which contained the single most abundant bin (Table 5-1), also did not cluster with any described, sequenced genomes within the phylum. Unlike the *Chloroflexi* bin, the *Acidobacteria* only formed two bins, Bin 1 which had the highest mean coverage of any single bin and Bin 20 which had one of the lowest coverages and was closely related to Bin 1. The high abundance but low diversity of the *Acidobacteria* bins suggests that this particular group of *Acidobacteria* are highly adapted to this ecosystem, but this adaptation is not prevalent throughout the *Acidobacteria* phylum. The *Acidobacteria* are commonly found in culture-independent studies of soil microbial communities, but relatively few strains have been isolated and characterized (Ward et al., 2009). To date, no members of the *Acidobacteria* have been found to fix inorganic carbon, however all genes associated with the Calvin Cycle were found in the Bin 1 sequences suggesting that this organism is capable of autotrophic growth. The only energy-yielding lithotrophic metabolism found in this bin was hydrogen oxidation, which can lead to the logical conclusion that the *Acidobacteria* can sustain chemoautotrophic growth using hydrogen oxidation coupled with the Calvin Cycle.

The Actinomycete bins also contained genes for lithoautotrophic growth. Bin 9 is most closely related to the acidophillic iron oxidizing bacterium *Acidimicrobium ferrooxidans*, however most housekeeping genes shared <70% identity with any cultured

organism. Bin 9 contains a complete Calvin Cycle suggesting that it is capable of autotrophic carbon fixation. The bin also contains both 'Form I' and 'Form II' CODH genes along with a copy of the high affinity hydrogenase gene. Bin 12 is more closely related to *Conexibacter woesei*, which is a soil bacterium deeply rooted in the Actinomycete phylum (Monciardini et al., 2003). This bin also contains all of the genes for the Calvin Cycle and the high affinity hyrogenase gene, however it lacks 'Form I' CODH genes and only contains 'Form II' genes. Bin 18 was more closely related to the *Amycolatopsis* group of *Actinomycete*, and also had a complete Calvin cycle and a high affinity hydrogenase gene, however this bin did not contain CODH genes.

The *Alphaproteobacteria* bins found in the metagenome appear to be heterotrophs with the exception of Bin 19 which contains a complete Calvin Cycle and the sox genes (Friedrich et al., 2005) necessary for sulfur oxidation. Bin 8 and Bin 13 are related to methylotrophic bacteria, however no methylotrophic genes such as methanol dehydrogenase were found in either of the bins, and the only C1 metabolizing gene that was found was formate dehydrogenase. Bin 7 appears to be a megaplasmid that is associated with Bin 8. The bin contains a single essential housekeeping gene, the large subunit ribosomal protein L31 which is missing in Bin 8. There are also numerous genes associated with plasmids, such as the plasmid partitioning gene *parA*, resolvase genes, and a copy of beta-lactamase. Many of the remaining, mostly low-coverage bins also appear to be heterotrophic bacteria.

Many of the Rubisco sequences found in the metagenome cluster with the “Form 1E” group, as named using the method of Tabita (1999). This group consists of non-proteobacterial phyla which are able to fix carbon using diverse metabolisms. Cultured

organisms capable of autotrophic growth found in this group include the *Actinobacteria Pseudonocardia dioxanivorans* which can grow autotrophically on molecular hydrogen (Grostern and Alvarez-Cohen, 2013), *Oscillochoris trichoides* is an anaerobic non-sulfur photoautotroph belonging to the *Chloroflexi* phylum (Ivanovsky et al., 1999), *Sulfobacillus acidophilus* and *Kyrpidia tusciae* are acidophillic *Firmicutes* capable of autotrophic growth using iron (Clark and Norris, 1996) and molecular hydrogen (Bonjour and Aragno, 1984), respectively. The group also contains the acidophilic *Verrucomicrobia* isolates that are capable of autotrophic methanotrophic growth (Dunfield et al., 2007). The bins that group within the Form IE Rubisco sequences only include the *Chloroflexi* and *Actinobacteria* bins. Two sequences from Bin 2 cluster near the firmicute sequences and Bin 18 clusters with *Pseudocardia sp*. P1, while the rest of the sequences form a monophyletic group that does not include any described isolates, however they cluster with uncultured sequences from arid and agricultural soils. This cluster appears to represent a large group of autotrophic soil bacteria that has not been previously cultured but may be environmentally relevant members of the carbon cycle in these environments.

All other Rubisco sequences clustered within the 'Form ID' clade, including a sequence from Bin 1. The sequence from this bin is the first Rubisco gene found in an *Acidobacteria* genome which are very common soil bacteria, but have only recently been cultivated (Ward et al., 2009). Other *Acidobacteria* genomes contain a form II CODH gene, but do not contain any known carbon fixation pathways which has led investigators to suggest that carbon monoxide oxidation is utilized as a mixotrophic metabolism utilized under carbon starvation, but this has never been experimentally shown (Ward et

al., 2009). Bin 1 does not include a CODH gene, but does encode a high affinity hydrogenase which may provide energy for autotrophic growth. Curiously, the Rubisco gene from Bin 1 clusters closely with the chemoautotrophic soil-associated *Alphaproteobacteria* which suggests lateral gene transfer has occurred between these two phyla as has been seen in other studies (Mazón et al., 2006).

The presence of multiple copies of the 'form I' CODH gene suggests that carbon monoxide oxidation is a common and important metabolism in this carbon-starved environment. Carbon monoxide has never been measured at elevated concentrations in Warren Cave using spot-measurements, but the elevated CO_2 emissions from the fumarole (Fischer et al., 2013) along with periodic pulses of volcanic activity of Mt. Erebus (Rowe et al., 1998) suggests that there may be ephemeral carbon monoxide pulses within the cave that can be utilized by these microbes. Sequences from Bins 2, 4, 5, and 9 cluster within the characterized 'form I' CODH clade with several known carboxydotrophs, such as *Oligotropha carboxidovorans*, which utilize carbon monoxide oxidation coupled with carbon fixation as their sole source of energy and carbon (King and Weber, 2007). Several bacteria termed carboxydovores have also been found with a form I CODH gene that can oxidize CO at atmospheric concentrations, but do not appear to be capable of autotrophic growth (King, 2003). It is currently not possible to differentiate between carboxydotrophs and carboxydovores using a purely genomic approach, however the presence of a complete Calvin Cycle in Bins 2, 4, and 9 suggest these bins are carboxydotrophs rather than carboxydovores.

Many other sequences cluster with the 'form II' clade of CODH genes, however the function of this CODH is still under debate (King and Weber, 2007). Lorite (2000)

showed that autotrophic growth on CO by Bradyrhizobium japonicum USDA 110 using only a 'form II' CODH gene was possible, although the strain grew at a much reduced rate than other carboxydovore bacteria. Another study showed that multiple strains of *Roseobacter sp*. with only a 'form II' CODH were unable to oxidize CO, while isolates with both both forms of the CODH large subunit were able to remove CO from a sealed culture tube (Cunliffe, 2011). It should be noted however that this study only used a single concentration of CO (1000 ppm) in the headspace that would not likely be found in the natural environment. Bins 2, 5, and 9 have both forms of CODH, while Bins 6, 11, 12, 15, 16, 19, and 20 contain only a 'form II' CODH gene.

The other lithotrophic metabolism found in the metagenome was the 'type 5' [NiFe]-hydrogenase genes. This newly-discovered group of hydrogenases have a high affinity for H_2 and are the only known hydrogenases capable of oxidizing H_2 down to atmospheric concentrations (~0.054 ppm) (Constant et al., 2010). The enzyme has only been characterized in a few isolates and has not yet been shown to be capable of fueling autotrophic growth, however inactivation of the gene in *Mycobacterium smegmatis* resulted in a growth defect that that resulted in a 40% decrease in growth yield which suggests the enzyme is providing the organism with energy for growth (Berney et al., 2014).

Survival using molecular hydrogen has been postulated before as a way for bacteria to gain enough energy while surviving under anabiosis to repair DNA and proteins while surviving over years or decades (Morita, 1999). It has also been theorized that organisms with both high affinity CODH and hydrogenase genes may provide enough energy to grow autotrophically with both active, rather than mixotrophically with

only one as has been shown in previous studies (Berney et al., 2014). Carbon monoxide and hydrogen oxidation genes are often found in the same organisms, and they may share an electron transport system that is specialized for these metabolisms (Kim and Hegeman, 1981). The level of activity that the microorganisms are respiring in the cave unfortunately remains unknown as the biomass found in the caves is too low to extract useful mRNA for transcriptomics using current methods. The metagenome contains bins, however, that appear to be adapted to the carbon starvation conditions found in the cave. Five of the six bins with the highest coverage (and therefore most abundant) all contain a complete Calvin Cycle and appear to be capable of autotrophic growth using carbon monoxide or molecular hydrogen as an energy source (Figure 5-11). While the low organic biomass and low abundance of heterotrophs in the community indicates a slow-growing community, biomass as assayed by QPCR (Tebo et al., 2014) is within an order of magnitude of deep-ocean basalts (Einen et al., 2008) and many active subsurface environments (Kallmeyer et al., 2012). This suggests that the community is actively growing, either steadily utilizing consistently low concentrations of CO and/or H_2 or in bursts mirroring the volcanic activity of Mt. Erebus which could increase both the concentrations of reduced gases as well as heat in the cave.

Genes for the ribosomal small subunit (16S) rRNA, CODH, and type 5 [NiFe]-hydrogenases that are closely related to these bins are often found in soil and rock environments; however, closely related isolates or enrichments from these environments have not been characterized or described. This metagenome provides insight into the metabolisms that are utilized by these microbes surviving in oxic oligotrophic soil or rock communities and may be common in desert sand, alpine tundra, and volcanic sediments.

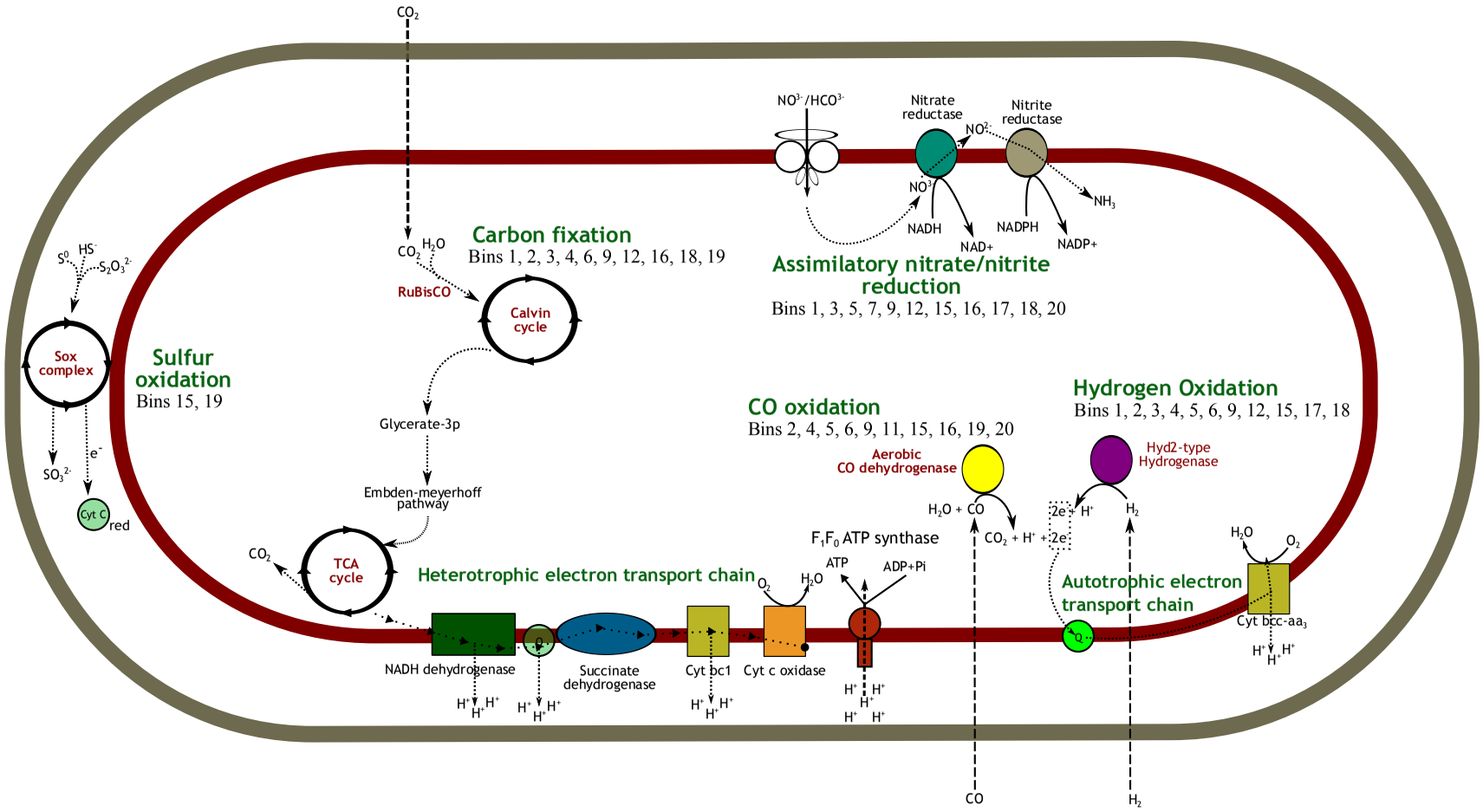

Figure 5-11. Conceptual pathway model of a composite Warren Cave microbe showing the major energy generating and carbon/nitrogen assimilating pathways.

These biomes cover significant portions of the Earth, yet their microbial communities are relatively unexplored, especially in the context of oligotrophic starvation and survival.

Works Cited

Alikhan, N.-F., Petty, N.K., Zakour, N.L.B., and Beatson, S.A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics *12*, 402.

Barnett, D.W., Garrison, E.K., Quinlan, A.R., Strömberg, M.P., and Marth, G.T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. Bioinformatics *27*, 1691–1692.

Belchik, S.M., Edwards, M.J., and Shi, L. (2012). Identification and characterization of MtoA: a decaheme c-type cytochrome of the neutrophilic Fe(II)-oxidizing bacterium Sideroxydans lithotrophicus ES-1. Front. Microbiol. Chem. *3*, 37.

Berney, M., Greening, C., Hards, K., Collins, D., and Cook, G.M. (2014). Three different [NiFe] hydrogenases confer metabolic flexibility in the obligate aerobe *Mycobacterium smegmatis*. Environ. Microbiol. *16*, 318–330.

Bonjour, F., and Aragno, M. (1984). *Bacillus tusciae*, a new species of thermoacidophilic, facultatively chemolithoautotrophic hydrogen oxidizing sporeformer from a geothermal area. Arch. Microbiol. *139*, 397–401.

Brown, C.T., Howe, A., Zhang, Q., Pyrkosz, A.B., and Brom, T.H. (2012). A reference-free algorithm for computational normalization of shotgun sequencing data. ArXiv Prepr. ArXiv12034802.

Cammarano, P., Palm, P., Creti, R., Ceccarelli, E., Sanangelantoni, A.M., and Tiboni, O. (1992). Early evolutionary relationships among known life forms inferred from elongation factor EF-2/EF-G sequences: Phylogenetic coherence and structure of the archaeal domain. J. Mol. Evol. *34*, 396–405.

Chang, Y., Land, M., Hauser, L., Chertkov, O., Del Rio, T.G., Nolan, M., Copeland, A., Tice, H., Cheng, J.-F., Lucas, S., et al. (2011). Non-contiguous finished genome sequence and contextual data of the filamentous soil bacterium *Ktedonobacter racemifer* type strain (SOSP1-21T). Stand. Genomic Sci. *5*, 97–111.

Chikhi, R., and Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. Bioinformatics *30*, 31–37.

Clark, D.A., and Norris, P.R. (1996). *Acidimicrobium ferrooxidans* gen. nov., sp. nov.: mixed-culture ferrous iron oxidation with Sulfobacillus species. Microbiology *142*, 785–790.

Constant, P., Chowdhury, S.P., Pratscher, J., and Conrad, R. (2010). *Streptomycetes* contributing to atmospheric molecular hydrogen soil uptake are widespread and encode a putative high-affinity [NiFe]-hydrogenase. Environ. Microbiol. *12*, 821–829.

Crusoe, M.R., Edvenson, G., Fish, J., Howe, A., McDonald, E., Nahum, J., Nanlohy, K., Pell, J., Simpson, J., Scott, C., et al. (2014). The khmer software package: enabling efficient sequence analysis. doi:10.6084/m9.figshare.979190

Cunliffe, M. (2011). Correlating carbon monoxide oxidation with cox genes in the abundant Marine Roseobacter Clade. ISME J. *5*, 685–691.

Curtis, A., and Kyle, P. (2011). Geothermal point sources identified in a fumarolic ice cave on Erebus volcano, Antarctica using fiber optic distributed temperature sensing. Geophys. Res. Lett. doi:/10.1029/2011GL048272

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Thomas, B.C., Yelton, A.P., and Banfield, J.F. (2009). Community-wide analysis of microbial genome sequence signatures. Genome Biol. doi:10.1186/gb-2009-10-8-r85.

Dunfield, K.E., and King, G.M. (2004). Molecular Analysis of Carbon Monoxide-Oxidizing Bacteria Associated with Recent Hawaiian Volcanic Deposits. Appl. Environ. Microbiol. *70*, 4242–4248.

Dunfield, P.F., Yuryev, A., Senin, P., Smirnova, A.V., Stott, M.B., Hou, S., Ly, B., Saw, J.H., Zhou, Z., Ren, Y., et al. (2007). Methane oxidation by an extremely acidophilic bacterium of the phylum *Verrucomicrobia*. Nature *450*, 879–882.

Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS Comput Biol. doi: 10.1371/journal.pcbi.1002195.

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

Einen, J., Thorseth, I.H., and Øvreås, L. (2008). Enumeration of Archaea and Bacteria in seafloor basalt using real-time quantitative PCR and fluorescence microscopy. FEMS Microbiol. Lett. *282*, 182–187.

Fischer, T.P., Curtis, A.G., Kyle, P.R., and Sano, Y. (2013). Gas discharges in fumarolic ice caves of Erebus volcano, Antarctica. AGU Fall Meet. Abstr. *-1*, 2706.

Friedrich, C.G., Bardischewsky, F., Rother, D., Quentmeier, A., and Fischer, J. (2005). Prokaryotic sulfur oxidation. Curr. Opin. Microbiol. *8*, 253–259.

Fry, J.C., Parkes, R.J., Cragg, B.A., Weightman, A.J., and Webster, G. (2008). Prokaryotic biodiversity and activity in the deep subseafloor biosphere. FEMS Microbiol. Ecol. *66*, 181–196.

Geszvain, K., McCarthy, J.K., and Tebo, B.M. (2013). Elimination of Manganese(II,III) Oxidation in *Pseudomonas putida* GB-1 by a Double Knockout of Two Putative Multicopper Oxidase Genes. Appl. Environ. Microbiol. *79*, 357–366.

Giggenbach, W. (1976). Geothermal ice caves on Mt Erebus, Ross Island, Antarctica. N. Z. J. Geol. Geophys. *19*, 365–372.

Grostern, A., and Alvarez-Cohen, L. (2013). RubisCO-based CO2 fixation and C1 metabolism in the actinobacterium *Pseudonocardia dioxanivorans* CB1190. Environ. Microbiol. *15*, 3040–3053.

Hall, T.A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser *41*, 95–98.

Hou, C.T., Laskin, A.I., and Patel, R.N. (1979). Growth and Polysaccharide Production by *Methylocystis parvus* OBBP on Methanol. Appl. Environ. Microbiol. *37*, 800–804.

Huson, D.H., Mitra, S., Ruscheweyh, H.-J., Weber, N., and Schuster, S.C. (2011). Integrative analysis of environmental sequences using MEGAN4. Genome Res. *21*, 1552–1560.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics *11*, 119.

Ivanovsky, R.N., Fal, Y.I., Berg, I.A., Ugolkova, N.V., Krasilnikova, E.N., Keppen, O.I., Zakharchuc, L.M., and Zyakun, A.M. (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.

Jones, R.T., Robeson, M.S., Lauber, C.L., Hamady, M., Knight, R., and Fierer, N. (2009). A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. ISME J. *3*, 442–453.

Kallmeyer, J., Pockalny, R., Adhikari, R.R., Smith, D.C., and D'Hondt, S. (2012). Global distribution of microbial abundance and biomass in subseafloor sediment. Proc. Natl. Acad. Sci. *109*, 16213–16216.

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27 –30.

Kim, Y.M., and Hegeman, G.D. (1981). Electron transport system of an aerobic carbon monoxide-oxidizing bacterium. J. Bacteriol. *148*, 991–994.

King, G.M. (2003). Uptake of Carbon Monoxide and Hydrogen at Environmentally Relevant Concentrations by *Mycobacteria*. Appl. Environ. Microbiol. *69*, 7266–7272.

King, G.M., and Weber, C.F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. Nat. Rev. Microbiol. *5*, 107–118.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lim, J.-M., Jeon, C.O., Lee, G.S., Park, D.-J., Kang, U.-G., Park, C.-Y., and Kim, C.-J. (2007). Leeia oryzae gen. nov., sp. nov., isolated from a rice field in Korea. Int. J. Syst. Evol. Microbiol. *57*, 1204–1208.

Lin, L.-H., Slater, G.F., Sherwood Lollar, B., Lacrampe-Couloume, G., and Onstott, T.C. (2005). The yield and isotopic composition of radiolytic H_2, a potential energy source for the deep subsurface biosphere. Geochim. Cosmochim. Acta *69*, 893–903.

Lloyd, A.T., and Sharp, P.M. (1993). Evolution of the *recA* gene and the molecular phylogeny of bacteria. J. Mol. Evol. *37*, 399–407.

Lorite, M.J., Tachil, J., Sanjuán, J., Meyer, O., and Bedmar, E.J. (2000). Carbon Monoxide Dehydrogenase Activity in *Bradyrhizobium japonicum*. Appl. Environ. Microbiol. *66*, 1871–1876.

Macalady, J.L., Jones, D.S., and Lyon, E.H. (2007). Extremely acidic, pendulous cave wall biofilms from the Frasassi cave system, Italy. Environ. Microbiol. *9*, 1402–1414.

Markowitz, V.M. (2006). The integrated microbial genomes (IMG) system. Nucleic Acids Res. *34*, D344–D348.

Mazón, G., Campoy, S., Erill, I., and Barbé, J. (2006). Identification of the *Acidobacterium capsulatum LexA* box reveals a lateral acquisition of the Alphaproteobacteria *lexA* gene. Microbiology *152*, 1109–1118.

McCollom, T.M., and Bach, W. (2009). Thermodynamic constraints on hydrogen generation during serpentinization of ultramafic rocks. Geochim. Cosmochim. Acta *73*, 856–875.

Monciardini, P., Cavaletti, L., Schumann, P., Rohde, M., and Donadio, S. (2003). *Conexibacter woesei* gen. nov., sp. nov., a novel representative of a deep evolutionary line of descent within the class *Actinobacteria*. Int. J. Syst. Evol. Microbiol. *53*, 569–576.

Morita, R.Y. (1999). Is H_2 the Universal Energy Source for Long-Term Survival? Microb. Ecol. *38*, 307–320.

Nealson, K.H., Inagaki, F., and Takai, K. (2005). Hydrogen-driven subsurface lithoautotrophic microbial ecosystems (SLiMEs): do they exist and why should we care? Trends Microbiol. *13*, 405–410.

Onstott, T.C., Moser, D.P., Pfiffner, S.M., Fredrickson, J.K., Brockman, F.J., Phelps, T.J., White, D.C., Peacock, A., Balkwill, D., Hoover, R., et al. (2003). Indigenous and contaminant microbes in ultradeep mines. Environ. Microbiol. *5*, 1168–1191.

Pati, A., LaButti, K., Pukall, R., Nolan, M., Rio, T.G.D., Tice, H., Cheng, J.-F., Lucas, S., Chen, F., Copeland, A., et al. (2010). Complete genome sequence of *Sphaerobacter thermophilus* type strain (S 6022^T). Stand. Genomic Sci. *2*, 49–56.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. *35*, 7188 –7196.

Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: Accurate High Throughput Multiple Sequence Alignment of Ribosomal RNA Genes. Bioinformatics *28*, 1823–1829.

Rowe, C.A., Aster, R.C., Kyle, P.R., Schlue, J.W., and Dibble, R.R. (1998). Broadband recording of Strombolian explosions and associated very-long-period seismic signals on Mount Erebus Volcano, Ross Island, Antarctica. Geophys. Res. Lett. *25*, 2297–2300.

Schippers, A., Neretin, L.N., Kallmeyer, J., Ferdelman, T.G., Cragg, B.A., John Parkes, R., and Jorgensen, B.B. (2005). Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. Nature *433*, 861–864.

Smith, D.C., Spivack, A.J., Fisk, M.R., Haveman, S.A., and Staudigel, H. (2000). Tracer-based estimates of drilling-induced microbial contamination of deep sea crust. Geomicrobiol. J. *17*, 207–219.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688 –2690.

Stevens, T.O., and McKinley, J.P. (1995). Lithoautotrophic Microbial Ecosystems in Deep Basalt Aquifers. Science *270*, 450–455.

Tabita, F.R. (1999). Microbial ribulose 1,5-bisphosphate carboxylase/oxygenase: A different perspective. Photosynth. Res. *60*, 1–28.

Takai, K., Moser, D.P., DeFlaun, M., Onstott, T.C., and Fredrickson, J.K. (2001). Archaeal Diversity in Waters from Deep South African Gold Mines. Appl. Environ. Microbiol. *67*, 5750–5760.

Takai, K., Gamo, T., Tsunogai, U., Nakayama, N., Hirayama, H., Nealson, K.H., and Horikoshi, K. (2004). Geochemical and microbiological evidence for a hydrogen-based, hyperthermophilic subsurface lithoautotrophic microbial ecosystem (HyperSLiME) beneath an active deep-sea hydrothermal field. Extremophiles *8*, 269–282.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. Mol. Biol. Evol. *24*, 1596–1599.

Tebo, B.M., Davis, R.E., Connell, L., Anitori, R., and Staudigel, H. (2014). Microbial communities in dark oligotrophic volcanic ice cave ecosystems (DOVEs) of Mt. Erebus, Antarctica. Prep.

Ultsch, A., and Mörchen, F. (2005). ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM (University of Marburg, Germany).

Ward, N.L., Challacombe, J.F., Janssen, P.H., Henrissat, B., Coutinho, P.M., Wu, M., Xie, G., Haft, D.H., Sait, M., Badger, J., et al. (2009). Three Genomes from the Phylum Acidobacteria Provide Insight into the Lifestyles of These Microorganisms in Soils. Appl. Environ. Microbiol. *75*, 2046–2056.

Weber, C.F., and King, G.M. (2010). Distribution and diversity of carbon monoxide-oxidizing bacteria and bulk bacterial communities across a succession gradient on a Hawaiian volcanic deposit. Environ. Microbiol. *12*, 1855–1867.

Whelan, S., and Goldman, N. (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. Mol. Biol. Evol. *18*, 691 –699.

Wu, D., Raymond, J., Wu, M., Chatterji, S., Ren, Q., Graham, J.E., Bryant, D.A., Robb, F., Colman, A., Tallon, L.J., et al. (2009). Complete Genome Sequence of the Aerobic CO-Oxidizing Thermophile *Thermomicrobium roseum*. PLoS ONE *4*, e4207.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. *18*, 821–829.

Chapter 6- General Considerations, Summary and Future Directions

General considerations about sample collection, processing, and inherent biases of methods used

This thesis has provided novel information about community dynamics and carbon fixation in dark oligotrophic volcanic ecosystems. These environments are understudied, yet important, ecosystems on Earth due to the role they play in geochemical cycles and carbon uptake on a global scale, and the numerous metabolic mechanisms and genetic diversity of these systems that are still poorly understood.

One of the reasons these systems have remained understudied is the inaccessibility to pristine samples coupled with the low biomass of the system compared to phototrophic systems. Samples with low biomass are at a greater risk of contamination from human, lab, and sampling contamination due to exposure to high biomass systems during collection. Deep-sea hydrothermal vent samples must be stored for hours after collection on an ROV and can be contaminated after sampling if they are not sealed from deep-seawater or high-biomass surface waters. Hydrothermal samples can be protected from post-sampling contamination by designing samplers which seal the samples from outside waters. Lightly sealed “bioboxes” are often used to store large samples such as rocks or animals and provide the least protection from contamination. These plastic boxes are generally sealed with bungee cords and some water exchange is generally accepted due to changing pressures as the ROV or submersible ascends. No samples used in this thesis were contained solely in a biobox, all hydrothermal vent microbial mat samples were collected with either a suction sampler or a PVC scoop sampler.

A microbial suction sampler is a device specifically designed to collect sediment

and microbial mats from the deep sea. These samplers generally consist of a hydraulic pump which generates a suction on a 3 inch hose. Material that is suction into the hose is collected in an in-line sampling "jar" where the heavy material, such as sediment and microbial mat, settles and light material smaller than 200 microns is ejected from a filtered outlet. Strengths of this device include ease of use by ROV/submersible pilots, being able to sample large areas rapidly, and minimal contamination issues during ascent. The sample is usually collected over the course of ten minutes, the jar is then sealed on a carousel and the sampler hoses are flushed into an empty jar. One disadvantage of this sampler is the disruptive action of the suction on the sampler. The structure of the microbial mat is generally destroyed during sample collection, preventing any fine dissection of the mat. The sampler also selects for microbial communities that are directly attached to the mat material and hundreds of liters of deep seawater is pumped through the sample. Any cells that are unattached or weakly attached to particles within the microbial mat are lost in the course of sampling. The suction sampler is also incapable of sampling small, targeted regions of microbial mat due to the volume of material that must pass through the hose for each sampling, and it also cannot sample microbial communities attached to solid surfaces such as rocks. The suction sampler is also not hermetically sealed against cross contamination and is difficult to clean and sanitize before use.

To alleviate many of these issues, PVC scoop samples were designed by our lab to sample microbial mats and deep-sea sediments. These samplers are not dependent on mechanical pumps which can fail in the field, and are easily cleaned, sanitized, and sealed prior to use. The scoops consist of a sampling chamber approximately 0.4 m in

length consisting of transparent 3 in PVC pipe which is attached to a ball valve (Figure 6-1). The scoop is cleaned and washed with deionized water before filling it with 70% ethanol and allowing it to soak for 10-15 minutes. The ethanol is then emptied from the scoop and is rinsed and filled completely with filter-sterilized ultrapure water before being sealed for use. The sampler is opened immediately prior to sampling, and is sealed directly after the sampling is completed. The sampler is therefor hermetically sealed during deployment and recovery of the ROV/submersible and the sample chamber is never exposed to shallow seawater. This sampler is effective at sampling material much more dense than seawater, such as sediments or metalliferous microbial mats, however it was not effective in collecting buoyant material or thin samples on a rocky surface. Samples that have been collected with these scoops appear to be pristine, with no obvious contamination from surface seawater phototrophs or human-associated microbes.

A second source of contamination of these samples can occur during sample processing and in the lab. The sample from Warren Cave was collected using aseptic techniques and a sterile 50 ml conical tube. This sample was then partitioned in a laminar flow hood using sterile instruments while maintaining aseptic technique to prevent contamination. Deep-sea hydrothermal samples collected with a suction sampler were allowed to settle for 2-3 hours in a cold room, and were partitioned into sterile 50 ml conical tubes using a turkey baster that had been washed in 70% ethanol. Samples collected using the scoop sample were allowed to settle for a maximum of 1 hour before being partitioned into sterile 50 ml conical tubes using either a sterile disposable 50 ml pipette or a stainless steel spoon that has been washed with 70% ethanol. All samples were immediately frozen at -80°C after being partitioned into conical tubes, transported

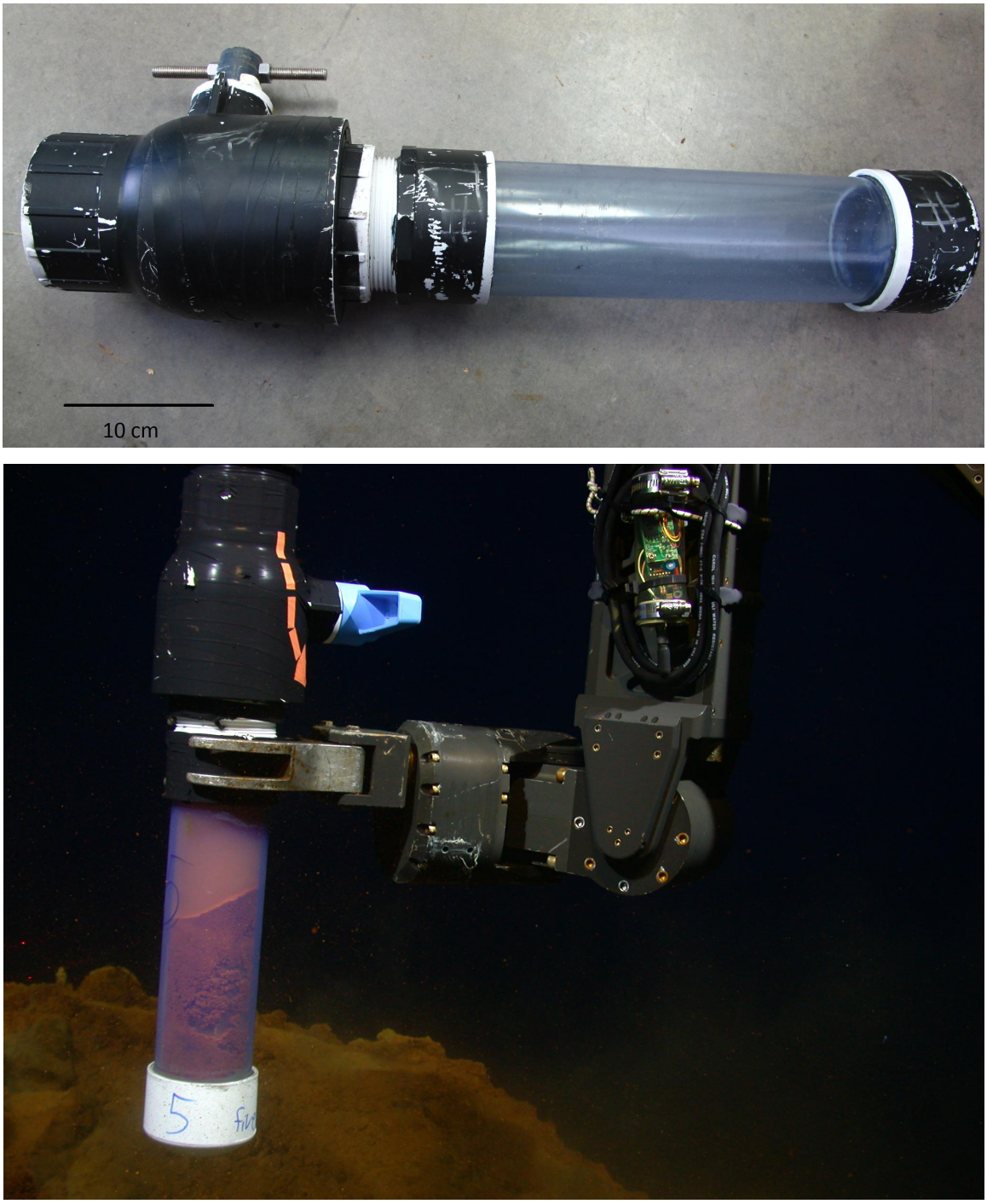

Figure 6-1. PVC Scoop used to sample microbial mat communities. The scoop (top) is constructed using transparent PVC for the sampling body and a 3 inch PVC ball valve to seal the sample. Microbial mat material is collected using a scooping motion and the ball valve is immediately closed after sampling. Each scoop sampler can collect approximately 0.75 L of microbial mat material after settling (bottom).

frozen to the lab, and stored at -80°C until extraction of nucleic acids was performed.

All lab reagents were either purchased as sterile solutions or were filter sterilized using 0.2 μm polycarbonate syringe filters, and only aerosol-resistant pipette tips were used for DNA extraction and PCR amplification. All DNA extractions were also performed with a control tube containing only sterile glass beads. No PCR amplification was seen in any of these negative controls. Samples were also routinely assayed for complete cell lysis after bead beating by staining the pulverized material with the DNA stain SYTO-13 and visually examining the sample for intact cells using an epifluorescent microscope.

The samples collected for this thesis did not show any obvious contamination from humans or other sources, however other experimental biases may affect the results. Biases in PCR have been well documented and every effort was made to limit these biases. The concentration of all genomic DNA was measured and each PCR performed contained 1 ng DNA / μL of PCR mixture to normalize the exponential phase of the PCR between samples. PCR primers were never the theoretical limiting reagent in the PCR reactions to prevent amplification biases due to saturation of one degenerate primers causing biases of the community. All PCR primers designed for this thesis were limited to priming sites which allow inclusive amplification of a group, while minimizing degenerate nucleotides and sequences forming primer dimers. All QPCR reactions were optimized using a temperature gradient for the annealing step of the reaction, and all reactions were subjected to a melt curve and visualized on an agarose gel to verify that only a single amplicon was the product of the reaction.

Biases associated with metagenomic libraries are less understood than those of

PCR reactions. Possible biases may include prejudice in favor of some phylotypes during library construction, especially using the Nextera kit. This kit uses transposons to fragment and tag the genomic DNA for library preparation. This kit may have GC biases for DNA fragmentation and a PCR step during library construction often generates many duplicate sequences. Assembly of metagenomes may also have many unknown biases. Different assembly programs and algorithms often produce radically different final assemblies. This bias was minimized in this thesis by using the most commonly used de novo assembler Velvet and merging several assemblies to form a final assembly. The robust results from each of the chapters would not be possible with the careful sample collection, handling, and preservation of each sample used in this thesis.

Concluding remarks

Chapter 2 shows the transition of a hydrothermal vent system after an eruptive event. The pre-eruption vent was dominated by Zetaproteobacteria and smaller populations of Epsilonproteobacteria. The community rapidly changed to a community based on volatile gases such as hydrogen and sulfide after the eruption. As the heat flux at the vents decreased, the volatile gases also decreased while the iron concentration increased. The microbial communities were then dominated by Zetaproteobacteria, which are currently an indicator group for iron oxidation. This transition from a community dominated by autotrophic hydrogen- and sulfur-oxidizing microbes to a community dominated by iron-oxidizing bacteria closely mirrors that of the Butterfield Model of hydrothermal vent evolution which shows vent fluids transitioning from a “vapor” phase dominated by volatile gases to a “brine” phase dominated by increased dissolved iron and

chloride in the vent fluids.

Future efforts must be made to determine if this model is also seen at other erupting systems. This study represents the longest temporal community succession time series, only a single other study has shown a community transition at an actively erupting volcano. Microbial mat communities at Iceberg Vent Field on NW Rota-1 transitioned from a hydrogen-oxidizing community to a sulfide-oxidizing community between when the eruption was discovered in 2004 to when it was next visited in 2006 (Davis and Moyer, 2008). Future visits to this site and to other actively erupting deep-sea volcanoes, such as West Mata, will provide insight into community succession at these ephemeral systems as well as provide a time line for community succession after an eruptive event.

Chapter 3 shows the pathways of all currently known carbon fixation pathways as well as the diversity of the microbes that utilize each pathway. As more isolates are sequenced, new carbon fixation pathways and alternative routes in existing pathways have been discovered. Additionally, novel autotrophic bacteria are isolated and sequenced at a highly accelerated rate potentially changing the phylogenetic distribution used for each pathway and even changing the carbon budgeting on Earth.

Future work following on this chapter should include maintaining a database of sequenced autotrophic isolates, including the carbon fixation pathway they utilize. While the chapter only utilizes cultured organisms, the phylogenetic information of these isolates may aid in characterizing or culturing environmental sequences. Incorporation of uncultured sequences from PCR studies, single-cell amplification genomes, and metagenomes may provide new insight into the diversity of microbes that utilize each carbon fixation pathway as well as provide information of what carbon fixation pathways

are used in different environments.

Chapter 4 presents a comprehensive survey of carbon fixation pathways used at Loihi Seamount, HI. This study showed that the dominant carbon fixation pathway used at Loihi was the Calvin cycle, with most of the organisms using form II Rubisco enzymes to fix carbon. This form of Rubisco has low affinity for CO_2, and is therefore only used under anaerobic or microaerophilic conditions. This information coupled with the presence of dissimilatory nitrate reductase genes in all of the Zetaproteobacteria genome groups of the metagenome strongly suggests that the Zetaproteobacteria living at these vents oxidize iron anaerobically coupled with nitrate reduction. The presence of large populations of methanogenic archaea and sulfur- or iron-reducing Deltaproteobacteria strongly suggests that hydrogen oxidation is also important in this vent community. The hydrogen may be provided to these organisms in a syntrophic relationship between the iron-oxidizing bacteria, which produce large quantities of hydrogen ions that can be reduced to form molecular hydrogen.

Future work should focus on confirming the syntrophic relationship between iron-oxidizing bacteria and hydrogen-oxidizing bacteria and archaea. Large populations of archaea have not been found in other microbial mats at Loihi, however this may be due to sampling biases associated with a suction sampler. If the hydrogen oxidizing bacteria and archaea are not physically attached to the iron oxide material, they would be flushed out of the suction sampling device and would not be seen in future analysis. It is possible that the scoop sampler used to collect the metagenome sample preserved a more accurate representation of the microbial community at the vents by preserving the populations that are not strongly attached to particles in the mat. Quantitative PCR primers may be

designed for either the SSU rRNA genes for these groups or for functional genes found in the metagenome. Samples can then be screened for hydrogen-oxidizing communities in a high-throughput manner using QPCR to find the ratio of methanogenic archaea to zetaprotobacteria in samples collected using scoops and suction samplers.

Another future direction for the research presented in this chapter is to determine where the hydrogen is being formed. The hydrogen could simply be a component of the vent fluid at Pohaku Vent Field, however the vent has never been measured at a temperature over 27°C and does not appear to contain any other volatile gases. Sensitive hydrogen measurements in freshly sampled vent fluids collected with a titanium gas-tight sampler would be needed to answer this. A second possibility is the hydrogen is being produced biologically via hydrogenase enzymes operating in the reductive direction to maintain a redox balance in the cell or to maintain a circumneutral pH of the microbial mat. This mechanism of hydrogen production presents the exciting possibility of syntrophic chemoautotrophic feeding between the Zetaproteobacteria population and the hydrogen-oxidizing populations of the microbial mat communities. This cross-feeding could stabilize the microbial mat by maintaining the pH of the system while eliminating one of the waste products of iron oxidation. The Zetaproteobacteria genome bins contain multiple potential hydrogenase genes. Future work could focus on which of these genes is actively expressed in the microbial mat, and whether it could run in the reverse direction reducing hydrogen ions and producing hydrogen gas. Protein extractions from natural samples could also be assayed for hydrogen production and pH change when an excess of hydrogen ions is added to the solution.

Chapter 5 examines a metagenome of Warren Cave, a pristine, dark oligotrophic

ice cave that was formed by fumarolic gases. This microbial community appears to be supported by primary production using the Calvin Cycle fueled by the aerobic oxidation of hydrogen and carbon monoxide, primarily by Chloroflexi, Acidobacteria, and Actinobacteria. The microbes in this cave appear to grow slowly based on the low numbers of heterotrophic bacteria in the metagenome. Nitrogen appears to be assimilated from nitrate or nitrite, possibly from rock weathering. These communities may be similar to communities found in oligotrophic soil communities, such as desert and alpine soils.

Future studies of microbial communities in Warren Cave should focus on the temporal variability of the microbial communities as well as the chemical components of the fumarolic gases. Accurate measurements of carbon monoxide and hydrogen gas in the fumaroles over the course of weeks or months would provide information about the content and temporal variability of these gases in the caves. This would be vital information for determining the microbial growth potential inside the cave community. Utilizing the metagenome for targeted culturing could also be a future direction. Utilizing nitrate for a nitrogen source and trace concentrations of hydrogen and carbon monoxide may allow previously uncultured organisms from Warren Cave to be cultured.

Future studies of Antarctic fumarolic ice caves should focus on other cave habitats, and contrasting those ecosystems with Warren Cave. For instance, shallow caves with light filtering through the roof of the cave should have different energy requirements than dark caves. These caves could utilize photosynthesis to generate energy, and therefore those communities should be limited by another resource. Understanding these resource limitations as well as the biogeography of the caves would give critical new understanding of microbial communities in these oligotrophic environments. The results

of this research will also increase our understanding of the food web dynamics, biogeography, and colonization of microbial communities which are exploiting similar niches, but have no known mechanism for gene flow between communities. Ultimately, understanding the mechanisms of how these organisms interact and survive in this extremely oligotrophic environment can be applied to other oligotrophic environments not only in Antarctica, but across oligotrophic volcanic systems worldwide.

In summary, the research presented in this thesis has helped unravel the metabolic and phylogenetic diversity of microbial communities largely supported by chemoautotrophic primary producers. We have only begun to understand the community dynamics, mechanisms for primary productivity, and the metabolic strategies of these carbon fixing populations. While open questions remain as to the ubiquity of the metabolisms seen in the metagenomes and the temporal scale of community succession at other hydrothermal sites, these dynamic systems are truly important ecosystems on Earth and these findings illustrate the need for future studies.

Work Cited

Davis, R.E., and Moyer, C.L. (2008). Extreme spatial and temporal variability of microbial mat communities along the Mariana Island Arc/Backarc system. J Geophys Res *113*, B08S15.