Metagenomic investigation of microbial dark carbon *fixation*, viral interactions, and horizonal gene transfer within a convergent margin subsurface ecosystem

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Timothy Joseph Rogers

August 2023

Copyright © 2023 by Timothy Joseph Rogers

All rights reserved.

Dedication

To my loving parents, Jan and Tim Rogers; my beloved wife and best friend, Soriya Cooper; and to my elementary and high school science teacher, Cindy Mosses (Ms. Studdard)

Acknowledgements

First and foremost, I want to thank my best friend and wife, Soriya Cooper. You have been by my side ever since I began this race of chasing my dream of becoming a scientist. Without you staying up with me until the wee hours of the morning while I worked on procrastinated homework, your unrelenting support when I left for long field expedition even while you wanted nothing more than to have me home with you, and your unyielding strength to lifting me up and cheering me on when the weight and pressure of this project felt like it was too much, I could not have grown from the kid I was into the 'Billy goat' I am today. Thank you for being my rock and helping me weather this storm. I love you more than you will ever know.

My Mother and Father, Jan and Tim Rogers. Mom, without you paying attention to the way I learn, putting me in speech therapy to help my sentence structure, spending hours on end teaching me to read with *Hooked on Phonics*, and forcing my elementary school to hold me back in 2nd grade, I would have let my ADHD and learning disability discourage me from ever thinking I could pursue a PhD. Dad, without you constantly encouraging me to questioning everything, even the unquestionable no matter how uncomfortable it made me, I would currently be leading a flock instead of standing on the shoulders of the giants I stand on today. Thank you and I love you both for everything you have done to help sculpt me into who I am today.

My 2nd mom, 3rd grade, and high school science teacher, Cindy Moses (Mrs. Studdard). You are the quintessential example of how a teacher can truly teach a 'class clown' to love to learn. Though you knew my world view, at the time, prevented me from accepting established science, you were able to show me how these two perspectives could complement each other, enabling me to feel more comfortable engaging with the science. Later in life when I began to question that world view, you were there to encourage and help me process my doubts. And when I was considering leaving the military to pursue a BS in EEB, and later this PhD, you were one of the fist cheerleaders excitedly cheering me on. I thank you for seeing something in me that other teachers could not or refused to see. I love you for always being there and never giving up on this problematic child.

My first scientific mentor, Melissa Cregger. Coming from a similar background as myself, you were not only able to connect with me on an intellectual level, but on a cultural level as well. Your willingness to share openly about your life experiences, the obstacles you had to overcome, and your background growing up in a similar context as I, gave me the confidence that I needed to know that I belong and am welcome in the scientific world. Thank you for everything and especially for taking a chance on me by allowing me to write my first paper as an undergraduate.

My primary advisor, Karen Lloyd. If it was not for you taking me under your wing and making me feel at home in your lab, I don't know that I could have completed this program. Your open-door policy, hands-off approach, willingness to listen, and ability to find time in your schedule for me has allowed me to grow and freely explore avenues of this project that, at first glance, seemed like rabbit holes. If not for you, I would not have gotten to visit all the amazing places, meet all the amazing people, and eat all the amazing and strange food that I have over the last four and a half years. Your approach to science and leadership is awe inspiring. Thank you for the funding that has allowed me to focus on my research, for allowing me the intellectual freedom to sculpt this project into my own, for always being willing to listen when the imposter syndrome kicked in, and for your patience with, and ability to translate, my jumbled thoughts and ideas.

And finally, to all my collaborators in the field, my committee members, and lab mates. I could dedicate a page to each of you, but this is already becoming a chapter in and of itself. Thank y'all for guiding me through this whole process, for teaching me about your expertise, and for sharing your time with me. This has been one of the best experiences of my life and having y'all, and those written above, along for this journey has been a pleasure.

Abstract

Convergent margins are geological regions where two or more tectonic plates collide, and the denser "subducting slab" is pushed beneath the less dense overriding plate. As the slab descends, it devolatilizes under higher temperatures and pressures, allowing dissolved inorganic carbon (DIC) and redox active volatile rich fluids to cycle between the upper crust and Earth's mantle. These fluids migrate through cracks and fissures in the upper mantle and crust, fueling chemolithoautotrophy-based microbial ecosystems in the subsurface before they are expelled on the surface in the form of hydrothermal seeps and springs. Chemolithoautotrophic ecosystems, such as those in the Costa Rican convergent margin, generate substantial amounts of organic matter with up to 24 mM dissolved organic carbon (DOC), and may sequester 1.4×10^9 to 1.4×10^{10} molC/year. However, the rate at which DIC is biologically sequestered is dependent on the geochemistry of the system, the chemolithoautotrophic species present therein, the carbon fixation pathways (CFPs) they possess, as well as potential viral-host integrations and the effects of horizontal gene transfer (HGT).

In this work, we first review the current knowledge on the geochemical constraints on and the evolutionary adaptations of chemolithoautotrophs and the CFPs they contain within CO_2-rich and CO_2 limited terrestrial subsurface ecosystems. Next, using approaches that combine co-located metagenome sequences and geochemical data, we describe the distribution of chemolithoautotrophs across the Costa Rican convergent margin. Then, we use metagenomic data from the hot springs of Costa Rica and Panama to describe the viral-host interactions within the Central American subduction zone as well as the distribution of viral auxiliary metabolic genes. Finally, we describe a case of HGT in which it appears that uncultured members of the bacteria class *Syntrophales* have acquired a key gene of the reductive citric acid cycle, a micro-aerophilic CFP, from the archaeal clade AMNE-I. Together these studies provide a more nuanced view of the factors that are important to prokaryotic and viral communities fed by deeply-sourced volatiles across a convergent plate margin.

Table of Contents

CHAPTER 1 INTRODUCTION EVOLUTIONARY AND ENERGY CONSTRAINTS ON BIOLOGICAL CARBON FIXATION IN DEEP SUBSURFACE ECOSYSTEMS 1

INTRODUCTION 2
Energy efficiency limiting steps 3
Distribution of CFPs in terrestrial subsurface systems 6
CO_2-depleted Systems 10
CONCLUSION 17
A fish out of water: Terrestrial distribution of the marine-like Mariprofundus 17
Carbon fixation pathway distribution differences 18
REFERENCES 21

CHAPTER 2 CHEMOLITHOAUTOTROPH DISTRIBUTIONS ACROSS THE SUBSURFACE OF A CONVERGENT MARGIN 29

ABSTRACT 30
INTRODUCTION 31
METHODS 33
Sample collection, DNA extraction, and sequencing. 33
Metagenomic assembly, binning, annotation, and metabolic predictions. 33
MAG phylogenetic and statistical analysis. 33
RESULTS 34
General description of the MAGs. 34
Cluster and multivariate analysis. 35
Metabolic and taxonomic descriptions of the clusters. 36
Transformation-based canonical correspondence analysis. 39
DISCUSSION 39
Distribution of carbon fixation pathways across geological provinces. 41
Metabolic potential in the outer forearc. 42
Metabolic potential in the forearc and arc. 43
Delocalized clusters are surface-associated. 44
CONCLUSION 45
ACKNOWLEDGEMENTS 45
DATA AVAILABILITY 46
REFERENCES 47
APPENDIX I: TABLES AND FIGURES 56
SUPPLEMENTAL MATERIAL 63
Sample collection, DNA extraction, and sequencing. 63
Metagenomic assembly and binning. 63
MAG statistical analysis 64
Redox metabolic predictions 64
APPENDIX I: TABLES AND FIGURES 68

CHAPTER 3 VIRAL METABOLIC AND TAXONOMIC LANDSCAPE OF THE TERRESTRIAL SUBSURFACE ECOSYSTEM WITHIN THE CENTRAL AMERICAN CONVERGENT MARGIN 69

ABSTRACT 70
INTRODUCTION 71
METHODS 73
Sample collection, DNA extraction, and sequencing 73

Generation and metabolic predictions of metagenomic assembled genomes 73
Viral contig and auxiliary metabolic genes identification 74
Virus-host prediction 75
Removal of viral contamination 76
Viral taxonomic identification 76
Abundance profiles 76
Statistical analysis 76
RESULTS 77
Overview of microbial communities 77
Viral Community Composition 78
Viral-Host Matching 79
Viral Diversity 80
Viral AMGs 81
DISCUSSION 82
Drivers of viral diversity 83
Diversity of AMGs 85
CONCLUSION 86
REFERENCES 88
APPENDIX I: TABLES AND FIGURES 96
Supplemental figures and tables: 107

CHAPTER 4 EVIDENCE OF HORIZONTAL GENE TRANSFER OF ATP-CITRATE LYASE FROM METHANOGENS TO SYNTROPHALES IN HOT SPRING FLUIDS OF THE CENTRAL AMERICAN CONVERGENT MARGIN 112

ABSTRACT 113
INTRODUCTION 114
METHODS 115
RESULTS AND DISCUSSION 116
CONCLUSION 117
REFERENCES 118
APPENDIX I: FIGURES 121

CHAPTER 5 CONCLUSION 125

REFERENCES 129

VITA 130

List of Tables

Table S2.4. Marginal effects of terms in the tb-CCA model. 68

Table 3.1. Viral PERMANOVA results..... 104

Table 3.2. Microbial variance partitioning results 105

Table 3.3. Microbial PERMANOVA results 106

Table 3.4. Table S3.1. Top 5 most abundant phyla within each geophysical province. 107

List of Figures

Figure 2.1. General convergent margin geological structure (A) and sample sites spanning the Costa Rican convergent margin (B). 56

Figure 2.2. Heatmap of metagenomic read recruitment to each MAG and hierarchical clustering based on Spearman rank correlations between MAGs (vertical clusters) and sample sites (horizontal clusters) shows separation of MAGs into those found mostly in the forearc/arc, the outer forearc, or delocalized across geological provinces. 57

Figure 2.3. Average read abundance (GCPM) per site of MAGs with metabolic pathways for redox reactions and carbon fixation pathways. 58

Figure 2.4. Nitrogen, sulfur, and hydrogen cycling pathways in MAGs with carbon fixation pathways from the outer forearc and the forearc/arc 59

Figure 2.5. Transformation based Canonical Correspondence Analysis showing variables most closely correlated with MAG abundance distribution at each site and site distribution of each MAG (scaling = 2). 60

Figure 2.6. Schematic cross-section (to scale) of the 3D distribution of subsurface microbial communities and their metabolisms across the Costa Rica convergent margin, showing a shallowing of potential habitable area from the outer forearc to the arc. 62

Figure 3.1. Differences in prokaryotic community composition between groups. 96

Figure 3.2. Log_{10} relative abundance of vOTUs across all provinces. 97

Figure 3.3. Log_{10} relative abundance of vOTUs colored by their linked host trophic level. 98

Figure 3.4. Comparison of viral community α-diversity between geological provinces and between sample types. 99

Figure 3.5. Viral community β-diversity analysis 100

Figure 3.6. Microbial community β-diversity analysis. 101

Figure 3.7. Microbial vs viral community dissimilarities. 102

Figure 3.8. Microbial host vs linked viral dissimilarities 103

Figure S3.1. Comparison between mapped read counts and standardized mapped read counts of vOTUs. 108

Figure S3.2. Pearson correlations between the richness of both the viral and prokaryotic communities. 109

Figure S3.3. Pearson correlations between the richness of both the linked viruses and their hosts. 110

Figure S3.4. The results from the model-selection analysis for viral β-diversity with geochemistry and prokaryotic site dissimilarly. 111

Figure 4.1. Representation of enzymatic genes identified within the four Syntrophales MAGs. 121

Figure 4.2. RAxML phylogenetic tree placement of MAGs bin 162, bin 202, bin 51, and bin 351 based on conserved marker genes. 122

Figure 4.3. RAxML phylogenetic tree of ATP-citrate lyase amino acid sequences (concatenated α and β subunits). 123

Figure 4.4. RAxML phylogenetic tree placement of MAGs bin 162, bin 202, bin 51, and bin 351 based on conserved marker genes the most parsimonies scenario for the horizontal acquisition of ATP-citrate lyase by Syntrophales. 124

Chapter 1 Introduction: Evolutionary and energetic constraints on biological carbon fixation in deep subsurface ecosystems

A version of this original review is being prepared for future publication submission with the following authors:

Timothy J. Rogers and Karen G. Lloyd

My primary contributions to this paper include: (i) Literature review, (ii) synthesizing the results, and (iii) writing the manuscript.

Introduction

The biological process of carbon sequestration through carbon fixation or autotrophy is one of the most powerful tools available to mitigate damage from climate change (1–4). Autotrophy, also referred to as primary production, can be separated into two broad categories: photoautotrophy - using sunlight to power biosynthesis, and chemoautotrophy - using chemical redox reactions to power biosynthesis. While plants and microbes are capable of photoautotrophy, only microbes are capable of chemoautotrophy. Although most photoautotrophs use the Calvin-Benson-Bassham (CBB) for carbon fixation (5), chemolithoautotrophs can use a variety of carbon fixation pathways in addition to CBB. Some estimates of global primary production suggest that photoautotrophy contributes orders of magnitude more biomass than chemoautotrophy (6). However, the estimates for chemoautotrophy in these studies are primarily focused on ocean systems (0.27 Pg C y^{-1} in the euphotic zone, 0.11 Pg C y^{-1} dark oceans, and 0.37 Pg C y^{-1} in ocean sediments) where only an estimated 0.4% of this primary production is powered by chemical energy derived from water-rock reactions (6). This means that most chemoautotrophy in the ocean is powered by reduced compounds such as ammonium that are byproducts of heterotrophy of photosynthetic biomass, meaning that most deep-sea chemoautotrophy recycles biomass originally produced by photoautotrophy (6).

The situation is different in the terrestrial subsurface where the total biomass is an estimated 14 to 135 Pg C (7–9). The absence of sunlight and photosynthetically derived organic matter in the terrestrial subsurface means that these communities are often reliant on endogenous sources of energy (derived from within Earth; i.e. water-rock interactions) (7,10,11). Here we review the distribution of carbon fixation

pathways (CFPs) across phyla and ecosystems in the terrestrial subsurface where chemoautotrophy produces net biomass from inorganic carbon. Much of the work on chemolithoautotrophic-based ecosystems is performed with culture-independent methods, such as 16S rRNA gene amplicon sequencing and -omic analysis (culture-independent genomic, proteomic, and metabolic analysis of microbial communities). This review is focused on chemolithoautotrophs; their CFPs; and the geological, geochemical, and evolutionary constraints on their distribution across terrestrial subsurface environments. We first examine the energy-limiting factors of the six known CFPs. Then, we discuss the distribution patterns of chemolithoautotrophs and their CFPs across CO_2-saturated and CO_2-limited terrestrial subsurface environments.

Energy efficiency limiting steps

Five other CFPs in addition to the CBB-cycle are currently known: reductive tricarboxylic acid (rTCA) cycle, reductive acetyl-CoA (Wood-Ljungdahl (WL)) pathway, 3-hydroxypropionate (3-HP) bicycle, 3-hydroxypropionate/4- hydroxybutyrate (3-HP/4-HB) cycle and dicarboxylate/4-hydroxybutyrate (DC/4-HB) cycle. These six pathways can be divided into two groups: the anaerobic associated (WL, rTCA, and the DC/4-HB) pathways and the aerobic/microaerophilic associated (CBB cycle, 3HP bicycle, and the 3HP/4HB cycle) pathways (12), though a few exceptions exist for each group. The distribution of these pathways has been suggested to be determined by the energetic feasibility of the environment because CFPs are dependent on different numbers of ATP molecules for the reduction of CO_2, the reduction potential of enzymes used to reduce CO_2, and the intercellular CO_2 concentration (13). The concentration of CO_2 in a system can have profound effects on the efficiency and evolution of the enzymes involved in carbon fixation. In fact, in basic, highly reduced environments, low CO_2 levels may be more limiting than the limited availability of electron acceptors (14,15). As such, the level of CO_2 can have a selective effect on the carboxylating enzymes found within these systems (4,16,17).

The aerobic pathways are more energy demanding than the others due to their higher ATP requirement for the CO_2 assimilation steps (12,18,19). This is due, in part, to whether carbon is fixed via a carboxylation and/or carboxyl reduction or by CO_2 reduction. Carboxylation and carboxyl reduction are

thermodynamically unfavorable process for all CFPs (20) requiring nearly all such reactions to be coupled directly or indirectly to ATP hydrolysis, resulting in more reducing power being diverted toward oxidative phosphorylation and away from carbon fixation (21). This means that in environments under more reduced conditions (e.g., terrestrial serpentinization sites) ATP demanding pathways (such as the CBB-pathway's requirement of 7 ATP (19)) may be at a disadvantage due to a limitation of highly oxidized electron acceptors.

The anaerobic associated pathways rTCA and DC/4-HB cycles lower their ATP requirement (2 ATP for rTCA and 5 for DC/4-HB (19)) by coupling their carboxylation steps to the reduction of their products via NAD(P)H and ferredoxin. However, this requires that anaerobic associated pathways have a lower (more negative) reduction potential (ferredoxin or NAD(P)H equivalent) to keep their oxidoreductase reactions flowing in the direction of carbon-fixation (13), more so than the aerobic associated pathways. For example, under moderate CO_2 concentrations, rTCA must maintain a reduction potential below -300 mV for carbon fixation. Even so, the carboxylating enzymes 2-oxoglutarate synthase (rTCA) and isocitrate dehydrogenase (rTCA and DC/4-HB) preferentially work in the oxidative direction (18), resulting in a total thermodynamic barrier of +42 kJ/mol for rTCA (13), as opposed to Ribulose-1,5-bisphosphate carboxylase/oxygenase's (RuBisCO) -30 kJ/mol for that step of CBB. Some prokaryotes using rTCA can keep their product-to-substrate ratio (i.e. acetyl-CoA or pyruvate relative to CO_2) low in environments with moderate to high CO_2 concentrations, enabling them to lower the Gibbs energy required for these steps (13). However, under low CO_2 conditions, these enzymes tend to be expressed at higher levels to make up for their slow kinetics (20). In fact, within low CO_2 concentrations, such as those of ultrabasic serpentinizing systems (22), the kinetics of these enzymes are impractically slow as the need to produce more of these enzymes quickly becomes a major carbon sink within these cells, diverting resources away from cell growth (13,21). This limitation is even harsher within the WL-pathway. Because of its requirement for < 1 ATP for pyruvate formation, the thermodynamic barrier of +32 kJ/mol of CO dehydrogenase, and the additional +18 kJ/mol of formate dehydrogenase used by methanogens, the WL-pathway requires the lowest reduction potential of all the known carbon fixation pathways,

potentially limiting its use to highly reduced systems with high concentrations of inorganic carbon (20,23). Even so, prokaryotes using the WL-cycle are abundant within several serpentinizing systems (24,25), as will be discussed below.

Carbon-concentrating mechanisms (CCMs) provide a way of lowering these energy constraints in systems with low-CO_2. Such a mechanism may exist within the acetogen *Thermoanaerobacter kivui* in the form of bundles of hydrogen-dependent CO_2 reductase nanowires that line the interior membrane (25). However, a type of CCMs, known as carboxysomes, have been studied extensively within CCB-containing chemolithoautotrophs (13,26). RuBisCO is inefficient as a carboxylase, as oxygen can easily take the place of CO_2 within its active site and create the toxin 2-Phosphoglycolate, wasting significant amounts of carbon in the process (13,18,27). To deal with this, various forms of RuBisCO have arisen over evolutionary time. The carboxylation forms of RuBisCO are type I, type II, and type III. Type I and II are normally associated with bacterial CBB and type III are in other metabolic processes within archaea (18,26). Type II functions best under high CO_2/low O_2 concentrations while type I is better at lower CO_2/higher O_2 concentrations (18,26). However, while type I has a higher specificity factor for CO_2 over O_2 ($S_{c/o}$) it has a much slower carboxylase reaction. This means that under low CO_2 concentrations significantly more ATP and NAD(P)H are used due to an increase in phosphoglycolate salvage (term for photorespiration proposed by Claassens et al (28)). To avoid this issue, RuBisCO can be stored in carboxysomes where CO_2 can be concentrated significantly increasing carboxylation reactions and decreasing photorespiration (26). Additionally, because the ratio of CO_2/O_2 is high within the carboxysomes, type I RuBisCOs within have been freed to decrease their $S_{c/o}$ allowing for a much higher product release rate constant for carboxylation. This has given rise to carboxysome associated variants of RuBisCO type I: type IAc found in proteobacteria and α-cyanobacteria, and IBc found in β-cyanobacteria (26).

Distribution of CFPs in terrestrial subsurface systems.

CO_2-rich Systems

Few studies have been conducted on the microbial ecology of natural CO_2-rich terrestrial systems (3,4,16,29,30) (defined here as ground water with extensive active degassing of CO_2 dominated gases), known as the "Deep Carbonated Biosphere" (30). Those that have been conducted have mainly focused on sites within the western USA (24,31,32) where elevated levels of CO_2 have led to surface expressed rich carbonic springs or long-term CO_2 storage (400 kya to 1 mya) in deep subsurface aquifers (3,4,16,29,30). These sites include Crystal Geyser in Utah (3,4,16,17,29), Tierra Amarilla (TA) anticline spring group in New Mexico (33,34), McElmo Dome in Colorado (30), as well as a site outside of the US context, the cold-water Andernach Geyser in Germany (29).

The gasses within the CO_2-rich systems of the western US (33–35), and the cold-water geyser Andernach near Koblenz, Germany (29) are largely derived from mantle input. However, the CO_2 sources of Crystal Geyser and the McElmo CO_2 reservoir are derived from decarbonation of Paleozoic carbonate rocks (Leadville Limestone) and Pennsylvanian Paradox Formation (36) with ≤ 20% of Crystal Geyser's CO_2 being mantle derived (37). Carbonate dissolution of another marine formation from the Pennsylvanian period, the Madera, supplies the TA anticline spring group with an average of 42% of their CO_2 (33). Average temperature varies between systems: The TA spring group has an average temperature of 37°C, McElmo Dome has and average temperature of 66°C, and both cold geysers have average temperatures of 17°C. The pH in Crystal Geyser fluctuates between eruption cycles from 4.5 and 6.4 (36). All other systems have a circumneutral pH (pH= 6.5-7.3) (1,29,35) while the major ionic composition varies across these systems. Prominent metabolic processes within these systems are N_2 fixation; anaerobic respiration; H_2, sulfur, and Fe oxidation; and fermentation (1,29,30,34). The available data suggests that WL and CBB are the dominant CFPs in CO_2-rich terrestrial systems respectively (1–3,29,30,33) with rTCA being in relatively low abundance if present at all (3,30).

Crystal Geyser (Utah). Crystal Geyser is a cold-water CO_2-geyser in Utah that began to erupt shortly after an 800 m well was dug during a 1936 hydrocarbon expedition (Heath et al., 2009). A time series

geochemical analysis from 2007 to 2014 found that Crystal Geyser had three eruption phases with each phase being fed predominantly by a different aquifer: Entrada aquifer (100-150 m), Navajo aquifer (320 to 480 m), and the Wingate Sandstone aquifer or the Paradox brine (500 to 800 m) (36). Each of these phases had distinct geochemical characteristics. Different CFPs were associated with each of these depths.

Metagenomic studies suggest that CBB and WL are the dominant CFPs in Crystal Geyser with rTCA present but at a relatively low abundance. The CBB-pathway is found at its highest abundance in the eruption fluids associated with the intermediate depth aquifer within members of *Gallionellales* (*Betaproteobacteriota*), marine-like *Mariprofundus* (*Zetaproteobacteria*), *Thiobacillus*-like *Hydrogenophilales*, and *Thiomicrospira crunogena* (4). Both *Gallionellales* and the marine-like *Mariprofundus* were predicted to be Fe(II)-oxidizers. *Gallionellales*, *Thiomicrospira crunogena*, and *Thiobacillus*-like *Hydrogenophilales* have the genetic potential for sulfur oxidation. *Thiobacillus*-like *Hydrogenophilales*, marine-like *Mariprofundus*, and *Gallionellales* appear to be capable of nitrogen reduction. Both *Gallionellales* and *Thiomicrospira crunogena* have the genetic potential for H_2 metabolism (4). All these autotrophs contain type II RuBisCO only, whose low $S_{c/o}$ (compared to type I) limits it to high-CO_2 (>1.5% (26))/low-O_2 (<1% (38,39)) conditions. These limitations could explain why CBB-containing autotrophs seem to be restricted to the intermediate depth aquifers which experience relatively low atmospheric influence compared to the shallow Entrada aquifer (3). The WL-pathway is found in the most abundant organism in the system, the archaea *Candidatus* Altiarchaeum within the Alti-1 clade. *Ca.* Altiarchaeum is a widespread uncultivated *Euryarchaeon* that has been identified in a variety of anaerobic groundwaters (40). *Ca.* Altiarchaeum may use H_2 and/or CO as electron donors (41). It is likely that the WL-pathway thrives at deeper depths due to the energy-limiting conditions of these depths (17). Additionally, *Ca.* Altiarchaeum in Crystal Geyser appear to lack a mechanism for oxygenic stress, likely limiting it to the deeper aquifers (16,17). The rTCA-cycle is predominately used by *Sulfurimonas* (*Epsilonproteobacteria*) (16). KEGG annotations indicated that the *Sulfurimonas* within this system are capable of N_2-fixation as well as sulfide oxidation through O_2-respiration (17). Unlike other known

Epsilonproteobacteria isolates, *Sulfurimonas* are tolerant of relatively high concentrations of O_2 (5-10%) (42–46). Further, genomic analysis of other *Sulfurimonas* strains demonstrate the variant of the rTCA-cycle within these organisms contains specific adaptations within the 2-oxaglutarate and pyruvate:ferredoxin oxidoreductase enzymes that allow for continuous carbon fixation under oxic conditions (47,48). These adaptations would allow *Sulfurimonas* to thrive in the shallower aquifer depths that experience atmospheric influence.

The TA anticline spring group (New Mexico). The TA anticline spring group is associated with the 1.6 km long TA anticline in north-central New Mexico. These springs are characterized by their travertine mounds caused from deeply derived actively venting CO_2-rich fluids. Uranium-series dating of the travertine suggests that these springs have been actively venting for around 210 to 270 ka (34). Two 16S rRNA gene amplicon 454 pyrosequencing studies have been conducted in the TA anticline, revealing the presence of same type of CBB-containing autotrophs found in Crystal Geyser (33,34). They found that the springs separated into three groups; those with lower, intermediate, and higher endogenic (deeply derived) fluid input. Those with a higher endogenic fluid input contained a higher abundance of marine-like *Mariprofundus* (33). Another spring within this group (endogenic level not given) was dominated by *Gallionellales* (*Gallionella capsiferriformans*) (33). Though no metagenomic analysis was done, the fact that higher abundances of marine-like *Mariprofundus* were found in the higher endogenic fluid suggests that these CBB-containing autotrophs are adapted to the high CO_2/low O_2 of these fluids as is seen in Crystal Geyser. There were no reports of WL or rTCA associated with these springs. However, these two studies mark the first microbiological studies of this system, so further research focused on -omics approaches could help resolve the metabolic potential of the microbial community.

McElmo Dome (Colorado). McElmo Dome CO_2 reservoir is the largest commercial supplier of CO_2 worldwide. This 800km^2 supercritical CO_2 reservoir began accumulating CO_2 40 to 72 million years ago. The wells here were dug down to the CO_2 that is trapped below the Leadville limestone formation with the low-permeability Paradox Salt Formation above it at a depth of 1800 to 2600 m (30). Geochemical analysis clustered the wells into two groups: Group 1 was characterized as having a low water to CO_2

ratio and were enriched in Mg, Fe, Ca, Al, Cr and Mn while Group 2 were more concentrated and enriched in Na, B, and As (30). Both the WL-pathway and the CBB-cycle were found associated with the fluids, but no sequences related to the rTCA-cycle were reported. A full WL-pathway was found in members of *Acetobacterium* while partial sequences were found in members of *Desulfovibrio*. Metagenomic sequencing suggests that *Acetobacterium* are capable of N_2-fixation and H_2 oxidation. It also contains phosphoenolpyruvate carboxylase (catalyzes the irreversible addition of bicarbonate) that is hypothesized to increase tolerance to the CO_2-rich environment (30). *Desulfovibrio* has the genetic potential for sulfate and nitrite reduction as well as using H_2 and Fe(II) as electron donors. It also contained a partial denitrification pathway (30). A full CBB-pathway was found within a *Rhizobium petrolearium* sequence and contained a full sulfur-oxidation pathway. Interestingly, this sequence did not contain a type II RuBisCO and instead contained types IC and ID typically associated with mixotrophs (30). *R. petrolearium* is normally associated with oil-contaminated soil. Type IC is adapted for medium to high CO_2 with O_2 present while type ID is associated with low CO_2 with O_2 present (26), suggesting that *R. petrolearium* may be adapted for fluctuating levels of O_2 in this system.

The Andernach Geyser (Germany). Like Crystal Geyser, Andernach Geyser is a cold-water CO_2-geyser. However, its CO_2 is predominantly derived from mantle input (29). A metagenomic study found that WL was by far the most abundant CFP followed by CBB (29). The rTCA-cycle was found in only one assembled genome belonging to *Caldisericales*, but it was the second most abundant genome in the study (29). As with the deeper aquifers of Crystal Geyser, the autotrophic WL-containing archaea *Ca.* Altiarchaeum within the Alti-1 clade (29,41,49) was the most abundant, recruiting 42% of the metagenomic reads (29). Genomic analysis suggests it can use H_2 as an electron donor. Other autotrophs with WL included *Clostridiales*, *Rubrobacteridae* (*Actinobacter*), and *Desulfovibrionales* (*Deltaproteobacteria*) (29). The CBB-cycle was the next most abundant CFP (29). It was found predominantly in *Cyanobacteria* (*Pleurocapsales* and *Nostocales*). Both have the genetic potential for aerobic respiration and sulfur oxidation, while *Pleurocapsales* can reduce nitrogen species. This pathway was also present in Alpha- (*Rhizobiales*), Beta- (*Hydrogenophiales*), and Gammaproteobacteria

(*Xanthomanadales*). The rTCA-pathway was only found in *Caldisericales* which had no genetic predictions for any redox reactions given (29).

CO_2-depleted Systems

Compared to their CO_2-rich counterparts, terrestrial CO_2-depleted systems have a wealth of studies characterizing the microbial community due to their analogous nature to environments that are expected to be found on other planetary bodies (50). These studies have largely focused on terrestrial serpentinizing systems. Most of these systems are the result of ophiolite emplacement; sections of ancient oceanic crust and upper mantle rock, rich in the minerals olivine and pyroxene, that have been thrust onto the continental shelf. These include the Coast Range Ophiolite (including the Cedars and the Microbial Observatory (51,52) in California-USA, the Tablelands Ophiolites in Newfoundland-Canada (53,54), as well as the volcanic outer for arc and the Santa Elena Ophiolite in Costa Rica (55). Fluids in these systems are usually meteoric in origin, causing them to be low in SO_4^{-2} concentrations relative to their seawater fed marine counterparts. Other terrestrial serpentinizing systems are the result of hydration of the mantle wedge at temperatures < 700°C (56) as seen on the Nicoya Peninsula in Costa Rica (57). This is the result of seawater and serpentinized fluids being released from the subducting oceanic plant and migrating up through the mantle of the continental plate.

What these systems have in common is the interaction of water with low-silica ultramafic rock rich in olivine and pyroxene leading to the oxidation of ferrous iron within these minerals, converting it into ferric iron and releasing copious H_2 (58). This process increases the pH of the system, often reaching a pH >10 in systems with temperatures < 150°C (58). Under high pH conditions, CO_3^{2-} becomes the dominant carbon species and aqueous CO_2 is low if not depleted. CO_3^{2-} interactions with Ca^{2+} liberated during the water-rock interaction leads to it precipitating out as calcium carbonate ($CaCO_3$), depleting the DIC within the system (50). These characteristics mean that microbes living in such environments must contend with systems that, while rich in electron donors (ie. CH_4 and H_2), are significantly limited in both electron acceptors and carbon substrates.

The Coast Range Ophiolite. Research into the geochemistry and geomicrobiology of the Coast Range Ophiolite has been conducted in two locations within Northern California: the Cedars in Sonoma County and the Coast Range Ophiolite Microbial Observatory (CROMO) at the UC-Davis McLaughlin Natural Reserve in Lower Lake. Within the Cedars, a shallow moderately alkaline (pH 8.3 to 8.6) Mg-HCO_3 meteoric source formed from weathering of serpentinites and a deep anoxic ultrabasic (ph > 11 (59)) Ca-OH source discharging from serpentinizing aquifers contributes to the fluid of this system (60). Geochemical analysis suggests that the shallow source only comes in contact with the peridotite body while the deep source interacts with the marine sediments associated with the Franciscan Subduction Complex before percolating through the peridotite body. During their migration to the surface, these two end sources mix with each other and are expressed as ultrabasic springs. When expressed on the surface, the pH of the fluids range between 11 and 12 with a reducing potential between -656 to -585 mV and a temperature range of 17.2 and 17.4°C (60). Additionally, while DIC concentrations were not given, these fluids are inorganic carbon limited with TIC between 6 µmol/L to 70 µmol/L and CO_3^{2-} being the dominant carbon species present (61). SO_4^{-2} concentrations are very low, with a range from below the limit of detection to 8 µmol/L with CH_4 concentrations between 3.23 mol/L to 9.63 mol/L. Cell counts of the Cedars are low and differ based on the degree of end member mixing of the spring fluids. Springs influenced mostly by the deep source have higher cell counts of 10^2 to 10^3 ml^{-1} compared to the 10 ml^{-1} found in springs fed by mixed sourced fluid (61).

Unlike the fluids from the Cedars, the fluids from wells within CROMO are believed to be primarily from trapped Cretaceous seawater with some degree of meteoric input (62). Wells have been dug at varying depths allowing researchers to examine different levels of intermixing as well as see a transition between a terrestrial serpentinization system to a more marine-like serpentinization system as depths of the wells penetrate into the ancient seabed (63). These wells can be separated into three groups. Group 1 are the shallower wells (8.8 m to 19.2 m) and represent a meteoric-alkaline water mixture with a pH between 7.8 to 9.5. Group 2 are classified as medium depth (between 19.5 m to 26.2 m) and are a mix of serpentinized meteoric fluids as well as serpentinized ancient seawater. Group 3 (27.4 m to 76.20 m) is

predominantly composed of serpentinized ancient seawater, likely representing a transition between a terrestrial serpentinizing system and that of a marine serpentinizing system (63). For the purposes of this paper, we will focus on the parameters and findings within the medium (group 2) and deep (group 3) depth wells as they are less influenced by fresh meteoric dilution of serpentinized fluids. pH within the medium depth wells is between 10.15 and 12.32 while with the deep depth wells it drops to between 9.68 and 10.89. Temperature is with the same range between the well groups (15°C to 18°C). SO_4^{-2} concentrations (27.69 to 317.41 µmol/L) generally increase with depth (63), testifying to the increased influence of the ancient seawater. DIC concentrations are low (20 to 593.43 µmol/L) though not as low as seen in the Cedars (62,64). The reduction potential of these wells is between -297 to -198 mV, though measurements of fluids within the CROMO have reported as low as -611 mV (62). Further, cell counts are higher compared to the Cedars ($2x10^5$cells/ml), though microbial diversity is still low.

While evidence of possible methanogenesis has been found in the CROMO, the methane isotopologue data (variation in the isotopic composition of methane) were unable to resolve the source of methane but suggested that it might be produced thermogenically and/or microbiologically (51,63). Additionally, low levels of *mcr* were expressed in the metatranscriptomes, though no evidence of the *mcrA* gene was found in the metagenomes (63). It seems that if methanogenesis is happening in the CROMO wells, it is likely under stress induced conditions such as H_2 limitation (63). An -omics study found complete CBB- and rTCA-cycles as well as the WL-pathway associated with homoacetogens in the metagenomes and metatranscriptomes for all the wells. However, genes for the WL-pathway had very low abundance/expression in the least alkaline wells. Genes for the rTCA cycle were found to have their highest abundance/expression in the well with the lowest DIC concentrations while the abundance/expression of CBB associated genes was consistent throughout the wells sampled (63).

Within the Cedars and CROMO, dominant phylotypes include *Betaproteobacteriota* (closely related to the genus *Hydrogenophaga*) and *Clostridia* (52,64). The *Hydrogenophaga*-like MAGs found at the Cedars are associated with springs influenced by the shallow meteoric (52) fluid source while those found in CROMO are associated with the most extreme (in terms of highest pH and most negative reduction

potential) wells with the least microbial diversity (64). An isolation study found that *Betaproteobacteriota* isolates from the Cedars are capable of autotrophic growth with H_2 oxidation and O_2 reduction while fixing $CaCO_3$ $CaCO_3$, likely through the liberation of carbonate, through the CBB-pathway with a type I RuBisCO (59). Members of other *Hydrogenophaga* isolates and sequences have been shown to contain RuBisCO type IAq, a variant that is not known to be associated with carboxysomes (26). Given that is the case, it is interesting to note that the Cedars *Hydrogenophaga* isolates contain shell proteins associated with carboxysomes, suggesting adaptation to the CO_2-poor environment (59). Further, another member of the *Betaproteobacteriota*, *Thiobacillus denitrificans*, has been found to contain type IAq RuBisCO with a a-carboxysome gene nearby (26), suggesting that IAq may actually be associated with carboxysomes in one form or another. Additionally, the Cbb3-type cytochrome used to deduce O_2 reduction in the Cedars *Betaproteobacteriota* isolates was expressed significantly higher during mixotrophic growth than under autotrophic conditions (26). Furthermore, Cbb3-type cytochrome has been suggested to used O_2 as an electron acceptor under low O_2 concentrations or as a means to protect against O_2 toxicity (65). In either case, it demonstrates that the cytochrome within the *Hydrogenophaga*-like *Betaproteobacteriota* of this system is likely adapted to low O_2 levels. While the authors note that no other studies have demonstrated fixation of $CaCO_3$, they propose that one explanation for this observation is that the *Hydrogenophaga*-like *Betaproteobacteriota* may aggregate on $CaCO_3$ and decrease the pH in the microenvironment around the $CaCO_3$, converting it to bicarbonate to be taken into the cell (65). *Betaproteobacteriota* from the CROMO likely have a similar capability given that they share 100% 16S rRNA gene sequence similarity with these Cedars isolates (64).

Clostridia are associated with both the shallow and deep fluid sources in the Cedars, indicating that the shallow fluids were likely inoculated with *Clostridia* from the deep fluid source (52). At the CROMO, while present in all samples, they were the highest abundant organism in the CSWold samples. The most abundant member of this group in both the CROMO and the Cedars has >97% sequence similarity to *Dethiobacter alkaliphilus* (52,64). This organism is capable of autotrophic growth using H_2 as an electron donor while using thiosulfate and polysulfide as electron acceptors. Carbon is fixed via the WL-pathway.

In addition to fixing CO_2, this organism has been demonstrated to use acetate as a carbon source, but not as an energy source (64,66). Interestingly, formate dehydrogenase, which converts CO_2 to formate as the first step of the WL-pathway in acetogens, was missing from every MAG with the WL-pathway recovered from the CROMO (51). Metagenomic analysis suggests that homoacetogenesis via the WL-pathway is present in both the metagenomes of the CROMO (51) and the Cedars (22). It has been suggested that formate dehydrogenase may not be necessary if formate is being produced through abiotic production via Fischer-Tropsch-like reactions (51). In this case, WL-containing prokaryotes could directly use formate as a carbon source for the methyl branch of the pathway (67). This would further reduce the WL-pathway's requirement of <1 ATP and would remove the need for external electron acceptor apart from co-assimilated CO_2 (51,67). Additionally, it has been found that abiotic CO can be used as a carbon source for the carbonyl branch in the Cedars (53,68).

Tablelands Ophiolites. The Tablelands are part of the Bay of Islands Ophiolite which were obducted onto the Laurentian continental margin from the ancient Iapetus Ocean during the Ordovician period (485 m.y.a.). They are located within the Gros Morne National Park, Newfoundland, Canada where multiple serpentinizing springs have been spatiotemporally studied (54). These springs have a recorded reducing potential between -468 to -690mV, pH > 11, and a temperature range from 8.3°C to 24.3°C (53,54). Cell counts range from not detectable to $4x10^5$ (54). While DIC concentrations are not given, the springs had a total inorganic carbon (TIC) range of 91.6 to 233 µmol/L (53). SO_4^{-2} concentrations are between 8.54 to 13.6 µmol/L with CH_4 concentrations between 20 to 26.2 µmol/L. Carbon isotopic measurements looking for evidence of methanogenesis support previous observations that CH_4 within the system is likely derived thermogenically and/or abiotically (53,69). Cell counts range from non-detectable to $4x10^5$.

As with the CROMO and the Cedars, it seems that the most abundant phylum in the Tablelands is the *Hydrogenophaga*-like *Betaproteobacteriota* (25,53). As with those found in the CROMOS and the Cedars, they also contain both the large and small subunits of type 1 RuBisCO. A 16S rRNA gene amplicon analysis found that while these *Hydrogenophaga*-like *Betaproteobacteriota* are the most abundant taxa in the emanating fluids (determined by recharge of spring pool) they are less abundant than

the community within surface and subsurface fluid interface (the spring pool fluid). The *Betaproteobacteriota* are rarely found in upstream surface water. The authors suggest that these *Hydrogenophaga*-like *Betaproteobacteriota* thrive in the mixing zone of the two fluid types and others have suggested that they are not likely subsurface in nature (22,53). However, it should be noted that 16S rRNA gene sequences with 100% identity to these *Betaproteobacteriota* are the most abundant sequences in the most extreme (in terms of highest pH and most negative reduction potential) wells with the least microbial diversity from CROMO (64). It is possible these *Betaproteobacteriota* may not be specifically adapted to the mixing zone but may be members of the subsurface community that experience a bloom-like response when coming in contact with the mixing zone. Interestingly, these *Betaproteobacteriota* seem to increase their growth with the addition of methane, a metabolic capability that has scant evidence in this system according to studies using either taxonomic or function gene approaches (64). Acetyl-CoA synthase (ACS), the key enzyme involved in the WL-pathway found within Tableland metagenomes clusters with the ACS of *D. alkaliphilus* (25). The presence of rTCA has not been reported in the Tablelands.

The Santa Elena Ophiolites and Costa Rica outer forearc. At the time of writing, two separate serpentinizing sites have been identified in Costa Rica: the Santa Elena ophiolites and Nicoya Peninsula within the outer forearc of Costa Rica. The hydrological mechanisms driving these serpentinization sites differ. While not the first potential serpentinization reported in a tropical climate (57), it was the first to be comprehensively characterized (55). This system is unique compared to other serpentinization systems as the increased rainfall in the tropical climate causes increased infiltration of meteoric water into the system, potentially driving further water-rock interactions (55). The Santa Elena ophiolite is a large complex within the Santa Elena peninsula of the Area Conservacion Guanacaste covering a 250 km^2 area composed of mafic and ultramafic lithologies (70). It has been suggested that this system likely originated during the Upper Cretaceous due to the complex being covered in rudist-rich reef limestones (55,70,71). Like the deep fluid source seen in the Cedars (see above and ref (60)), the fluids here are ultrabasic (pH > 11.01-11.18) and representative of Ca-OH waters. Average temperature of the spring fluids is 29°C (55),

about 10° warmer than those of the CROMO and the Cedars. No CO_2 concentrations were given, but the DIC concentrations are very low (126.3 to 254.3 µmol/L) (50). Cell counts ranged from $2.0x10^4$ to $1.51x10^5$ (55), similar to those seen in the CROMO.

In contrast to Santa Elena ophiolites, serpentinization on the Nicoya Peninsula is largely driven by seawater and serpentinized fluid hydration of the mantle wedge as the oceanic plate subducts below the continental plate. The pH of the fluids is lower here compared to the other serpentinizing systems (pH 8.53-9.99; median = pH 9.89) while the temperature is higher (26.4°C to 48.7; median = 33°C). CO_2 concentrations are depleted as well (all sites at 0 µmol/L, except for El Sitio which registered 1,530 µmol/L) while DIC concentrations are somewhat higher than other systems (all sites between 100 to 550 µmol/L except El Sitio which registered 2240 µmol/L). Interestingly, SO_4^{-2} concentrations are high relative to other terrestrial serpentinizing systems (30 to 948 µmol/L) (24,31,32). The elevation in SO_4^{-2} is likely due to inmixing of SO_4^{-2} -rich seawater within this system. Cell counts can be higher than those found in the Santa Elena ophiolites with a range from $1.88x10^4$ to $3.04x10^6$ (median = $3.61x10^5$) (31,32).

Information on carbon fixation within the Santa Elena ophiolites is very limited. One study looking for the contribution of methanogens within the microbial community found that 90% of archaeal 16S sequences were members of known methanogens (50). They identified all the key genes associated with the WL-pathway; the carbon fixation pathway associated with methanogenesis (50). In fact, this system contains a higher abundance of methanogens than has been reported in other terrestrial serpentinizing systems (50). While the $\delta^{13}C$-CH_4 signatures of the springs suggest that a majority of the methane is likely abiotically produced, the identification of all genes associated with methanogenesis and the high abundance of known methanogens suggests that biological signals of methanogenesis may be overshadowed by abiotic signals (50). One of most dominant phyla in the Santa Elena ophiolite springs is a similar autotrophic hydrogenotrophic denitrifying *Hydrogenophaga*-like *Betaproteobacteriota* found in the Cedars and CROMO (55), suggesting that a low-CO_2 adapted CBB-pathway may be active in this system as well, utilizing $CaCO_3$ as a direct carbon source. The presence of rTCA has not been reported in the Santa Elena ophiolite.

In a metagenomic analysis of the Costa Rican convergent margin, it was found that 39% of the read abundance in the outer forearc samples mapped to chemolithoautotrophs (Chapter 2). 74% of these were mapped to WL-containing metagenomic assembled genomes (MAGs) belonging to members of *Thermodesulfovibrionia*, *Ca.* Bipolaricaulia, and CSP1-3 (an unclassified order in the phylum *Armatimonadetes*). Genomic analysis predicted that the *Thermodesulfovibrionia* MAGs are hydrogenotrophic sulfate reducers, *Ca.* Bipolaricaulia MAGs can grow hydrogenotrophicly, and CSP1-3 are capable of sulfide oxidation (24). However, unlike Santa Elena, there was no evidence for methanogenesis in any of these MAGs, nor was the *mcr* gene located in any of the outer forearc metagenomes. Further, a 16S rRNA gene amplicon sequencing study of the same metagenomes did not uncover any sequences associated with known methanogens (32). The CBB- and the rTCA-cycle were also present in the outer forearc, though they only accounted for 5% of the total mapped reads each. The CBB-cycle is predicted in unclassified members of the genus *Thiobacillus* (*Gammaproteobacteria*) capable of sulfide oxidation with N_2 and/or O_2 reduction (microaerophilic). Cultured members of this genus are known to possess both a type II and type IAq RuBisCO (26), suggesting adaptation to changes in CO_2 levels. The rTCA-cycle is predicted in *Syntrophia* (*Desulfobacterota*) capable of SO_4^{-2} reduction. Interestingly, the key gene ATP-citrate lyase of the rTCA-pathway within these *sytrophia* MAGs seems to have been horizontally transferred from methanogenic archaea (the subject of chapter 4), a characteristic that seems to be specific to *Syntrophia* in this system.

Conclusion

A fish out of water: Terrestrial distribution of the marine-like *Mariprofundus*

The marine-like *Mariprofundus* reported in high CO_2 semi-saline systems is an interesting observation. Unlike their marine counterparts, this terrestrial variant of *Mariprofundus* lacks the potential for assimilatory nitrate reduction to ammonium but has the potential for N_2-fixation (73). Marine-like *Mariprofundus* are not commonly expected to be found in terrestrial environments due to their need for micro-oxic brackish to hypersaline (≥ 5 ppt salinity) environments with an ample supply of Fe(II) (73). A recent genomics study exploring the distribution of *Zetaproteobacteria* found that the strains discovered

in Crystal Geyser and TA anticline are likely endemic or specifically adapted to terrestrial CO_2 and Fe(II)-rich springs where they make up between 10-35% of the community structure (73). However, both the TA anticline and Crystal Geyser are fed by aquifers associated with ancient marine limestone deposits belonging to the Paradox formation. Recent research using Krypton-81 dating suggests brines associated with the Paradox formation contain remnants of Paleozoic age (>250 Ma) seawater mixed with relatively recent meteoric water from 16-35 ka (74). Based on these findings, it's possible that the marine-like *Mariprofundus* found in these sites are remnants of an ancient marine community that was present during sea level rise of the Pennsylvanian time period.

Carbon fixation pathway distribution differences

In both CO_2-rich and depleted systems, the WL-pathway and the CBB-cycle seem to be the most abundant and expressed carbon fixation pathways while the rTCA-cycle seems to appear more often in CO_2-rich than CO_2-depleted systems (though in low abundance). However, the taxa that possess these pathways are more similar within system types than between system types and show potential adaptations to the CO_2 concentration of the system in which they are found. In CO_2-rich systems, the CBB-pathway is predominantly found in members of marine-like *Mariprofundus* within the phylum *Zetaproteobacteriota* and *Gallionella* within the phylum *Betaproteobacteriota* which both contain the high CO_2 adapted type II RuBisCO. However, in CO_2-depleted systems the CBB-pathway is predominantly found in members of *Hydrogenophaga*-like *Betaproteobacteriota* (55) which contain a reduced CO_2 adapted type I RuBisCO and shell proteins that suggest the use of carboxysomes; characteristics of an organism adapted to low levels of CO_2. Additionally, the *Hydrogenophaga*-like *Betaproteobacteriota* have the ability to use $CaCO_3$, the dominant carbon species in calcium containing serpentinization systems, for carbon fixation while using acetate and/or methane for growth.

Similar CO_2 and O_2 level dependent adaptations and taxa distributions can be found in WL- and rTCA- containing chemolithoautotrophs. When rTCA was present in a system it tended to be associated with a single low abundance type of bacteria. In the CO_2-rich environment of Crystal Geyser rTCA containing *Sulfurimonas* were specially adapted to deal with elevated O_2 levels. rTCA was reported in

low abundance in CO_2-rich systems, suggesting that it is not a big player in terrestrial serpentizing systems. WL-containing *Dethiobacter alkaliphilus* seems to be endemic within most of the serpentinizing systems reviewed. Interestingly, it can use acetate as a carbon source, potentially giving it a competitive advantage in CO_2-depleted systems. *Ca.* Altiarchaeum seems to be abundant, if not the dominant taxa, in a number of deep anoxic subsurface systems (29,40,41). The WL-containing *Ca.* Altiarchaeum within the Alti-1 clade was reported as the most dominant taxon in the cold CO_2 geysers, but not reported at all in the CO_2-depleted systems. Further, they seemed to be restricted to the deep anoxic aquifers of these geysers. However, there is no indication in their metagenomes which would suggest that their distribution should depend on CO_2 concentrations. One explanation for their absence in the CO_2-depleted systems may be the lack of extreme fluid flow seen in cold CO_2 geysers and other high flow environments in which they are found. It has been suggested that this high flow rate may be necessary for the delivery of substrates and/or the removal of metabolites (72), such as those rates seen in cold CO_2-geysers but not seen in serpentinization systems. *Ca.* Altiarchaeum within the Alti-1 clade, which are recovered from high flow environments, contain nanoscale 'grappling hooks' that might allow for the production of strong biofilm formations as well as anchoring into biomats (41). *Ca.* Altiarchaeum found in calm saline environments do not display the 'grappling hook' phenotype, further suggesting its function in high flow rate environments (41). Another possible explanation for their absence in CO_2-depleted systems is the high pH found in these systems. The environments in which *Ca.* Altiarchaeum have been recovered have a circumneutral pH (29,40,41), suggesting that the ultrabasic conditions found in serpentinizing systems may not be conducive to growth of *Ca.* Altiarchaeum.

The pH of both systems might explain the near absence of rTCA reported in both CO_2 rich and depleted systems. As mentioned earlier, operation of 2-oxoglutarate synthase and isocitrate dehydrogenase has a thermodynamic cost of + 42 kJ/mol. To overcome this limitation, the organism could keep the cellular pH well below 7, bringing the thermodynamic cost of the two enzymes down to ~28 kJ/mol (20). Multiple studies have confirmed the presence and the degree of community contribution of the rTCA within Crystal Geyser. This carbon fixation pathway was also found in the second most

abundant MAG from the Andernach cold CO_2-geyser in Germany. Both systems have the lowest pH ranges measured in the studies reviewed here. For organisms with rTCA, it might be more energetically feasible to maintain an intracellular pH < 7 in an environment with a pH < 7, which could explain the lack of reports of this pathway in terrestrial serpentinizing systems which tend to have pH > 9.

References

1. Shelton JL, Andrews RS, Akob DM, DeVera CA, Mumford AC, Engle M, et al. Compositional Analysis of Formation Water Geochemistry and Microbiology of Commercial and Carbon Dioxide-Rich Wells in the Southwestern United States. Report, Scientific Investigations. 2020;
2. Wicaksana A. Science of Carbon Storage in Deep Saline Formations: Process Coupling across Time and Spatial Scales. Https://Medium.Com/. 2016.
3. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO 2 concentrations. Environmental Microbiology. 2017;19(2):459–74.
4. Emerson JB, Thomas BC, Alvarez W, Banfield JF. Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla. Environmental microbiology. 2016;18(6):1686–703.
5. Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. Function, Structure, and Evolution of the RuBisCO -Like Proteins and Their RuBisCO Homologs. Microbiology and Molecular Biology Reviews. 2007;71(4):576–99.
6. Middelburg JJ. Chemoautotrophy in the ocean. Geophysical Research Letters. 2011;38(24):94–7.
7. Momper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. ISME Journal. 2017;11(10):2319–33.
8. Mcmahon S, Parnell J. Weighing the deep continental biosphere. FEMS Microbiology Ecology. 2014;87(1):113–20.
9. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. Global distribution of microbial abundance and biomass in subseafloor sediment. Proceedings of the National Academy of Sciences. 2012 Oct;109(40):16213–6.
10. James Fredrickson, Onstott TC. Microbes Deep inside the Earth. Scientific America. 1996;275:68–73.
11. Stevens T. Lithoautotrophy in the subsurface. FEMS Microbiology Reviews. 1997;20(3–4):327–

37.

12. Zhao T, Li Y, Zhang Y. Biological carbon fixation: a thermodynamic perspective. Green Chemistry. 2021;23(20):7852–64.

13. Bar-Even A, Noor E, Milo R. A survey of carbon fixation pathways through a quantitative lens. Journal of Experimental Botany. 2012;63(6):2325–42.

14. Kraus EA, Nothaft D, Stamps BW, Rempfert KR, Ellison ET, Matter JM, et al. Molecular Evidence for an Active Microbial Methane Cycle in Subsurface Serpentinite-Hosted Groundwaters in the Samail Ophiolite, Oman. Applied and Environmental Microbiology. 2021;87(2):1–18.

15. Fones EM, Colman DR, Kraus EA, Stepanauskas R, Templeton AS, Spear JR, et al. Diversification of methanogens into hyperalkaline serpentinizing environments through adaptations to minimize oxidant limitation. ISME Journal. 2021;15(4):1121–35.

16. Probst AJ, Elling FJ, Castelle CJ, Zhu Q, Elvert M, Birarda G, et al. Lipid analysis of CO_2-rich subsurface aquifers suggests an autotrophy-based deep biosphere with lysolipids enriched in CPR bacteria. ISME Journal. 2020;1547–60.

17. Probst AJ, Ladd B, Jarett JK, Geller-Mcgrath DE, Sieber CMK, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nature Microbiology. 2018;3(3):328–36.

18. Berg IA. Ecological aspects of the distribution of different autotrophic CO 2 fixation pathways. Applied and Environmental Microbiology. 2011;77(6):1925–36.

19. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. Nature Reviews Microbiology. 2010;8(6):447–60.

20. Bar-Even A, Flamholz A, Noor E, Milo R. Thermodynamic constraints shape the structure of carbon fixation pathways. Biochimica et Biophysica Acta - Bioenergetics. 2012;1817(9):1646–59.

21. Cotton CA, Edlich-Muth C, Bar-Even A. Reinforcing carbon fixation: CO_2 reduction replacing and supporting carboxylation. Current Opinion in Biotechnology. 2018;49:49–56.

22. Twing KI, Brazelton WJ, Kubo MDY, Hyer AJ, Cardace D, Hoehler T, et al. Serpentinization-influenced groundwater harbors extremely low diversity microbial communities adapted to high pH. Frontiers in Microbiology. 2017;8:308.

23. Drake HL, Gößner AS, Daniel SL. Old acetogens, new light. Annals of the New York Academy of Sciences. 2008;1125(January):100–28.

24. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, et al. Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME Journal. 2022;(October):1–11.

25. Brazelton WJ, Nelson B, Schrenk MO. Metagenomic evidence for h(2) oxidation and h(2) production by serpentinite-hosted subsurface microbial communities. Frontiers in microbiology. 2012;2:268.

26. Badger MR, Bek EJ, Plant M, Group P, Australian T. Multiple Rubisco forms in proteobacteria : their functional significance in relation to CO 2 acquisition by the CBB cycle. 2008;59(7):1525–41.

27. Badger MR, Sharwood RE. Rubisco , the imperfect winner : it ' s all about the base. 2022;1–19.

28. Claassens NJ, Scarinci G, Fischer A, Flamholz AI, Newell W. Phosphoglycolate salvage in a chemolithoautotroph using the Calvin cycle. 2020;

29. Bornemann TLV, Adam PS, Turzynski V, Schreiber U, Figueroa-Gonzalez PA, Rahlff J, et al. Genetic diversity in terrestrial subsurface ecosystems impacted by geological degassing. Nature Communications. 2022;13(1):1–12.

30. Freedman AJE, Tan BF, Thompson JR. Microbial potential for carbon and nutrient cycling in a geogenic supercritical carbon dioxide reservoir. Environmental Microbiology. 2017;19(6):2228–45.

31. Barry PH, de Moor JM, Giovannelli D, Schrenk M, Hummer DR, Lopez T, et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. Nature. 2019;568(7753):487–92.

32. Fullerton KM, Schrenk MO, Yücel M, Manini E, Basili M, Rogers TJ, et al. Effect of tectonic

processes on biosphere–geosphere feedbacks across a convergent margin. Nature Geoscience. 2021;14(5):301–6.

33. Crossey LJ, Karlstrom KE, Schmandt B, Crow RR, Colman DR, Cron B, et al. Continental smokers couple mantle degassing and distinctive microbiology within continents. Earth and Planetary Science Letters. 2016;435:22–30.

34. Cron B. Geochemical characteristics and microbial diversity of CO_2-rich mound springs of the Tierra Amarilla anticline, New Mexico. 2012;

35. McGibbon C, Crossey LJ, Karlstrom KE, Grulke T. Carbonic springs as distal manifestations of geothermal systems, highlighting the importance of fault pathways and hydrochemical mixing: Example from the Jemez Mountains, New Mexico. Applied Geochemistry. 2018;98:45–57.

36. Han WS, Watson ZT, Kampman N, Grundl T, Graham JP, Keating EH. Periodic changes in effluent chemistry at cold-water geyser: Crystal geyser in Utah. Journal of Hydrology. 2017;550:54–64.

37. Fabiana Meijon Fadul. Plumbing the depths – testing natural tracers of subsurface CO_2 origin and migration, Utah, USA. 2019;

38. Robinson JJ, Scott KM, Swanson ST, O'Leary MH, Horken K, Tabita FR, et al. Kinetic isotope effect and characterization of form II RubisCO from the chemoautotrophic endosymbionts of the hydrothermal vent tubeworm Riftia pachyptila. Limnology and Oceanography. 2003;48(1 I):48–54.

39. Léniz B, Murillo AA, Ramírez-Flandes S, Ulloa O. Diversity and transcriptional levels of RuBisCO form II of sulfur-oxidizing γ-proteobacteria in coastal-upwelling waters with seasonal anoxia. Frontiers in Marine Science. 2017;4(JUL):1–17.

40. Probst AJ, Weinmaier T, Raymann K, Perras A, Emerson JB, Rattei T, et al. Biology of a widespread uncultivated archaeon that contributes to carbon fixation in the subsurface. Nature Communications. 2014;5(May).

41. Bird JT, Baker BJ, Probst AJ, Podar M, Lloyd KG. Culture independent genomic comparisons

reveal environmental adaptations for altiarchaeales. Frontiers in Microbiology. 2016;7(AUG):1–14.

42. Hügler M, Sievert SM. Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Annual Review of Marine Science. 2011;3(1):261–89.

43. Inagaki F, Takai K, Kobayashi H, Nealson KH, Horikoshi K. Sulfurimonas autrotrophica gen. nov., sp. nov., a novel sulfur-oxidizing E-proteobacterium isolated from hydrothermal sediments in the Mid-Okinawa Trough. International Journal of Systematic and Evolutionary Microbiology. 2003;53(6):1801–5.

44. Inagaki F, Takai K, Nealson KH, Horikoshi K. Sulfurovum lithotrophicum gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the E-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. International Journal of Systematic and Evolutionary Microbiology. 2004;54(5):1477–82.

45. Holden JF, Scott K, Wang S, Jiang L, Hu Q, Cui L, et al. Characterization of Sulfurimonas hydrogeniphila sp . nov ., a Novel Bacterium Predominant in Deep-Sea Hydrothermal Vents and Comparative Genomic Analyses of the Genus Sulfurimonas. 2021;12(February):1–20.

46. Hu Q, Wang S, Lai Q, Shao Z, Jiang L. Sulfurimonas indica sp. Nov., a hydrogen-and sulfur-oxidizing chemolithoautotroph isolated from a hydrothermal sulfide chimney in the Northwest Indian Ocean. International Journal of Systematic and Evolutionary Microbiology. 2021;71(1):1–7.

47. Molari M, Hassenrueck C, Laso-Pérez R, Wegener G, Offre P, Scilipoti S, et al. A hydrogenotrophic Sulfurimonas is globally abundant in deep-sea oxygen-saturated hydrothermal plumes. Nature Microbiology. 2023;8(April).

48. Bayer B, Saito MA, McIlvin MR, Lücker S, Moran DM, Lankiewicz TS, et al. Metabolic versatility of the nitrite-oxidizing bacterium Nitrospira marina and its proteomic response to oxygen-limited conditions. ISME Journal. 2021;15(4):1025–39.

49. Ortiz M, Leung PM, Shelley G, Goethem MW Van, Bay SK, Jordaan K, et al. A genome

compendium reveals diverse metabolic adaptations of Antarctic soil microorganisms. 2020;

50. Crespo-Medina M, Twing KI, Sánchez-Murillo R, Brazelton WJ, McCollom TM, Schrenk MO. Methane dynamics in a tropical serpentinizing environment: The Santa Elena Ophiolite, Costa Rica. Frontiers in Microbiology. 2017;8:916:1–14.

51. Seyler LM, Brazelton WJ, McLean C, Putman LI, Hyer A, Kubo MDY, et al. Carbon Assimilation Strategies in Ultrabasic Groundwater: Clues from the Integrated Study of a Serpentinization-Influenced Aquifer. mSystems. 2020;5(2).

52. Suzuki S, Ishii S, Wu A, Cheung A, Tenney A, Wanger G, et al. Microbial diversity in the Cedars, an ultrabasic, ultrareducing, and low salinity serpentinizing ecosystem. Proceedings of the National Academy of Sciences of the United States of America. 2013;110(38):15336–41.

53. Morrill PL, Brazelton WJ, Kohl L, Rietze A, Miles SM, Kavanagh H, et al. Investigations of potential microbial methanogenic and carbon monoxide utilization pathways in ultra-basic reducing springs associated with present-day continental serpentinization: The Tablelands, NL, CAN. Frontiers in Microbiology. 2014;5(NOV):1–13.

54. Brazelton WJ, Morrill PL, Szponar N, Schrenk MO. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. Applied and environmental microbiology. 2013 Jul;79(13):3906–16.

55. Sánchez-Murillo R, Gazel E, Schwarzenbach EM, Crespo-Medina M, Schrenk MO, Boll J, et al. Geochemical evidence for active tropical serpentinization in the Santa Elena Ophiolite, Costa Rica: An analog of a humid early Earth? Geochemistry, Geophysics, Geosystems. 2014 May;15(5):1783–800.

56. Deschamps F, Godard M, Guillot S, Hattori K. Geochemistry of subduction zone serpentinites: A review. Lithos. 2013;178:96–127.

57. DeShon HR, Schwartz SY. Evidence for serpentinization of the forearc mantle wedge along the Nicoya Peninsula, Costa Rica. Geophysical Research Letters. 2004;31(21):2–6.

58. Schrenk MO, Brazelton WJ, Lang SQ. Serpentinization, Carbon, and Deep Life. Reviews in

Mineralogy and Geochemistry. 2013;75(1).

59. Suzuki S, Kuenen JG, Schipper K, Van Der Velde S, Ishii S, Wu A, et al. Physiological and genomic features of highly alkaliphilic hydrogen-utilizing Betaproteobacteria from a continental serpentinizing site. Nature Communications. 2014;5(May).

60. Barnes I, Lamarche VC, Himmelberg G, Science S, Series N, May N. Geochemical Evidence of Present-Day Serpentinization Published by : American Association for the Advancement of Science Stable URL : https://www.jstor.org/stable/1722022. 1967;156(3776):830–2.

61. Morrill PL, Kuenen JG, Johnson OJ, Suzuki S, Rietze A, Sessions AL, et al. Geochemistry and geobiology of a present-day serpentinization site in California: The Cedars. Geochimica et Cosmochimica Acta. 2013;109:222–40.

62. Cardace D, Hoehler T, McCollom T, Schrenk M, Carnevale D, Kubo M, et al. Establishment of the Coast Range ophiolite microbial observatory (CROMO): Drilling objectives and preliminary outcomes. Scientific Drilling. 2013;(16):45–55.

63. Sabuda MC. Biogeochemistry of environmental gradients in serpentinization-influenced groundwater at the Coast Range Ophiolite Microbial Observatory, California. ProQuest. 2017;136.

64. Crespo-Medina M, Twing KI, Kubo MDY, Hoehler TM, Cardace D, McCollom T, et al. Insights into environmental controls on microbial communities in a continental serpentinite aquifer using a microcosm-based approach. Frontiers in Microbiology. 2014 Nov;5:604.

65. Sievert SM, Scott KM, Klotz MG, Chain PSG, Hauser LJ, Hemp J, et al. Genome of the Epsilonproteobacterial Chemolithoautotroph. 2008;74(4):1145–56.

66. Sorokin DY, Tourova TP, Mußmann M, Muyzer G. Dethiobacter alkaliphilus gen. nov. sp. nov., and Desulfurivibrio alkaliphilus gen. nov. sp. nov.: Two novel representatives of reductive sulfur cycle from soda lakes. Extremophiles. 2008;12(3):431–9.

67. Bar-Even A. Formate Assimilation: The Metabolic Architecture of Natural and Synthetic Pathways. Biochemistry. 2016;55(28):3851–63.

68. Suzuki S, Ishii S, Hoshino T, Rietze A, Tenney A, Morrill PL, et al. Unusual metabolic diversity

of hyperalkaliphilic microbial communities associated with subterranean serpentinization at the Cedars. ISME Journal. 2017;11(11):2584–98.

69. Szponar N, Brazelton WJ, Schrenk MO, Bower DM, Steele A, Morrill PL. Geochemistry of a continental site of serpentinization, the Tablelands Ophiolite, Gros Morne National Park: A Mars analogue. Icarus. 2013;224(2):286–96.

70. Gazel E, Denyer P, Baumgartner PO. Magmatic and geotectonic significance of Santa Elena Peninsula, Costa Rica. Geologica Acta. 2006;4(1–2):193–202.

71. Denyer P, Gazel E. The Costa Rican Jurassic to Miocene oceanic complexes: Origin, tectonics and relations. Journal of South American Earth Sciences. 2009;28(4):429–42.

72. Probst AJ, Moissl-Eichinger C. "Altiarchaeales": Uncultivated archaea from the subsurface. Life. 2015;5(2):1381–95.

73. McAllister SM, Moore RM, Gartman A, Luther GW, Emerson D, Chan CS. The Fe(II)-oxidizing Zetaproteobacteria: Historical, ecological and genomic perspectives. FEMS Microbiology Ecology. 2019;95(4):1–18.

74. Kim JH, Ferguson G, Person M, Jiang W, Lu ZT, Ritterbusch F, et al. Krypton-81 Dating Constrains Timing of Deep Groundwater Flow Activation. Geophysical Research Letters. 2022;49(11).

Chapter 2 Chemolithoautotroph distributions across the subsurface of a convergent margin

This chapter is a revised version of published work:

Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, Ramírez CJ, Barry PH, Yücel M, Selci M, Cordone A. "Chemolithoautotroph distributions across the subsurface of a convergent margin." The ISME Journal. 2023;17(1):140-50.

My primary contributions to this paper include: (i) Investigations and data acquisition, (ii) data synthesis and statistical analysis, (iii) figure and table generation, (iv) and writing the manuscript.

Abstract

Subducting oceanic crust releases fluids rich in biologically relevant compounds into the overriding plate, fueling subsurface chemolithoautotrophic ecosystems. To understand the impact of subsurface geochemistry on microbial communities, we collected fluid and sediments from 14 natural springs in a ~200 km transect across the Costa Rican convergent margin and performed shotgun metagenomics. The resulting 404 metagenome-assembled genomes (MAGs) cluster into geologically distinct regions based on MAG abundance patterns: outer forearc-only (25% of total relative abundance), forearc/arc-only (38% of total relative abundance), and delocalized (37% of total relative abundance) clusters. In the outer forearc, *Thermodesulfovibrionia*, *Candidatus* Bipolaricaulia, and Firmicutes have hydrogenotrophic sulfate reduction and Wood Ljungdahl (WL) carbon fixation pathways. In the forearc/arc, *Anaerolineae*, *Ca*. Bipolaricaulia, and *Thermodesulfovibrionia* have sulfur oxidation, nitrogen cycling, and microaerophilic respiration and WL, while *Aquificae* have aerobic sulfur oxidation and reverse tricarboxylic acid (rTCA) carbon fixation pathways. Canonical transformation-based correspondence analysis shows that MAG distribution corresponds to concentrations of aluminum, iron, nickel, dissolved inorganic carbon and phosphate. While delocalized MAGs appear surface-derived, the subsurface chemolithoautotrophic, metabolic, and taxonomic landscape varies by availability of mineral/metals and volcanically derived inorganic carbon. However, the WL pathway persists across all samples, suggesting that this versatile, energy-efficient carbon fixation pathway helps shape convergent margin subsurface ecosystems.

Introduction

Convergent margins are regions where two or more tectonic plates collide, and the denser "subducting slab" is pushed beneath the less dense overriding plate. As the slab descends, it devolatilizes under higher temperatures and pressures, allowing dissolved inorganic carbon (DIC) and redox active volatiles to cycle between the upper crust and Earth's mantle (1) (Fig. 2.1A). These volatiles are brought to the surface by hydrothermal fluids migrating through cracks and fissures in the upper mantle and crust, possibly stimulating microbial activity (2). In addition to providing volatiles, convergent margins offer a large potential space for a subsurface habitat since the 122°C sterilizing isotherm (Fig. 2.1A) can be many kilometers deep (3,4).

At the Costa Rican convergent margin, the Cocos Plate subducts beneath the Caribbean Plate, releasing fluids entrained with volatiles such as sulfur, iron, hydrogen, and DIC (5,6). These fluids fuel chemolithoautotrophy-based microbial ecosystems in the subsurface before they are expelled at the surface at hydrothermal seeps and springs (7). These chemolithoautotrophic ecosystems generate substantial amounts of organic matter with up to 24 mM dissolved organic carbon (DOC), that may sequester 1.4×10^9 to 1.4×10^{10} molC/year (ref 7). This is similar to the amount of carbon estimated to be sequestered by calcite deposition in the overriding plate (8). However, carbon sequestration by microbial autotrophy is linked to other elemental cycles in a different way than calcite precipitation is, since chemolithoautotrophs derive energy for carbon fixation by catalyzing a wide variety of exergonic redox reactions (9).

Using an approach that combines co-located metagenome sequences and geochemical data, we investigated whether chemolithoautotrophic organisms, their encoded carbon fixation pathways (CFP), and/or the types of redox reactions they potentially use to fuel to carbon fixation vary across the major geological provinces of the Costa Rican convergent margin (outer forearc, forearc, and volcanic arc). The outer forearc is the province closest to the trench where the down going oceanic plate subducts beneath the overriding plate. Here, the subducting slab is relatively shallow and seawater along with pore fluids and serpentinization products are released from subducted sediments (10–12). In the forearc, farther

inland from the trench, deeply derived volatiles that flux to the surface in natural seep fluids have likely undergone some hydrothermal alteration and mixing with mantle-derived fluids. The arc consists of volcanic edifices farther inland parallel to the trench (5) typically coinciding with the point where the slab is approximately 90-110 km deep. Here, slab fluids and melts interact with the mantle wedge to generate magmas (13,14).

Subsurface microorganisms in other convergent margins have been shown to perform carbon fixation as well as hydrogen, sulfur, and nitrogen cycling, but this has only been studied within single provinces of a convergent margin such as the backarc (15–17) or outer forearc (18). Landscape-scale (defined here as tens to hundreds of kilometers) distributions of chemolithoautotrophic metabolisms have been observed in intraplate natural springs in Yellowstone National Park, USA (19–21), and Tengchong, Yunnan, China (22), but not across a subducting margin. Here, we expand our understanding of the geobiological underpinnings of carbon release and sequestration across a convergent margin.

We report 404 metagenome-assembled genomes (MAGs) from 14 natural deeply sourced springs (27 samples total) distributed across a ~200 km transect of the Costa Rican convergent margin spanning the outer forearc, forearc, and arc provinces (Figs. 2.1B and S2.1). These springs are deeply sourced as their high helium-3/helium-4 values indicate substantial mantle contributions to all samples (8). A previous study of the unassembled metagenomes from these springs indicated the presence of genes related to the reductive acetyl-CoA pathway (Wood-Ljungdahl; WL), the reductive tricarboxylic acid cycle (rTCA), and the reductive pentose phosphate cycle (Calvin Benson-Bassham Cycle; CBB) for carbon fixation (7). However, unassembled metagenomic reads cannot be used to directly relate CFPs to the taxa that have them in their genomes, nor their redox-active metabolic pathways. Here, we show that subsurface- and surface-associated MAGs are distinguished by their distributions and metabolic pathways. Genes encoding for WL and rTCA are abundant across the margin but are encoded by MAGs with different taxonomic affiliations. In addition, these CFP are potentially powered by different redox metabolisms across different geological provinces. Together this suggests a strong geological control over the distribution of different metabolic strategies across the convergent margin landscape.

Methods

Sample collection, DNA extraction, and sequencing. Details on collection, DNA extraction, and sequencing of samples from actively venting fluids of 14 natural springs (Fig. 2.1B) have been described in detail previously (7,8).

Metagenomic assembly, binning, annotation, and metabolic predictions. Raw reads were trimmed and mate-paired by Trimmomatic (version 0.38 (23)) using default settings. Reads were then assembled de novo in metaSPAdes (24) to create site-specific assemblies with a minimum contig length of 1.5 kb. Assemblies were then used to construct 404 metagenomic assembled genomes (MAGs/bins) with 70% completeness and < 5% contamination (Table S2.1 (refs 25-30)) through the MetaWRAP pipeline (31). Prokka (32), FeGenie (33), GhostKOALA (34), and KEGGdecoder (35) were used for open reading frame prediction, gene annotation, KEGG orthology (KO) identifier linkage, and metabolic predictions. The genetic potential to use hydrogen, sulfur, nitrogen, and/or iron species as electron donors or accepters was assessed based on the presents of the enzymatic components detailed in the Supplementary Methods (Table S2.2). Elemental sulfur oxidation and sulfate reduction were determined by phylogenetic analyses of the dissimilatory (bi)sulfite reductase alpha subunit (*dsrA* (Fig. S2.2)). Additionally, a MAG was considered capable of autotrophy if 60% of the genes and all key enzymes for a CFP were present. As with *dsrA*, phylogenetic analysis was conducted on the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) large subunit to delineate the different RuBisCO forms of CBB (Fig. S2.3). In-depth details for metagenomic assembly, binning, annotation, and metabolic predictions are given in Supplementary Methods.

MAG phylogenetic and statistical analysis. The Genome Taxonomy Database Toolkit (GTDB-TK (36); version 1.4.0) was used for taxonomic identification and GTDB nomenclature has been used throughout this manuscript, although we refer to other commonly used names when it increases clarity. MAG abundance was quantified in units of GCPM (Genome Copy Per Million Reads). An in-depth description on how GCPM was calculated is given in Supplementary Methods. Abundance correlations across and within samples were determined using hierarchical clustering of Spearman rank correlation

distance matrices (Figs. 2.2 and S2.1). MAG abundance distributions across the convergent margin and their correlations with relevant environmental variables (7) were then determined via transformation based canonical correspondence analysis (tb-CCA). See Supplemental Methods for detailed information about the statistical analysis.

Results

General description of the MAGs. In total, 404 medium to high quality MAGs spanning 5 archaeal and 43 bacterial phyla were recovered from the 27 samples of hydrothermal fluids and sediment. The 372 bacterial and 32 archaeal MAGs are ≥70% complete with <5% multi-genomic contamination. Of these, 221 MAGs are >90% complete, with <5% contamination and 43 of these are high-quality based on MIMAGs criteria (25 (Table S2.1)). Most MAGs belong to lineages with no cultured relatives at the family level or higher (69% of archaea and 64% of bacteria). One archaeal MAG was classifiable only at the domain level.

Most MAGs (92%) have predictions for at least one metabolic pathway (Figs. 2.2 and S2.1). Of the MAGs with none of the genes of interest for the electron donors or acceptors, 13 are archaea (*Crenarchaeota* = 4, *Micrarchaeota* = 4, *Nanoarchaeota* = 4, and *Thermoplasmatota* =1) and 19 are bacteria (*Actinobacteriota* = 1, *Chloroflexota* =2, *Desulfobacterota* = 1, *Nitrospirota* = 1, *Omnitrophota* = 1, *Patescibacteria* = 8, and WOR-3 = 4). Neither complete methanogenic pathways nor the methyl coenzyme M reductase gene are present in any of the MAGs. Photosystems are present only in the four Cyanobacteria MAGs, accounting for 0.2% of total GCPM.

CBB, rTCA, and WL pathways are present in about one third of the MAGs (n=117, ~35% of total GCPM). The most abundant is WL (n=67, 19% of total GCPM), followed by CBB (n=42, 8% of total GCPM), and rTCA (n=12, 7% of total GCPM) (Fig. S2.4). Although three MAGs (bins 112, 253, and 91) contain ≥ 60% of the 3-hydroxypropionate/4-hydroxybutyrate cycle, none contain the key enzymes for this pathway (37, 38). Three *Anaerolineae* MAGs (bins 55, 251, and 386) contain both WL and CBB, and have ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO (for CBB)) and Carbon monoxide

dehydrogenase/acetyl-CoA synthase (CODH/ACS (for WL)). These genes are closely related to others from *Chloroflexi*, suggesting the co-occurrence of these pathways is not due to genomic contamination. Another MAG in the class *Syntrophia* (bin 351) contains genes for both WL and rTCA. Genes for the WL pathway are found in MAGs from 22 classes across 12 different phyla (*Actinobacteria*, *Bipolaricaulota*, *Chloroflexota*, *Deferrisomatota*, *Desulfobacterota*, Firmicutes, MBNT15, *Acidobacteriota*, *Coprothermobacterota*, *Nitrospirota*, *Omnitrophota*, and *Planctomycetota*). Genes for the rTCA cycle are found in MAGs from 4 classes across 4 phyla (*Aquificota*, *Chloroflexota*, *Desulfobacterota*, and *Nitrospirota*). Genes for the CBB cycle are found in MAGs from 8 classes across 6 phyla (*Chloroflexota*, *Proteobacteria*, *Actinobacteriota*, *Cyanobacteria*, *Deinococcota*, and *Firmicutes*), with RuBisCO Form I (25 MAGs), Form II (9 MAGs), both Form I and II (3 MAGs), or an unknown form (4 MAGs) (Fig. S2.3).

Together, WL-containing MAGs have the most diverse metabolic predictions, containing all the metabolic pathways that were identified in this study in some combination. Among the WL-containing MAGs, the most abundant MAGs have the genetic potential for the oxidation of hydrogen and sulfide, and some possess bd and aa3-type cytochromes. rTCA-containing MAGs have the genetic potential for every redox pathway except nitrogen cycling and iron (III) reduction. In this category, the most abundant MAGs have the genetic potential for sulfide and thiosulfate oxidation and the reduction of oxygen. CBB-containing MAGs have the genetic potential for every redox pathway except sulfate reduction. These MAGs have the genetic potential for hydrogen, sulfide, and thiosulfate oxidation and the reduction of oxygen (microaerophilic and/or aerobic) and nitrate.

Cluster and multivariate analysis. Hierarchical clustering of sites based on Spearman correlations of MAG abundance resulted in all the outer forearc sites clustering together and most of the forearc and arc sites clustering together (Figs. 2.2 and S2.1, horizontal sorting). The only exception is the sediment sample from Borinquen, as discussed below. Of the 10 sites where metagenomes were sequenced from both fluids and sediments, four have their fluid and sediment samples clustering together (El Tucano, Quebrada Naranja, Ranchero el Salitral, and Quepos). Four other sites in the forearc/arc (Blue River 1 and

2, Santa Lucia, and Finca Asociación Nacional de Educadores y Educadoras (ANDE)), have their fluid MAGs clustering together and their sediments clustering adjacent to the fluid MAGs (Figs. 2.2 and S2.1). This suggests that fluids and sediments from the same site are generally more similar to each other than to fluids and sediments in different pools, in agreement with 16S rRNA gene results (7). MAGs cluster into five groups, based on their distribution across the sites, with a linkage height of 10 (Figs. 2.2 and S2.1, vertical sorting). Cluster 1 MAGs are mainly in the forearc/arc, cluster 2 MAGs are abundant in the forearc/arc but are also common elsewhere, cluster 3 MAGs are mainly in the outer forearc, and clusters 4 and 5 MAGs are either not localized to any geologic province or are present in only a few sites with no clear distribution pattern across the convergent margin. Based on these clustering patterns, MAGs separate into three groups: two province-specific groups (cluster 1 from forearc/arc and cluster 3 from outer forearc) and delocalized (clusters 2, 4, and 5).

Metabolic and taxonomic descriptions of the clusters. *Outer forearc (Cluster 3).* Cluster 3 comprises 94 MAGs, 19 phyla, and 25% of total GCPM, almost exclusively from outer forearc sites. *Bacteroidota*, *Nitrospirota*, and *Chloroflexota* dominate numbers of MAGs (n = 21, 16, and 13 respectively), and 19%, 21%, and 9% of cluster 3 GCPM, respectively. The majority of cluster 3 MAGs (73%) are from uncultured families or higher taxonomic levels (Fig. S2.5). Cluster 3 MAGs metabolic pathway predictions are dominated by hydrogen (79% of cluster 3 GCPM) and sulfur (26% of cluster 3 GCPM) oxidation, as well as sulfate reduction (29% of cluster 3 GCPM), microaerophilic respiration (22% of cluster 3 GCPM) and nitrogen reduction (13% of cluster 3 GCPM) (Fig. 2.3). Genomic material from the top 10 classes is predominantly found in geothermal areas, aquifers, and other deep subsurface systems (Table S2.3).

WL is the most abundant and diverse CFP, encompassing 23% of MAGs and 29% of cluster 3 GCPM. *Thermodesulfovibrionia* MAGs dominate WL-containing MAGs and have the genetic potential for hydrogenotrophic sulfate reduction. A single MAG classified as Cat. *Bipolaricaulia* (bin 327) is the second most abundant WL-containing member which also has the genetic potential for hydrogenotrophic growth (39). Four MAGs classified as CSP1-3 (in phylum *Armatimonadetes*), the third most abundant

WL-containing group, have the genetic potential for sulfur oxidation (Fig. 2.4). Together, these cluster 3 WL-containing members encompass 11 MAGs and 19% of cluster 3 GCPM. In this cluster, genes for rTCA and CBB are in lower abundance (5% each of cluster 3 GCPM). *Syntrophia* MAGs dominate the rTCA-containing MAGs and have the genetic potential for sulfate reduction (Fig. 2.4). Most CBB-containing MAGs in cluster 3 classify as *Gammaproteobacteria* and encode RuBisCO Form IA. This form is the primary form in chemolithoautotrophs from a diversity of subsurface ecosystems (40). These MAGs have the genetic potential for the sulfide oxidation and reduction of nitrate/nitrite (41) and/or oxygen (microaerophilic) (Fig. 2.4).

Forearc/arc (Cluster 1). Cluster 1 comprises 102 MAGs, 34 phyla, and 38% of total GCPM, mainly in forearc/arc sites. *Chloroflexota, Acidobacteriota* and *Aquificota* dominate this cluster (~15%, 14%, and 12% of cluster 1 GCPM, respectively). Chloroflexota MAGs are the most diverse of these (n=15), then *Acidobacteriota* MAGs (n=8), while there is only one MAG classified as *Aquificota* (bin 303, *Sulfurihydrogenibium* sp.). This cluster has most of the archaeal MAGs (n=19) and 88% of the MAGs are taxonomically identified as uncultured families or higher taxonomic levels (Fig. S2.5). The cluster contains MAGs with many redox metabolisms, predominantly nitrogen (60% of cluster 1 GCPM) and oxygen reduction (microaerophilic = 32% and aerobic = 26% of cluster 1 GCPM), as well as sulfur (39% of cluster 1 GCPM) and hydrogen (16% of cluster 1 GCPM) oxidation. Forearc sites contain the only MAGs capable of ammonia oxidation (n=4; 5% of cluster 1 GCPM) within the archaea class *Nitrososphaeria* (Figs. 2.2 and 2.3). DNA sequences from previously sequenced representatives of the top 10 classes are predominantly found in geothermal areas, aquifers, and other deep subsurface systems as well as soils (Table S2.3).

As with cluster 3, WL is the most abundant CFP in cluster 1 (n=24, 25% of cluster 1 GCPM), and is mostly found in *Anaerolineae* MAGs with the genetic potential for S-dependent reduction of nitrate and/or nitrite. *Ca.*Bipolaricaulia MAGs, the second most abundant group of WL-containing MAGs, have the genetic potential for hydrogen oxidation and the reduction of nitrogen species. All *Thermodesulfovibrionia* MAGs, the next most abundant group of WL-containing MAGs, contain all

genes for sulfate reduction (*dsrAB*, *aprAB*, and *sat*). Together, the WL-containing members of *Anaerolineae*, Ca. *Bipolaricaulia*, and *Thermodesulfovibrionia* account for 16 MAGs and 20% of cluster 1 GCPM. A single *Sulfurihydrogenibium* MAG accounts for the prominence of rTCA as the second most abundant CFP (12% of cluster 1 GCPM). This MAG has the genetic potential for sulfur oxidation and microaerophilic reduction of oxygen (Fig. 2.4), similar to cultured relatives (42,43). It recruited no reads from the outer forearc and <0.05% total GCPM found in the arc sites, suggesting that it is specialized to the forearc (Fig. S2.1). CBB is present in 5 MAGs from the *Deinococci*, *Anaerolineae*, UBA4738 (in phylum *Actinobacteriota*), and *Cyanobacteria* classes with only 8% of cluster 1 GCPM. The *Deinococci* and *Anaerolineae* MAGs contain the genetic potential to oxidize sulfur and reduce nitrogen species (Fig. 2.4).

Delocalized (Clusters 2, 4, and 5). MAGs in clusters 2 (78 MAGs, 24% of total GCPM), 4 (18 MAGs, 1% of total GCPM), and 5 (112 MAGs, 12% of total GCPM) are common across multiple provinces. Cluster 2 is more concentrated in the outer forearc and forearc, but these MAGs are much more common in other provinces than those of cluster 1, and they share the characteristics discussed below with clusters 4 and 5, so we include them in the delocalized cluster. The top 10 classes in the delocalized clusters are commonly found in humans, animals, and/or plants (44); human-influenced environments (45); soil (46-51); marine/freshwater (47,48,50,52-57); and hot spring microbial mats (58). These clusters contained three of the only four *Cyanobacteria* MAGs in the whole dataset, all of the common lab contaminant genera (*Sphingomonas*, *Pelomonas*, and *Aquabacterium*) in the whole dataset (59), and high abundances of human commensals (*Chromobacterium haemolyticum*, *Acinetobacter junii*, and *Acinetobacter indicus* (60,61)) in hot spring spas at resort sites (Blue River and Recreo Verde). Borinquen is the only site dominated by MAGs from clusters 4 and 5. This could be because the low biomass at Borinquen allowed laboratory and human contaminants to comprise a larger proportion of the extracted and amplified DNA, causing it to cluster away from other forearc/arc sites. Most delocalized MAGs have the genetic potential for sulfur and/or hydrogen oxidation as well as oxygen and/or nitrate reduction. CBB dominates the CFPs

in these clusters, mostly with Form I RuBisCO, which is typically found in organisms that respire atmospheric levels of oxygen (Fig. S2.1).

Transformation-based canonical correspondence analysis. The tb-CCA of MAGs from province-specific clusters (clusters 1 and 3) explains 36.3% of the variation in the samples and recovers the same province-specific clusters predicted in Spearman-based hierarchical clusters. The important variables are, in order of most marginal effect significance: concentrations of sediment derived aluminum, aqueous iron concentrations, DIC, temperature, aqueous phosphate concentrations, and aqueous nickel concentrations (Fig. 2.5 and Table S2.4). Based on strong correlations, the concentration of aqueous iron is also a proxy for aqueous zinc concentrations. MAGs at outer forearc sites correlate with sediment associated aluminum; MAGs at forearc sites correlate with DIC, phosphate concentrations, and to a lesser extent aqueous iron and nickel concentrations, while arc sites are tightly associated with iron and nickel concentrations. Distribution of WL-containing MAGs is well-explained by concentrations of aluminum, nickel, and iron, whereas rTCA- and CBB-containing MAGs are more influenced by DIC and phosphate concentrations. Heterotrophs are distributed alongside autotrophs, with the highly abundant heterotrophs co-occurring with rTCA- and CBB-containing MAGs. For MAGs in the delocalized clusters, a tb-CCA plot explains only 14.2% of the between-site variation, and results in no discernable clustering (Fig. S2.6), suggesting that local geochemical variables are not major determinants of the distribution of the delocalized MAGs.

Discussion

Each site contains MAGs with multiple carbon fixation and metabolic pathway predictions, suggesting taxonomically diverse populations with carbon fixation and metabolic pathways that differ systematically across the three geological provinces (outer forearc, forearc, and arc) (Fig. 2.6). Approximately one-third of the metagenomic reads map to MAGs containing CFPs, consistent with the proportion of chemolithoautotrophs found in other subsurface lithoautotrophic ecosystems (27,62,63). Even though temperatures are low enough for photosynthesis (< 70°C at >90% of sites), only four MAGs contain

photosystems, and all are *Cyanobacteria* (<1% total GCPM). Only one was found in cluster 1 (bin 379) (0.04% of total GCPM) while the other three were found in the delocalized group. These findings support the conclusions reached by Barry et al. (2019) and Fullerton et al. (2021) that chemolithoautotrophy is the main driver of primary productivity in the subsurface of these natural spring communities.

While most MAGs (63% of total GCPM) separate by Spearman-based hierarchical clustering into those abundant in outer forearc sites and those abundant in forearc/arc sites (Fig. 2.2 and S2.1), forearc and arc sites only became differentiated from each other in the tb-CCA analysis (Fig. 2.5). The tb-CCA analysis shows separation of sites by province (outer forearc, forearc, and arc), based on relationships between their associated MAGs' correlations with metals, slab-derived carbon, and phosphate concentrations as well as temperature. Differences in community composition are significantly correlated with changes in concentrations of aqueous iron, nickel (with zinc by proxy), and sediment associated aluminum ($p < 0.01$). Correlations with aqueous iron and nickel concentrations are strongest in the more acidic forearc/arc samples where reactions between the magmatic fluid and the underlying mafic rocks lead to higher concentrations of dissolved iron and iron-sulfides, also yielding pyrite framboids (7). In particular, MAGs from Blue River (arc) and Finca ANDE (forearc) have the greatest association with aqueous iron and nickel concentrations, likely reflecting the greater influence of metals from the large hydrothermal system.

MAGs in outer forearc sites correlate with lower concentrations of aqueous phosphate and higher concentrations of sediment associated aluminum. These fluids contain the lowest concentrations of aqueous phosphate (<.001 mmol/L) and are partially equilibrated feldspars and clays (7), likely derived from the dehydration of mineral-bound water in the slab material 12-15 km below the surface (64,65). The isostructural similarity of aluminum phosphate with polymorphs of silica allows for extensive aluminum phosphate substitution of silica in feldspar (66,67). Microorganisms in soil and environments low in aqueous phosphate can liberate this phosphate through biological weathering of feldspar (68,69), increasing the concentration of solid and aqueous aluminum species (70,71). Therefore, the correlation of outer forearc MAGs with aluminum concentrations may reflect an increase in the liberation of phosphate

from feldspars. Relative to the other provinces, the outer forearc has a higher abundance of *Thermodesulfovibrionia*, *Actinobacteriota*, *Desulfobacterota*, and *Gammaproteobacteria*; the forearc has a higher abundance of *Sulfurihydrogenibium* sp., *Thaumarchaeota*, and *Acidobacteriota*; and the arc has a higher abundance of Firmicutes. *Chloroflexi* and Ca. *Bipolaricaulota* are abundant everywhere, although their individual MAGs differ between provinces.

Distribution of carbon fixation pathways across geological provinces. WL is the dominant CFP across the outer forearc, forearc and arc, while rTCA is abundant only in the forearc, and CBB is in very low abundance throughout the provinces (Figs. 2.2, S2.1, and S2.4). The WL pathway is highly oxygen-sensitive, the rTCA cycle is moderately oxygen-sensitive, and both tolerate high temperatures (38,72), making them suitable for subsurface hydrothermal systems. In addition to being more abundant and widespread, WL-containing MAGs are from more diverse taxa and contain a wider range of redox pathway predictions than rTCA and CBB. Since WL has the lowest energy requirements of all known CFPs (38), it may provide autotrophs with a selective advantage. rTCA is abundant only in the forearc, mostly in *Sulfurihydrogenibium* sp. The four MAGs (class *Anaerolineae*: bins 55, 251, & 386 and class *Syntrophia*: bin 351) that contain genes encoding two CFPs each may be adapted to variable carbon dioxide and oxygen concentrations (27,38,73).

WL-containing MAGs contribute most to the correlations between the microbial composition of the sites and higher iron and nickel concentrations (Fig. 2.5). WL has high metal requirements, especially iron and nickel, since nickel is required for CODH and iron is required for ferredoxin (38). In addition, higher iron, nickel, and zinc concentrations are proxies for more reducing, low pH, and high temperature environments, which would be favored by WL-containing organisms. Abundances of CBB- and rTCA-containing MAGs, on the other hand, are more associated with DIC and phosphate in the arc and forearc, suggesting that the availability of carbon and phosphorous may be limiting for these chemolithoautotrophic communities, rather than pH, redox metabolic elements, or other nutrients. This is supported by a positive correlation between DOC and DIC concentrations across these sites, suggesting that more DOC can be produced when more DIC is available (7). DIC in this system is almost entirely

derived from the deep slab/mantle mixture and it increases in concentration at the arc due to higher volcanic inputs and less removal of carbon (through the formation of calcite) in the crust (8). The low abundant rTCA-containing autotrophs contain a low carbon dioxide affinity pyruvate synthase (74) that may be disadvantageous in the outer forearc where alkaline conditions favor calcite precipitation and low carbon dioxide concentrations.

Outer forearc WL-containing MAGs have a higher association with aluminum and a lower association with nickel and iron than their counterparts in the arc and forearc. In the outer forearc, iron and nickel concentrations remain constant through the sites (0.7 mg/g and 0.8 µmol/L respectively), except for the Quepos samples (1.83 mg/g and 2.5 µmol/L respectively). However, as mentioned above, aqueous phosphate here is limited. Nutrient-limited environments, such as those limited in phosphate (71), have been shown to encourage lithoautotrophic communities to enhance the weathering of minerals (70,71,75,76), such as feldspar, leaving behind an aluminum rich residual (70,71). The distribution of heterotrophic MAGs is similar to that of autotrophic MAGs, suggesting that they are either similarly influenced by the geochemical variables, or that autotroph distribution drives heterotroph distribution.

Metabolic potential in the outer forearc. In the outer forearc, hydrogenotrophic sulfate reduction is the most abundant predicted metabolic pathway and occurs in the widest range of MAGs, whether or not they have CFPs. In the outer forearc, dominant WL-containing MAGs classifying as *Thermodesulfovibrionia* and Ca. *Bipolaricaulia* have the genetic potential for hydrogen oxidation and *Thermodesulfovibrionia* MAGs also have genes for sulfate reduction, in agreement with other studies of this strictly anaerobic class (77). Two of the four dominant rTCA-containing MAGs classify as *Syntrophia* and contain hydrogenases that have been demonstrated to perform hydrogen uptake (supplemental methods), while all four contain *dsrAB* genes for sulfate reduction. The genetic propensity to perform carbon fixation via hydrogenotrophic sulfate reduction and hydrogen oxidation has been found in subsurface organisms associated with analogous submarine serpentinization sites (e.g., the Lost City marine hydrothermal field) terrestrial hydrothermal systems (78,79), and a ~3 km deep borehole in South Africa (80).

Serpentinization-sourced hydrogen and low dissolved oxygen concentrations in the outer forearc create amenable conditions for WL and rTCA pathways, hydrogen oxidation, and sulfate reduction. These results contrast sharply with the high prevalence of methanogens, absence of sulfate-reducers, and evidence for very limited biogenic sulfate reduction just to the north in the Santa Elena serpentinizing system (81-83). The Santa Elena system is ophiolite-hosted (82), whereas outer forearc sites in the Nicoya peninsula are more influenced by the mantle wedge (84). Santa Elena also has lower sulfate concentrations, which may allow methanogens to compete for electron donors (82). This suggests that differences in underlying geology, or possibly different degrees of inmixing of subducted sulfate-rich seawater, can drive differences in microbial metabolic pathways, even though both systems are serpentinizing and have similar pH and temperatures.

Metabolic potential in the forearc and arc. The most abundant and widespread predicted metabolic pathways in the forearc are sulfur oxidation and the reduction of oxygen and oxidized nitrogen compounds. Arc MAGs have fewer genes encoding for fully aerobic oxygen reduction (i.e. aa3-and/or bo-type cytochromes) and lack ammonium oxidation genes. Acidic water-rock interactions in the arc and forearc enable higher concentrations of aqueous carbon dioxide, which may alleviate DIC limitation in these chemosynthetic communities. Additionally, the high pressure of deep magmatic degassing, cooling as these gases rise, and gas-fluid-rock interactions in local hydrothermal systems favor the more reduced sulfide over oxidized sulfate (85,86) creating the ideal conditions for sulfide oxidation with high carbon dioxide availability as an autotrophic substrate. In the forearc and arc, the most abundant WL-containing MAGs (class *Anaerolineae*) and the lower abundant CBB-containing MAGs (classes *Deinococci* and *Anaerolineae*) have the potential to power carbon fixation with sulfur-dependent denitrification. The second most abundant WL-MAGs (class *Ca*. Bipolaricaulia) have the genetic potential to power carbon fixation with hydrogen oxidation using nitrate/nitrite as terminal electron acceptors. In the forearc, the rTCA-containing MAG (genus *Sulfurihydrogenibium* sp.) has the genetic potential for aerobic oxidation of sulfide and thiosulfate. While sulfide oxidation is the most abundant energy pathway in the arc and forearc, the resulting oxidized sulfate can be used by the third most abundant WL-containing MAGs

(class *Thermodesulfovibrionia*) to power carbon fixation with sulfate reduction. Although CBB was less prevalent, the CBB-containing MAGs (class *Deinococci*) contains form IC of RuBisCO, which may be advantageous in medium to high carbon dioxide suboxic environments (87). MAGs classified as *Anaerolineae* in the forearc/arc contain an unknown form of RuBisCO that may be the recently discovered Form I' or Form I-α RuBisCO which have been proposed to be the ancestral form of all RuBisCO forms (88).

As distance increases from the trench toward the arc, the geothermal gradient to the 122°C isotherm steepens (3), potentially compressing the overall volume of available subsurface habitat. The greater presence of respiratory pathways in the forearc and arc may therefore reflect their greater proximity to atmospheric recharge and faster recharge with oxygenated meteoric water driven by the steep geothermal gradient and high rainfall in these mountainous areas.

Delocalized clusters are surface-associated. We explored the possibility that the delocalized MAGs in clusters 2, 4, and 5, constitute a low abundance surface-associated community. The tb-CCA analysis showed that the distribution of the delocalized MAGs across sites is poorly explained by local geochemical variables (Fig. S2.6) and their ten most abundant classes are common soils or animal inhabitants. The delocalized clusters also contain the only MAGs from the genera *Aquabacterium*, *Alishewanella*, *Rheinheimera*, and *Exiguobacterium* whose 16S rRNA genes were previously found to lack correlation with geological parameters (7). Most MAGs in clusters 2, 4, and 5 (51%) are in cultured family to species (Fig. S2.5), a feature common in human-influenced environments (89). The presence of common soil inhabitants and human commensals, prevalence of fully aerobic metabolisms including CBB, and lack of correlations with local geochemical and geological parameters suggests that the delocalized clusters are composed mostly of surface-associated MAGs and a few laboratory/in situ contaminants.

Conclusion

We conclude that the distribution of province-specific MAGs across the subsurface of the Costa Rican convergent margin correlates with aluminum, iron, nickel, DIC, and phosphate concentrations, which reflect the underlying geology, fluid sources, and fluid-rock interactions. Hydrogenotrophic sulfate reduction dominates the outer forearc, whereas more oxidized substrates such as nitrate and oxygen can oxidize sulfide in the forearc and arc. This suggests that, unlike individual volcanic systems (i.e., Yellowstone National Park (19,90)), pH and temperature are not the driving variables that control the microbial distribution at a landscape scale in a convergent margin. MAGs with chemolithoautotrophic pathways comprise a third of this subsurface ecosystem. The WL pathway is found in a wide diversity of taxa, which collectively also contain genes encoding for all available metabolic pathways, in every part of the subsurface landscape, with metal availability associated distributions. rTCA, CBB, and sulfur oxidation are more prominent in the forearc and arc which has high carbon dioxide concentrations. We conclude that these subsurface chemolithoautotrophic ecosystems support a diverse set of metabolic processes and associated autotrophic pathways, tightly connected to deep subsurface geological and geochemical parameters such as dissolved metal content and the degree to which deep volcanic inorganic carbon and other volatile components are available for microbial usage.

Acknowledgements

The authors acknowledge the Biology Meets Subduction Project, funded by the Alfred P. Sloan Foundation and the Deep Carbon Observatory (G-2016-7206) to P.H.B, J.M.d.M, D.G., and K.G.L., with DNA sequencing from the Census of Deep Life. Additional support came from NSF FRES (Award# 21211637) to P.H.B., J.M.d.M and K.G.L. U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomic Science Program (DE-SC0020369 to K.G.L) and FONDECYT Grant 11191138 (ANID Chile) and COPAS COASTAL ANID FB210021 to G.L.J. D.G. A.C. and M.S. were partially supported by funding from the European Research Council (ERC) under the European Union‘s Horizon 2020 research and innovation program Grant Agreement No. 948972—COEVOLVE—ERC-2020-STG.

Data Availability

Metagenomic data are available in the NCBI SRA with project ID PRJNA627197. All MAGs analyzed in this study are also included under the same BioProject. BioSample accessions for each MAG are included in Table S2.1. The full environmental dataset along with the R scripts for this analysis is available at https://github.com/TJrogers86/Costa-Rica-2017.git and released as a permanent version (v1.0.0) using Zenodo under https://doi.org/10.5281/zenodo.6259294 .

References

1. Kelemen PB, Manning CE. Reevaluating carbon fluxes in subduction zones, what goes down, mostly comes up. Proc Natl Acad Sci U S A. 2015; 112:E3997-4006.
2. Vitale Brovarone A, Sverjensky DA, Piccoli F, Ressico F, Giovannelli D, Daniel I. Subduction hides high-pressure sources of energy that may feed the deep subsurface biosphere. Nat Commun. 2020; 11:1-1.
3. Harris RN, Wang K. Thermal models of the middle America trench at the Nicoya Peninsula, Costa Rica. Geophys Res Lett. 2002; 29:6-1.
4. Plümper O, King HE, Geisler T, Liu Y, Pabst S, Savov IP, et al. Subduction zone forearc serpentinites as incubators for deep microbial life. Proc Natl Acad Sci U S A. 2017; 114:4324-9.
5. Lee H, Fischer TP, de Moor JM, Sharp ZD, Takahata N, Sano Y. Nitrogen recycling at the Costa Rican subduction zone: The role of incoming plate structure. Sci Rep. 2017; 7:1-10.
6. Stern RJ. Subduction zones. Rev Geophys. 2002; 40:3-38.
7. Fullerton KM, Schrenk MO, Yücel M, Manini E, Basili M, Rogers TJ, et al. Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. Nat Geosci. 2021; 14:301-6.
8. Barry PH, de Moor JM, Giovannelli D, Schrenk M, Hummer DR, Lopez T, et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. Nature. 2019; 568:487-92.
9. Moore EK, Jelen BI, Giovannelli D, Raanan H, Falkowski PG. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. Nat Geosci. 2017; 10:629-36.
10. Barnes JD, Cullen J, Barker S, Agostini S, Penniston-Dorland S, Lassiter JC, et al. The role of the upper plate in controlling fluid-mobile element (Cl, Li, B) cycling through subduction zones: Hikurangi forearc, New Zealand. Geosphere. 2019; 15:642-58.
11. Clift P, Vannucchi P. Controls on tectonic accretion versus erosion in subduction zones: Implications for the origin and recycling of the continental crust. Rev Geophys. 2004; 42:1-31.
12. Rüpke LH, Morgan JP, Hort M, Connolly JA. Serpentine and the subduction zone water cycle.

Earth Planet Sci Lett. 2004; 223:17-34.

13. Carr MJ, Feigenson MD, Bennett EA. Incompatible element and isotopic evidence for tectonic control of source mixing and melt extraction along the Central American arc. Contrib Mineral Petrol. 1990; 105:369-80.

14. Gazel E, Carr MJ, Hoernle K, Feigenson MD, Szymanski D, Hauff F, et al. Galapagos-OIB signature in southern Central America: Mantle refertilization by arc–hot spot interaction. Geochem Geophys Geosyst. 2009; 10:1-32.

15. Trembath-Reichert E, Butterfield DA, Huber JA. Active subseafloor microbial communities from Mariana back-arc venting fluids share metabolic strategies across different thermal niches and taxa. ISME J. 2019; https://doi.org/10.1038/s41396-019-0431-y.

16. Power JF, Carere CR, Lee CK, Wakerley GL, Evans DW, Button M, et al. Microbial biogeography of 925 geothermal springs in New Zealand. Nat Commun. 2018; 9:1-2.

17. Acocella V, Spinks K, Cole J, Nicol A. Oblique back arc rifting of Taupo Volcanic zone, New Zealand. Tectonics. 2003; 22:1-18.

18. Curtis AC, Wheat CG, Fryer P, Moyer CL. Mariana forearc serpentinite mud volcanoes harbor novel communities of extremophilic archaea. Geomicrobiol J. 2013; 30:430-41.

19. Inskeep WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. Front Microbiol. 2013; 4:1-21.

20. Colman DR, Lindsay MR, Amenabar MJ, Boyd ES. The intersection of geology, geochemistry, and microbiology in continental hydrothermal systems. Astrobiology. 2019; 19:1505-22.

21. Inskeep WP, Jay ZJ, Tringe SG, Herrgård MJ, Rusch DB, YNP Metagenome Project Steering Committee and Working Group Members. The YNP metagenome project: environmental parameters responsible for microbial distribution in the Yellowstone geothermal ecosystem. Front Microbiol. 2013; 4:1-15.

22. Hou W, Wang S, Dong H, Jiang H, Briggs BR, Peacock JP, et al. A comprehensive census of

microbial diversity in hot springs of Tengchong, Yunnan Province China using 16S rRNA gene pyrosequencing. PloS One. 2013; 8:1-15.

23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114-20.

24. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017; 27:824-34.

25. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TB, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017; 35:725-31.

26. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015; 25:1043-55.

27. Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO_2 concentrations. Environ Microbiol. 2017; 19:459-74.

28. Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CM, Emerson JB, et al. Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface. Nat Microbiol. 2018; 3:328-36.

29. He C, Keren R, Whittaker M, Farag IF, Doudna J, Cate JH, et al. Huge and variable diversity of episymbiotic CPR bacteria and DPANN archaea in groundwater ecosystems. *bioRxiv*. 2020; https://doi.org/10.1101/2020.05.14.094862.

30. Grettenberger CL, Hamilton TL. Metagenome-Assembled Genomes of Novel Taxa from an Acid Mine Drainage Environment. Appl Environ Microbiol. 2021; https://doi.org/10.1101/2020.07.02.185728.

31. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018; 6:1-3.

32. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014; 30:2068-9.

33. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol. 2020; https://doi.org/10.3389/fmicb.2020.00037.

34. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016; 428:726-31.

35. Graham ED, Heidelberg JF, Tully BJ. Potential for primary productivity in a globally distributed bacterial phototroph. ISME J. 2018; 12:1861–6.

36. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020; 36:1925-27.

37. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, et al. Autotrophic carbon fixation in archaea. Nat Rev Microbiol. 2010; 8:447-60.

38. Berg IA. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways. Appl Environ Microbiol. 2011; 77:1925-36.

39. Youssef NH, Farag IF, Hahn CR, Jarett J, Becraft E, Eloe-Fadrosh E, et al. Genomic characterization of candidate division LCP-89 reveals an atypical cell wall structure, microcompartment production, and dual respiratory and fermentative capacities. Appl Environ Microbiol. 2019; 85:1-19.

40. Nigro LM, King GM. Disparate distributions of chemolithotrophs containing form IA or IC large subunit genes for ribulose-1, 5-bisphosphate carboxylase/oxygenase in intertidal marine and littoral lake sediments. FEMS Microbiol Ecol. 2007; 60:113-25.

41. Aminuddin M, Nicholas DJ. Electron transfer during sulphide and sulphite oxidation in *Thiobacillus denitrificans*. Microbiology. 1974; 82:115-23.

42. Giovannelli D, Sievert SM, Hügler M, Markert S, Becher D, Schweder T, et al. Insight into the evolution of microbial metabolism from the deep-branching bacterium, *Thermovibrio ammonificans*. eLife. 2017; 6:1–31.

43. Nakagawa S, Shataih Z, Banta A, Beveridge TJ, Sako Y, Reysenbach AL. *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum*. Int J Syst Evol Microbiol. 2005; 55:2263–8.

44. Leclerque A, Kleespies RG. A *Rickettsiella* bacterium from the hard tick, *Ixodes woodi*: molecular taxonomy combining multilocus sequence typing (MLST) with significance testing. PloS One. 2012; https://doi.org/10.1371/journal.pone.0038062.

45. Quatrini R, Johnson DB. *Acidithiobacillus ferrooxidans*. Trends Microbiol. 2019; 27:282-3.

46. Spang A, Poehlein A, Offre P, Zumbrägel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing *Candidatus* Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. Environ Microbiol. 2012; 14:3122-45.

47. Chen CY, Chen PC, Weng FC, Shaw GT, Wang D. Habitat and indigenous gut microbes contribute to the plasticity of gut microbiome in oriental river prawn during rapid environmental change. PloS One. 2017; https://doi.org/10.1371/journal.pone.0181427.

48. Garcia R, Müller R. The family *Myxococcaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F, (eds). The prokaryotes: *Deltaproteobacteria* and *Epsilonproteobacteria*. Berlin: Springer, 2014. pp 191-212.

49. Garcia R, Müller R. *Simulacricoccus ruber* gen. nov., sp. nov., a microaerotolerant, non-fruiting, myxospore-forming soil myxobacterium and emended description of the family *Myxococcaceae*. Int J Syst Evol Microbiol. 2018; 68:3101-10.

50. Iino T. The family *Ignavibacteriaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). The Prokaryotes: Other Major Lineages of Bacteria and the Archaea. Springer Science + Business Media, 2014. pp 701-3.

51. Petrie L, North NN, Dollhopf SL, Balkwill DL, Kostka JE. Enumeration and characterization of iron (III)-reducing microbial communities from acidic subsurface sediments contaminated with

uranium (VI). Appl Environ Microbiol. 2003; 69:7467-79.

52. Fincker M, Huber JA, Orphan VJ, Rappé MS, Teske A, Spormann AM. Metabolic strategies of marine subseafloor *Chloroflexi* inferred from genome reconstructions. Environ Microbiol. 2020; 22:3188-204.

53. Chen YJ, Leung PM, Wood JL, Bay SK, Hugenholtz P, Kessler AJ, et al. Metabolic flexibility allows bacterial habitat generalists to become dominant in a frequently disturbed ecosystem. ISME J. 2021; 15:2986-3004.

54. Flieder M, Buongiorno J, Herbold CW, Hausmann B, Rattei T, Lloyd KG, et al. Novel taxa of *Acidobacteriota* implicated in seafloor sulfur cycling. ISME J. 2021; 15:3159-80.

55. Kim M, Wilpiszeski RL, Wells M, Wymore AM, Gionfriddo CM, Brooks SC, et al. Metagenome-Assembled Genome Sequences of Novel Prokaryotic Species from the Mercury-Contaminated East Fork Poplar Creek, Oak Ridge, Tennessee, USA. Microbiol Resour Announc. 2021; https://doi.org/10.1128/MRA.00153-21.

56. Santos-Júnior CD, Logares R, Henrique-Silva F. Microbial population genomes from the Amazon River reveal possible modulation of the organic matter degradation process in tropical freshwaters. Mol Ecol. 2022; 31:206-19.

57. Youssef NH, Farag IF, Hahn CR, Jarett J, Becraft E, Eloe-Fadrosh E, et al. Genomic characterization of candidate division LCP-89 reveals an atypical cell wall structure, microcompartment production, and dual respiratory and fermentative capacities. Appl Environ Microbiol. 2019; 85:1-19.

58. Yamada T, Sekiguchi Y. Cultivation of uncultured *Chloroflexi* subphyla: significance and ecophysiology of formerly uncultured *Chloroflexi* 'subphylum I' with natural and biotechnological relevance. Microbes Environ. 2009; 24:205-16.

59. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, Schrenk MO, et al. Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. Front Microbiol. 2018; https://doi.org/10.3389/fmicb.2018.00840.

60. Doughari HJ, Ndakidemi PA, Human IS, Benade S. The Ecology, Biology and Pathogenesis of *Acinetobacter* spp.: An Overview. Microbes Environ. 2011; 26:101-12.

61. Han XY, Han FS, Segal J. *Chromobacterium haemolyticum* sp. nov., a strongly haemolytic species. Int J Syst Evol Microbiol. 2008; 58:1398-403.

62. Lau MC, Kieft TL, Kuloyo O, Linage-Alvarez B, Van Heerden E, Lindsay MR, et al. An oligotrophic deep-subsurface community dependent on syntrophy is dominated by sulfur-driven autotrophic denitrifiers. Proc Natl Acad Sci U S A. 2016; 113:E7927-36.

63. Momper L, Jungbluth SP, Lee MD, Amend JP. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. ISME J. 2017; 11:2319-33.

64. Worzewski T, Jegen M, Kopp H, Brasse H, Taylor Castillo W. Magnetotelluric image of the fluid cycle in the Costa Rican subduction zone. Nat Geosci. 2011; 4:108-11.

65. Hensen C, Wallmann K, Schmidt M, Ranero CR, Suess E. Fluid expulsion related to mud extrusion off Costa Rica—A window to the subducting slab. Geology. 2004; 32:201-4.

66. Simpson DR. Aluminum phosphate variants of feldspar. Am Mineral. 1977; 62:351-5.

67. London DA, Cerny P, Loomis J, Pan JJ. Phosphorus in alkali feldspars of rare-element granitic pegmatites. Can Mineral. 1990; 28:771-86.

68. Petrillo C, Castaldi S, Lanzilli M, Selci M, Cordone A, Giovannelli D, et al. Genomic and physiological characterization of *Bacilli* isolated from salt-pans with plant growth promoting features. Front Microbiol. 2021; https://doi.org/10.3389/fmicb.2021.715678.

69. Ghiorse WC, Wilson JT. Microbial ecology of the terrestrial subsurface. Adv Appl Microbiol. 1988; 33:107-72.

70. Barker WW, Welch SA, Chu S, Banfield JF. Experimental observations of the effects of bacteria on aluminosilicate weathering. Am Mineral. 1998; 83:1551-63.

71. Bennett PC, Rogers JR, Choi WJ, Hiebert FK. Silicates, silicate weathering, and microbial ecology. Geomicrobiol J. 2001; 18:3-19.

72. Hügler M, Sievert SM. Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. Ann

Rev Mar Sci. 2011; 3:261-89.

73. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, et al. Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science. 2007; 315:247-50.

74. Bar-Even A, Noor E, Milo R. A survey of carbon fixation pathways through a quantitative lens. J Exp Bot. 2012; 63:2325-42.

75. Stevens TO, McKinley JP. Lithoautotrophic microbial ecosystems in deep basalt aquifers. Science. 1995; 270:450-5.

76. Barker WW, Welch SA, Banfield JF. Biogeochemical weathering of silicate minerals. Rev Mineral Geochem. 1997; 35:391-428.

77. Frank YA, Kadnikov VV, Lukina AP, Banks D, Beletsky AV, Mardanov AV, et al. Characterization and genome analysis of the first facultatively alkaliphilic *Thermodesulfovibrio* isolated from the deep terrestrial subsurface. Front Microbiol. 2016; https://doi.org/10.3389/fmicb.2016.02000.

78. Woycheese KM, Meyer-Dombard DA, Cardace D, Argayosa AM, Arcilla CA. Out of the dark: transitional subsurface-to-surface microbial diversity in a terrestrial serpentinizing seep (Manleluag, Pangasinan, the Philippines). Front Microbiol. 2015; 6:1-12.

79. Brazelton WJ, Morrill PL, Szponar N, Schrenk MO. Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. Appl Environ Microbiol. 2013; 79:3906-16.

80. Moser DP, Gihring TM, Brockman FJ, Fredrickson JK, Balkwill DL, Dollhopf ME, et al. *Desulfotomaculum* and *Methanobacterium* spp. dominate a 4-to 5-kilometer-deep fault. Appl Environ Microbiol. 2005; 71:8773-83.

81. Schwarzenbach EM, Gill BC, Gazel E, Madrigal P. Sulfur and carbon geochemistry of the Santa Elena peridotites: Comparing oceanic and continental processes during peridotite alteration. Lithos. 2016; 252:92-108.

82. Sánchez-Murillo R, Gazel E, Schwarzenbach EM, Crespo-Medina M, Schrenk MO, Boll J, et al.

Geochemical evidence for active tropical serpentinization in the Santa Elena Ophiolite, Costa Rica: An analog of a humid early Earth?. Geochem Geophys Geosyst. 2014; 15:1783-800.

83. Crespo-Medina M, Twing KI, Sánchez-Murillo R, Brazelton WJ, McCollom TM, Schrenk MO. Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. Front Microbiol. 2017; https://doi.org/10.3389/fmicb.2017.00916.

84. DeShon HR, Schwartz SY. Evidence for serpentinization of the forearc mantle wedge along the Nicoya Peninsula, Costa Rica. Geophys Res Lett. 2004; https://doi.org/10.1029/2004GL021179.

85. Delmelle P, Stix J. Volcanic Gases. In Sigurdsson H, Houghton B, McNutt S, Rymer H, Stix J, (eds). Encyclopedia of Volcanoes. Elsevier. 2000. pp 803–15.

86. Kharaka YK, Mariner RH. Geothermal systems. In Sigurdsson H, Houghton B, McNutt S, Rymer H, Stix J, (eds). Encyclopedia of Volcanoes. Elsevier. 2000. pp 817–34.

87. Badger MR, Bek EJ. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO_2 acquisition by the CBB cycle. J Exp Bot. 2008; 59:1525-41.

88. West-Roberts JA, Carnevali PB, Scholmerich MC, Al-Shayeb B, Thomas A, Sharrar AM, et al. The *Chloroflexi* supergroup is metabolically diverse and representatives have novel genes for non-photosynthesis based CO_2 fixation. *bioRxiv*. 2021; https://doi.org/10.1101/2020.05.14.094862.

89. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. mSystems. 2018; 3:1-12.

90. Colman DR, Lindsay MR, Boyd ES. Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. Nat Commun. 2019; 10:1-3.

Appendix I: Tables and Figures

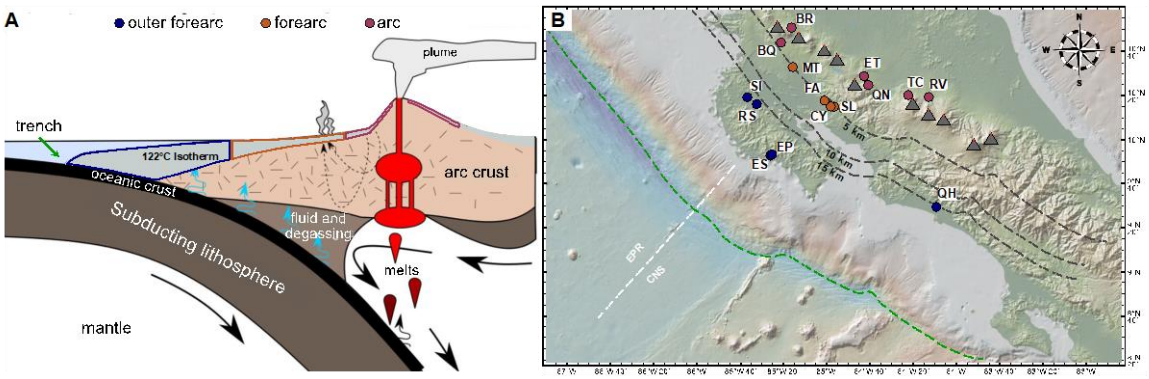


Figure 2.1. General convergent margin geological structure (A) and sample sites spanning the Costa Rican convergent margin (B). Geological provinces (A) and sample sites (B) are colored by their province location of outer forearc (blue), forearc (orange), and arc (red). A) Green arrow indicates location of trench, black arrows show direction of plate movements, blue arrows show fluids and degassing throughout subduction. B) Green dotted line indicates the trench while gray dotted lines mark the depth of the 122°C isotherm in 5 km increments, based on data for this region of Costa Rica from Harris and Wang 2010. Location of volcanoes are shown with volcano icons. Dotted white line demarcates the plates emerging from the East Pacific Rise (EPR) and the Cocos Nazca Spreading center (CNS).

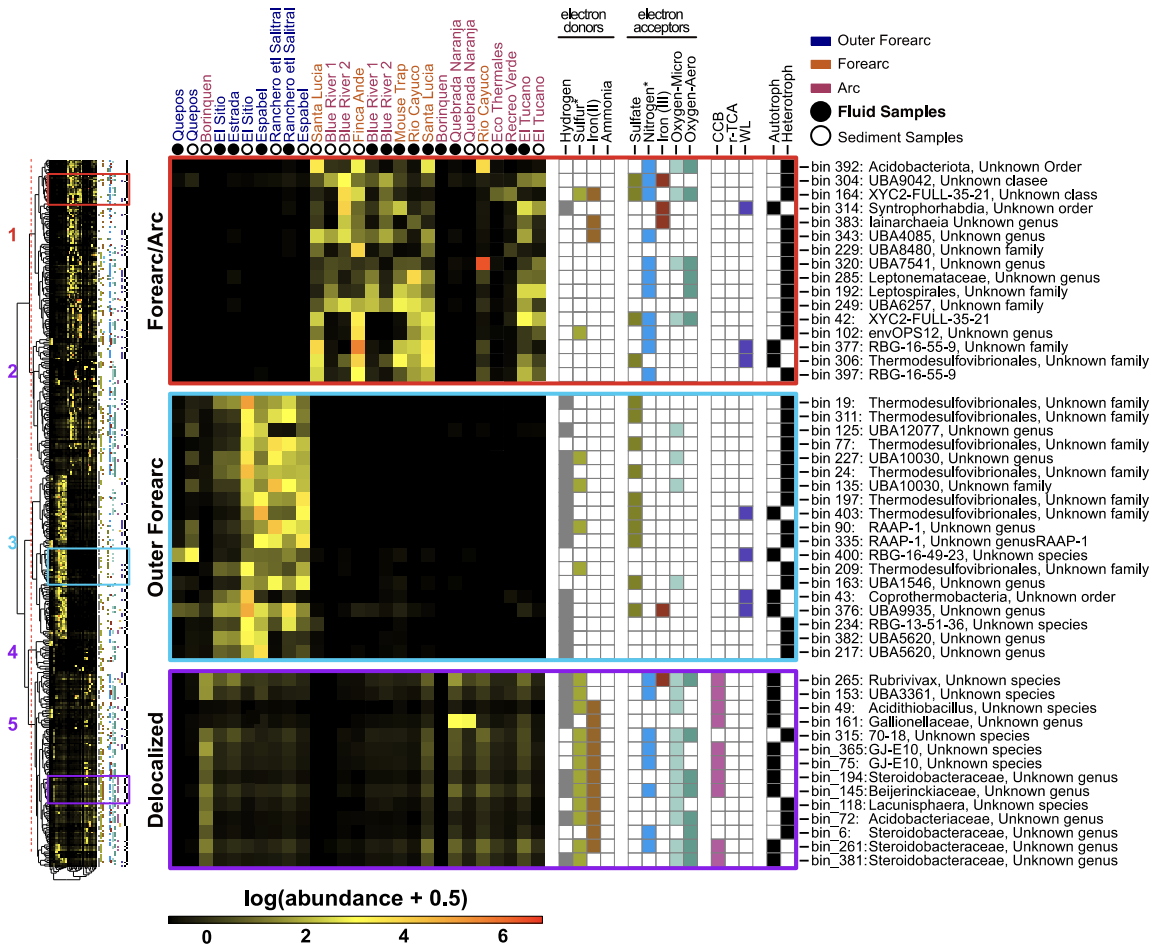


Figure 2.2. Heatmap of metagenomic read recruitment to each MAG and hierarchical clustering based on Spearman rank correlations between MAGs (vertical clusters) and sample sites (horizontal clusters) shows separation of MAGs into those found mostly in the forearc/arc, the outer forearc, or delocalized across geological provinces. Entire heat map is shown to the left, with numbers corresponding to clusters; and representative enlarged sections are shown to the right, with metabolic pathways for redox couples and carbon fixation pathways listed. Subsections were chosen based on inclusion of the key metabolisms and taxa that characterize province specific clusters (clusters 1 and 3) and delocalized clusters (clusters 2, 4, and 5). Sulfur* and nitrogen* refer to the metabolic capability of metabolizing multiple oxidation states. The full heatmap in high resolution is in Figure S2.1.

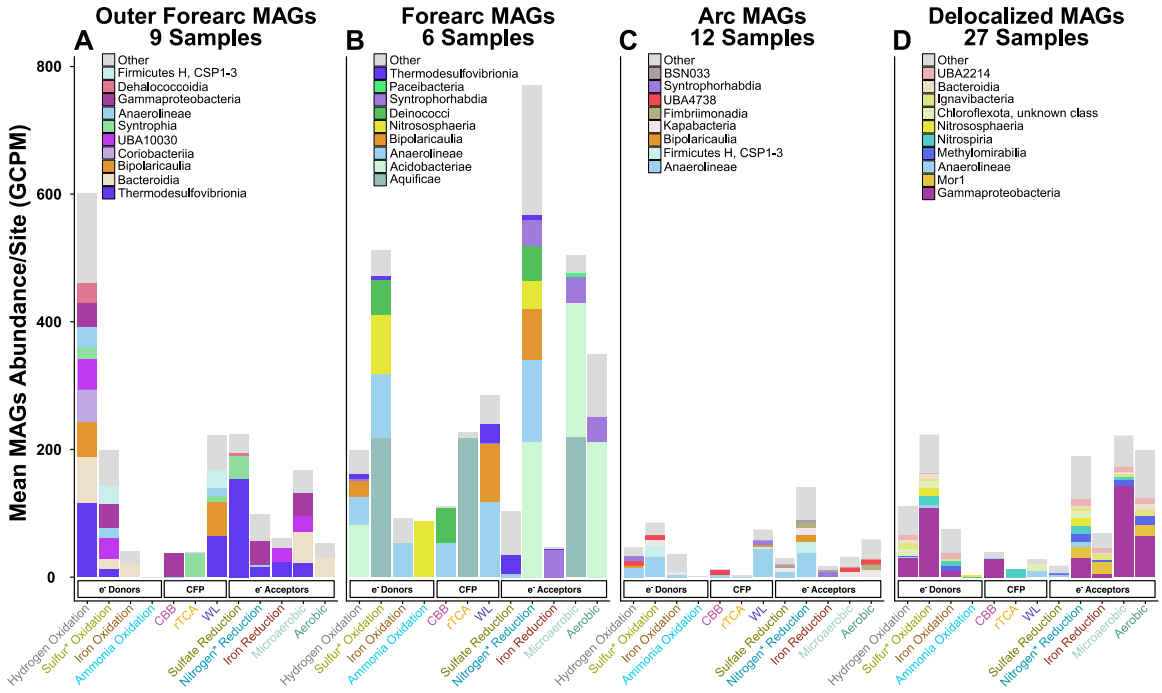


Figure 2.3. Average read abundance (GCPM) per site of MAGs with metabolic pathways for redox reactions and carbon fixation pathways. A, B, and C) show the province-specific average read recruitment of MAGs with these metabolic pathways at all A) outer forearc sites, B) forearc sites, and C) arc sites, after removing MAGs in the delocalized cluster, which are in panel D. Colors correspond to the top 10 most abundant classes for each province and the delocalized community. Panels B and C show top nine and eight most abundant classes, respectively, since MAGs lacking redox genes and carbon fixation pathways were in the top 10 most abundant classes. For bars with multiple MAGs, the GCPM of is the average of the sum GCPM for those MAGs at each site.

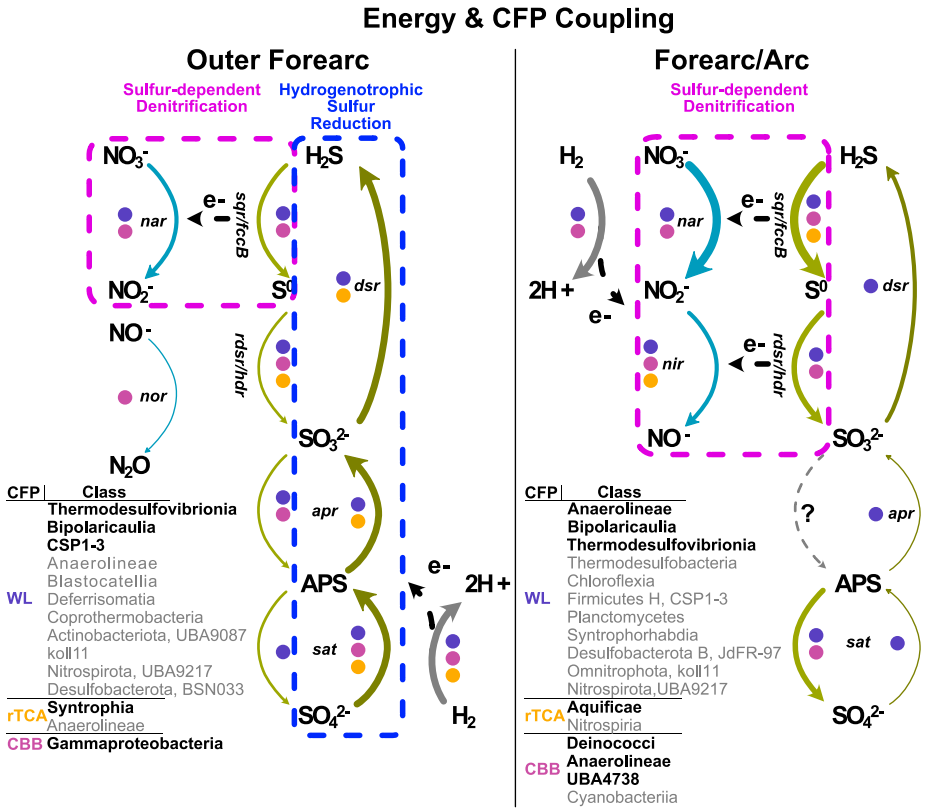


Figure 2.4. Nitrogen, sulfur, and hydrogen cycling pathways in MAGs with carbon fixation pathways from the outer forearc and the forearc/arc. Arrow width corresponds to abundance of MAGs with that gene and dots show which CFP pathways were present in the MAGs that had that gene, with blue for WL, yellow for rTCA, and pink for CBB. Class names are listed from most to least abundant (top to bottom), and MAGs in the top three CFP-containing classes are in bold. Pink dashed box highlights sulfur-associated chemoautotrophic nitrogen reduction. Blue dashed box highlights hydrogen-associated sulfur reduction. Gray dotted line with question mark represents missing *aprBA* predictions for all sulfide-oxidizing MAGs.

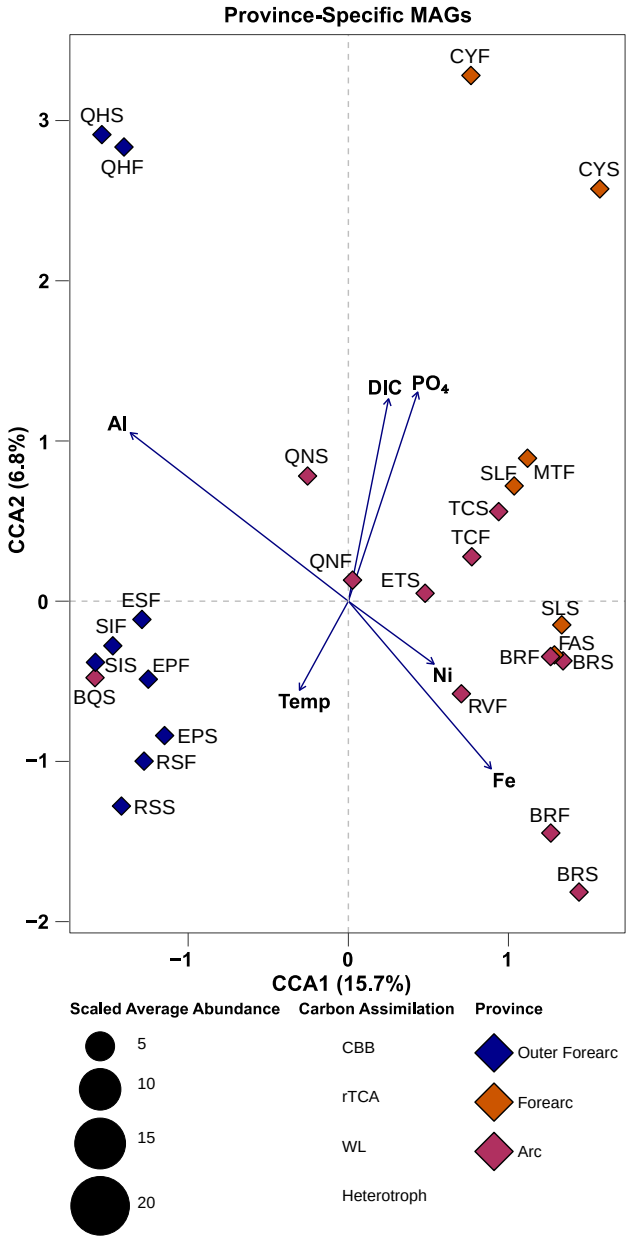


Figure 2.5. Transformation based Canonical Correspondence Analysis showing variables most closely correlated with MAG abundance distribution at each site and site distribution of each MAG (scaling = 2). Distances between MAGs are chi-square distances. A 90° projection of a MAG marker on an environmental vector represents the maximum abundance of that MAG along that vector. Site markers

are distributed around the weighted centroid of the MAGs. The model explains 36% of the total variation in community structure and the axes that are displayed contributed 62% of the total explained variation. The four MAGs with multiple carbon fixation pathways have those two colors overlain.

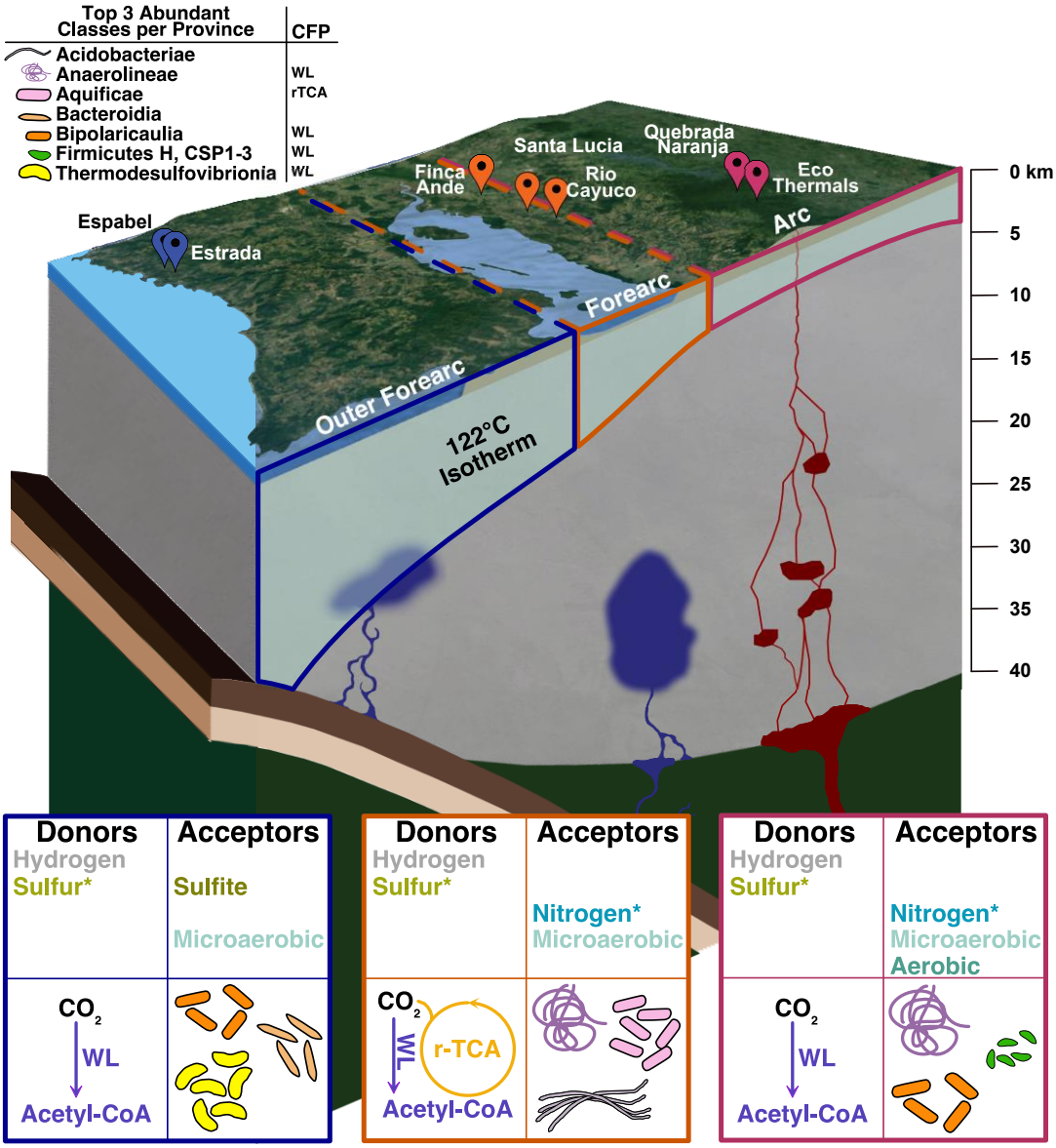


Figure 2.6. Schematic cross-section (to scale) of the 3D distribution of subsurface microbial communities and their metabolisms across the Costa Rica convergent margin, showing a shallowing of potential habitable area from the outer forearc to the arc. A subset of sites is shown. Crustal thickness, angle of subduction, and 122°C isothermal depth are from Harris and Wang (3); while aquifer depth, location, and size are from Worzewski T et al. (64). Top three classes per province and their associated metabolisms are listed. Cell morphology displayed are based on morphologies of cultures from each class. Site colors follow those of Figure 2.1.

Supplemental Material

Sample collection, DNA extraction, and sequencing. Samples were collected from actively venting fluids of 14 natural springs (Fig. 2.1B). One to two liters of fluids were filtered (0.2 µm, Sterivex, MilliporeSigma) at the source. About 10 ml sediments that were constantly saturated by the upwelling fluid were placed into sterile 15 ml tubes, and samples were immediately frozen in a liquid nitrogen-cooled dry shipper (ThermoFisher Scientific, Arctic Express 20). After transport to the University of Tennessee, Knoxville samples were placed into a -80°C freezer. Total genomic DNA was extracted using a phenol-chloroform extraction optimized for low biomass samples (1,2). Metagenomic libraries were sequenced through Census of Deep Life (CoDL) at the Marine Biological Lab (Woods Hole, MA) using an NextSeq platform (Illumina; 2 × 150 nt paired-end reads).

Metagenomic assembly and binning. Raw reads were trimmed and mate-paired by Trimmomatic ((3); version 0.38) using default settings then assembled *de novo* in metaSPAdes (4) to create site-specific assemblies with a minimum contig length of 1.5 kb. Binning, Bin_refinement, and Quant_bins modules of MetaWRAP (5) were used to construct the metagenomic assembled genomes (MAGs/bins) and calculate their abundances. The Binning module used MaxBin2 (6), MetaBAT2 (7), and CONCOCT (8) which resulted in 1217, 1009, and 366 bins respectively. Bin sets were refined using the Bin_refinement module into one consensus bin set of 404 MAGs with ≥ 70% (refs 9-14) completeness and < 5% contamination (Table S2.1). The Bin_refinement module uses CheckM (15) v1.1.3 to evaluate completeness and contamination based on lineage-specific marker genes. MAG quantification across all samples was calculated using the metaWRAP Quant_bins module (default settings). First, this module uses Salmon (16) to calculate the transcripts per million reads (TPM) for each contig at each site. These values are then used to calculate the average read coverage of MAGs in units of GCPM (Genome Copy Per Million Reads). The formula is $GCPM_m = \frac{\sum_{c=1}^{n} TPM_c \times length_c}{\sum_{c=1}^{n} length_c}$ where $GCPM_m$ is the GCPM value of MAG m, TPM_c is the TPM of contig c in MAG m, $length_c$ is the length of contig c in MAG m, and n is

the number of contigs in MAG m (5). The Genome Taxonomy Database Toolkit (GTDB-TK (26); version 1.4.0) was used for taxonomic assignment.

MAG statistical analysis. Correlations between MAGs based on their relative abundance patterns across sites, and correlations between sites based on their MAG compositions, were determined by hierarchical clustering of Spearman correlation distance matrices using Rstudio (17) with R (18) packages ggdendro (method = "complete", metric = "euclidean" (19)) and ggplot2 (20) for visualization through geom_tile (Figs. 2.2 and S2.1). A cutoff of ≥0.03 GCPM was applied to filter out false positives due to small amounts of read recruitments to evolutionarily conserved sequences. This cutoff is equal to the highest abundance of the least abundant MAG. Environmental variables previously shown to correlate with 16S rRNA gene amplicon distributions (2) were chosen for hypothesis testing via transformation based canonical correspondence analysis (tb-CCA) in the R package vegan (2.5-7; ref 21)). These variables include temperature; pH; concentrations of dissolved inorganic and organic carbon (DIC and DOC respectively); aqueous iron, nickel, zinc, and phosphate; sediment associated aluminum, nickel, and chromium. Forward stepwise selection using the "*ordistep*" function from vegan was used to determine significant variables to finalize the model. Statistical hypothesis testing (alpha=0.05) was done using 9 999 permutations of the data.

Redox metabolic predictions

Open reading frame prediction and gene annotation were performed with Prokka (22). KEGG Orthology (KO) identifiers were linked to protein sequence predictions using GhostKOALA (23). KEGGdecoder (24) used KO identifiers to construct putative metabolic pathway predictions within the MAGs. Iron oxidation and reduction were predicted by FeGenie (25). In the MAGs with metabolic pathway predictions, the GhostKOALA output was then used to identify KO numbers of the catalytic subunits in the case of hydrogen, elemental sulfur, sulfide, and iron(II) oxidation and the reduction of sulfate, nitrogen species, iron(III), and oxygen; or at least two subunits in the case of thiosulfate and ammonia oxidation ((26) Table S2.2).

Hydrogen oxidation and production. KEGGdecoder metabolic pathway predictions were used to identify MAGs that contained hydrogenases specifically capable of H_2 oxidation, based on our review of the literature. The KO identifiers were used to identify of the catalytic subunits of the predicted hydrogenases: *hydA* and *hydB* for [NiFe]-hydrogenase (27); *hydA3* and *hydB3* for hydrogen:quinone oxidoreductase (24), *hoxHY* for NAD+-reducing hydrogenase (28); *hyaAB* for [NiFe]- hydrogenase Hyd-1 (29,30); and *hndD* for NADP-reducing hydrogenase. These annotations were verified against the Prokka annotations. No other H_2 oxidizing hydrogenases were identified within the MAGs.

Nitrogen reduction and oxidation. Ammonia oxidation was predicted based on the presence of at least two subunits of *amoABC* (26). Nitrogen reduction was predicted based on the presence of *nosZ* for nitrous-oxide reduction; *norBC* for nitric oxide reduction; and *narGH* or *napAB* for dissimilatory nitrate reduction (24,26). Nitrite reduction was predicted based on the presence of genes encoding nirK and/or nirS (24,26).

Oxygen reduction. Aerobic oxygen reduction predictions were based on the presence of the genes coding for the three core subunits of cytochrome *c* oxidase aa_3 (*coxABC*) (31-33) and cytochrome *o* (*cyoB*) (34). Microaerobic oxygen reduction predictions were based on the presence of the genes coding for the catalytic subunits of cytochrome cbb_3 type oxidase (*ccoN*) (35) and for the cytochrome *bd* complex (*cydAB*) (36). The presence of cytochromes within anaerobes and WL-containing organisms has been demonstrated previously and these genes have been interpreted to function as residual O_2 scavengers for O_2/reactive oxygen species (ROS) detoxification to protect O_2-sensitive proteins (37).

Iron oxidation and reduction. Iron oxidation and reduction predictions were predicted by FeGenie. FeGenie looks for the presences of *cyc2*, *mtoAB*, *pioABC*, *foxEYZ*, and sulfocyanin genes for iron oxidation, while its iron reduction predictions were based on the presence of *mtrCAB*, *omcFZ*, and polymerized *omcS*.

Sulfur species oxidation and reduction. Hydrogen sulfide oxidation was inferred from the use of *spr* for sulfide:quinone oxidoreductase and *fccB* for sulfide dehydrogenase (38). At least two genes of the sox pathway (*soxABCXYZ*), *tsdA* for thiosulfate dehydrogenase, and/or *doxAD* for thiosulfate:quinone

dehydrogenase were required for putative thiosulfate oxidation predictions (24,26). Elemental sulfur oxidation and sulfate reduction were determined by phylogenetic analyses of the alpha subunit of dissimilatory (bi)sulfite reductase (*dsr*). First, KEGGdecoder metabolic pathway prediction was used to identify MAGs that contained *dsr*. The GhostKOALA annotations were then used to identify MAGs that contained the KO identifiers for both the alpha and beta subunits. Annotations were verified against Prokka annotations. In total, 73 MAGs were determined to have both subunits. Fifty-two known oxidative (*r*)*dsrA* and 45 known *dsrA* sequences were selected from the ARB *dsrAB* reference database (39) to use as reference sequences. All identified *dsrA* sequences were aligned against the reference sequences and used to construct Randomized Axelerated Maximum Likelihood (RAxML) trees as described below.

Autotrophy. A MAG was considered capable of autotrophy if 60% of the genes and all key enzymes for a CFP were present (key enzymes: Carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) complex for WL, ATP-citrate lyase or citryl-CoA synthase and citryl-CoA lyase for rTCA, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) and phosphoribulokinase for CBB (40,41). MAGs that did not meet these CFP criteria were categorized as heterotrophs. The gene sequences of the RuBisCO large subunit (*rbcL*) were aligned against those of previously described major lineages (42) and used to construct RAxML trees as described below.

Dissimilatory (bi)sulfite reductase and rbcL phylogenetic analysis. Using MUSCLE through the MEGA X 10.1.8 software with default parameters, the identified *dsrA* and *rbcL* genes were aligned their respective representative sequences (39,42) and used to construct RAxML trees (Figs. S2.2 and S2.3). Phylogeny was inferred through CIPRES Science Gateway cluster (V.3.3: https://www.phylo.org/index.php/portal/v33) using RAxML-HPC BlackBox with the following parameters:

> datatype = protein, mlsearch = true, no_bfgs = false, prot_matrix_spec = JTT, number of bootstraps = 1000

The trees were visualized with iTOL (https://itol.embl.de/), formatted in Affinity Designer 1.10.1 (Figs. S2.2 and S2.3). The *dsrA* sequences that grouped with the reference (*r*)*dsrA* sequences were used to

predict elemental sulfur oxidation. Sulfate reduction predictions were based on those sequences that fell in with the reference *dsrA* sequences (Fig. S2.2). The RuBisCO form was determined based on *rbcL* sequence placement within the RuBisCO RAxML tree (Fig. S2.3).

Appendix I: Tables and Figures

Table S2.4. Marginal effects of terms in the tb-CCA model. Statistical significance was determined with 9999 permutations of the data. Significant environmental constraints are sediment derived aluminum, aqueous iron, DIC, temperature, aqueous phosphate, and aqueous nickel concentrations.

Variable	Df	ChiSquare	F	p-value
Temp	1	0.22441	2.2292	0.011
DIC	1	0.22138	2.1991	0.0056
Fe (ac)	1	0.2511	2.4943	0.0013
Ni (ac)	1	0.17321	1.7206	0.0479
PO_4	1	0.18647	1.8523	0.0285
Al (s)	1	0.31694	3.1484	0.0001
Residual	19	1.9127		

For further supplemental tables and figures referenced within this chapter, please refer to online version of this publication. Online version contains all supplemental materials to maintain brevity of this document: https://doi.org/10.1038/s41396-022-01331-7

Chapter 3 Viral metabolic and taxonomic landscape of the terrestrial subsurface ecosystem within the Central American convergent margin

This original research is being prepared for future publication submission with at least the following authors:

Timothy J. Rogers, Joy Buongiorno, Gerdhard L. Jessen, Matthew Schrenk, James A. Fordyce, J. Maarten de Moor, Carlos Ramirez, Peter H. Barry, Mustafa Yücel, Matteo Selci, Angela Cordone, Donato Giovannelli, Karen G. Lloyd

My primary contributions to this paper include: (i) Investigations and data acquisition, (ii) data synthesis and statistical analysis, (iii) figure and table generation, (iv) and writing the manuscript.

Abstract

Chemosynthetic microbial communities within hydrothermal fluids of terrestrial convergent margins are tied to deep tectonic processes, however the effect of viruses on these communities has not been explored. Here we explore viral distribution, host interaction, and characterize viral auxiliary metabolic genes (AMGs) across the geological and volcanic provinces of the Central American convergent margin. We find that most viruses are unclassified. Of those that could be taxonomically identified, *Myoviridae*, *Siphoviridae*, and *Peduoviridae* are the most abundant. Just 15% of the viral community are able to be linked to 33% of the potential hosts identified through metagenomic-assembled genomes (MAGs). *Proteobacteria* and *Bacteroidota* are the most frequently predicted hosts. Of those viruses linked to hosts, 90% are predicted to be lysogenic with < 1% predicted to have a broad host range. We find that viral community richness is positively correlated with both the prokaryotic community and host richness. Additionally, viral community dissimilarity is systematically lower than and positively correlated with the prokaryotic community dissimilarity, suggesting that lysogeny is in fact the dominant lifestyle in the hydrothermal fluids of this terrestrial convergent margin. Metabolic analysis suggests a high diversity of AMGs, with viral communities within each geophysical province possessing their own distinct repertoire of AMGs. It seems that lysogenic viruses play a key role in the ecology and biogeochemical cycling of the prokaryotic community with the hydrothermal ecosystem of the Central American convergent margin.

Introduction

Recent studies have demonstrated that the distribution of terrestrial subsurface chemosynthetic communities and their metabolic potential are intimately linked to deep tectonic processes within hydrothermal fluids of convergent margins (1,2, Basili et al. in prep). Moreover, several studies have demonstrated that chemosynthetic communities within hydrothermal systems provide opportunities to investigate viral-host relationships and the effects that viruses have on the metabolic interactions between their host and hydrothermally-derived volatiles (3–6).One way that viruses influence these interactions is through auxiliary metabolic genes (AMGs) which they use to modulate and/or augment their hosts' metabolisms (6–8). However, viral distributions, viral-host interactions, as well as the types and distributions of AMGs within the terrestrial subsurface of convergent margins have not been explored.

Convergent margins allow for biologically relevant volatile rich fluids to be released from a dense subducting slab into the less dense overriding plate. As the slab subducts, devolatilization increases due to the increases in temperature and pressure allowing for the release of these volatile rich fluids. These rich fluids migrate through cracks and fissures in the overriding plate where they encounter deep subsurface chemosynthetic microbial communities before reaching the surface (1,2). Distinct variations in the community membership, redox-active metabolic pathways, and carbon fixation pathways within volcanic provinces of convergent margin were linked to deep tectonic processes (1,2). These volatiles directly support the community and their metabolic potential, and chemolithoautotrophs, in turn, directly influence the amount of carbon delivered back to the mantle; sequestering an estimated $1.4x10^{9}$ to $1.4x10^{10}$ mol of carbon released from the slab/mantle mixture per year (2). To accomplish this, chemolithoautotrophs can use a variety of redox-metabolic processes such as the oxidation and/or reduction of hydrogen, sulfur, and nitrogen species within these volatile rich fluids. Similar interactions between chemosynthetic communities and geo-hydrothermal processes have been observed in marine (4,5) and freshwater (3) hydrothermal vents. The discovery of viral AMGs associated with processes such as sulfur cycling (3,5,6), nitrogen cycling (4,5), as well as phosphorus scavenging (4) suggest that viruses might play a critical role in biogeochemical cycling within these systems. In this manner, viruses can

directly influence the rate of nutrition uptake and directly or indirectly affect carbon sequestration rates (3,4). While most studies of viral-host interactions within hydrothermal systems have focused on hydrothermal vent systems and intraplate hot springs, convergent margin hydrothermal systems offer an overlooked and unique opportunity to explore the interplay between viruses and their prokaryotic host as well as the mechanisms by which viruses might alter the rate of host volatile uptake.

Using geochemical data and co-located metagenomic sequences, we investigate viral-host pairings, characterize viral AMGs, and determine whether the distribution of viruses and their AMGs vary across geological and volcanic provinces of the Central American convergent margin (Southern Panama, Cordillera Talamanca, Costa Rica outer forearc, active volcanic arc, and active volcanic back arc). Southern Panama is a break in the subducting Nazca plate as it subducts below the Colombian plate allowing for a mantle 'wind' from the Galapagos mantle plume to be laterally transported at shallow mantle depths 1200 km up through the Costa Rican back arc (9). This results in geothermal fluids within Central Panama containing the highest mantle input ever reported within an area far from volcanic activity (9), as well as explains the lack of arc volcanism and low seismic activity within Western Panama (10). Cordillera Talamanca, between southeastern Costa Rica and western Panama, is characterized by a lack of subduction-related volcanic activity due to a lack of dehydration of the subducting slab (11). This is a result of the shallow horizontal subduction of the Cocos plate in this area, causing it to underride the Caribbean plate and preventing it from reaching the depths needed for slab dehydrations (11). In the Costa Rica outer forearc the subduction of the Cocos plate is relatively shallow and releases pore fluids, seawater, and serpentinization products from the subducted material into the overriding Caribbean plate (12–14). The slab reaches a depth of 90-110 km at the active volcanic arc in Costa Rica, where slab fluids and flux-derived melts (15) interact with the mantle wedge to generate magmas (16,17). Between 202 to 207 km inland from the Costa Rica trench is the active volcanic backarc, where the subducted slab reaches a depth of $\geq$ 120 km (18) and decompression melts are largely responsible for magma generation (15). As viruses depend on their host for replication, we expect that geophysical province will have a indirect effect on viral distribution patterns. The analysis was conducted on 31 sediment and 36 fluid

metagenomes from actively venting geochemically distinct hot springs from across Costa Rica and Panama.

Methods

Sample collection, DNA extraction, and sequencing

Samples were collected during two separate field seasons (2017 and 2018) from geochemically distinct natural springs and wells. One to two liters of actively venting fluids were filtered at the source (0.2 µm, Sterivex, MilliporeSigma). About 10 ml sediments that were constantly saturated by the upwelling fluids were placed into sterile 15 ml tubes and samples were immediately frozen in a liquid nitrogen-cooled dry shipper (ThermoFisher Scientific, Arctic Express 20). After transport to the University of Tennessee, Knoxville samples were placed into a -80°C freezer. Total genomic DNA was extracted using a phenol-chloroform extraction optimized for low biomass samples (2,19). Metagenomic libraries were sequenced through Census of Deep Life (CoDL) at the Marine Biological Lab (Woods Hole, MA) using an NextSeq platform (Illumina; 2 × 150 nt paired-end reads).

Generation and metabolic predictions of metagenomic assembled genomes

Metagenomic assembly, binning, and taxonomic identification of the 2018 Panama/Costa Rica metagenomic data set followed the protocol laid out in Rogers et al. (1) with minor changes. Changes include using bin quality cutoffs of ≥ 50% complete and ≤ 10% contamination (redundancy) as well as only using MaxBin2 (20) and MetaBAT2 (21) for the initial binning step. MAGs from the 2017 dataset (1,2) were included in the clustering step described below. These MAGs were cutoff at ≥70% completeness with ≤5% redundancy.

After viral contig identification (see *Viral contig and auxiliary metabolic genes identification*) within the metagenomic assembled genomes (MAGs), dRep (22) was used to dereplicate bins at 95% ANI as this percentage has been demonstrated to delineate the bacterial species boundaries (23). The dRep primary clustering step using MASH was deactivated due to MASH's inaccuracy when clustering incomplete genomes (22). This increases computational time but is more accurate. fastANI was used for the secondary clustering step. This algorithm has been demonstrated to be just as accurate as other ANI

algorithms but at a much faster rate (23). MAGs were clustered at 95% ANI and the highest quality MAGs were automatically selected as cluster representatives. However, MAGs that were ≤ 75% complete were ignored by the clustering algorithm and retained as their own cluster. The Genome Taxonomy Database Toolkit (GTDB-TK (24); version 1.4.0) was used for taxonomic identification and GTDB nomenclature has been used throughout this manuscript.

To evaluate the viral effect on chemolithoautotrophic members, metabolic predictions of the MAGs were carried out through the Anvi'o (version 7.1 (25)) program *anvi-estimate-metabolism*. The classification of autotrophs was given to MAGs that contained at least 60% of the genes and all key enzymes for a carbon fixation pathway. MAGs that fell outside of these standards were classified as heterotrophs.

Viral contig and auxiliary metabolic genes identification

VIBRANT (8) was used to identify viral contigs as well as any auxiliary metabolic genes (AMGs) within the contig assemblies and MAGs on default settings. Both checkV (26) and VIBRANT were used to categorize whether a viral contig was likely to be lysogenic or lytic in the unbinned assemblies as well as provide a quality score for all identified viral contigs. Viral contigs that were not characterized as lysogenic were manually labeled as unknown due to the error rate of falsely identified lytic viruses. Only viral contigs with a length ≥ 3kb were retained.

Viral quality was assessed by VIBRANT and CheckV. Those with a quality score of "undetermined" were left out of further analysis. VIBRANT has four quality categories: high, medium, and low, each of which can have an additional classification of circular. These categories are assigned based on circularization of scaffold sequence, virus orthologous group (VOG) annotations, total VOG nucleotide replication proteins, and total VOG viral hallmark proteins (6). CheckV has five quality categories: complete, high quality (>90% complete), medium quality (50-90% complete), low quality (0-50% complete), and undetermined (no completeness estimate) (26). These categories are generally assigned based on a ratio between the length and the expected length of the viral contig (26). If there was a disagreement in the quality score between checkV and VIBRANT, the best quality classification of the

two was chosen. This led to four quality categories: high, medium, low, and complete. To remove viral contig redundancy, viral contigs were clustered at ≥ 95% ANI over 85% of the contig length using the *anicalc* and *aniclust* scripts provided by checkV.

Virus-host prediction

A total of 5 methods were used to identify viral-host matches. First, viral contigs that were identified within the MAGs were automatically categorized as lysogenic and host-matched due to their inclusion in the MAG. VirMatcher (27) was then used to further link viral contigs to potential host MAGs. VirMatcher predicts microbial hosts using viral sequence matches to (i) host CRISPR-spacers, (ii) integrated prophages in host genomes, (iii) host tRNA genes, (iv) host k-mer signatures. MinCED (28) is used to identify CRISPR-spacers in the MAGs which are then blasted against the vOTU sequences to look for matches using BLASTn. To identify prophages (lysogenic viruses) BLASTn is used to blast the vOTU sequences against the MAGs. vOTU and MAG tRNA are identified using tRNA-scan (29) and then a BLASTn is performed between these tRNA groups. Promiscuous tRNAs are removed before BLASTn is performed. Host-viral k-mer signature matches are predicted by WiSH after masking tRNA sequences on the viral genome (30). VirMatcher then assigns a score of high, medium, or low to each prediction made by each method. To be accepted as virus-host pairing, we required that the pairing had one or more of the following:

1. Prophage predictions required ≥ 85% ANI ≥ 80% of the sequence length.
2. CRISPR match score of ≥ 3 (3 is a high score). This score indicates that there were no mismatches in one or more sequences (27).
3. WiSH score of ≥ 3 (high score) and non-promiscuous tRNA match score of 3 (high score). A high WiSH or tRNA match score alone was not included as a viral-host match due to uncertainties inherent in both methods.

Viral contigs whose lifestyles were previously categorized as unknown and were matched to host(s) based on prophage predictions were then categorized as lysogenic.

Removal of viral contamination

vOTUs that were host matched to known microbial lab contaminants as well as those who's max abundance was found in the extraction blank were removed from further processing.

Viral taxonomic identification

For taxonomic identification of vOTUs, the NCBI Viral RefSeq database was installed locally. Prodigal (31) predicted ORFs of the vOTU sequences were aligned against this database using tblastn (E-value ≤ .0001, bit score ≥ 50). Accession numbers of top hits were then searched against the NCBI taxonomy database using the R package rentrez (version 1.2.3 (32)). Viral taxonomy was assigned based on the most top hit taxonomic agreements between ORFs at the family level.

Abundance profiles

For abundance counts, quality-controlled reads from each metagenome were mapped to identified vOTUs and MAGs with bowtie2 (33) on default settings. For vOTUs, mapped reads were then scaled by 10^6 and finally standardized by dividing the scaled counts by the average viral contig length based on vOTU membership (Fig. S3.1). These standardized counts were rarefied to 12450 sequences per sample. Read counts for MAGs were converted to transcripts per million reads (TPM) using coverM ((v0.5.0) https://github.com/wwood/CoverM).

Statistical analysis

α-diversity. For both MAGs and vOTUs, differences in species evenness and total richness were compared using Pielou's evenness and Chao1 indexes respectively. Observed species richness and diversity were compared using Hill numbers (34) of the rarefied abundances at orders of $q = 0$, $q = 1$, and $q=2$. The parameter q determines the relative weighting of rare species. At $q = 0$, all species are weighted equally (observed richness); at $q= 1$, species are weighted proportionally to their relative abundance (exponential Shannon's index); and at $q = 2$, rare species are down-weighted (inverse Simpson's index). Hill numbers were calculated with the R package vegetarian (v1.2 (35)). We compared α-diversities across volcanic provinces with a t-test and reported only the adjusted p-values. Comparisons between the observed species richness of the MAGs and vOTUs were made through linear regression.

β-diversity. The vegan (v2.6.2 (36)) package was used to conduct distance-based redundancy analyses (db-RDA) of the MAGs and vOTUs using Bray-curtis dissimilarity matrices (β-diversity). First, the vOTU rarefied counts and retained MAG TPMs were converted to relative abundance counts and scaled to the vOTU rarefied library size (12450 sequences) for comparisons. To evaluate environmental influence on community dissimilarity, a total of 22 environmental variables were checked for coverability. This resulted in 15 non-covarying (covariance ≤ 0.6) variables that were used for hypothesis testing via db-RDA. These variables include geophysical province, pH, temperature, sample type (fluid and sediment), SO_4^{2-}, Cl^-, CO_2, H_2, O_2, CH_4, NH_4^+, Mg^{2+}, Ca^{2+}, and Al^{3+}. For both the MAG and vOTU based db-RDAs forward double stop step selection using the "ordiR2step" function from vegan was used to determine significant variables to finalize the model. Statistical hypothesis testing via permutational multivariate analysis of variance (PERMANOVA) (alpha=0.05) was done using 9 999 permutations of the data. To check the effect of viral and prokaryotic community on each other's β-diversity, the Bray-curtis dissimilarity matrices were regressed against each other as well as both were plotted against the physical distance (km) between sites within pairwise comparisons. Geographical coordinates were used to calculate the km between sites using the R package geosphere v1.5.18 (37). A linear regression model was constructed using the viral dissimilarity matrix as a dependent variable and the prokaryotic dissimilarity matrix as an explanatory variable to identify any effect prokaryotic species turnover may have on viral species turnover. The calc.relimp and boot.relimp functions of the R package relaimpo (38) were used to find the variation explained by this model via the Lindeman, Merenda, and Gold (LMG) method of R^2 decomposition with 1 000 bootstraps.

Results

Overview of microbial communities

To identify potential viral hosts that contain lysogenic viruses, we constructed MAGs. Metagenomic binning of the 2018 Costa Rica and Panama fluid and sediment metagenomes resulted in a total of 1220 medium to high quality MAGs based on MIMAGs criteria ((39) ≥ 50% complete with < 10% redundancy). We then added the 404 MAGs from the 2017 data set (1) and removed MAGs whose

highest abundance was in the extraction blank. This left a total of 1592 medium to high quality MAGs (Fig. 3.1). MAGs that were taxonomically classified as known animal, human, and/or laboratory contaminants (40,41) were removed from downstream analysis, leaving a total of 1359 MAGs (Fig S3.2). *Proteobacteria* is the predominant phylum across all provinces (mean relative abundance: 38% ± 5.5% standard error (SE); Table S3.1 and S3.2). The second most abundant phyla very across geophysical provinces between *Bacteroidota* in the active volcanic backarc, Costa Rica outer forearc, and the Panama slab (11% ± 2% SE; Table 3.1 and S3.1) and *Chloroflexota* in the active volcanic arc and Cordillera Talamanca (10% ± 2% SE; Table S3.1 and S3.2).

Viral Community Composition

General description of vOTU's. A total of 30,337 putative viral contigs were identified within the binned and unbinned contigs. Clustering at 95% ANI over 80% of the sequence and removal of sequences that were < 3 kb generated 9,374 low to high quality vOTUs with mean and median sequence lengths of 13.1 kb and 6.8 kb, respectively. Low quality vOTUs were retained for analysis since novel viruses rarely contain known viral genes and are usually categorized as low quality (6). Of the vOTUs recovered, 23% (n=2,122) were predicted to be lysogenic, with the other 77% having an unknown lifestyle, since determining that a virus is primarily lytic is difficult from primary DNA sequence data alone.

Viral Taxonomic assignment. Taxonomic assignment revealed that the majority of vOTUs were unidentified at the order (94% of vOTU sequences) and family (77%) level. The remainder were predominantly assigned to six families: *Siphoviridae* (3.3%), *Myoviridae* (2.2%), *Peduoviridae* (2.0%), *Casjensviridae* (1.5%), *Autographiviridae* (1.4%), *Kyanoviridae* (1.2%), *Herelleviridae* (1.2%), *Phycodnavirida* (1.1%), and *Mesyanzhinovviridae* (1%). All were unassigned at the order level. The rest of the sequences were classified into 67 other families each representing ≤ 1.0% of the vOTUs.

The most abundant viruses at the family level across all provinces were the taxonomically unknown (78% median relative abundance), *Myoviridae* (2% median relative abundance), *Siphoviridae* (2% relative abundance), and *Peduoviridae* (1%) (Fig. 3.2).

Viral-Host Matching

Based on viral sequences identified within MAGs, prophage sequence similarities, oligonucleotide frequencies, CRISPR spacers, and tRNA sequences, 33% (n=524) of the MAGs were host-matched to 15% (n =1,359) of the vOTUs. Host predictions spanned 6 archaeal and 48 bacterial phyla, with *Proteobacteria* and *Bacteroidota* (36% and 13% of total MAG hosts, respectively) being the most frequently predicted. Archaea comprised only 4% of host matches. Most vOTUs had a very narrow host range with only two vOTUs predicted to have a broad host range across multiple prokaryotic phyla. A small host range has been observed for environmental viruses in previous studies (42–45). Interestingly, while only 22% of total viruses were predicted to be lysogenic, 90% (n=1228) of those linked to hosts were predicted to be lysogenic. Some of the methods we used to link viruses to hosts looked specifically for lysogeny, for instance by identifying viruses binned or assembled into MAGs. However, our other methods were sufficient to identify lytic viruses as well. Another caveat is that our fluid samples were filtered at 0.2 µm, which would allow viral particles to pass. However, about half of our samples were sediment samples, which were not subjected to filtering and would have contained free viral particles as well.

To understand what effect viruses might have on primary producers within this system, MAGs that contained key enzymes for carbon fixation pathways and ≥ 60% of all genes associated with those pathways were examined. Examination of annotated genes of all MAGs revealed < 0.5% of the MAGs contained photosystems. Chemolithoautotrophs made up 22% (n=118) of the MAGs linked to vOTUs with 145 containing CBB, 140 containing WL, 24 containing rTCA, 5 containing D/H, 3 containing D/H and WL, 3 containing CBB and WL, and one containing both CBB and D/H. Of these, 106 chemolithoautotroph MAGs were host matched to vOTU's: 73 containing CBB, 27 containing WL, 4 containing rTCA, one containing D/H, and one containing both CBB and D/H. vOTU's within the viral families of *Myoviridae, Siphoviridae, Mesyanzhinovviridae, Casjensviridae*, and *Matshushitaviridae* were primarily paired to CBB-containing MAGs. Most viruses linked to both WL- and rTCA-containing MAGs belonged to novel unclassified viral families (Fig. 3.3). The *Halomagnusviridae* family was linked

to D/H only in the active volcanic arc. Other viral family level matches to autotrophic hosts were below the limit of detection.

Viral Diversity

α-diversity. vOTU total species richness (Chao1 index) was significantly less in the Costa Rica outer forearc when compared to Cordillera Talamanca (t-test: adjusted p-value < 0.01), the active volcano back arc (adjusted p-value < 0.01), and Southern Panama (adjusted p-value < 0.01). Cordillera Talamanca also had a significantly higher Chao1 index than the active volcanic arc (adjusted p-value < 0.05). However, the species evenness (Pielou's evenness index) was only significantly higher in Cordillera Talamanca when compared to the Costa Rica outer forearc and the active volcanic arc (adjusted p-value < 0.01) (Fig 3.4). While there was no significant difference in Pielou's evenness index between the fluid and sediment samples, the Chao1 index was significantly lower in the fluids than in the sediments (*p*-value < 0.05).

To evaluate the possible effect of the microbial community diversity on viral community diversity, vOTU richness was plotted against the MAG richness within each sample. vOTU species richness has a significant positive correlation with MAG richness within the fluids and sediments (Fig. S3.2). These interactions are stronger when only considering matched host and virus richness (Fig. S3.3). Further, significant correlations were found between the abundance of both linked viruses and their hosts within sample type and across geophysical provinces (Fig. S3.4).

β-diversity. The db-RDA of the vOTUs presented clear dissimilarity in the viral communities, explaining 12% ($p < 0.001$, $R^2_{adj} = 0.122$) of the variation in the samples (Fig. 3.5A and B). The important variables are, in order of most marginal effect significance: geophysical province, temperature, sample type, and the geochemical variables Cl^-, pH, Mg^{2+}, O_2, CH_4, and Ca^{2+}. Variance partitioning analysis reveals that most of the constrained variation between sites is explained by geophysical province ($p < 0.001$, adj-R2 = 0.04) as well as geochemistry ($p < 0.001$, $R^2_{adj} = 0.04$) with temperature and sample type explaining smaller portions (temperature: $p < 0.001$, $R^2_{adj} = 0.019$; sample type: $p < 0.01$, $R^2_{adj} = 0.012$) (Fig. 3.5C and D). A pairwise PERMANOVA analysis reveals that observed dissimilarities were significant between provinces and sample types (Fig. 3.5A and B; Table 3.1).

Strong dissimilarities are also seen in the microbial community as revealed in the db-RDA of the MAGs which explains 16% ($p < 0.001$, $R^2_{adj} = 0.161$) of the variation. Here, the important variables vary slightly from the vOTU db-RDA. As with the vOTU, the variables with the most significant marginal effect are geophysical provinces as well as temperature. However, fluids and sediments are not important environmental variables. Other variables with a significant marginal effect are Cl^-, pH, H_2, O_2, Ca^{2+}, CH_4, and SO_4^{2-}. This db-RDA reveals tighter clustering of the sites within each geophysical province than in the vOTU db-RDA (Fig. 3.6A). Most of the constrained variation between sites is explained by geochemistry ($p < 0.001$, $R^2_{adj} = 0.056$) and geophysical province ($p < 0.001$, $R^2_{adj} = 0.049$) with temperature explaining the least amount of variation ($p < 0.001$, $R^2_{adj} = 0.023$) (Fig. 3.6B and Table 3.2). As with the vOTU analysis, a pairwise PERMANOVA analysis revealed that observed dissimilarities were significant between provinces (Fig. 3.6B; Table 3.3).

The viral community dissimilarity between samples was systematically lower than and had a strong correlation with that of the prokaryotic community in both the sediments and the fluids (Mantel test: Sediments $r = 0.558$, $p < 0.001$; Fluids $r = 0.709$, $p < 0.001$; Fig. 3.7A-D). This continuity was even stronger when considering the linked viral dissimilarity and their host dissimilarity (Mantel test: Sediments $r = 0.665$, $p < 0.001$; Fluids $r = 0.726$, $p < 0.001$; Fig. 3.8A-D). The prokaryotic community dissimilarity explained 34% of the variation in the viral community dissimilarity within and across geophysical provinces as revealed by the LMG of the linear regression (Fig. S3.4).

Viral AMGs

A total of 1661 vOTUs contained AMGs. The most abundant AMGs were DNA cytosine methyltransferase (*dcm*) (28% mean relative abundance), phosphopantothenoylcysteine decarboxylase/phosphopantothenate-cysteine ligase (*dfp*) (6%), 2-polyprenyl-6-hydroxyphenyl methylase/3-demethylubiquinone-9 3-methyltransferase (*ubiG*) (6%), D-beta-D-heptose 7-phosphate kinase/D-beta-D-heptose 1-phosphate adenosyltransferase (*gmhC*) (5%), ureidoacrylate peracid hydrolase (*rutB*) (5%), 7-cyano-7-deazaguanine reductase (*queF*) (5%), Ureidoglycine aminohydrolase (*ylbA*) (5%), phosphoadenosine phosphosulfate reductase (*cysH*) (5%), 3-oxoadipate enol-lactonase (*pcaD*) (5%), and

UDP-glucose:(heptosyl)LPS alpha-1,3-glucosyltransferase (*waaG*) (5%). The most numerous viruses with each of these pathways were taxonomically unclassified at the order and family level.

Of the vOTUs that are linked to hosts, 13% (n=176) contain AMGs and are matched to 162 hosts (Fig. 3.4). Each geophysical province has a distinct abundance pattern of AMGs from linked viruses. *dcm* is the most abundant within all geophysical provinces except in the active volcanic backarc (19% to 48% mean relative abundance). Within the active volcanic backarc the most abundant are genes associated with "7-cyano-7-deazaguanine biosynthesis pathway" (preQ0: *queE*, *queC*, *folE*, and *queD* and preQ1: *queF*; 23%) with the second most abundant being *dmc* (20%). In the outer forearc the next four most abundant are AdoMet-dependent heme synthase (*ahbD*) (12%), molybdopterin-synthase adenylyltransferase (*moeB*) (7%), *cysH* (6%), and phosphoribosylformylglycinamidine cyclo-ligase (*purM*) (5%). In Cordillera Talamanca they are phosphoribosyl 1,2-cyclic phosphate phosphodiesterase (*phnP*) (7%), CysO sulfur-carrier protein]-S-L-cysteine hydrolase (*mec*) (4%), *purM* (4%), and riboflavin deaminase (*ribD*) (4%). In the active volcanic arc, they are *ylbA* (14%), *cysH* (10%), *gmhC* (10%), and *rutB* (10%). In Southern Panama, they are *ubiG* (17%), *queF* (9%), *dfp* (6%), and *pcaD* (5%).

Discussion

Out of the viruses that could be taxonomically identified, the most abundant were members of the *Caudoviricetes* class, which has been observed in many other ecosystems including seawater, cold seeps, human gut biome, and others (46–49). The *Siphoviridae, Myoviridae,* and *Peduoviridae* are the most abundant viral members within this class. These are the most abundant viral families in marine systems as well (48), including the marine water column, sediment (46), and marine-associated mud volcanoes (47). However, the vast majority of vOTUs were unclassified, which has been reported previously from a mud volcano (47) but appears not to be true for the majority of other environmental viromes (48). This suggests that our hot spring system contains more unexplored viral diversity than the more commonly studied marine environments.

Of the 1359 MAGs and 9374 vOTUs, only 39% and 15%, respectively, are virus-host matched. Most of the linked hosts are from the bacterial phyla *Proteobacteria* and *Bacteroidota* (36% and 13% of total MAG hosts, respectively). *Proteobacteria* and *Bacteroidota* were also found to be two of the most common viral hosts in pelagic and benthic regions of the Baltic Sea (46). All the top 10 most abundant classified viral families in our dataset, as well as the unclassified viruses (Fig. 3.3) are linked to *Proteobacteria* and 5 of the most abundant classified viral families are linked to *Bacteroidota*. The fact that the highest viral abundance, whether from classified or unclassified taxa, are linked to the *Proteobacteria* and *Bacteroidota* may be due to the relatively fast growth rates of these two phyla (0.464 h^{-1} and 0.284 h^{-1} mean growth rates, respectively (50). Faster growing members of a microbial population are more susceptible to the predation through viral infection than their slower growing counterparts due to a tradeoff between the ability to grow fast or having increased resistance to viral infection (51). Our results show that the *Proteobacteria* and *Bacteroidota* are linked to 19 and 22 distinct viral families, respectively, whereas members of slower-growing *Desulfobacterota* (maximum mean growth rates 0.063 h^{-1} (52)) are linked to just 4 viral families and the phylum *Actinobacteriota* (maximum growth rate of 10 to 15 h (53)) are linked to only 6 viral families, supporting this conclusion.

Drivers of viral diversity

Richness in the Costa Rican outer forearc viral community was significantly lower than that of the other geophysical provinces. Viral diversity can be constrained either by their own sensitivity to environmental conditions or by the sensitivities of their hosts to environmental conditions (54). The outer forearc is a serpentinizing environment, characterized by high pH (> 9) conditions and limited nutrient availability (55,56). This suggests that the richness of the viral community may be directly (virus sensitivity) or indirectly (host sensitivity) limited due to these potentially stressful environmental conditions. Significant positive correlations are found between the richness of both the viral and the microbial communities (Fig S3.2). The correlations between viruses and their linked host are stronger than the correlations between the two communities (Fig S3.3). Host richness has been shown to drive viral richness in reservoir species systems (57) as well as in the human gut microbiome (58). These

observations suggest that within the Central American convergent margin, host availability drives viral richness while environmental characteristics, such as serpentinizing conditions, limits the diversity of potential hosts.

Evenness of viral communities within all of the geophysical provinces is high (Chao1 ~0.75). High evenness is associated with increases in specialized consumers which decrease interspecific competition and encourage higher rates of niche partitioning (59–61). Less than 1% of host-linked vOTUs exhibit a broad host range across multiple host phyla, reflecting the high degree of viral host specificity (specialized consumers) often seen within diverse ecosystems (43,47). These observations suggest that the high viral evenness observed across all geophysical provinces is the result of a high degree of host specificity (specialist consumers).

Interestingly, the β-diversity of the microbial community explained more of the variation in the β-diversity of the viral community (34% of the variation) than did temperature, geophysical province, and/or geochemistry (12% of the variation). Additionally, the viral β-diversity is systematically lower than, and significantly correlated with, the microbial β-diversity (median dissimilarity difference 0.031). These patterns are maintained when accounting for only viruses and their linked hosts, though the median difference in dissimilarity is smaller (0.023). This is in stark contrast to observations made of viral-host interactions in the eutrophic coastal marine system of Osaka Bay where community dissimilarity was higher for the viral population as opposed to the prokaryotic population (62). This difference might be explained by differences in the preferred lifestyle (lytic vs lysogenic) of viruses in hydrothermal systems vs those in marine systems. Viral infections may proceed with one of two strategies: the lytic strategy of "plunder and pillage" or the lysogenic strategy of "batten down the hatches" (63,64). When bacteria hosts are in high density, as seen in coastal marine waters (62), viruses tend to choose the "plunder and pillage" strategy by quickly hijacking and augmenting the host metabolic and immune systems to divert resources to the production of viral progeny. However, when bacterial host densities are low, as seen in marine and terrestrial hydrothermal systems (4,48,65–67), it may benefit viruses to "batten down the hatches" by integrating into the host genome, potentially increase host fitness by augmenting the host metabolism

through AMGs (discussed below), and relying on host duplication for its own duplication (though there are other methods of lysogenic viral duplication (see Warwick-Dugdale et al. (63)). Our analysis found that 91% of the vOTUs that were linked to hosts were predicted to be lysogenic. Together, this suggests that lysogeny is more common in these viral populations than a lytic lifestyle.

Diversity of AMGs

Interestingly, 11 of the most abundant AMGs associated with viruses that are linked to MAGs in our data set have been found in diverse metagenomes and are thought to constitute a set of 14 conserved core AMGs (6,46). These include *dcm*, *cysH*, *folE*, *phnP*, *ubiG*, *moeB*, *ahbD*, *mec*, *queE*, *queD*, and *queC*. Kieft, Zhou, & Anantharaman (6) proposed that these core genes are likely central to viral replication regardless of host or environment. For example, the most abundant AMG in the viral community and within viruses linked to hosts is *dcm* across all provinces. Within prokaryotes, *dcm* is associated with restriction-modification systems (RM) (68). RM systems act as a type of immune system within prokaryotes by cleaving foreign DNA and also preventing the cleavage of the prokaryotic DNA by mobile genetic elements (68). However, RM expression in Δ*dcm* mutants results in extensive cell genomic degradation and severe loss in cell viability (69). This suggests that *dcm* serves to protect against genomic degradation by modulating RM activity. Viruses that contain *dcm* could protect themselves from host-induced cleavage (70,71).

Another possible mechanism to protect viruses from host immune systems is the presence of genes associated with the 7-cyano-7-deazaguanine biosynthesis pathway (preQ0 and preQ1). This biosynthesis pathway constructs queuosine, a hyper-modified guanosine analogue, which is used to modify tRNA by replacing guanine (72–74). This is hypothesized to allow improved reading frame maintenance within prokaryotes (75). Some evidence suggests that within the viral context the preQ0 biosynthesis pathway is used to modify the viral genome to protect itself against host restriction endonucleases (76). However, it has also been suggested that this pathway may benefit the host through relieving a metabolic bottleneck by supplementing (preQ0 present only) or altogether replacing (preQ0 and preQ1 present) a host biosynthetic function (73). While the genes associated with this pathway are found throughout all

geophysical provinces, within viruses linked to hosts, their highest abundance (31% of AMGs) is in the active volcanic backarc, compared to 2-13% elsewhere. An unclassified virus contains all the genes for preQ0 biosynthesis is highly abundant in the backarc, and it has a Proteobacteria host. However, it is unknown whether the higher abundance in the backarc reflects a particular environmental driver, or if it is just random.

In the Costa Rica outer forearc *moeB* and *cysH* are two of the most abundant AMGs. Within prokaryotes, *moeB* is involved in the sulfur relay system while *cysH* is involved in assimilatory sulfate reduction (77). These genes have also been found to be abundant viral AMGs in coastal algae blooms (78). The fact that we find them in such a different setting suggests that they may be generally advantageous to viral processes. *cysH* encodes an enzyme involved in the oxidation of inorganic sulfur compounds (79) and is involved in assimilatory sulfate reduction. *moeB* participates in the biosynthesis of molybdopterin, a cofactor involved in many cellular metabolic processes (80). *moeB* viral AMG of the algae *Ulva prolifera* is found to be more abundant after a blooming event while *cysH* is found to be more abundant during the bloom (78). This suggests that *cysH* may be involved in helping the host with energy production, while *moeB* helps in viral replication when the viruses are ready to go lytic. These may be particularly important in the outer forearc, due to the greater abundance of sulfate metabolisms in this province relative to the others (1).

Conclusion

While regional geochemistry and temperature did explain some of the variation in the viral communities, this variation was better explained by variation in the prokaryotic community. Further, we found that viral community richness and dissimilarity are positively correlated with the prokaryotic community. Viral community evenness was high across all geophysical provinces, potentially explained by their high host-specificity which would allow for a higher rate of niche partitioning and less intraspecific composition. Additionally, 91% of all viruses that were paired to host were predicted to be lysogenic, suggesting a largely lysogenic lifestyle for viruses in the terrestrial hydrothermal systems of

Central America. *Proteobacteria* and *Bacteriodetes* were the most often phylum that was host matched to the vOTUs, likely due to their fast growth rate relative to other members of the community. Surprisingly we found that WL- and rTCA-containing host were only linked to unclassified viruses, suggesting our knowledge of the viruses that target these types of chemolithoautotrophs is very limited. We found that only 13% of viruses that were linked to host contained AMGs. Interestingly, the highest abundant AMGs across all provinces were associated with evading host restriction systems through avoidance of genomic degradation or utilization of $preQ_0$. The genes associated with these host avoidance systems have been observed across a diversity of environments, lending support to the hypothesis of globally conserved core viral AMGs (6).

References

1. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, et al. Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME Journal. 2022;(October):1–11.
2. Fullerton KM, Schrenk MO, Yücel M, Manini E, Basili M, Rogers TJ, et al. Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. Nature Geoscience. 2021;14(5):301–6.
3. McKay LJ, Nigro OD, Dlakić M, Luttrell KM, Rusch DB, Fields MW, et al. Sulfur cycling and host-virus interactions in Aquificales-dominated biofilms from Yellowstone's hottest ecosystems. ISME Journal. 2022;16(3):842–55.
4. Anderson RE, Brazelton WJ, Baross JA. Is the genetic landscape of the deep subsurface biosphere affected by viruses? Frontiers in Microbiology. 2011;2(NOV):1–16.
5. Castelán-Sánchez HG, Lopéz-Rosas I, García-Suastegui WA, Peralta R, Dobson ADW, Batista-García RA, et al. Extremophile deep-sea viral communities from hydrothermal vents: Structural and functional analysis. Marine Genomics. 2019;46(January):16–28.
6. Kieft K, Zhou Z, Anantharaman K. VIBRANT: Automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. Microbiome. 2020 (June) 10;8(1).
7. Bragg JG, Chisholm SW. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. PLoS ONE. 2008;3(10):1–9.
8. Mann NH, Cook A, Bailey S, Clokie M, Amanullah A, Azam N, et al. Bacterial photosynthesis genes in a virus. Nature. 2003;424(August):741–2.
9. Bekaert D V, Gazel E, Turner S, Behn MD, Moor JM De, Zahirovic S, et al. High 3He/4He in central Panama reveals a distal connection to the Galapagos plume David. PNAS. 2021;118(47):1–8.
10. de Boer JZ, Defant MJ, Stewart RH, Restrepo JF, Clark LF, Ramirez AH. Quaternary calc-

alkaline volcanism in western Panama: Regional variation and implication for the plate tectonic framework. Journal of South American Earth Sciences. 1988;1(3):275–93.

11. Protti M, Gündel F, McNally K. The geometry of the Wadati-Benioff zone under southern Central America and its tectonic significance: results from a high-resolution local seismographic network. Physics of the Earth and Planetary Interiors. 1994;84(1–4):271–87.
12. Barnes JD, Cullen J, Barker S, Agostini S, Penniston-Dorland S, Lassiter JC, et al. The role of the upper plate in controlling fluid-mobile element (Cl, Li, B) cycling through subduction zones: Hikurangi forearc, New Zealand. Geosphere. 2019;15(3):642–58.
13. Clift P, Vannucchi P. Controls on tectonic accretion versus erosion in subduction zones: Implications for the origin and recycling of the continental crust. Review of Geophysics. 2004;42(2):1–31.
14. Rüpke LH, Morgan JP, Hort M, Connolly JAD. Serpentine and the subduction zone water cycle. Earth and Planetary Science Letters. 2004;223(1–2):17–34.
15. Patino L, Carr M, Feigenson M. Petrology and geochemistry of lavas. Central America. 2007;1–26.
16. Carr MJ, Feigenson MD, Bennett EA. Incompatible element and isotopic evidence for tectonic control of source mixing and melt extraction along the Central American arc. Contributions to Mineralogy and Petrology. 1990;105(4):369–80.
17. Gazel E, Carr MJ, Hoernle K, Feigenson MD, Szymanski D, Hauff F, et al. Galapagos-OIB signature in southern Central America: Mantle refertilization by arc-hot spot interaction. Geochemistry, Geophysics, Geosystems. 2009;10(2).
18. Harris RN, Wang K. Thermal models of the Middle America Trench at the Nicoya. Geophysical Research Letters. 2010;29(21):1–4.
19. Vetriani C, Jannasch HW, Macgregor BJ, Stahl DA, Reysenbach AL. Population structure and phylogenetic characterization of marine benthic Archaea in deep-sea sediments. Applied and Environmental Microbiology. 1999;65(10):4375–84.

20. Yu-Wei Wu, Yung-Hsu Tang, Susannah G Tringe BAS and SWS. MaxBin: an automated binning method to recover individual genomes from metagenomes using. Microbiome. 2014;2(1):1–18.

21. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ. 2019;7:e7359.

22. Olm MR, Brown CT, Brooks B, Banfield JF. DRep: A tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME Journal. 2017;11(12):2864–8.

23. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications. 2018;9(1):1–8.

24. Chaumeil P, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics. 2020;36(6):1925–1927.

25. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: An advanced analysis and visualization platformfor 'omics data. PeerJ. 2015;2015(10):1–29.

26. Nayfach S, Camargo AP, Schulz F, Eloe-Fadrosh E, Roux S, Kyrpides NC. CheckV assesses the quality and completeness of metagenome-assembled viral genomes. Nature Biotechnology. 2021;39(5):578–85.

27. Gregory AC, Zablocki O, Zayed AA, Howell A, Bolduc B, Sullivan MB. The Gut Virome Database Reveals Age-Dependent Patterns of Virome Diversity in the Human Gut. Cell Host and Microbe. 2020;28(5):724-740.e8.

28. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, et al. CRISPR Recognition Tool (CRT): A tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics. 2007;8:1–8.

29. Lowe TM, Eddy SR. TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research. 1996;25(5):955–64.

30. Galiez C, Siebert M, Enault F, Vincent J, Söding J. WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. Bioinformatics. 2017;33(19):3113–4.

31. Doug Hyatt, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, , Frank W Larimer LJH. Prodigal: prokaryotic gene recognition and translation initiation site identification. Nature Communications. 2010;11(1):1–8.

32. Winter DJ. rentrez: An R package for the NCBI eUtils API. R Journal. 2018;3(31):520–6.

33. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature Methods. 2012;9(4):357–9.

34. Jost L. Entropy and diversity. Oikos. 2006;113(2):363–75.

35. Charney N, Record S. Package "Vegetarian." 2016.

36. Oksanen J. Vegan: an introduction to ordination. 2015.

37. Robert A, Hijmans J, Karney C, Williams E, Vennes C, Hijmans MRJ. Package ' geosphere .' 2022.

38. Grömping U. Relative importance for linear regression in R: The package relaimpo. Journal of Statistical Software. 2006;17(1):1–27.

39. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature Biotechnology. 2017;35(8):725–31.

40. Sheik CS, Reese BK, Twing KI, Sylvan JB, Grim SL, Schrenk MO, et al. Identification and removal of contaminant sequences from ribosomal gene databases: Lessons from the Census of Deep Life. Frontiers in Microbiology. 2018;9(840):1–14.

41. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF, et al. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biology. 2014;12(1):1–12.

42. Li Z, Pan D, Wei G, Pi W, Zhang C, Wang JH, et al. Deep sea sediments associated with cold seeps are a subsurface reservoir of viral diversity. ISME Journal. 2021;15(8):2366–78.

43. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. Nature. 2016;536(7617):425–30.

44. Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: Mining viral signal from microbial genomic data. PeerJ. 2015;2015(5):1–20.

45. Zhao J, Wang Z, Li C, Shi T, Liang Y, Jiao N, et al. Significant Differences in Planktonic Virus Communities Between "Cellular Fraction" (0.22 ~ 3.0 μm) and "Viral Fraction" (< 0.22 μm) in the Ocean. Microbial Ecology. 2022;1–8.

46. Heyerhoff B, Engelen B, Bunse C. Auxiliary Metabolic Gene Functions in Pelagic and Benthic Viruses of the Baltic Sea. Frontiers in Microbiology. 2022;13(July):1–16.

47. Li Z, Moskowitz BM, Zheng J, Xiong Q, Zhou X, Yang J, et al. Petromagnetic Characteristics of Serpentinization and Magnetite Formation at the Zedang Ophiolite in Southern Tibet. Journal of Geophysical Research: Solid Earth. 2020;125(9):1–15.

48. Breitbart M. Marine Viruses: Truth or Dare. Annual Review of Marine Science. 2012;4(1):425–48.

49. Gulyaeva A, Garmaeva S, Kurilshikov A, Vich Vila A, Riksen NP, Netea MG, et al. Diversity and Ecology of Caudoviricetes Phages with Genome Terminal Repeats in Fecal Metagenomes from Four Dutch Cohorts. Viruses. 2022;14(10).

50. Ramin KI, Allison SD. Bacterial Tradeoffs in Growth Rate and Extracellular Enzymes. Frontiers in Microbiology. 2019;10(2956):1–10.

51. Haven N. Bacteriophage Ecology. Bacteriophage Ecology. 2008;42–57.

52. Qatibi AI, Bennisse R, Jana M, Garcia JL. Anaerobic degradation of glycerol by Desulfovibrio fructosovorans and D. carbinolicus and evidence for glycerol-dependent utilization of 1,2-propanediol. Current Microbiology. 1998;36(5):283–90.

53. Eichorst SA, Breznak JA, Schmidt TM. Isolation and characterization of soil bacteria that define Terriglobus gen. nov., in the phylum Acidobacteria. Applied and Environmental Microbiology. 2007;73(8):2708–17.

54. Mojica KDA, Brussaard CPD. Factors affecting virus dynamics and microbial host-virus interactions in marine environments. FEMS Microbiology Ecology. 2014;89(3):495–515.

55. Plümper O, King HE, Geisler T, Liu Y, Pabst S, Savov IP, et al. Subduction zone forearc serpentinites as incubators for deep microbial life. Proceedings of the National Academy of Sciences of the United States of America. 2017 Apr;114(17):4324–9.

56. Barry PH, de Moor JM, Giovannelli D, Schrenk M, Hummer DR, Lopez T, et al. Forearc carbon sink reduces long-term volatile recycling into the mantle. Nature. 2019;568(7753):487–92.

57. Mollentze N, Streicker DG. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. Proceedings of the National Academy of Sciences of the United States of America. 2020;117(17):9423–30.

58. Moreno-Gallego JL, Chou SP, Di Rienzi SC, Goodrich JK, Spector TD, Bell JT, et al. Virome Diversity Correlates with Intestinal Microbiome Diversity in Adult Monozygotic Twins. Cell Host and Microbe. 2019;25(2):261-272.e5.

59. Abbott KC, Eppinga MB, Umbanhowar J, Baudena M, Bever JD. Microbiome influence on host community dynamics: Conceptual integration of microbiome feedback with classical host–microbe theory. Ecology Letters. 2021;24(12):2796–811.

60. Finke DL, Snyder WE. Niche partitioning increases resource exploitation by diverse communities. Science. 2008;321(5895):1488–90.

61. Sonkoly J, Kelemen A, Valkó O, Deák B, Kiss R, Tóth K, et al. Both mass ratio effects and community diversity drive biomass production in a grassland experiment. Scientific Reports. 2019;9(1):1–10.

62. Tominaga K, Ogawa-Haruki N, Nishimura Y, Watai H, Yamamoto K, Ogata H, et al. Prevalence of Viral Frequency-Dependent Infection in Coastal Marine Prokaryotes Revealed Using Monthly Time Series Virome Analysis. mSystems. 2023;8(1):1–17.

63. Warwick-Dugdale J, Buchholz HH, Allen MJ, Temperton B. Host-hijacking and planktonic piracy: How phages command the microbial high seas. Virology Journal. 2019;16(1):1–13.

64. Luo XQ, Wang P, Li JL, Ahmad M, Duan L, Yin LZ, et al. Viral community-wide auxiliary metabolic genes differ by lifestyles, habitats, and hosts. Microbiome. 2022;10(1):1–18.

65. Williamson SJ, Cary SC, Williamson KE, Helton RR, Bench SR, Winget D, et al. Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. ISME Journal. 2008;2(11):1112–21.

66. Schoenfeld T, Patterson M, Richardson PM, Wommack KE, Young M, Mead D. Assembly of viral metagenomes from Yellowstone hot springs. Applied and Environmental Microbiology. 2008;74(13):4164–74.

67. Sharma A, Schmidt M, Kiesel B, Mahato NK, Cralle L, Singh Y, et al. Bacterial and Archaeal Viruses of Himalayan Hot Springs at Manikaran Modulate Host Genomes. Frontiers in Microbiology. 2018;9(December):1–15.

68. Militello KT, Simon RD, Qureshi M, Maines R, van Horne ML, Hennick SM, et al. Conservation of Dcm-mediated cytosine DNA methylation in Escherichia coli. FEMS Microbiology Letters. 2012;328(1):78–85.

69. Takahashi N, Naito Y, Handa N, Kobayashi I. A DNA methyltransferase can protect the genome from postdisturbance attack by a restriction-modification gene complex. Journal of Bacteriology. 2002;184(22):6100–8.

70. Kęsik-Szeloch A, Drulis-Kawa Z, Weber-Dąbrowska B, Kassner J, Majkowska-Skrobek G, Augustyniak D, et al. Characterising the biology of novel lytic bacteriophages infecting multidrug resistant Klebsiella pneumoniae. Virology Journal. 2013;10:1–12.

71. Murphy J, Mahony J, Ainsworth S, Nauta A, van Sinderen D. Bacteriophage orphan DNA methyltransferases: Insights from their bacterial origin, function, and occurrence. Applied and Environmental Microbiology. 2013;79(24):7547–55.

72. Reader JS, Metzgar D, Schimmel P, De Crécy-Lagard V. Identification of Four Genes Necessary for Biosynthesis of the Modified Nucleoside Queuosine. Journal of Biological Chemistry. 2004;279(8):6280–5.

73. Sazinas P, Redgwell T, Rihtman B, Grigonyte A, Michniewski S, Scanlan DJ, et al. Comparative genomics of bacteriophage of the genus seuratvirus. Genome Biology and Evolution. 2017;10(1):72–6.

74. Wang R, Lin J, Liu X, Pang X, Zhang C, Yang C, et al. Sulfur Oxidation in the Acidophilic Autotrophic Acidithiobacillus spp . 2019;9(January):1–20.

75. Urbonavičius J, Qian Q, Durand JMB, Hagervall TG, Björk GR. Improvement of reading frame maintenance is a common function for several tRNA modifications. EMBO Journal. 2001;20(17):4863–73.

76. Kulikov EE, Golomidova AK, Letarova MA, Kostryukova ES, Zelenin AS, Prokhorov NS, et al. Genomic sequencing and biological characteristics of a novel Escherichia Coli bacteriophage 9g, a putative representative of a new siphoviridae genus. Viruses. 2014;6(12):5077–92.

77. Henriques AC, De Marco P. Methanesulfonate (MSA) catabolic genes from marine and estuarine bacteria. PLoS ONE. 2015;10(5):1–15.

78. Han M, Sun J, Yang Q, Liang Y, Jiang Y, Gao C, et al. Spatiotemporal Dynamics of Coastal Viral Community Structure and Potential Biogeochemical Roles Affected by an Ulva prolifera Green Tide. mSystems. 2023;8(2):1–14.

79. Hesketh-Best PJ, Bosco-Santos A, Garcia SL, O'Beirne MD, Werne JP, Gilhooly WP, et al. Viruses of sulfur oxidizing phototrophs encode genes for pigment, carbon, and sulfur metabolisms. Communications Earth and Environmentm. 2023;4(1):1–12.

80. Leimkühler S, Wuebbens MM, Rajagopalan K V. Characterization of Escherichia coli MoeB and Its Involvement in the Activation of Molybdopterin Synthase for the Biosynthesis of the Molybdenum Cofactor. Journal of Biological Chemistry. 2001;276(37):34695–701.

Appendix I: Tables and Figures

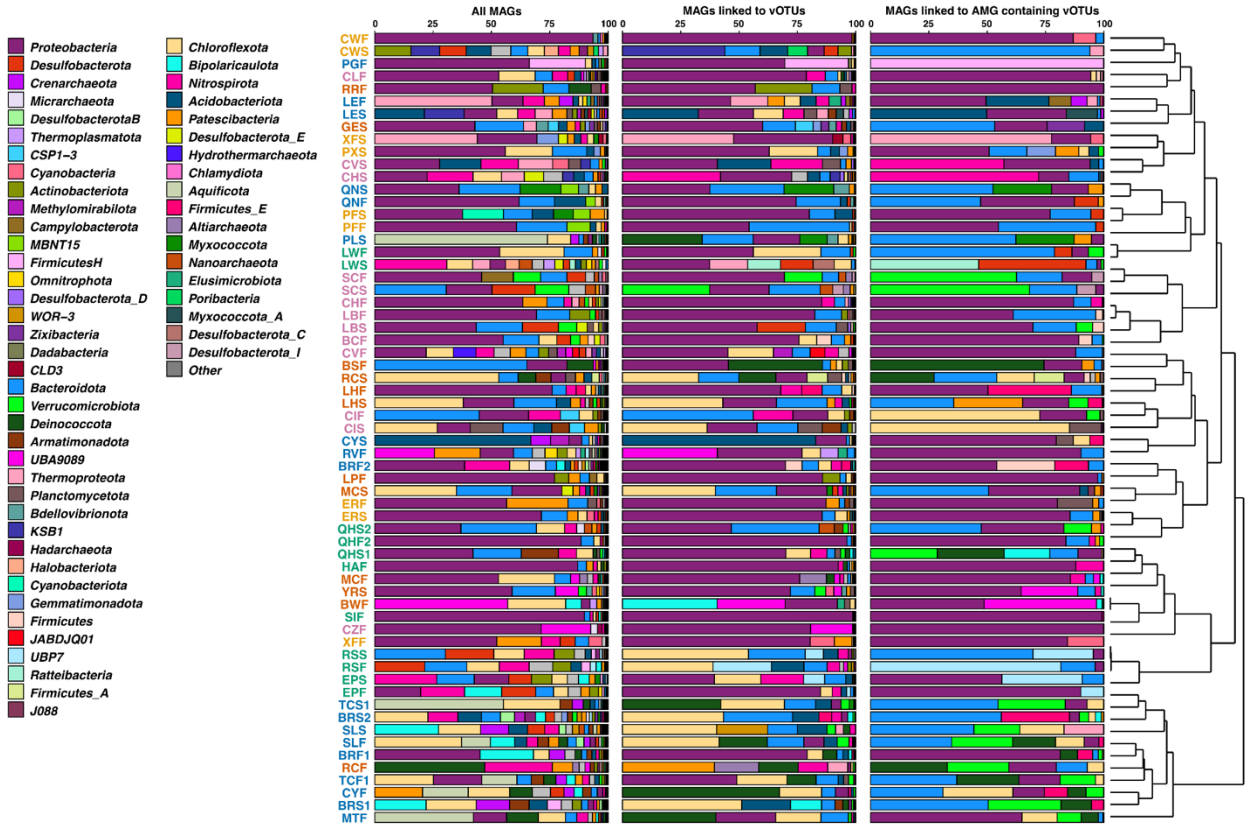


Figure 3.1. Differences in prokaryotic community composition between groups. The right most panel contains all MAGs, the middle panel contains all MAGs that have been linked to vOTUs, and the left most panel contains MAGs that are linked to vOTUs that contain AMGs. Hierarchical clustering of all panels is based on Spearman rank correlation-distances sites with MAGs linked to vOTUs that contain AMGs. Phylum colors were chosen based on the top five most abundant phylum within each site.

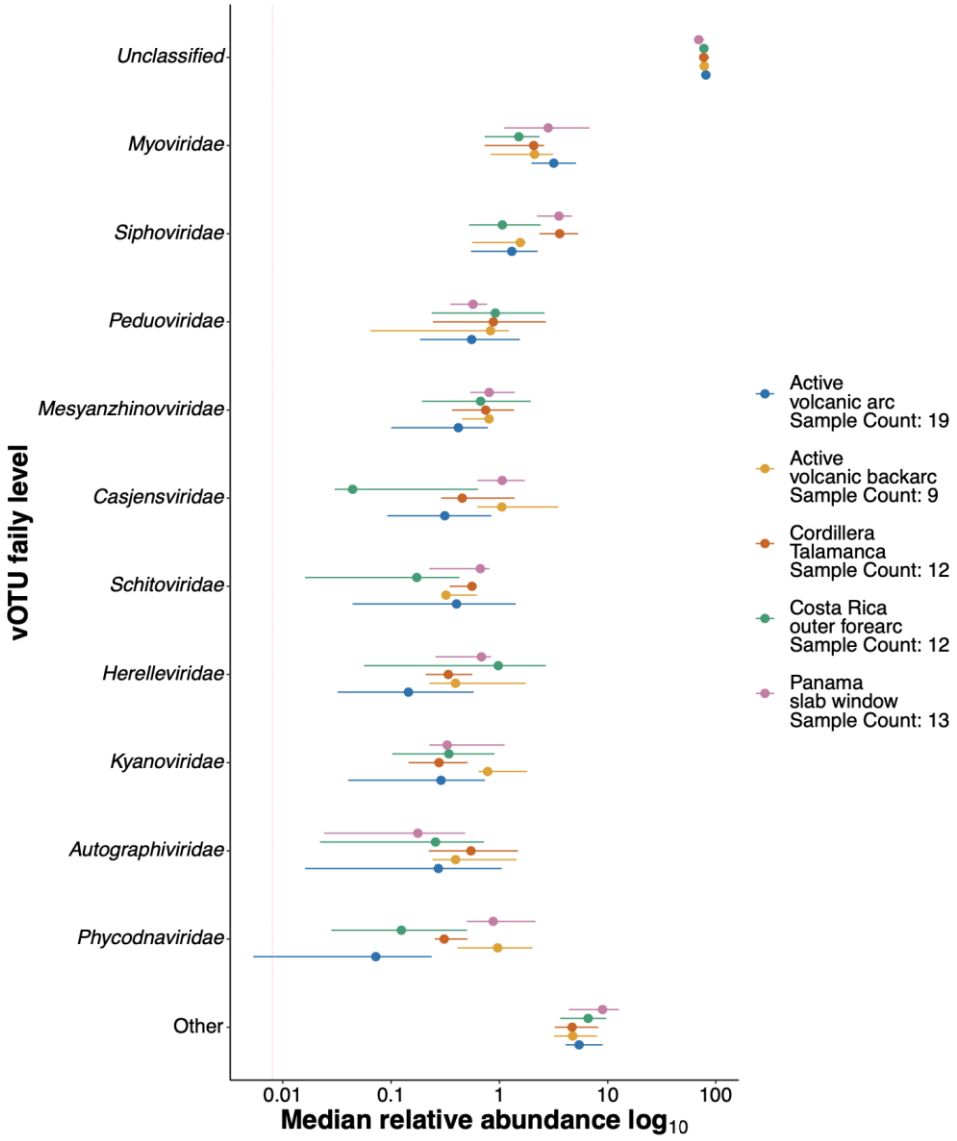


Figure 3.2. Log_{10} relative abundance of vOTUs across all provinces. Top 10 viral families were chosen based on their median relative abundance across all sites.

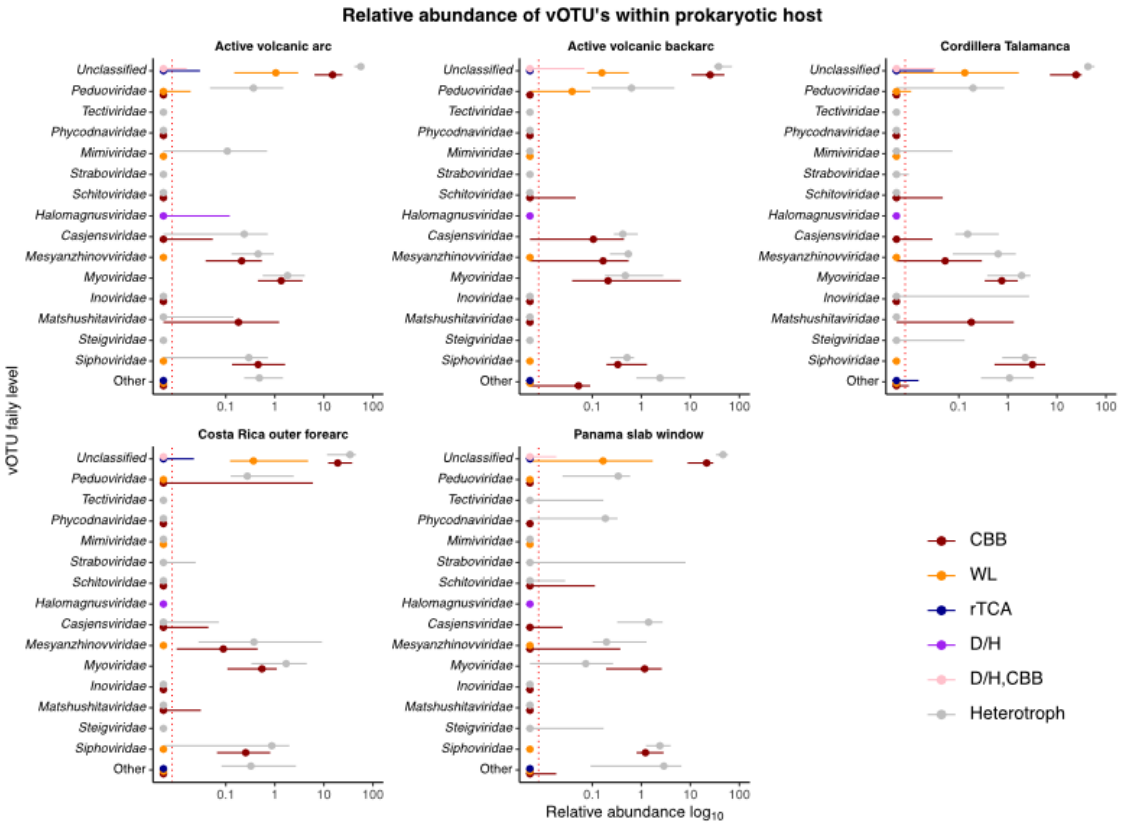


Figure 3.3. Log_{10} relative abundance of vOTUs colored by their linked host trophic level.

Represented are the top 15 most abundant viral families across all provinces. “Other” represents vOTUs that were < 0.01% abundant.

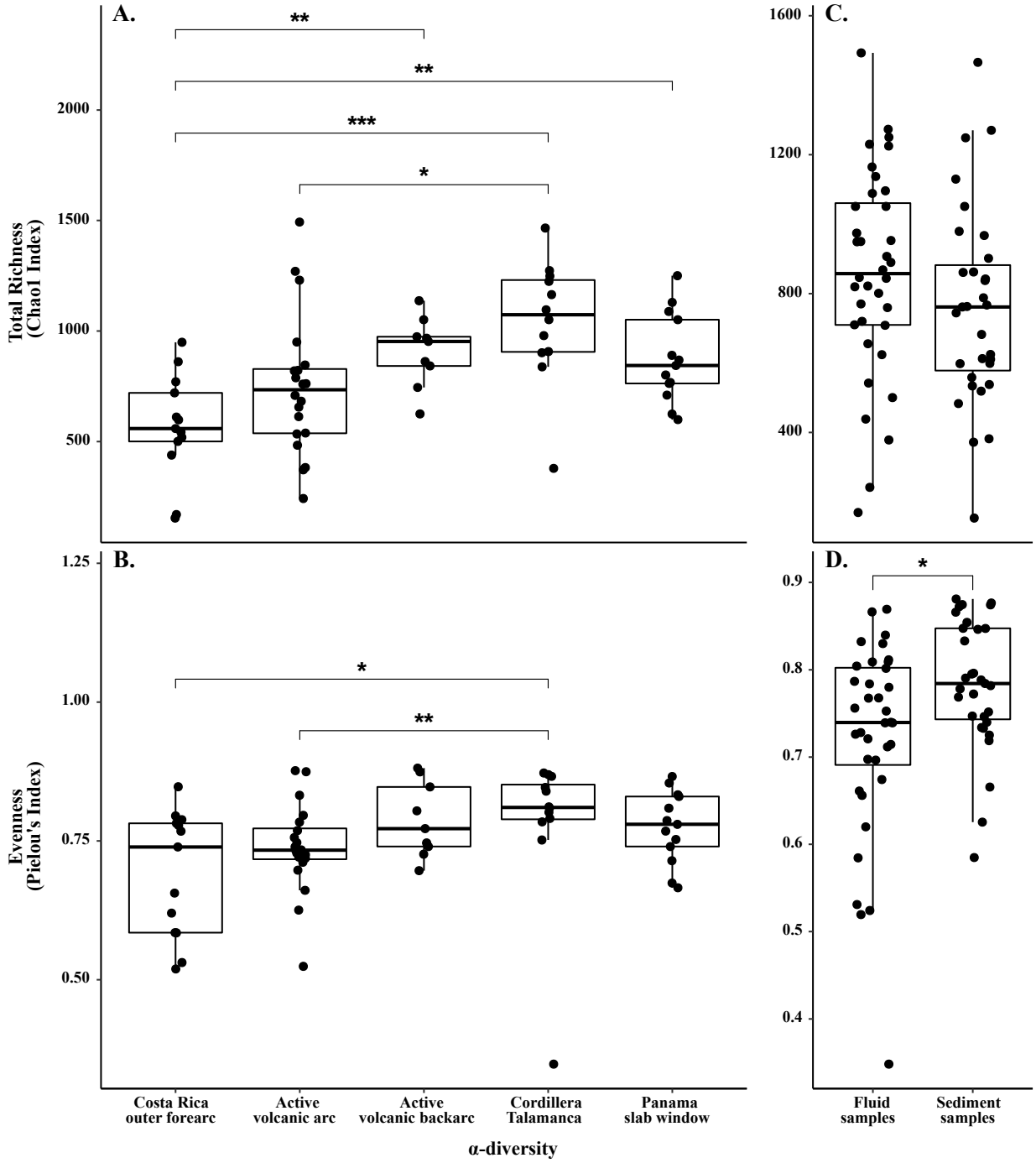
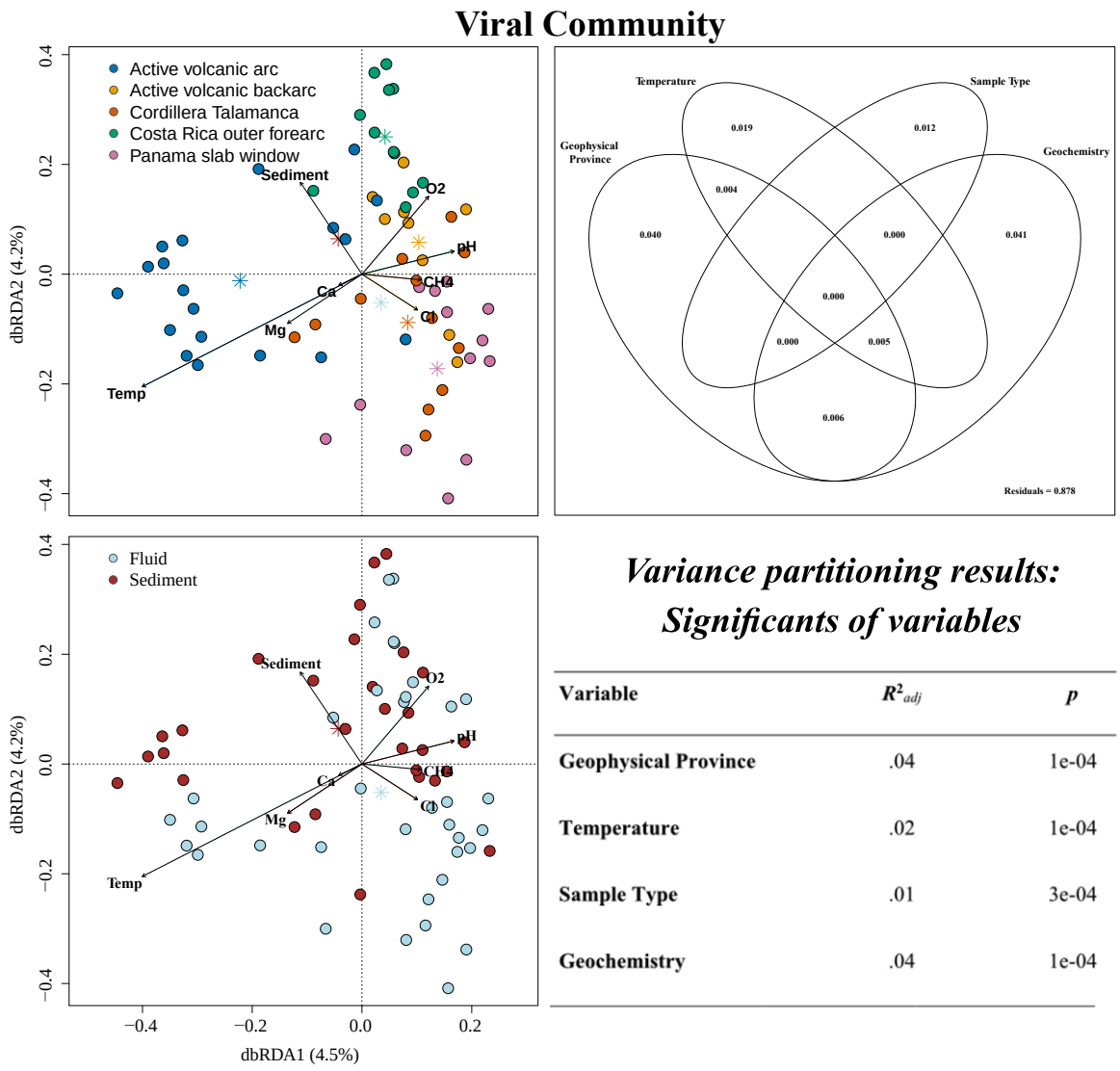


Figure 3.4. Comparison of viral community α-diversity between geological provinces and between sample types. **A** and **B** are the Pleiou's evenness and Chao1 total richness indexes between the geological provinces respectively. **C** and **D** are the Pleiou's evenness and Chao1 total richness indexes between sample types (fluids and sediments) respectively. Asterisks denote significance, with * indicating $p<0.05$, ** indicating $p<0.01$, and *** indicating $p<0.001$. Significance was tested using t-test.



Variable	R^2_{adj}	p
Geophysical Province	.04	1e-04
Temperature	.02	1e-04
Sample Type	.01	3e-04
Geochemistry	.04	1e-04

Figure 3.5. Viral community β-diversity analysis. Panels A and B are distance-based RDA analysis colored by A. geophysical province and B. sample type (fluids and sediments).

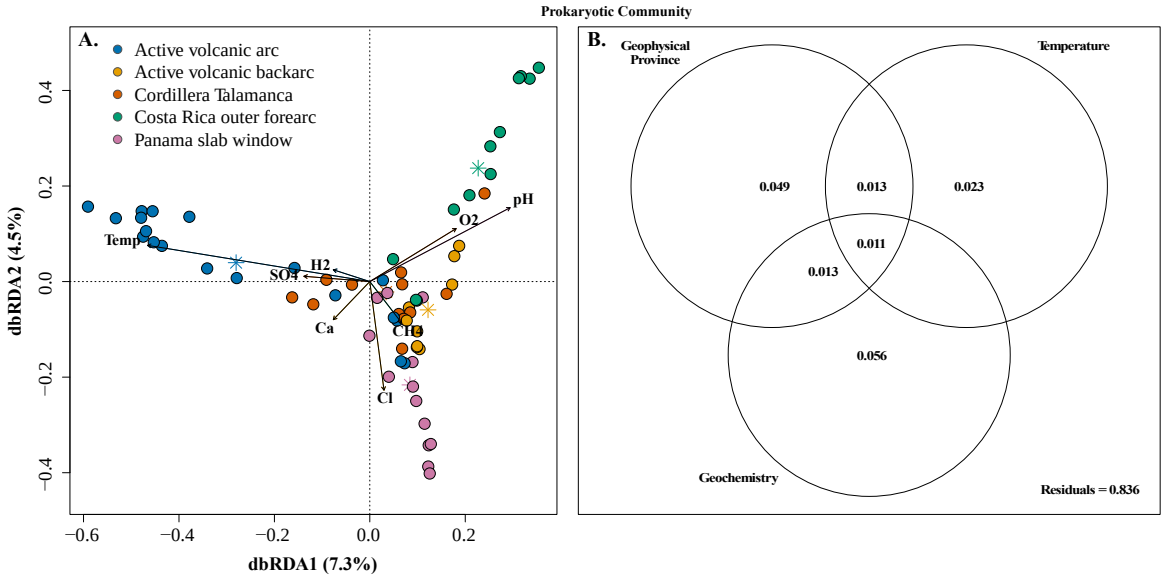


Figure 3.6. Microbial community β-diversity analysis. Panel **A** is a distance-based RDA analysis colored by geophysical province.

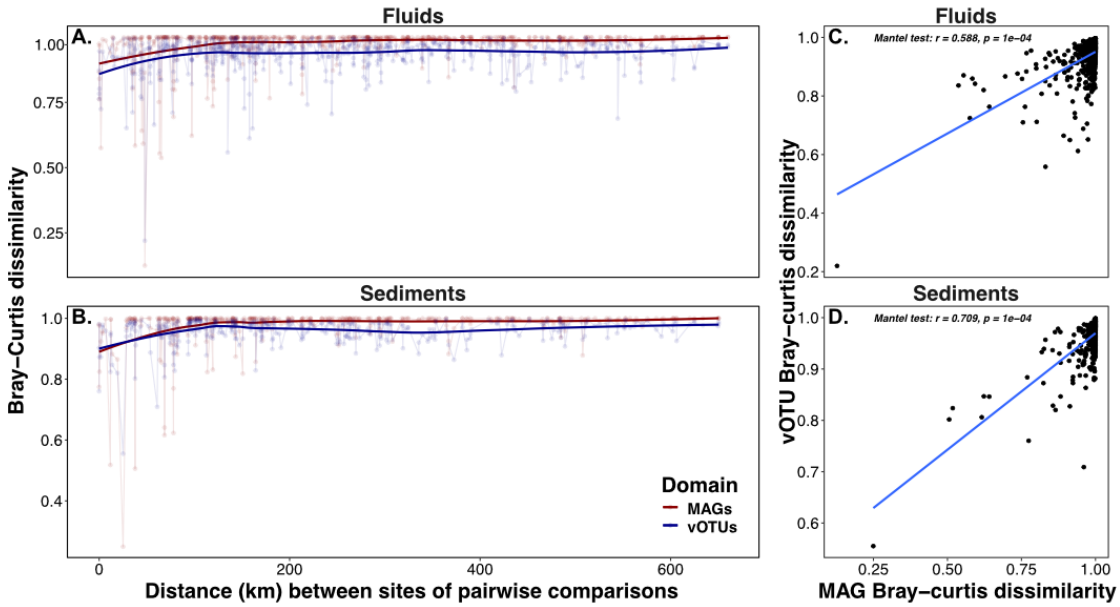


Figure 3.7. Microbial vs viral community dissimilarities. Panels **A** and **B** are the dissimilarity comparisons of the fluids and sediments. Panels **C** and **D** show the regressions of the two dissimilarity matrices against each other for fluids and sediments. A Mantel test was conducted on the fluid and sediment matrices to test significance of the correlations.

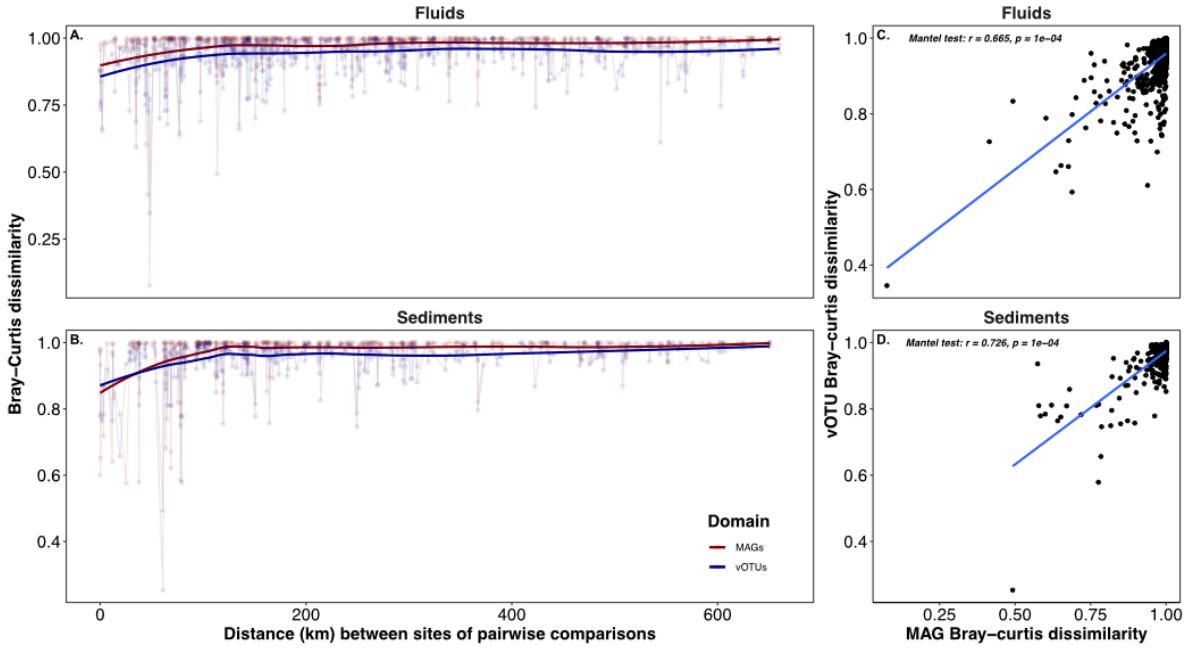


Figure 3.8. Microbial host vs linked viral dissimilarities. Panels **A** and **B** are the dissimilarity comparisons of the fluids and sediments. Panels **C** and **D** show the regressions of the two dissimilarity matrices against each other for fluids and sediments. A Mantel test was conducted on the fluid and sediment matrices to test significance of the correlations.

Table 3.1. Viral PERMANOVA results

Comparisons	Df	Sum. Sq.	R^2	F	P	P_{adj}
Fluid vs Sediments	1	0.73	0.03	1.62	0.001	0.001
Active volcanic arc vs Cordillera Talamonca	1	0.82	0.06	1.91	0.001	0.001
Active volcanic arc vs Panama	1	1.05	0.08	2.50	0.001	0.001
Active volcanic arc vs Active volcanic backarc	1	0.82	0.07	1.90	0.001	0.001
Active volcanic arc vs Costa Rica outer forearc	1	1.00	0.07	2.31	0.001	0.001
Cordillera Talamonca vs Panama	1	0.56	0.06	1.35	0.001	0.039
Cordillera Talamonca vs Active volcanic backarc	1	0.56	0.06	1.29	0.001	0.017
Cordillera Talamonca vs Costa Rica outer forearc	1	0.80	0.08	1.83	0.001	0.001
Panama vs Active volcanic backarc	1	0.71	0.08	1.69	0.001	0.001
Panama vs Costa Rica outer forearc	1	1.05	0.10	2.47	0.001	0.001

Table 3.2. Microbial variance partitioning results

Variable	R^2_{adj}	P
Geophysical Province	0.05	0.0001
Temperature	0.02	0.0001
Geochemistry	0.01	0.0001

Table 3.3. Microbial PERMANOVA results

Comparisons	Df	Sum. Sq.	R^2	F	P	P_{adj}
Active volcanic arc vs Cordillera Talamonca	1	1.03	0.08	2.38	0.00	0.001
Active volcanic arc vs Panama	1	1.39	0.10	3.27	0.00	0.001
Active volcanic arc vs Active volcanic backarc	1	1.17	0.10	2.78	0.00	0.001
Active volcanic arc vs Costa Rica outer forearc	1	1.67	0.12	4.08	0.00	0.001
Cordillera Talamonca vs Panama	1	0.78	0.07	1.67	0.00	0.001
Cordillera Talamonca vs Active volcanic backarc	1	0.68	0.07	1.44	0.00	0.003
Cordillera Talamonca vs Costa Rica outer forearc	1	0.89	0.08	1.99	0.00	0.001
Panama vs Active volcanic backarc	1	0.81	0.08	1.77	0.00	0.001
Panama vs Costa Rica outer forearc	1	1.17	0.10	2.68	0.00	0.001

Supplemental figures and tables:

Table 3.4. Table S3.1. Top 5 most abundant phyla within each geophysical province.

Geophysical Province	Phyla	Mean Relative Abundance
Active volcanic arc	*Proteobacteria*	18.16
Active volcanic arc	*Chloroflexota*	13.06
Active volcanic arc	*Aquificota*	11.68
Active volcanic arc	*Acidobacteriota*	8.32
Active volcanic arc	*Bipolaricaulota*	6.65
Active volcanic backarc	*Proteobacteria*	50.78
Active volcanic backarc	*Bacteroidota*	9.15
Active volcanic backarc	*Patescibacteria*	7.46
Active volcanic backarc	*Thermoproteota*	5.09
Active volcanic backarc	*Chloroflexota*	4.20
Cordillera Talamonca	*Proteobacteria*	35.76
Cordillera Talamonca	*Chloroflexota*	15.24
Cordillera Talamonca	*Bacteroidota*	15.23
Cordillera Talamonca	*Deinococcota*	6.47
Cordillera Talamonca	UBA9089	6.16
Costa Rica outer forearc	*Proteobacteria*	43.65
Costa Rica outer forearc	*Bacteroidota*	12.85
Costa Rica outer forearc	*Nitrospirota*	10.07
Costa Rica outer forearc	*Chloroflexota*	8.89
Costa Rica outer forearc	*Desulfobacterota*	6.19
Panama	*Proteobacteria*	40.65
Panama	*Bacteroidota*	13.62
Panama	*Chloroflexota*	6.45
Panama	*Nitrospirota*	5.29
Panama	*Desulfobacterota*	4.44

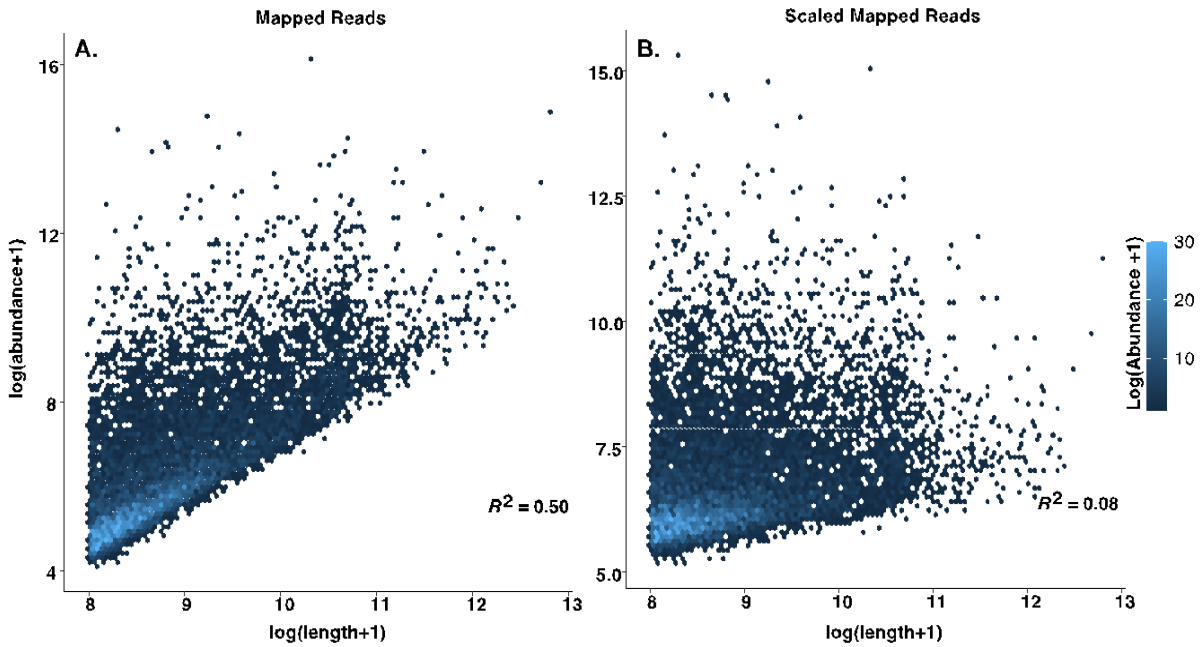


Figure S3.1. Comparison between mapped read counts and standardized mapped read counts of vOTUs. Pannal **A** demonstrates that mapped read counts are influenced by contig length. **B)** To account for this, mapped reads were scaled to the average viral contig length based on vOTU membership.

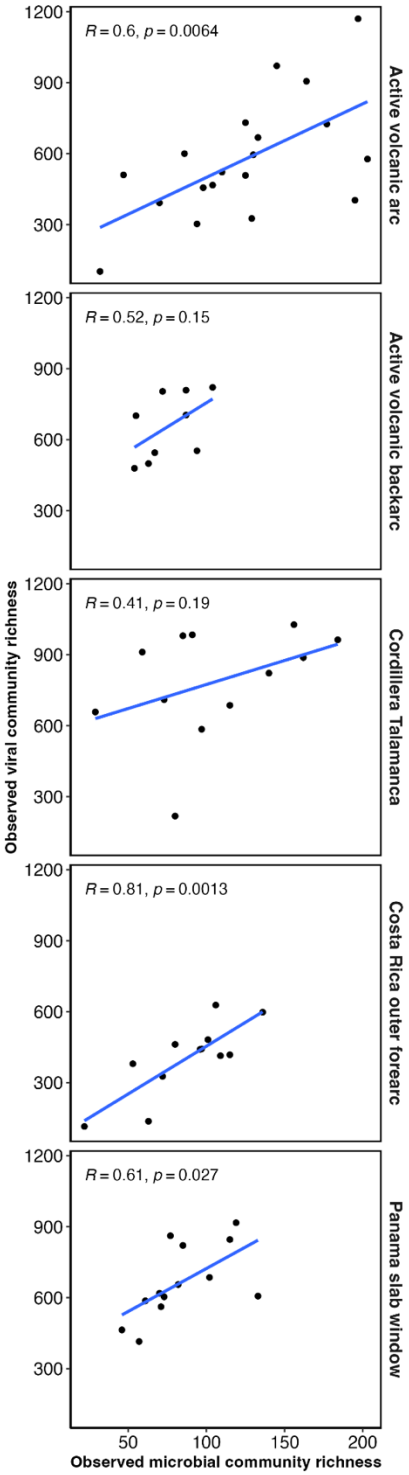


Figure S3.2. Pearson correlations between the richness of both the viral and prokaryotic communities.

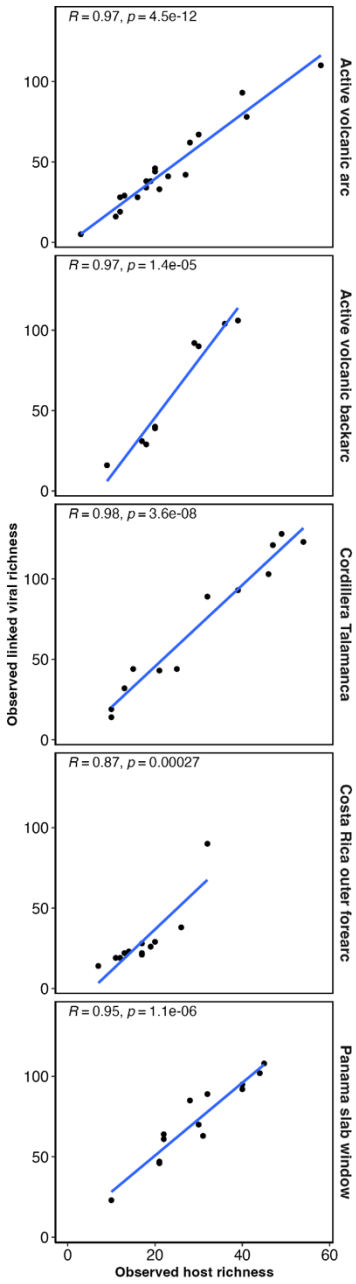


Figure S3.3. Pearson correlations between the richness of both the linked viruses and their hosts.

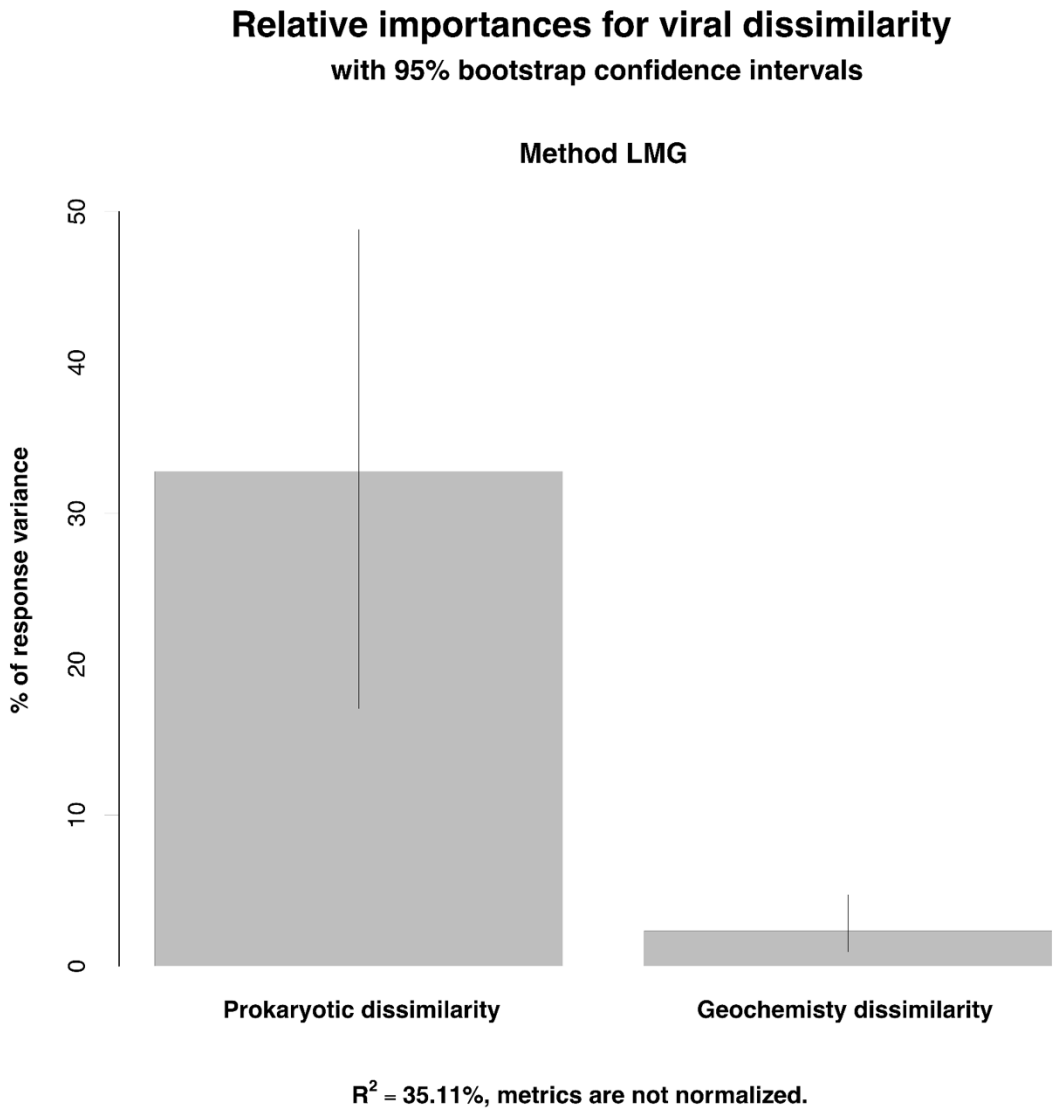


Figure S3.4. The results from the model-selection analysis for viral β-diversity with geochemistry and prokaryotic site dissimilarly. Variance explained was determined using the Lindeman, Merenda, and Gold (LMG) method of R^2 decomposition.

Chapter 4 Evidence of horizontal gene transfer of ATP-citrate lyase from methanogens to *Syntrophales* in hot spring fluids of the Central American Convergent margin

This original research is being prepared for future publication submission with at least the following authors:

Timothy J. Rogers and Karen G. Lloyd

My primary contributions to this paper include: (i) Investigations and data acquisition, (ii) data synthesis and statistical analysis, (iii) figure and table generation, (iv) and writing the manuscript.

Abstract

Syntrophales is an order of strictly anaerobic bacteria within *Thermodesulfobacteriota*. Cultured members within this order participate in syntrophic acetate oxidation (SAO), providing their syntrophic partner with CO_2 and H_2, making them obligate anaerobic chemoorganoheterotrophs. No member of this group, either from cultures or from culture-independent genomic research, is known to contain a carbon fixation pathway. We identified four metagenomic assembled genomes (MAGs) taxonomically classified as *Syntrophales* which contain the full reverse citric acid (rTCA) cycle. Through BLAST and phylogenic analysis, we describe the horizontal gene transfer of the rTCA's key enzyme ATP-citrate lyase from *Methanophagales* to *Syntrophales*. To our knowledge, this is the first report of rTCA within *Syntrophales* and the first description of horizontal gene transfer of ATP-citrate lyase from *Methanophagales* to *Syntrophales*.

Introduction

In chapter 2, we reported four *Thermodesulfobacteriota* (formally a class in the phylum *Desulfobacterota*) metagenomic assembled genomes (MAGs) were identified in deeply-sourced serpentinizing hot springs, classified at the order level as *Syntrophales* (1). Cultured members of *Syntrophales* are obligate anaerobic fermenters that can perform syntrophic acetate oxidation (SAO) and supply their H_2/formate-utilizing syntrophic partners (2) with CO_2 and H_2 (3). Being strictly chemoorganoheterotrophs (2), *Syntrophales* are not known to contain carbon fixation pathways. While members have been found to contain an incomplete citric acid cycle (TCA) (4,5), we reported that these MAGs contained the reductive TCA (rTCA) (1). While reactions are mostly conserved between the TCA and the rTCA, there are a few reactions that are specific to rTCA. One of the key reactions, the cleavage of citrate, is achieved via one of two paths: through the combination of citryl-CoA synthetase (CCS) and citryl-CoA lyase (CCL) or the use of ATP-citrate lyase (ACL) is used (6,7). It is currently understood that ACL likely resulted from the fusion of CCS and CCL (8). All four of our MAGs contain both ACL subunits (*aclA* and *aclB*).

Due to rTCA genes being found in bacteria and archaea (9), it has been proposed to be a potential primordial pathway, possibly found in the Last Universal Common Ancestor (LUCA) (10). This is supported by multiple reactions within the rTCA being experimentally demonstrated to occur abiotically with non-enzymatic catalysts (using Zn^{2+}, Cr^{3+} and Fe^0 (10,11)). However, evidence of extensive horizonal gene transfer (HGT) of the rTCA's key enzyme ATP-citrate lyase (ACL) and its homologs calls into question its presence in LUCA (9).

HGT allows prokaryotes, which largely reproduce asexually, to maintain genetic diversity by allowing for the exchange of genetic information within and between species (Soucy et al., 2015). Culture-independent and -omics approaches have shown that HGT events are a major force in the evolution of prokaryotic genomes (Boucher & Doolittle, 2000; Calteau et al., 2005). Such events are facilitated through transduction of mobile genetic elements, transformation of environmental DNA, as well as conjugation in which DNA is passed directly from one prokaryotic community member to the other

(Soucy et al., 2015). The latter of the three has been found extensively in syntrophic interactions between methanogenic archaea and the bacteria phylum *Thermodesulfobacteriota*. For example, the orange protein (ORP) gene complex in *Desulfovibrio vulgaris* was acquired via HGT from *Methanosarcina barkeri* (in the order *Methanophagales*) (Scholten et al., 2007). This gene complex is implicated in cell division of *D. vulgaris* (Fiévet et al., 2011; Neca et al., 2015) and is downregulated during syntrophic growth and upregulated during sulfate reduction (Fiévet et al., 2011). While, outside of our previous study (Chapter 2), no report of rTCA has been reported in *Syntrophales*, rTCA has been reported in its syntrophic partners (12). We thus hypothesize that the ACL found in these *Syntrophales* was acquired through a HGT event from a syntrophic partner. We used a phylogenetic approach to test this hypothesis.

Methods

ACL α and β subunit sequences from Rogers et al. (1) were retrieved from four MAGs taxonomically identified as *Syntrophales* by GTDB-Tk. Blast searches of the ACLs α and β subunits were conducted using BLASTP on NCBI's Web BLAST. Published and known isolate representatives of the top BLAST hits were downloaded from the NCBI database and retained for phylogenetic analysis. Also included in the phylogenetic analysis were 33 sequences from Giovannelli et al. (6). Conserved marker genes of these representatives, along with an additional seven *Thermodesulfobacteriota* representatives, were also used for phylogenetic placement of the four MAGs to verify taxonomic assignment. Conserved gene identification, alignment, and concatenation was preformed using the anvi'o (13,14) phylogenomics workflow.

The amino acid fasta files of the ACL α and β subunits were imported into Rstudio using the Biostrings package (v 2.66.0 (15)). Multiple amino acid sequence alignments were performed on both the α and β subunits using ClustalW through the msa R package (v 1.30.1 (16)). The aligned subunits were then converted to phyDat objects, concatenated together, and saved in fasta format using the phangorn package (v 2.11.1 (17)). The CIPRES Science Gateway (18) was used for final phylogentic tree construction of both the conserved marker gene and ACL trees using the RAxML-HPC BlackBox with Le

and Gascuel (LG) gamma replacement matrix for the amino acid replacement model (19) with 1,000 bootstraps. Both trees were then imported into R and midpoint rooted using the Biostrings and phangorn package.

Results and discussion

Two of the *Syntrophales* MAGs (bin 51 and bin 351) contain the full rTCA pathway while the other two (bins 162 and 202) are only missing isocitrate dehydrogenase [EC: 1.1.1.42] (Fig. 4.1), likely due to their lower quality. Both bin 51 and 351 are 94% complete with < 3% contamination (redundancy), whereas bins 162 and 202 are 93% and 78% complete respectively with <2% redundancy.

The top 10 hits (published and known isolates) to the *Syntrophales* ACL α and β subunits are members of the archaeal order *Methanophagales*. While *Methanophagales* are not known to preform carbon fixation using the rTCA pathway, they have been found to contain a partial rTCA that is used for 2-oxoglutarate synthesis (12). Horizontal transfer of ACL between bacteria and archaea has been previously reported (9). It's possible that ACL was horizontally acquired by *Syntrophales* from *Methanophales*.

A phylogenic approach to characterizing past HGT events of ACL with the phylum *Aquificota* has released the occurrence of two separate HGT events within this phylum (6). All cultured members of *Aquificota* are chemolithoautrophs and are thought to be the earliest bacterial lineage (6,7,20,21). Interestingly, both the CCS/CCL (in *Aquificaceae*) and ACL (in *Hydrogenothermaceae* and *Desulfurobacteriaceae*) variants of the rTCA have been found within this phylum (6,7). As this phylogeny has been well tested, the sequences used in both Giovannelli et al (6) and Hügler et al. (7) were used to see where *Syntrophales* fall in relation to this phylogeny. This included 10 *Chlorobiota*, 12 *Canpylobacterota* (formally *Epsilonproteobacteria*), and 11 *Aquificota* ACL sequences. Additionally, as suggest by both Giovannelli et al (6) and Hügler et al. (7), I constructed an "ancestral" ACL by concatenating CCS and CCL from *Hydrogenobacter thermophilus* and *Aquifex aeolicus* (*Aquificaceae*).

Before conducting a phylogenetic analysis on the ACL sequence, I first wanted to verify taxonomic placement originally based on GTDB-TK (22) (from Chapter 2; (1)) of the *Sytrophales* MAGs. For this, I

downloaded the genomes of the top 10 hits to the ACL sequences, the genomes of from both Giovannelli et al (6) and Hügler et al. (7), as well as 7 *Thermodesulfobacteriota* genomes from NCIB. I found that the four *Syntrophales* MAGs fall with the *Syntrophales* clade (Fig. 4.2), confirming their taxonomic placement.

The phylogenetic analysis of the ACL sequences reveled that the *Syntrophales* ACL sequences fall within the *Methanophagales* clade, which forms a monophyletic group to the *Aquificota* and *Campylobacterota* (formally *Epsilonbacteraeota*) clade (Fig. 4.3). Hügler et al. (7) previously predicted that *Hydrogenothermaceae* likely acquired their ACL from *Campylobacterota* while Giovannelli et al (6) predicted that *Desulfurobacteriaceae* acquired theirs from *Hydrogenothermaceae* (Fig. 4.4). Based on our analysis, the most parsimonious interpretation of the data is that once ACL had been acquired by *Methanophagales* it was then transferred to *Syntrophales* (Fig. 4.4).

Conclusion

Syntrophales is a syntrophic organism that performs SAO to provide their hydrogenotrophic methanogenic partner with its needed CO_2 and H_2. In this paper we describe the HGT for ACL from methanogenic *Methanomicrobia* to *Syntrophales*. This suggests that the *Syntrophales* in the Costa Rica subduction zone are potentially facultative autotrophs. However, further research is needed to establish if these pathways are functioning for the purposes of carbon fixation.

References

1. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, et al. Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME Journal. 2022;(October):1–11.
2. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, et al. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia , and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. 2020;
3. Schöcke L, Schink B. Membrane-bound proton-translocating pyrophosphatase of Syntrophus gentianae, a syntrophically benzoate-degrading fermenting bacterium. European Journal of Biochemistry. 1998;256(3):589–94.
4. McInerney MJ, Rohlin L, Mouttaki H, Kim U, Krupp RS, Rios-Hernandez L, et al. The genome of Syntrophus aciditrophicus: Life at the thermodynamic limit of microbial growth. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(18):7600–5.
5. Metagenomic Characterization of Candidatus Smithella cisternae Strain M82 _ 1 , a Syntrophic Alkane-Degrading Bacteria , Enriched from the Shengli Oil Field. 2017;32(3):234–43.
6. Giovannelli D, Sievert SM, Hügler M, Markert S, Becher D, Schweder T, et al. Insight into the evolution of microbial metabolism from the deep-branching bacterium, thermovibrio ammonificans. eLife. 2017;6:1–31.
7. Hügler M, Huber H, Molyneaux SJ, Vetriani C, Sievert SM. Autotrophic CO_2 fixation via the reductive tricarboxylic acid cycle in different lineages within the phylum Aquificae: Evidence for two ways of citrate cleavage. Environmental Microbiology. 2007;9(1):81–92.
8. Aoshima M, Igarashi Y. Nondecarboxylating and decarboxylating isocitrate dehydrogenases: Oxalosuccinate reductase as an ancestral form of isocitrate dehydrogenase. Journal of Bacteriology. 2008;190(6):2050–5.
9. Sokolskyi T, Dassarma S. Role of horizontal gene transfers and microbial ecology in the evolution of fluxes through the tricarboxylic acid cycle. 2022;1–34.

10. Muchowska KB, Chevallot-Beroux E, Moran J. Recreating ancient metabolic pathways before enzymes. Bioorganic and Medicinal Chemistry. 2019;27(12):2292–7.
11. Muchowska KB, Varma SJ, Chevallot-Beroux E, Lethuillier-Karl L, Li G, Moran J. Metals promote sequences of the reverse Krebs cycle. Nature Ecology and Evolution. 2017;1(11):1716–21.
12. Chadwick GL, Skennerton CT, Laso-Pérez R, Leu AO, Speth DR, Yu H, et al. Comparative genomics reveals electron transfer and syntrophic mechanisms differentiating methanotrophic and methanogenic archaea. Vol. 20, PLoS Biology. 2022.
13. Eren AM, Esen OC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: An advanced analysis and visualization platformfor 'omics data. PeerJ. 2015;2015(10):1–29.
14. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology. 2021;6(1):3–6.
15. Delhomme N. Package ' Biostrings .' 2023;
16. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. Msa: An R package for multiple sequence alignment. Bioinformatics. 2015;31(24):3997–9.
17. Daru BH, Karunarathne P, Schliep K. phyloregion: R package for biogeographical regionalization and macroecology. Methods in Ecology and Evolution. 2020;11(11):1483–91.
18. Miller MA, Pfeiffer W, Schwartz T. Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. 2010;
19. Le SQ, Gascuel O. An Improved General Amino Acid Replacement Matrix. 2001;
20. Di Giulio M. The universal ancestor lived in a thermophilic or hyperthermophilic environment. Journal of Theoretical Biology. 2000;203(3):203–13.
21. Pitulle C, Yang Y, Marchiani M, Moore ERB, Siefert JL, Aragno M, et al. Phylogenetic position of the genus Hydrogenobacter. International Journal of Systematic Bacteriology. 1994;44(4):620–6.
22. Chaumeil P, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome

Taxonomy Database. Bioinformatics. 2020;36(6):1925–1927.

Appendix I: Figures

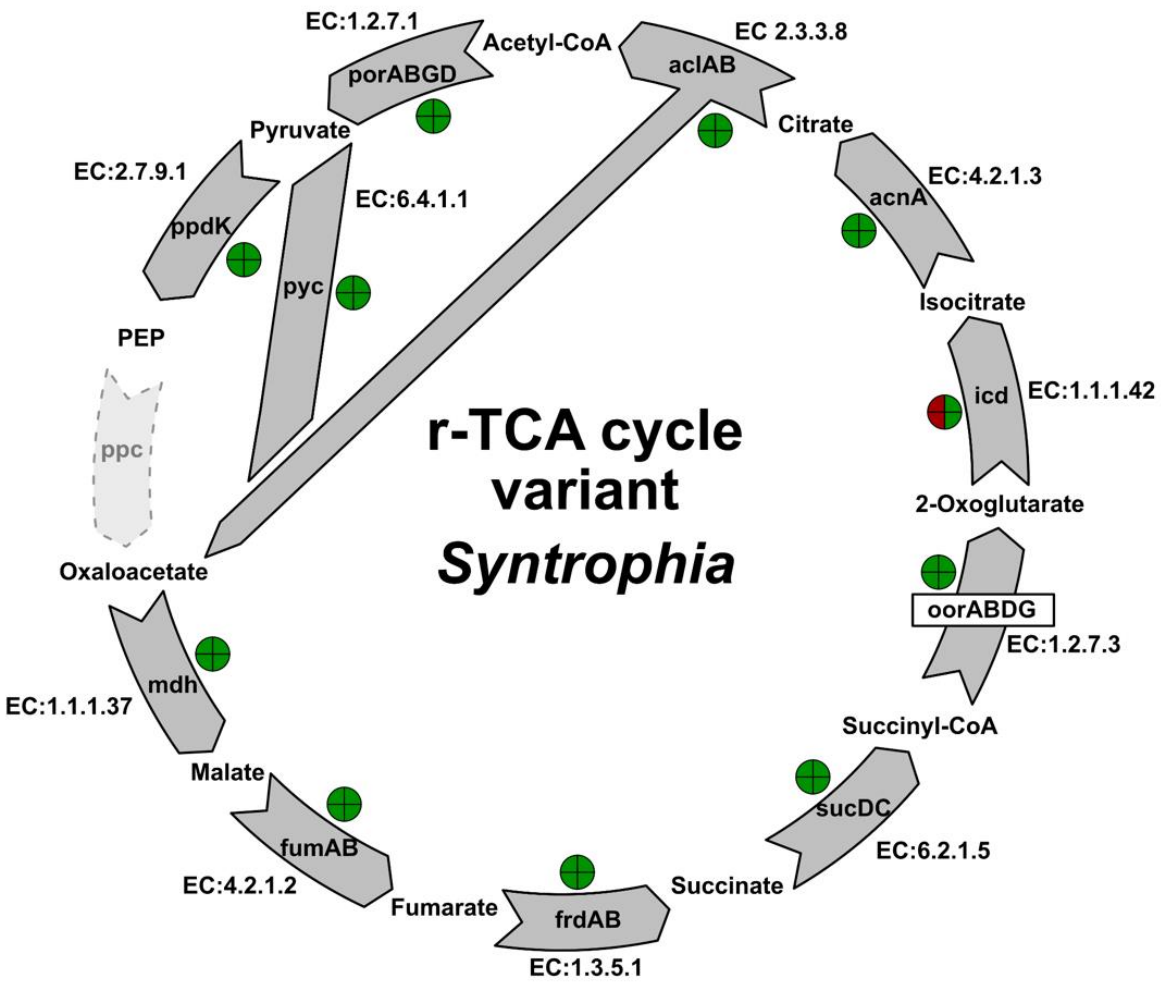


Figure 4.1. Representation of enzymatic genes identified within the four *Syntrophales* MAGs.

Pie charts represent the number of MAGs the enzymes were identified in. The presence of an enzyme is indicated by green while the absence is indicated by red.

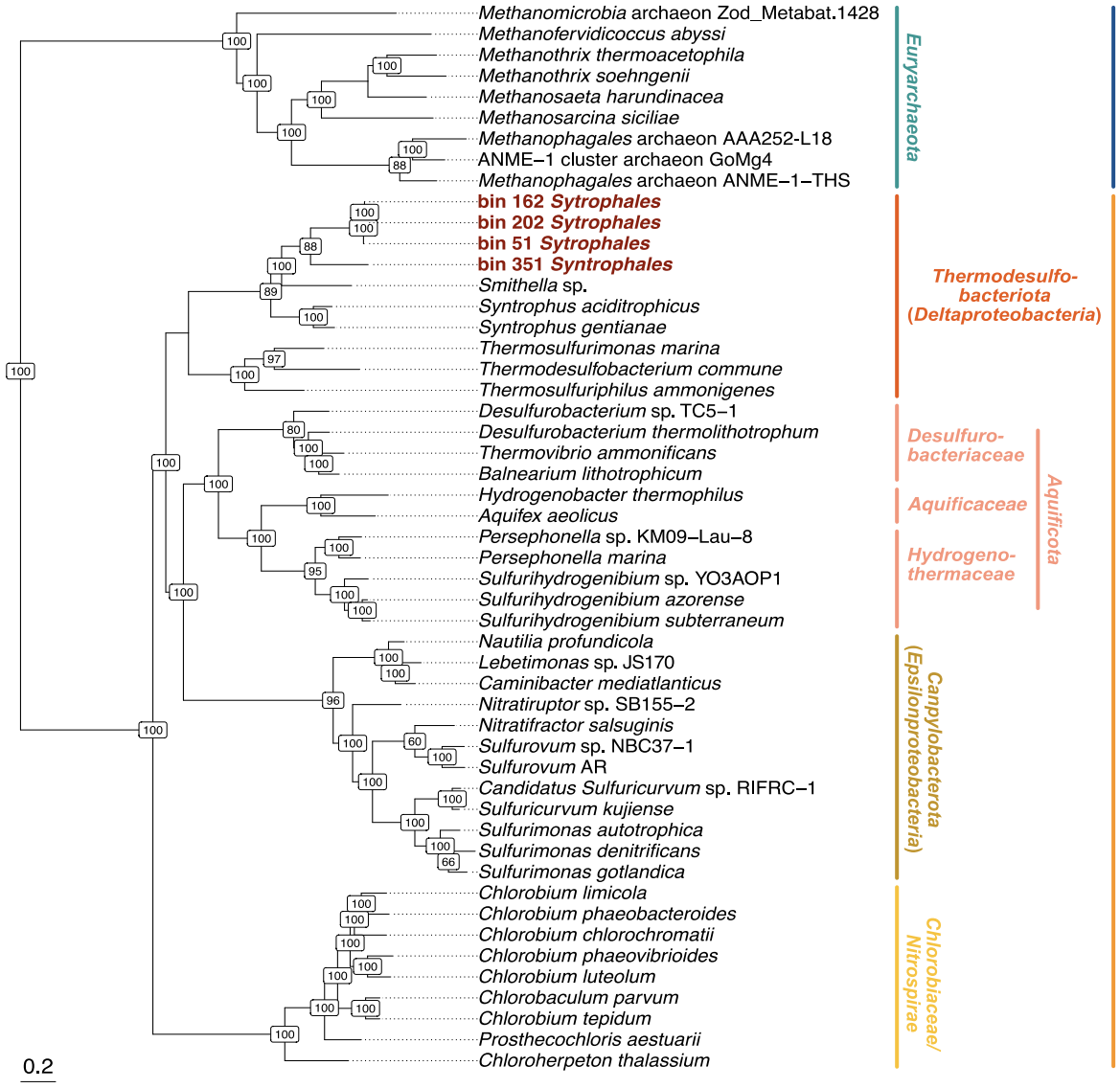


Figure 4.2. RAxML phylogenetic tree placement of MAGs bin 162, bin 202, bin 51, and bin 351 based on conserved marker genes. Bar, 20% substitution estimations. Substitution model is the LG (Le and Gascuel) gamma model. Reference sequences from Chlorobiaceae/Nitrospirae, Canpylobacterota, and Aquificota are sequences from Hügler et al. 2007 and Giovannelli et al. 2017. Sequences from Euryarchaeota are the top hits to the ATP-citrate lyase sequences found in the four MAGs. The Thermodesulfobacteriota sequences were chosen to verify GTDB-TK taxonomic placement of the four MAGs.

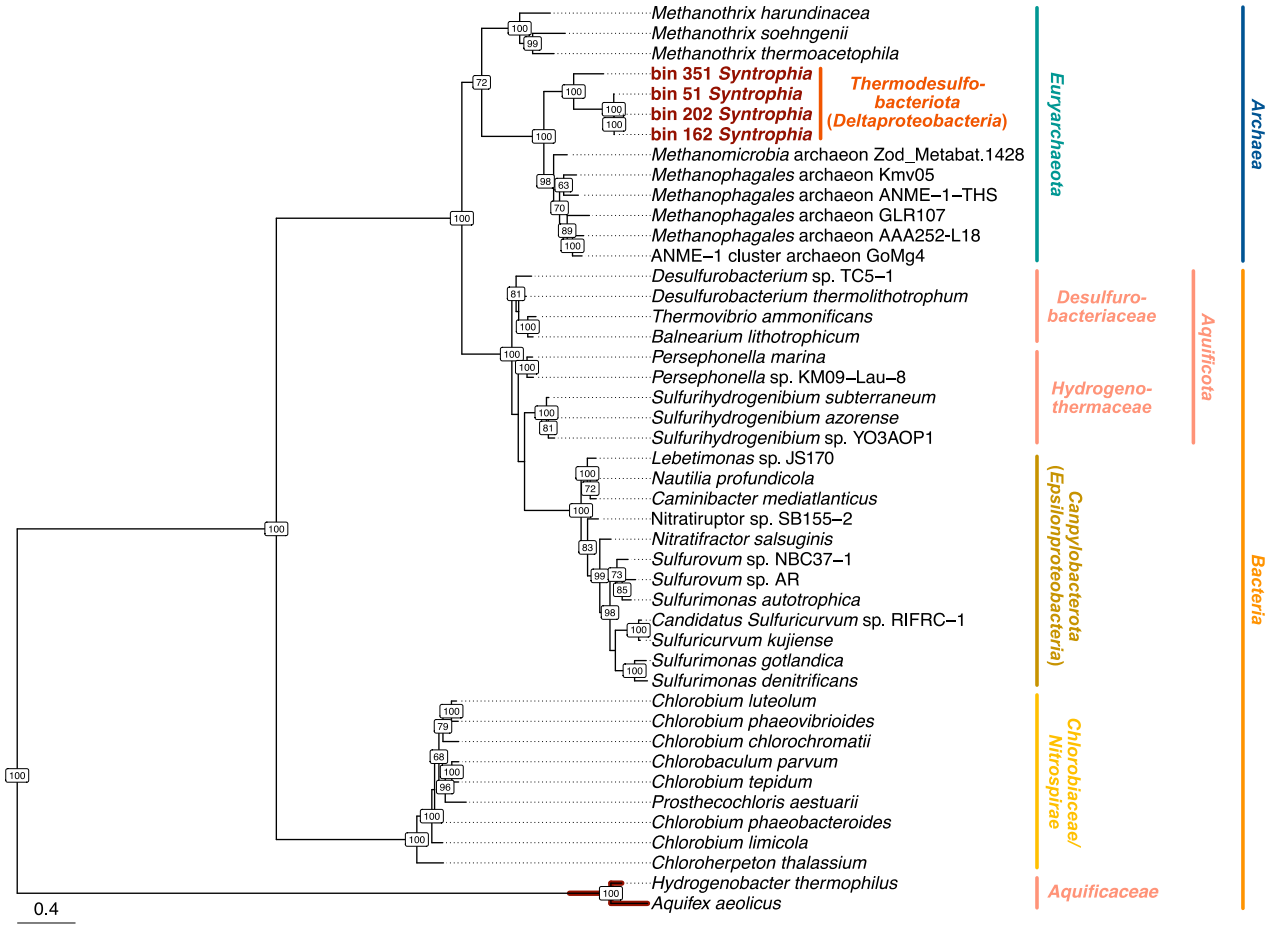


Figure 4.3. RAxML phylogenetic tree of ATP-citrate lyase amino acid sequences (concatenated α and β subunits). The tree was constructed with 1 000 bootstraps. Numbers on the nodes represent bootstrap values. Bar, 40% substitution estimations. Substitution model is the LG (Le and Gascuel) gamma model. Red branches represent ancestral ATP-citrate lyase. Ancestral ATP-citrate lyase was manually constructed by constructed by concatenating the citryl- CoA synthetase and citryl-CoA lyase of H. thermophilus and A. aeolicus, respectively as suggested by Hügler et al. 2007 and Giovannelli et al. 2017. Reference sequences are from the same taxa used in Figure 4.1.

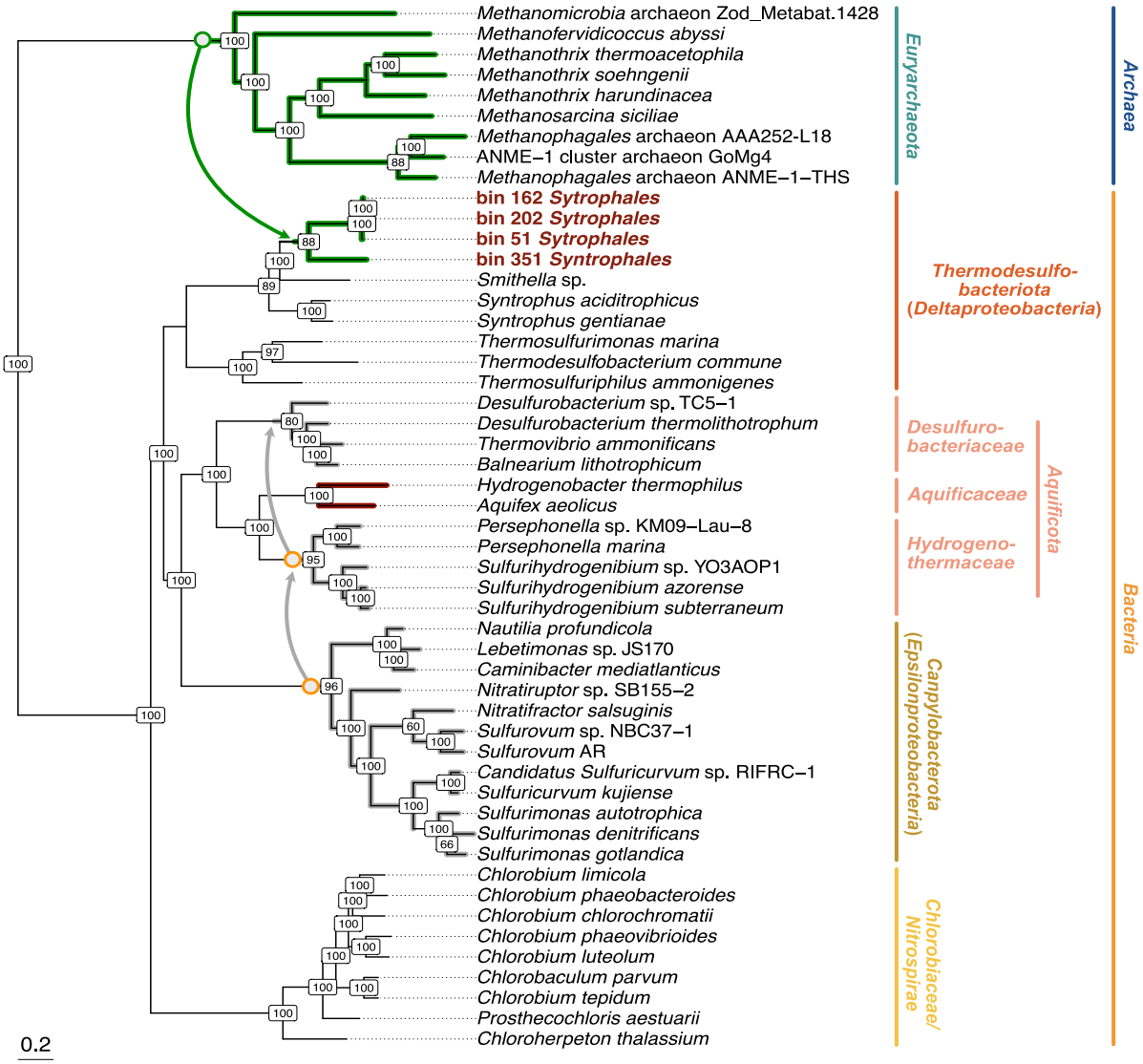


Figure 4.4. RAxML phylogenetic tree placement of MAGs bin 162, bin 202, bin 51, and bin 351 based on conserved marker genes the most parsimonies scenario for the horizontal acquisition of ATP-citrate lyase by *Syntrophales*. Bar, 20% substitution estimations. Substitution model is the LG (Le and Gascuel) gamma model. Red branches represent ancestral ATP-citrate lyase. Ancestral ATP-citrate lyase was manually constructed by constructed by concatenating the citryl-CoA synthetase and citryl-CoA lyase of Hydrogenobacter thermophilus and Aquifex aeolicus, respectively, as suggested by Hügler et al. 2007 and Giovannelli et al. 2017. Reference sequences are from the same taxa used in Figure 4.1.

Chapter 5 Conclusion

Within convergent margins, the devolatilization of a subducting oceanic slab leads to the release and migration of fluids rich in dissolved inorganic (DIC) and redox active volatiles through the overriding plate and onto the surface in the form of hot springs and other hydrothermal features. This dissertation has focused on describing how chemosynthetic communities interact with these volatiles, the role viruses might play in shaping these interactions, and how syntrophic interactions may have allowed *Syntrophales*, an otherwise fermentative order of bacteria, to acquire the ability to autotrophically fix carbon.

Historically, chemolithoautotrophy has been viewed as a small player in global primary production. However, in Chapter 1 I demonstrate that this is in large part due to the focus given to photoautotrophs within marine systems. Through our literature review, we show that chemolithoautotrophs within the terrestrial subsurface are largely isolated from surface photoautotrophic processes and are dependent on water rock interactions for redox active volatiles to power biomass production through chemosynthesis. Both the Wood-Ljungdahl (WL) pathway and Calvin-Benson-Bassham (CBB) cycle were found to be the most abundant carbon fixation pathways (CFPs) in both CO_2 rich and depleted systems with the reductive tricarboxylic acid (rTCA) cycle being in low abundance or absent. Interestingly, the archaeal specific CFPs 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) and dicarboxylate/4-hydroxybutyrate (DC/4-HB) cycles were rarely found. However, we show that while the WL, CBB, and rTCA are found in both system types, the chemolithoautotrophic species that bear them as well as the subtypes of these CFPs are linked to the amount of DIC, the dominant DIC species, as well as O_2 concentrations within the system.

In Chapter 2, I show that these observations are largely reflected in the distribution of CFPs in deeply-sourced hot springs across and along the Costa Rica convergent margin. For example, in the CO_2 depleted anoxic fluids of the outer forearc the WL pathway was the dominant CFP due to its low energy requirement and its low O_2 tolerance. We found that in all volcanic provinces, the CBB has a very low abundance, and is only the dominant CFP in the surface-associated communities. As in Chapter 1, rTCA is in low abundance in nearly all provinces. Nevertheless, rTCA is found to be very abundant in the forearc due to one highly abundant MAG identified as *Sulfurihydrogenibium*. Interestingly, some of the chemolithoautotrophs contain multiple carbon fixation pathways, an adaptation that has been previously

suggested to be a means to deal with fluctuations in DIC concentrations (1). Further, the respiratory metabolisms of the prokaryotic communities are adapted to the distinct biogeochemical conditions found within each of these volcanic provinces. Overall, these results demonstrate that the chemosynthetic communities within the Costa Rica convergent margin are associated with deeply-derived tectonically associated volatiles.

Next, we considered the roles viruses play within these chemosynthetic communities. The distribution of viral communities within the Central American convergent margin is largely driven by the distribution patterns of prokaryotic communities. Viral-host analysis revealed that fast growing well cultivated prokaryotes, such as members of *Proteobacteria* and *Bacteriodetes*, are the most often host matched. Nearly all of the viruses are predicted to have very narrow host range, with < 1% of viruses linked to hosts from multiple phyla. While highly abundant taxonomically identified viruses tend to target chemolithoautotrophs with CBB, unclassified viruses, the largest portion of viruses in this system, are almost exclusively the only viruses that target both WL and rTCA containing chemolithoautotrophs. Only 13% of the viruses that are host matched contain auxiliary metabolic genes (AMGs). The distribution of these genes is largely specific to each geophysical province, with those associated with evading host restriction systems being the most abundant.

Finally, we described the horizontal gene transfer (HGT) of ATP-citrate lyase from *Methanophagales* to *Syntrophales*. *Syntrophales* is a member of the phylum *Thermodesulfobacteriota* (formally a class in the phylum *Desulfobacterota*) which is known to form syntrophic interactions with methanogenic archaea (2,3). During these interactions, members of this order ferment benzoate and/or acetate, supplying their hydrogenotrophic methanogen partners with CO_2 and H_2. All four of the MAGs contain the α and β subunits for dissimilatory sulfite reductase (*dsrAB*) while the most complete MAG also contains the α and β subunits for adenylylsulfate reductase (*aprAB*). Phylogenetic analyses of *dsrA* revealed that all four MAGs contain the reductive variant. It was found that all four contain hydrogenases that have been demonstrated to perform hydrogen uptake (4). While not reported in previous literature, the presence of key genes of rTCA pathways in all four of these *Syntrophales* MAGs implies that they are capable of

carbon fixation. Phylogenetic analysis revealed that the key enzyme for rTCA, ATP-citrate lyase, has been horizontally acquired from *Methanophagales*.

In this dissertation I have shown that the chemosynthetic world of the terrestrial subsurface is shaped by deep geological and geochemical processes. Chemolithoautotrophy is the only means of primary production within the deep-subsurface of the Central American convergent margin, and the distinct CFPs used by these primary producers are distributed based on their CO_2 concentration requirements as well as their O_2 tolerance. Viruses of these chemosynthetic community are largely lysogenic with their distribution tied to that of the prokaryotic community. Further, these viruses employ the use of AMGs to protect themselves against host restriction, to help host energy production, and/or to help in their own replication. Additionally, syntrophic interactions within this system have led to the fermentative *Syntrophales* gaining the potential for carbon fixation through a HGT event from its syntrophic partner *Methanophagales*. While I have been able to shine a dim light on this dark complex world, here are a few questions that future research may like to consider: 1. In what proportions and where are each of the CFPs being transcribed? 2. Are members of *Syntrophales* capable of using their "newfound" CFPs for carbon fixation or heterotroph metabolic processes? What measurable effect do viruses within this system have on prokaryotic carbon fixation rates? I leave these questions to future PhD students and/or undergraduates.

References

1. Berg IA. Ecological aspects of the distribution of different autotrophic CO 2 fixation pathways. Applied and Environmental Microbiology. 2011;77(6):1925–36.
2. Gray ND, Sherry A, Grant RJ, Rowan AK, Hubert CRJ, Callbeck CM, et al. The quantitative significance of Syntrophaceae and syntrophic partnerships in methanogenic degradation of crude oil alkanes. Environmental Microbiology. 2011;13(11):2957–75.
3. Schöcke L, Schink B. Membrane-bound proton-translocating pyrophosphatase of Syntrophus gentianae, a syntrophically benzoate-degrading fermenting bacterium. European Journal of Biochemistry. 1998;256(3):589–94.
4. Rogers TJ, Buongiorno J, Jessen GL, Schrenk MO, Fordyce JA, de Moor JM, et al. Chemolithoautotroph distributions across the subsurface of a convergent margin. ISME Journal. 2022;(October):1–11.

VITA

Timothy Joseph (TJ) Rogers was born on Upper Heyford the Royal Air Force base in near Upper Heyford, Oxfordshire England. However, his dad separated from the US Air Force two years later and they move back to his dad and mom's hometown of Athens, Tennessee where he spent the rest of his childhood and teenage years. Due to ADHD and a reading/learning disability, TJ found elementary school extremely challenging, spending hours on end trying to complete short homework assignments. Only managing to receive C's and D's on his report cards, his mom made the school hold him back in 2nd grade. This was great for TJ as he was able to become a mostly honor roll student throughout the rest of elementary school. Though he was diagnosed with ADHD, he was never medicated for it until later in life, so he still struggled through every assignment, and soon finding that most teachers found his constant questions and fidgeting bothersome. As a result of this, learning fell out of favor as TJ would rather spend his time after school in the creek across the road from his house catching frogs, salamanders, and crawdads.

After barely graduating high school, TJ joined the US Air Force as he had no desire to ever attend school again if he could help it. Three years later, he decided that he wanted to leave the military to become a young earth creationist preacher to preach against what he called "evil-ution", due to his fundamentalist Christian upbringing. But, in doing so he wanted to know as much about this subject as he could. He soon found himself pouring over websites for and against evolution, watching video after video on the topic, and digging into as much as the literature as he could understand. After a few months of this, he found that evolution wasn't just a fact, but that it was extremely fascinating. However, this fascination led to a three-year depressing spiral of questioning everything he grew up believing. Once he was able to reach the bottom of the rabbit hole and see there was nothing to fear, he decided not to pursue preaching, but a BS degree in evolutionary biology as he had become obsessed with the subject at this point. So, after 8.5 years, two deployments, and eight temporary duty assignments, he decided to separate from the Air Force. In 2014, at 28 years old, TJ left the military and started studying Ecology and Evolutionary Biology at the University of Tennessee-Knoxville (UTK). For the first time in his life, he found that he

was excelling in school. Yeah, the ADHD was still problematic, but his love for the subject gave him the motivation to stay up late into the night working on the required homework. Towards the end of his senior year, he felt he had learned as much about evolution as he needed and tuned his sights towards studying analogs of early life. In 2018, he joined Dr. Karen Lloyd's lab at UTK to study chemosynthetic communities within convergent margins. His research has afforded him the pleasure to travel to remote parts of Argentina, Chile, and Iceland. Currently, he looks forward to starting his new life with his wife, Soriya, in Charlotte North Carolina as he begins his new career in Dr. Elain Luo's lab studying viral interactions with chemosynthetic communities within marine and freshwater systems.