# The balance between prokaryotic autotrophy and heterotrophy in the oceans

Rachel K. Lange Spietz

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

Doctor of Philosophy

University of Washington

2017

Reading Committee:

Robert Morris, Chair

Gabrielle Rocap

Anitra Ingalls

Program Authorized to Offer Degree:
Oceanography

University of Washington

## Abstract

The balance between prokaryotic autotrophy and heterotrophy in the oceans

Rachel K. Lange Spietz

Chair of the Supervisory Committee:
Prof. Robert Morris
Oceanography

Bacteria play in integral role in the marine carbon cycle by mediating crucial transformations of organic and inorganic matter. Within the organic carbon pool, heterotrophic bacteria control the flux of labile and recalcitrant organic compounds, which has significant impacts on the marine food web and nutrient cycling. In this dissertation, I used a suite of culture-dependent and culture-independent techniques to investigate the functional roles and distributions of key microorganisms in the marine carbon cycle. The SUP05 clade of marine gammaproteobacteria is an abundant group of closely related sulfur-oxidizing bacteria. Members of the SUP05 clade have the genetic potential for autotrophic, heterotrophic, and mixotrophic growth. SUP05 is widely distributed throughout the world's oceans including oligotrophic gyres, eutrophic estuaries, hydrothermal vent plumes, and oxygen minimum zones. In chapter one, we used laboratory techniques including physiological growth experiments, proteomic surveys, and genomic reconstructions to find that a potentially mixotrophic strain of SUP05 predominantly functions as a heterotroph under oligotrophic and copiotrophic conditions. In chapter two, I sequenced the complete genome of a third isolate from the SUP05 clade and compared its genomic content to other complete and incomplete

SUP05 genomes. While the SUP05 strains were all very similar at the 16S rDNA level, striking differences in metabolic potential exist in the SUP05 clade. Notably, the newly sequenced genome does not contain the metabolic potential for carbon fixation, which was previously thought to be a defining characteristic of the SUP05 clade. Finally, in chapter three I expand my exploration of key carbon and energy-cycling microbes beyond the SUP05 clade to explore the heterogeneous distribution of microbes across fresh lava from a deep-sea volcanic eruption. Collectively, the results from this research show that closely related microorganisms may differ wildly in their role in the marine carbon and energy cycles. Further, fine-scale disturbances in geochemical parameters have the potential to alter the composition of microbial communities, potentially impacting the overall community function.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

Scientific research is best conducted through collaborative efforts. It is full of failures, successes, rejections, and breakthroughs. This dissertation is a reflection of the endless support and assistance provided by members of the School of Oceanography at the University of Washington. Understanding the value of collaboration has come from the wonderful years of working with my advisor, Dr. Bob Morris. I am forever grateful for the many opportunities to pursue independent research that Bob provided, and the positive light he has always been able to shine on frustrating data. I am grateful for the advice and wisdom from my committee members Dr. Gabrielle Rocap, Dr. Anitra Ingalls, Dr. Allan Devol, and Dr. Steven Roberts.

Thank you to the brilliant scientists and caring friends I have met. To my wonderful officemates Vega, Helena, Diane, and Rachelle, thank you for sharing your ideas with me and making graduate school so much more enjoyable and fun. Thank you to the scientists and students I have had the honor of mentoring or teaching, you have given my pursuit of a PhD reason and I have been fortunate to be able to share my love of science with you.

I would have been lost without the support and everlasting encouragement from my family and friends. Thank you to my parents for never questioning my life goals. To the Spietz/Maybell family, thank you for showing genuine interest in my research and even talking nerdy with me once and a while. My brother, Joe, who has repeatedly helped me through my computational struggles and has always been a phone call away from hours of conversation to cheer me up and make me laugh. Lastly, thank you to my best friend and husband, Alex, who has taught me how to live with purpose, given me confidence to succeed in reaching my dreams, and has possibly earned an honorary doctorate after listening to me talk about my research without end. Having you by my side makes life worth living, risks worth taking, and goals worth reaching.

# DEDICATION

To my parents.

# Chapter 1

# INTRODUCTION

Though diminutive at the individual level, the shear abundance and widespread distribution of microorganisms allow for their dominance of global biogeochemical cycling of carbon, nitrogen, and sulfur. In the sunlit surface waters of the world′s oceans phototrophic microbes capture sunlight to fuel carbon fixation, which supplies abundant organic carbon to the pelagic marine food web. In the absence of sunlight, chemoautotrophic microbes supply additional organic carbon to higher trophic levels through several carbon fixation pathways fueled by a diverse suite of reduced chemical species (Hügler and Sievert, 2011). Contrastingly, heterotrophic microorganisms throughout the water column utilize fixed organic carbon for both assimilatory and dissimilatory reactions resulting in the remineralization of organic carbon to inorganic forms. The tradeoff between autotrophic, both light and chemical driven, and heterotrophic microbial lifestyles determines the fate of carbon in the marine environment thus controlling the concentrations of $CO_2$ and $O_2$ in the atmosphere and Earth′s redox balance.

The global ocean is responsible for over half of net primary production on Earth (Falkowski et al., 2000), yet the fate of fixed organic carbon in different oceanic regions remains unknown. In mathematical terms, net community production (NCP) is equal to gross primary production (GPP) minus community respiration (R). If NCP is greater than 0, then the

autotrophic ecosystem is a sink for atmospheric $CO_2$. The converse is true when NCP is less than 0, thus defining a heterotrophic system. The nutrient-rich waters of coastal seas cover only 6% of sea surface area but account for over one-third of the global net primary production making these zones highly autotrophic-dominated systems (Gattuso et al., 1998; Ducklow and McCallister, 2005). Conversely, for decades scientists have debated whether the vast oligotrophic gyres of the open ocean are autotrophic or heterotrophic (reviewed in Ducklow and Doney (2013)). Accurate modeling of NCP in oligotrophic reaches of the open ocean is imperative because of the implications for predicting future global carbon cycle fluxes due to anthropogenic-induced change.

In a 2013 review, Duarte and colleagues argue that the oligotrophic regions of the world's oceans are a net source of $CO_2$ to the atmosphere (Duarte et al., 2013). From a compiled dataset of 1,623 estimates of NCP from oligotrophic gyres of the North and South Atlantic Subtropical Gyre, the North Pacific Subtropical Gyre, and the Mediterranean Sea over half (57%) showed net heterotrophic communities (Duarte et al., 2013). These estimates were generated from a variety of methods each with their own set of biases such as exclusion of zooplankton respiration and bottle effects in incubation-based experiments and assumptions in mixing rates and gas diffusivities for incubation-free approaches (reviewed in Robinson and Williams (2005)). While discrete estimates of NCP help to inform scientists of trends in autotrophic vs heterotrophic-dominated systems, we have not reached a consensus using these methods after nearly two decades of collection. Variability in NCP estimates could largely be due to heterogeneity in microbial communities across space and time that are not captured by discrete measurements. For this reason, investigations into the specific metabolisms of abundant oligotrophic microbes, which dominate carbon and nutrient cycling in the open ocean, are imperative to better constrain models of global carbon cycling.

Life for marine microbes in the oligotrophic open ocean is challenged by extremely dilute concentrations of organic carbon and other nutrient or energy sources. Despite low carbon flux and sustained nutrient concentrations near zero, microorganisms in oligotrophic regions make-up the majority of biomass (Schutl et al., 1997) with cell concentrations on the order of 0.5-5x$10^5$ cells $mL^{-1}$ of seawater (Whitman et al., 1998) owing to physiological adaptations to extreme nutrient limitation. A better understanding of microbial adaptations to oligotrophy will improve our ability to cultivate ecologically important marine microbes in order to resolve their physiological roles in the global carbon cycle. This dissertation aims to using cultivation dependent and independent approaches to understand the roles of oligotrophic marine microorganisms in the carbon cycling.

In 1979, during the Dahlem Conference on Strategies of Microbial Life in Extreme Environments, Hirsch and colleagues described five characteristics for microbial survival in oligotrophic conditions: i) cellular adaptation for low carbon requirements in regards to shape and size, ii) the ability to use metabolic energy for nutrient uptake, particularly during periods of starvation, iii) the ability to constantly take-up nutrients from the environment, iv) contain low-specificity, high-affinity transport systems to scavenge the dilute environment for non-specific organic compounds, and v) maintain the ability to store and reserve nutrients following uptake (Hirsch et al., 1979). Cultivation-dependent and -independent approaches have since tested the above requirements for oligotrophic growth and adaptations of marine microbes.

Until recent advances in cultivation techniques, marine oligotrophic bacteria eluded laboratory isolation because they do not survive in typical nutrient-rich microbiology media (Connon and Giovannoni, 2002; Rappé et al., 2002; Cho and Giovannoni, 2004). With the advent of using sterile filtered seawater as media to isolate novel oligotrophs, weve begun

to understand mechanisms of survival in low nutrient conditions. Arguably one of the most well-studied marine oligotrophic bacteria comes from the SAR11 clade, which accounts for up to half of surface microbial communities (Morris et al., 2002). SAR11 cells are among some of the smallest in size and maintain small, streamlined genomes, which reduces the cellular energy demand under nutrient deplete conditions (Giovannoni et al., 2014). In the nutrient-poor surface waters of the Sargasso Sea, a metaproteomics study highlighted that a dominant proportion of SAR11 proteins were periplasmic substrate-binding proteins (Sowell et al., 2009). The small genome and cell size combined with the dedication of cell volume and energy to periplasmic space and nutrient uptake systems lends to SAR11's survival in oligotrophic waters and long periods of starvation, which also supports the first condition described by Hirsch et al. (Hirsch et al., 1979). Under laboratory conditions, SAR11 continues to express nutrient transport proteins during stationary phase presumably to compete for sparse nutrients during periods of starvation (Sowell et al., 2008), in line with Hirsch's third condition for oligotrophic adaptation. Beyond the SAR11 clade, tonB-dependent transporters (TBDTs) from alpha- and delta-proteobacteria represent a significant fraction of membrane-bound proteins in the nutrient-deplete South Atlantic gyre (Morris et al., 2010). TBDTs utilize a proton motive force, possibly harvested from sunlight by rhodopsins, to transport nutrients across the membrane in gram-negative marine bacteria. This suggests that oligotrophic microbes may use alternative energy sources to fuel constant nutrient acquisition during daylight hours in oligotrophic regions of the ocean, further supporting the second and third conditions described by Hirsch et al. (1979).

Members of the chemoautotrophic SUP05 clade of gammaproteobacterial sulfur oxidizers (GSOs) inhabit oligotrophic regions of the worlds oceans and may contribute significantly to carbon and nutrient cycling on a global scale. The SUP05 clade is particularly abundant

in suboxic and anoxic regions (Walsh et al., 2009; Glaubitz et al., 2013; Louca et al., 2016; Rogge et al., 2017), which are currently expanding due to climate change. SUP05 bacteria possess genes required for carbon fixation, sulfur-oxidation, and nitrate respiration (Walsh et al., 2009; Swan et al., 2011; Anantharaman et al., 2013). Interestingly, some members of the SUP05 clade harbor genes for carbon and energy acquisition from organic carbon substrates suggesting a potentially mixotrophic lifestyle (Swan et al., 2011; Murillo et al., 2014). Given SUP05's cosmopolitan distribution and potentially globally significant roles in carbon fixation and respiration, members of this clade are of particular interest for the focus of physiological studies.

The SUP05 clade is diverse and contains both free-living and symbiotic members. Phylogenetic analysis of the 16S rRNA gene reveals two sister clades within SUP05: *sup05* and *arctic96BD-19* (Figure 2.1). The two sister clades differ in their metabolic potential as *sup05* appear to be strict chemoautotrophs while *arctic96BD-19* may be mixotrophic for both carbon and energy acquisition (Lalish, 2015). The first two chapters of this dissertation examine the physiological role of members of the *arctic96BD-19* clade in carbon and energy processing. My overarching hypothesis for chapters one and two is that members of the *acrtic96BD-19* subclade of SUP05 have a flexible role in the marine carbon cycle both on the individual and population-level depending on the availability of carbon and reduced energy sources.

While the metabolic state of the oceans is intimately linked with biological processes, large-scale events can locally alter the geochemical signatures of an ecosystem lending to shifts in overall metabolisms. One such example is an eruption event along a mid-ocean ridge (MOR) system. Eruptive events at MORs rapidly alter the geology and chemistry of hydrothermal fluids emitted at the seafloor and into overlying pelagic zones (Butterfield et al.,

1997). Abrupt changes in chemistry following eruptive events at MOR influence microbial communities (Huber et al., 2002, 2003; Meyer et al., 2013; Gulmann et al., 2015). Fresh basalt from eruptive events is rapidly colonized by microbes in as few as a hours to days as biofilms develop on the newly exposed surfaces, demonstrating predictable successional patterns over time (Gulmann et al., 2015). Geochemical models suggest that the enriched hydrothermal fluids that are released fuel high rates of microbial productivity (McCollom and Shock, 1997; McCollom, 2000; Shock and Holland, 2004). Primary production within hydrothermally-enriched plumes is only a fraction of photosynthetic production in surface waters, but the input of organic carbon may be significant on a global scale given the broad spatial extent of diffuse flow plumes, such as those emanating from fresh lava flows (McCollom, 2000). The results from the third chapter of my dissertation support the hypothesis that deep-sea eruptions along MORs alter the composition of microbial communities in the overlying water column, which has implications for carbon and energy cycling in the deep-sea.

Chapter 2

# HETEROTROPHY WITHIN A SULFUR-OXIDIZING LINEAGE OF CHEMOAUTOTROPHIC MARINE BACTERIA

Rachel Lange Spietz, Rachel Lundeen, Anitra Ingalls, and Robert M. Morris [In prep]

## *2.1 Abstract*

The balance between autotrophy and heterotrophy determines the size of the ocean's carbon sink. Shifts in the underlying metabolisms controlling carbon flux determine the contributions of major bacterial and archaeal lineages to carbon cycling in the oceans. Members of the SUP05 clade are often described as a single lineage of chemoautotrophic gammaproteobacterial sulfur-oxidizers (GSOs), but representatives from the *Arctic96BD-19* subclade have the genetic potential to both produce and consume organic matter. *Ca.* Thioglobus singularis strain PS1 is a member of the *Arctic96BD-19* subclade of SUP05 that was isolated from Puget Sound, WA, USA. Strain PS1 has the genetic potential to use light, reduced sulfur, or organic compounds for energy and to assimilate organic and inorganic carbon for biomass. We grew *Ca.* T. singularis strain PS1 on an artificial growth medium (ASW) and quantified protein expression under variable growth conditions to test metabolic predictions and elucidate the balance between autotrophy and heterotrophy. Here, we show that strain PS1 uses organic substrates to meet its energy and carbon demands. The addition of a phytoplankton lysate and individual phytoplankton-derived organic compounds significantly enhanced the growth of strain PS1. Proteins involved in methylotrophic pathways for carbon

and energy were significantly upregulated when lysate was added to growth media, suggesting that PS1 potentially uses methylated compounds derived from phytoplankton or other marine organisms. Although sulfur oxidation is a defining characteristic of the GSO clade, strain PS1 does not require a reduced form of inorganic sulfur for growth and there is little evidence of inorganic carbon fixation. These data indicate that members of the widespread *Arctic96BD-19* subclade of SUP05 have a predominantly heterotrophic role in the marine carbon cycle.

## *2.2 Introduction*

Phytoplankton are the major producers of organic matter in the surface ocean, and it is estimated that 30-50% of their production is released in the dissolved fraction as dissolved organic carbon (DOC) (Hansell and Carlson, 2015). Heterotrophic bacteria play a crucial role in the cycling of DOM in the oceans by consuming the DOM fixed through photosynthesis and returning some of the energy to higher trophic levels through the microbial loop (Azam et al., 1983). It is estimated that heterotrophic microbes consume 70-90% of the labile-DOM produced through photosynthesis (Hansell and Carlson, 2015). The DOM pool in the upper ocean is complex and phytoplankton community composition is linked to variability in the DOM pool (Becker et al., 2014; Sarmento and Gasol, 2012; Sarmento et al., 2013). In turn, changes in the composition of DOM constituents are important factors in the activity and composition of heterotrophic bacteria (Judd et al., 2006; Gómez-Consarnau et al., 2012). Therefore, understanding the specific interactions between heterotrophic bacteria and the DOM pool is essential for modeling and predicting global carbon flux in the face of climatic changes and alterations in the distribution and composition of phytoplankton and the organic substrates they produce.

The SUP05 clade of gammaproteobacteria are among the most widespread and abundant groups of marine bacteria and have been implicated in having a significant impact in the transformation of nitrogen, sulfur, and carbon in the global ocean. Since the clade was first described in a hydrothermal vent plume at Suiyo Seamount (Sunamura et al., 2004), several subsequent studies have helped to resolve the phylogeny of SUP05 into two subclades: *Sup05* and *Arctic96BD-19* (Walsh et al., 2009). The *Sup05* subclade has been the focus of many studies because it is often the dominant member of the microbial community at oxic-anoxic interfaces such as those in oxygen minimum zones, upwelling regions, and hydrothermal systems. Environmental genomic and proteomic studies have linked *Sup05* with chemoautotrophic processes. Recent laboratory experiments have demonstrated that a member of the *Sup05* subclade couples sulfur oxidation and nitrate reduction to fix inorganic carbon under anaerobic conditions (Shah et al., 2016).

The sister subclade of *Sup05*, *Arctic96BD-19*, is also widely distributed in hydrothermal vent plumes (Anderson et al., 2013; Mattes et al., 2013), the deep-ocean (Swan et al., 2011), oxygenated surface waters (Williams et al., 2012), and oxygen deficient zones (Walsh et al., 2009; Wright et al., 2012). Similar to *Sup05*, members of the *Arctic96BD-19* subclade contain genes for carbon fixation and sulfur oxidation (Swan et al., 2011). However, partial genome assemblies of *Arctic96BD-19* from an oxygen-deficient coastal environment identified genes for a complete TCA cycle and several transporters for organic substrates (Murillo et al., 2014). This extends the metabolic potential of *Arctic96BD-19* from a chemolithoautotroph to a versatile mixotroph that can utilize a variety of carbon and energy sources. Further, *Arctic96BD-19* expressed transport proteins for amino acids, dipeptides, sugars, and taurine in the Bedford Basin, a seasonally hypoxic inlet in the Atlantic Ocean, yet sulfur oxidation and carbon fixation proteins were not identified (Georges et al., 2014). This suggests that

members of the *Arctic96BD-19* sublcade of SUP05 may function primarily as a heterotroph.

The first cultured isolate from the *Arctic96BD-19* subclade was obtained from coastal surface waters in 2015 and named *Ca.* Thioglobus singularis, strain PS1 (Puget Sound 1) (Marshall and Morris, 2013). The complete genome sequence indicates that strain PS1 has the genetic potential for autotrophic and heterotrophic metabolisms, including the ability for lithotrophic, phototrophic, and organotrophic energy production. This suggests that PS1 is a dual-mixotroph for both carbon and energy (Marshall and Morris, 2015). We tested the hypothesis that *Ca.* T. singularis, strain PS1 grows as a chemolithoautotroph under low nutrient (nutrient-deplete) conditions and as an organoheterotroph under high nutrient (nutrient-replete) conditions. We combined laboratory physiology experiments with proteomic analyses of cells grown in media either depleted or rich in nutrients, including organic carbon, to identify shifts in the expression levels of autotrophic and heterotrophic proteins.

## *2.3 Results and discussion*

### *2.3.1 Growth on nutrient-deplete and nutrient-replete media*

We adapted an artificial seawater media (ASW) Carini et al. (2012) to evaluate the potential for Ca. T. singularis strain PS1 to use organic matter. We were unable to inhibit the growth of strain PS1 on blank ASW media with no amendments (Figure 2.1). A previous study of *Pelagibacter ubique* (SAR11) found that this oligotrophic marine microbe also grows on minimal ASW media that included the addition of vitamins but no other organic substrates, which led to the suggestion that SAR11 could potentially use vitamins as a carbon sources (Carini et al., 2012). We tested the growth of PS1 in ASW without the addition of vitamins or nitrogen and found that it grew to similar cell densities after several subsequent transfers (SI

Figure B.1). *Pelagibacter sp.* can use volatile organic compounds (VOCs) (Sun et al., 2011; Halsey et al., 2017), which could be present as a laboratory contaminant. However, strain PS1 lacks similar genes to metabolize VOCs. Surprisingly, strain PS1 grew to higher cell densities on ASW than on natural seawater media (NSW) (Figure 2.1), which suggests that trace amounts of carbon and other nutrients capable of supporting the growth of extremely oligotrophic marine bacteria are introduced as common laboratory contaminants. Regardless, the low growth yields obtained on ASW media were sufficient as a baseline to evaluate the potential for enhanced growth on organic substrates. After verifying purity in several transfer cultures, strain PS1 was grown with and without reduced sulfur (1mM thiosulfate) as a potential energy source (Marshall and Morris, 2013) and under nutrient-deplete and nutrient-replete growth conditions. Nutrient-replete media was prepared by adding a cellular lysate from the marine diatom, *Thalassiosira pseudonana* CCMP 1335. PS1 grew to similar cell densities in NSW and ASW media under nutrient-replete conditions ($2.36 \times 10^6$ and $2.35 \times 10^6$ cells $mL^{-1}$, respectively) (Figure 2.1). This indicates that PS1 can use a natural source of phytoplankton-derived organic matter to overcome nutrient limitation.

### *2.3.2 Comparison of protein expression on nutrient-deplete and nutrient-replete media*

Protein expression profiles were compared for cells grown under nutrient-deplete (thiosulfate only) and nutrient-replete (thiosulfate plus diatom lysate) conditions (Figure 2.2). We predicted that strain PS1 would upregulate ribulose bis-phosphate carboxylase (RuBisCO) and other Calvin-Benson-Basham (CBB) proteins required for carbon fixation under nutrient-deplete conditions. Using high-stringency criteria (>2 peptides in at least two replicate injections), we identified 214 proteins from $1.46 \times 10^8$ cells grown on nutrient-deplete media and 991 proteins from $1.49 \times 10^9$ cells grown on nutrient-replete media. These identifications

represent 13% and 59% of all predicted proteins, respectively (Figure 2.2). Differences in the total number of proteins detected are due to differences in biomass obtained under different growth conditions. Although total proteomic coverage was higher when a diatom lysate was added, the proteomic profiles were similar in both treatments (SI Figure B.2 and B.3). Two-thirds of the proteins detected in the nutrient-deplete proteome were among the top proteins in the nutrient-replete proteome, demonstrating that the most abundant proteins were similar between the two treatments. In both treatments, the 25 most highly expressed proteins included ABC transporters for organic substrates (SI Tables B.1 and B.2).

Analysis of the most highly expressed proteins suggests that *Ca.* T. singularis strain PS1 is assimilating organic carbon under both growth conditions (SI Tables B.1 and B.2). Of the 212 proteins detected in both proteomes, more than half (60.2%) had a two-fold increase in expression in the nutrient-replete treatment (SI Table B.3). Nearly a third (30.3%) of these belonged to the translation, ribosomal structure, and biogenesis functional category. Proteins involved in organic substrate transport and biosynthesis accounted for 27% of all up-regulated proteins in the nutrient-replete proteome and 12% were proteins involved in energy production and conversion. These data are consistent with higher growth rates and higher maximum cell densities under nutrient-replete conditions. An uncharacterized membrane-associated protein (A0A0M3T2F3) was the highest expressed protein in both proteomes but was proportionally higher in the nutrient-deplete treatment compared to nutrient-replete. This may suggest that the surface area to volume ratio of cells grown under nutrient-deplete conditions was higher than the nutrient-replete cells, therefore indicating smaller cell sizes under nutrient-deplete conditions. Given the higher growth rate of cells under nutrient-replete conditions and larger cell size, it is expected that the proteins detected in both treatments had higher concentrations in the nutrient-replete proteome (SI Figure B.3).

Only three proteins were expressed more than two-fold higher in the nutrient-deplete treatment relative to the nutrient-replete treatment; a C4-dicarboxylate ABC transporter substrate-binding, a succinate-semialdehyde dehydrogenase, and an uncharacterized protein associated with the cytoplasmic membrane (SI Table B.4). The C4-dicarboxylate ABC transporter and succinate-semialdehyde dehydrogenase are located next to each other on the strain PS1 genome. A BLAST search against the UniProtKB database revealed that PS1's succinate-semialdehyde dehydrogenase was 81.6% identical to a long-chain fatty acid-coA ligase from *Ca.* Thioglobus sp. TMED218, a genome assembled from a Mediterranean Sea metagenome. It is possible that the succinate-semialdehyde dehydrogenase may oxidize a fatty acid or another organic substrate transported by the neighboring C4-dicarboxylate transporter under extremely oligotrophic conditions, though further experiments are necessary to discern the functions of these two proteins.

### *2.3.3 Proteomic evidence of organoheterotrophic growth in Ca. T. singularis*

Carbon-fixing organisms express high-levels of the RuBisCO protein and the expression of RuBisCO is affected by environmental conditions, such as sunlight (Fuszard et al., 2012). Furthermore, many carbon-fixing marine organisms contain carbon concentrating mechanisms (CCMs) to utilize bicarbonate ($HCO_3^-$) in $CO_2$-depleted seawater. However, strain PS1 does not contain CCMs and would likely need to synthesize a surplus of RuBisCO to rely on carbon-fixation for biosynthesis. Only two of the six proteins required for carbon fixation via the CBB cycle were identified in the nutrient-replete proteome: the large subunit of Rubsico (CbbL) and a post-translational activator protein for RuBisCO (CbbQ). The small subunit of Rubsico (CbbS) and three accessory proteins (CbbR, CbbO, and CbbY) were not detected. Protein CbbR is considered a master regulator of carbon fixation in chemoau-

totrophs that use RuBisCO to fix carbon (Dangel and Tabita, 2015) and CbbY is thought to aid in alleviating RuBisCO feedback inhibition by xylulose-1,5-bisphosphate (Bracher et al., 2015). The absence of these two proteins could indicate a non-functioning carbon fixation pathway where unregulated CbbL and CbbQ are constitutively expressed at low levels or have other functions in the cell. No CBB proteins were identified in the nutrient-deplete proteome.

The specific mechanisms of organic substrate utilization were investigated further in the proteome of strain PS1 under nutrient-replete conditions because we obtained greater proteomic coverage and it was highly similar to the nutrient-deplete proteome (Figure 2.2). All proteins required for glycolysis (data not shown) and the TCA cycle (Figure 2.3) were expressed in the strain PS1 proteome. The second most highly expressed protein in the TCA cycle was isocitrate lyase which catalyzes the conversion of isocitrate to succinate and glyoxylate through the anapleurotic glyoxylate bypass. Through the glyoxylate bypass, organisms use TCA intermediates for biosynthesis and anabolic processes rather than oxidation of the intermediates for generation of ATP and NADPH. Isocitrate lyase was also one of the most highly expressed proteins in the nutrient-deplete treatment (Figure 2.3), further suggesting that strain PS1 conserves carbon through the glyxoylate bypass of the TCA cycle.

Only four of the seven proteins required for the anabolic pentose phosphate pathway (PPP) were detected under nutrient-replete conditions. Two of the three absent proteins, glucose-6-phosphate 1-dehydrogenase and 6-phosphogluconate dehydrogenase, are a required part of the oxidative branch of the PPP used to generate reducing equivalents for other energy-requiring cellular processes. Two of the three proteins that comprise the non-oxidative branch of the PPP, ribulose phosphate epimerase and transketolase, were present in the nutrient-replete proteome, as well as ribose phosphate pyrophosphokinase, which aids in the

production of important nucleotide precursors. Collectively, strain PS1 appears to use the PPP to generate pentose phosphates for nucleotide biosynthesis rather than for NADPH reducing power for other biosynthesis reactions in the cell.

Strain PS1 contains a reduced set of genes that may be involved in sulfur oxidation including *aprMAB*, *dsrCKLM*, three rhodaneses, sat, and sulfite reductase (Marshall and Morris, 2015). In the nutrient-replete proteome, AprAB were expressed at comparable levels to one another indicating a relationship between the activity of these two proteins. AprM, DsrC, DsrK, sat (sulfate adenylyltransferase), sulfite reductase, and two of the three rhondaneses had low levels of expression. A third rhodanese and DsrM, a membrane-bound sulfite reductase, was not detected in the nutrient-replete proteome.

Recent surveys of abundant marine bacteria both in the lab and in the field have suggested the importance of methylated amines and osmolytes for energy acquisition in heterotrophic bacteria (Lidbury et al., 2015, 2017; Sun et al., 2016). In fact, the genome of strain PS1 contains all the genes necessary for complete oxidation of trimethylamine (TMA), tetrahydroformate (THF)-linked substrates, and marine osmolytes glycine betaine and choline (Figure 2.4). Further, all proteins involved in this mode of energy acquisition were expressed under nutrient-replete conditions. A subset of these proteins was detected in the nutrient-deplete treatment as well (Figure 2.4), indicating that the degradation of methylated amines is a significant source of reducing equivalents for strain PS1 under nutrient-replete and nutrient-deplete conditions.

### *2.3.4 Growth on TCA intermediates and methyl-containing carbon compounds*

A complete citric acid cycle, or TCA cycle, is part of the central metabolism for strain PS1 as indicated by the expression of all required proteins (Figure 2.3). This is in contrast to an

incomplete, or horseshoe, TCA cycle utilized by the closely-related and strictly autotrophic *Ca.* T. autotrophicus (Shah et al., 2016). To further understand the role of strain PS1 in its environment, we tested whether utilization of individual TCA intermediates as substrates for growth. Pyruvate, a precursor to the TCA cycle, elicited a 70% increase in final cell density over the blank control media (Table 2.1). Three TCA-cycle intermediates, oxaloacetate, alpha-ketoglutarate, and malate, had no significant effect on final cell yield. Glucose, a common organic carbon substrate widely used by heterotrophic bacteria also had no significant effect on growth (Table 2.1). Malonate, a competitive inhibitor of succinate dehydrogenase, increased final cell density by 84% over the nutrient-deplete control (Table 2.1).

Methylotrophy is widespread among marine microbes and strain PS1 has the genetic potential for utilization of methylated compounds. Acetate and malonate elicited an increase in final cell density over nutrient-deplete control media. The addition of formate to ASW media resulted in nearly a three-fold increase in final cell density (Table 2.1). Strain PS1 expressed proteins involved in the degradation of methylated compounds including marine osmolytes and methylated amines, which are abundant in the marine realm (Figure 2.4). Further, the addition of 100 $\mu$M of glycine betaine, choline, TMA, and TMAO elicited a significant increase in maximum cell yield over nutrient-deplete control (Table 2.1).

### *2.3.5 Growth on reduced sulfur compounds*

No significant enhancement of growth was detected when 1 mM of thiosulfate was added to blank NSW or ASW (Figure 2.1). Because *Ca.* T. singularis is closely related to sulfur-oxidizing symbionts and hydrothermal vent plume microbes, alternative reduced sulfur sources common to those environments were tested for growth responses. At concentrations of 100 $\mu$M, thiosulfate, taurine, and thiotaurine had no significant effect on the growth of

*Ca.* T. singularis in ASW media (Table 2.2). Hypotaurine inhibited growth 56%, while sulfur-containing amino acids cysteine and methionine were toxic to growth (Table 2.2).

## *2.4 Conclusions*

Taxa from the gammaproteobacterial clade of SUP05 commonly share the genetic potential for sulfur-oxidation coupled to carbon fixation (Walsh et al., 2009), yet metabolic diversity in the consumption and utilization of organic carbon may vary by taxon. Taxa from the *Arctic96BD-19* sub-clade of SUP05, including *Ca.* T. singularis, inhabit oxygenated waters and have the enhanced genetic potential for transporting and processing DOC while maintaining the genes necessary for carbon fixation, including form 1Aq of RuBisCO Marshall and Morris (2013). Here, we show using physiologic growth experiments coupled with proteomic surveys that strain PS1 functions as an organic carbon consuming heterotroph without a requirement for reduced sulfur for chemosynthesis under nutrient-replete and -deplete conditions.

A defining characteristic of the SUP05 clade is its role in the sulfur cycle and as a sulfur-oxidizing chemoautotroph. As a member of the sulfur-oxidizing SUP05 clade, strain PS1 contains the genes that encode the enzymes APS reductase (*aprMAB*) and ATP sulfurylase (*sat*) which catalyze the oxidation of thiosulfate to sulfate, producing APS as a byproduct (Ghosh and Dam, 2009; Lalish, 2015). An earlier study reported a positive growth response with the addition of thiosulfate and glucose and detected globules within strain PS1 cells that were comprised of elemental sulfur (Marshall and Morris, 2013), supporting genomic predictions for incomplete oxidation of sulfur for energy when grown on natural seawater media supplemented with organic carbon. In this study, we detected low levels of expression of sulfur oxidation proteins (AprMAB and ATP sulfurylase) when strain PS1 was grown

on nutrient-replete media. However, strain PS1 did not require the addition of exogenous reduced sulfur when grown on ASW (Figure 2.1, Table 2.2), nor did we detect any sulfur oxidation proteins when grown on ASW amended with only thiosulfate. In a proteomic study of the sulfur-oxidizing bacterial endosymbiont for *Riftia pachyptila*, proteins essential for sulfur oxidation were among the most highly expressed under sulfur-rich conditions (Markert et al., 2007). While we cannot definitively state that strain PS1 did not express any sulfur oxidation proteins under the nutrient-deplete condition, it is unlikely that sulfur oxidation was a dominant metabolism given our proteomic coverage. This indicates that strain PS1 maintains the genetic potential for the oxidation of thiosulfate or other reduced sulfur source but it is not a significant or primary mode of energy acquisition.

By combining genomic predictions, physiological growth responses, and protein expression data, we propose a metabolic model for the growth of strain PS1 on methylated compounds primarily for energy acquisition (Figure 2.4). Methylated compounds are common constituents of marine DOC. Marine osmolytes glycine betaine, choline, and trimethylamine N-oxide (TMAO) produced by phytoplankton and other marine organisms are among the most abundant methylated compounds in the DOC mix and are readily metabolized by marine bacteria (Diaz et al., 1992; Barrett and Kwan, 1985; Sun et al., 2011; Lidbury et al., 2015, 2014). Methylotrophy has been well described for over a century; however, recent advances in genomic and physiological studies have shown that newly characterized microbes can use novel combinations of known enzymes or have entirely undiscovered pathways of using C1 compounds (Chistoserdova et al., 2009). When strain PS1 was grown in the presence of diatom lysate, that presumably contains diatom-derived methylated osmolytes, the bacterium expressed all the proteins necessary for obtaining energy from the oxidation of methyl groups. Further, strain PS1 demonstrated enhanced growth when supplemented with

individual osmolytes or C1-intermediates in the proposed methylotrophic metabolism (Figure 2.4). Recent culturing studies using SAR11, the most abundant marine heterotrophic bacterium, has led to the discovery of methylovory, through which the bacteria oxidize C1 compounds solely for cellular energy, distinguishing them from methylotrophs (Sun et al., 2011). Strain PS1 oxidized methylated compounds as indicated by an increase in growth and expression of proteins in the methylotrophy pathway, though future studies will be necessary to distinguish between complete oxidation for energy and biosynthetic incorporation of these relatively abundant marine organic carbon sources.

## *2.5 Experimental procedures*

### *2.5.1 Culturing conditions for T. singularis*

A defined, artificial seawater media (ASW) was used to test the carbon and energy requirements for growth of T. singularis (for ingredients and details of media preparation see Carini et al. (2012). All cultures were grown in acid washed (10% HCl), autoclaved 250-ml polycarbonate flasks containing 50-mL of ASW at in-situ temperature 13°C under 16-hr light and 8-hr dark cycles. Purity and identification of cultures were checked using terminal restriction fragment length polymorphism analysis and sequence of the 16S rRNA gene as previously described (Shah et al., 2016).

Lysate from the diatom *Thalassiosira pseudonana* CCMP1335 was prepared by growing the diatom on F/2 media to late exponential stage. Diatom cells were collected by concentrating the diatom culture via centrifugation at 3220 x G for 30 minutes at 20°C then subjecting the cell pellet to a series of freeze-thaw cycles (at -80°C/70°C) to lyse the cells. The lysed diatom cells were filtered through two in-line 0.2 m filters. Lysate was added to the ASW media to a final concentration equivalent to a ratio of pre-concentrated diatom culture

volume to ASW volume of 1:1. For example, to prepared 50 mL of ASW the equivalent of 50 mL of diatom culture (cell density  1.0x$10^6$ cells/mL) was concentrated, lysed, filtered, and added back to the ASW. Growth experiments testing carbon substrates were carried out in ASW amended with 1 mM thiosulfate and 100 $\mu$M of each carbon substrate and ASW with thiosulfate only as a control. Growth experiments testing reduced sulfur sources were carried out in ASW with 100 $\mu$M of each sulfur source and blank ASW as a control. Cell density in all bacterial cultures were enumerated by staining with SYBR Green and counted on a Guava EasyCyte flow cytometer (MilliporeSigma, Burlington, MA, USA).

### *2.5.2 Protein extraction and analysis*

Detailed materials and methods used for protein extraction and analysis are described in Appendix B (SI Methods). Briefly, exponential phase PS1 cells were lysed from filters using chemical and bead-beating techniques. Proteins were extracted from lysed cells through disulfide reduction, alkylation, and finally in-solution trypsin/Lys-C digestion. An internal standard containing synthetic peptides (Hi3 Escherichia coli Standard, Waters) was added to desalted protein samples. Samples were injected onto a Waters ACQUITY M-class UPLC coupled to a Thermo QExactive HF Orbitrap high-resolution mass spectrometer (HRMS) equipped with a nano-electrospray (NSI) source.

Data processing was conducted using software from the trans-proteomic pipeline (TPP v.4.8.1) (Nesvizhskii et al., 2007). Briefly, raw data was searched using COMET (Eng et al., 2013, 2015) against a FASTA protein database consisting of either PS1 (nutrient-deplete treatment) or PS1 and Thalassiosira pseudonana (nutrient-replete treatment), as well as the E. coli Hi3 standard. Following COMET searches, files were searched in PeptideProphet, iProphet (Shteynberg et al., 2011), and ProteinProphet (Nesvizhskii et al., 2003) within

the TPP. The final list of protein identifications was filtered using a 95% or greater protein probability as calculated from ProteinProphet corresponding to a false discovery rate <0.5%.

### *2.5.3 Identification of constitutively and differentially expressed proteins*

Differential protein expression between the thiosulfate-only and the lysate plus thiosulfate treatment was calculated as the $\log_2$ fold change of the average normalized spectral counts per cell. Differential expression between -0.5 and 0.5 were considered to be constitutively expressed in the two treatments. Differential expression greater than 0.5 represented those proteins that had at least a two-fold increase in expression in the thiosulfate-only treatment. Likewise, differential expression less than -0.5 represented proteins that had at least a two-fold increase in expression in the lysate plus thiosulfate treatment. Only proteins that were detected in at least two replicate injections of each treatment were considered in the differential protein analysis.

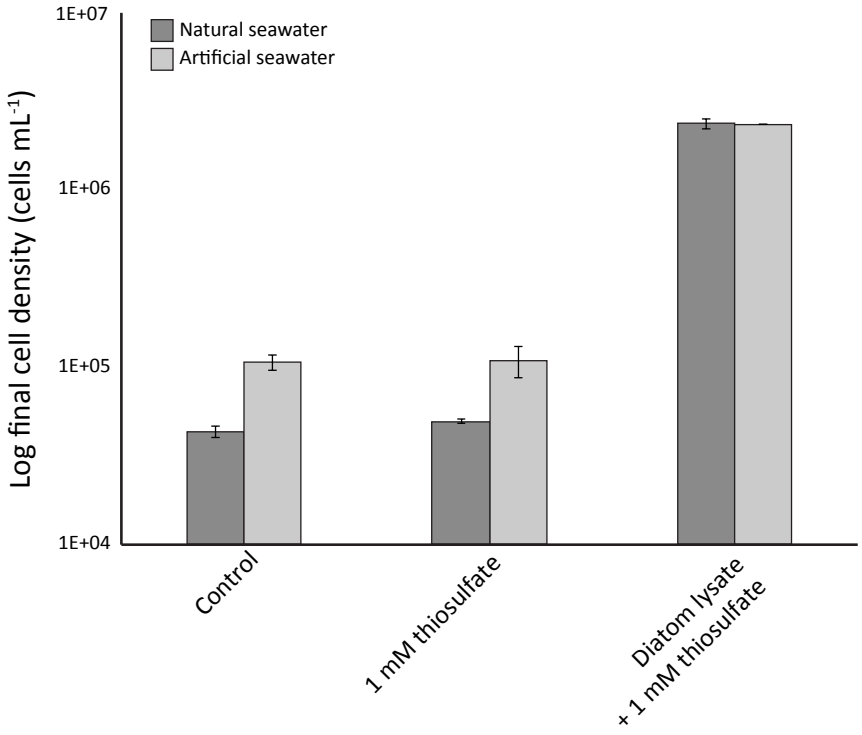


Figure 2.1: Final cell densities of *Ca.* T. singularis strain PS1 on natural seawater media and artificial seawater media amended with either 1 mM thiosulfate or with 1 mM thiosulfate and a dilute diatom lysate to simulate growth under nutrient-deplete or nutrient-replete conditions, respectively.

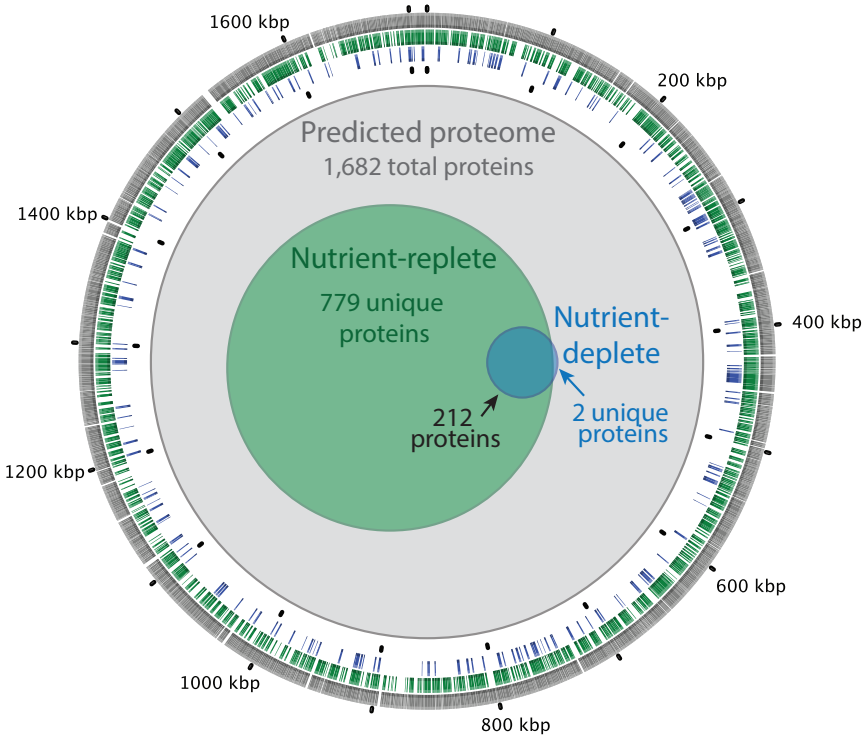


Figure 2.2: Proteomic coverage of *Ca.* T. singularis strain PS1 grown on nutrient-deplete and nutrient-replete media. Ring assignments from inner: (blue) proteomic coverage on nutrient-deplete media, (green) proteomic coverage on nutrient-replete media, (gray) predicted proteome from genome using UniProt. The inner Venn diagram shows the overlap of proteins detected in the nutrient-replete and nutrient-deplete proteomes with each other and within the predicted proteome of strain PS1.

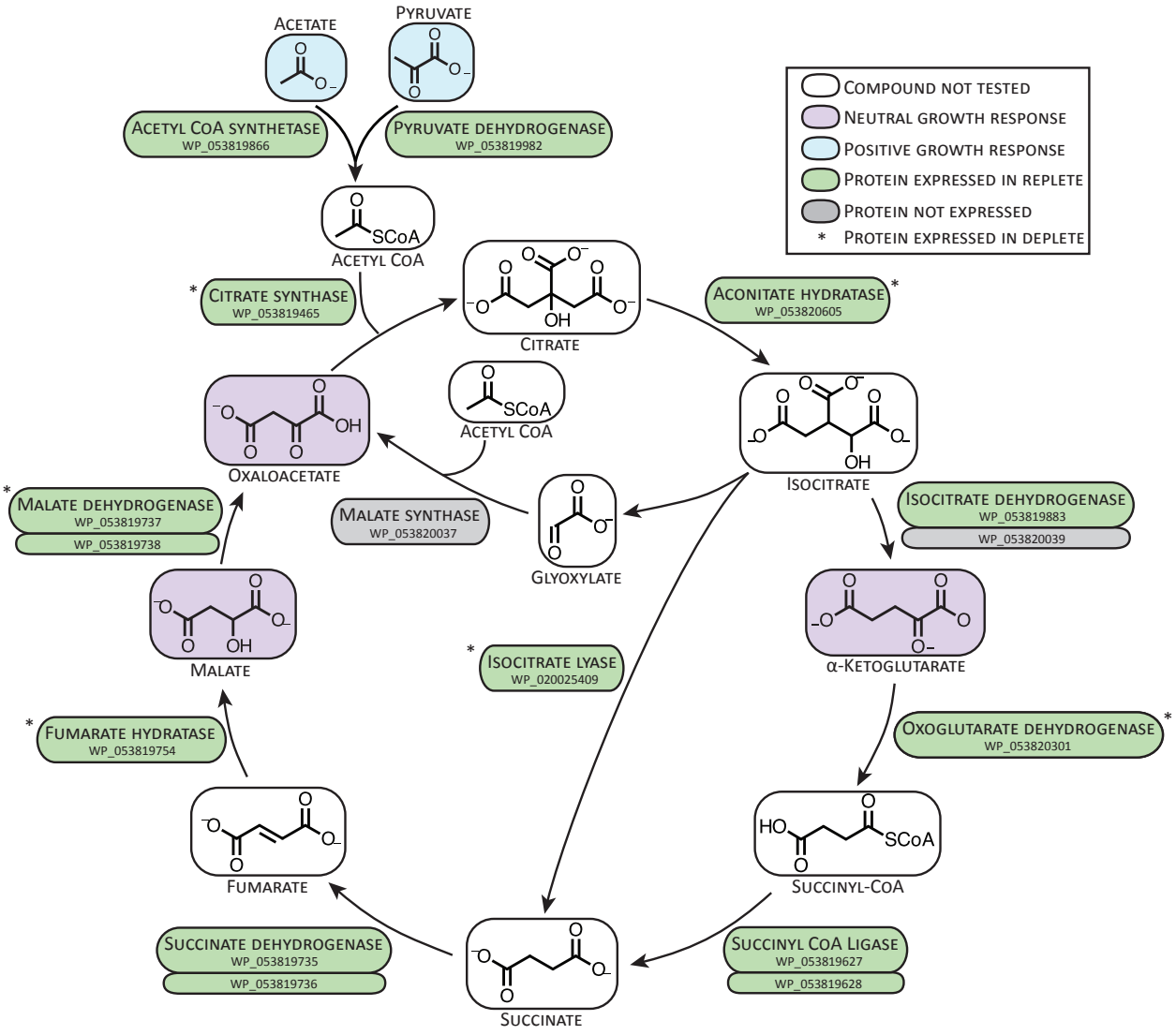


Figure 2.3: The TCA cycle in *Ca.* T. singularis strain PS1. NCBI protein IDs highlighted in green boxes were expressed in the nutrient-replete proteome, an (*) indicates expression also in the nutrient-deplete proteome, and proteins in a gray box were not detected in either proteome. Two proteins within the same box depict multiple subunits of an enzyme, while two stacked boxes represent multiple copies of the protein-encoding gene within the strain PS1 genome. Compounds highlighted blue enhanced the growth of strain PS1, compounds in purple had no significant effect on growth, and compounds in white were not tested.

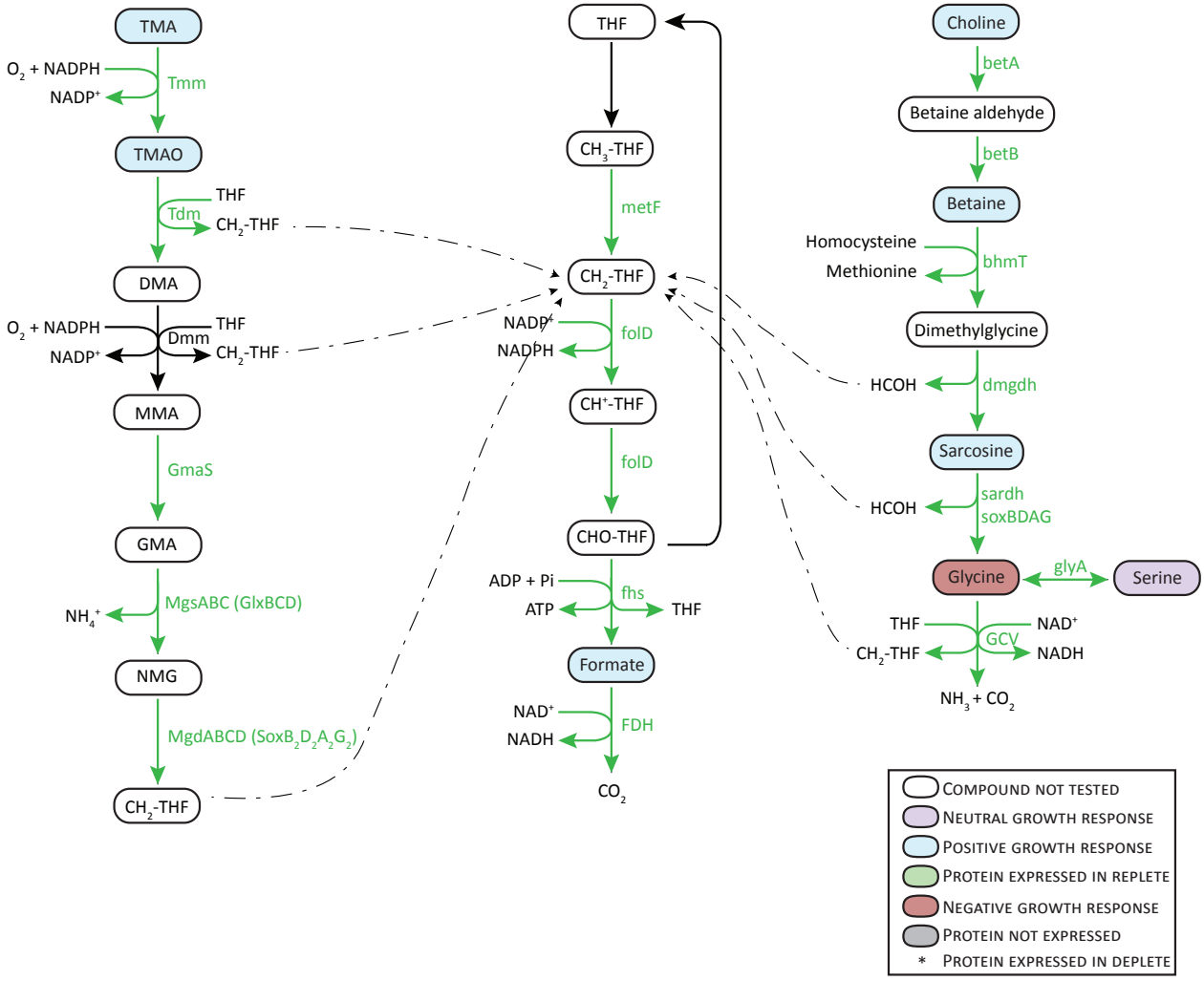


Figure 2.4: Proposed model for the metabolism of methylated organic substrates and marine osmolytes in *Ca.* T. singularis strain PS1. Substrates in blue text elicited an enhancement of growth, substrates in red inhibited growth, substrates in purple did not significantly affect growth, and substrates in black were not tested for a growth response. Proteins encoded by the genes in green text were expressed when *Ca.* T. singularis was grown on a dilute diatom lysate.

Table 2.1: Potential organic carbon sources for *Ca* T. singularis strain PS1 when grown in ASW with 1mM thiosulfate.

| Potential carbon source | Maximum cell density[a] |
|---|---|
| Diatom lysate | 13.70 ± 3.68 |
| Glycine betaine | 3.01 ± 0.55 |
| Formate | 2.86 ± 1.06 |
| Sarcosine[b] | 2.11 ± 0.10 |
| Choline | 2.12 ± 0.35 |
| TMAO | 1.80 ± 0.06 |
| Malonate | 1.71 ± 0.13 |
| Acetate | 1.71 ± 0.12 |
| Pyruvate | 1.58 ± 1.20 |
| TMA | 1.54 ± 0.01 |
| Oxaloacetate | 1.07 ± 0.29 |
| Nutrient-deplete (control) | 0.93 ± 0.20 |
| DHPS | 0.85 ± 0.16 |
| $\alpha$-ketoglutarate | 0.84 ± 0.33 |
| Serine[b] | 0.83 ± 0.16 |
| Malate | 0.80 ± 0.17 |
| Glucose | 0.78 ± 0.14 |
| DMSP | 0.59 ± 0.15 |
| Glycine[b] | 0.35 ± 0.03 |

[a] Displayed as $10^4$ cells $mL^{-1}$

[b] S.D. based on duplicates rather than triplicates

Table 2.2: Potential reduced sulfur sources for *Ca.* T. singularis strain PS1 when grown in ASW with no added organic carbon.

| Potential sulfur source | Maximum cell density[a] |
|---|---|
| Thiotaurine | 1.58 ± 0.23 |
| Taurine | 1.53 ± 0.11 |
| Blank control | 1.45 ± 0.10 |
| Thiosulfate | 1.43 ± 0.16 |
| Hypotaurine | 0.64 ± 0.10 |
| Cysteine | 0.03 ± 0.01 |
| Methionine | 0.03 ± 0.01 |

[a] Displayed as $10^4$ cells $mL^{-1}$

# Chapter 3

# THE COMPLETE GENOME SEQUENCE OF *CANDIDATUS* THIOGLOBUS SINGULARIS STRAIN NP1, REVEALS A STRICTLY HETEROTROPHIC REPRESENTATIVE FROM THE SUP05 CLADE

Rachel Lange Spietz, Katharine M. Lalish, Xiaowei Zhao, Daniela Nicastro, and Robert Morris [In prep]

## *3.1 Abstract*

*Candidatus* Thioglobus singularis strain NP1 is a Gammaproteobacterium that was isolated from the deep-chlorophyll maximum (45 m) in the Northeast Pacific oligotrophic gyre. The complete NP1 genome was sequenced using the PacBio platform, which produced a single circular chromosome 1.69 Mbp in length. We compared the NP1 genome to the genomes of closely related representatives from the SUP05 clade of symbiotic and free-living bacteria to identify differences in the metabolic potential of strains isolated from different environments. Strain NP1 is most closely related to a coastal isolate, *Ca.* T. singularis strain PS1 (76.7% identical at the nucleotide level), which was cultured from the Northeast Pacific Ocean in Puget Sound. Similar to strain PS1, strain NP1 codes for a complete TCA cycle and can synthesize all amino acids except histidine. Differences between strain NP1 and two other complete SUP05 genomes include the absence of key genes for glycolysis or gluconeogenesis (aldose 1-epimerase), the oxidative branch of the pentose phosphate pathway (glucose 6-phosphate dehydrogenase, 6-phosphogluconolactonase, 6-phosphogluconate dehydrogenase),

carbon fixation (RuBisCO), sulfur metabolism (*dsrCKLM*, APS kinase, and sulfite reductase), nitrogen acquisition (P-II regulatory protein and ammonium transporters), and type II pilus biosynthesis. The genome of strain NP1 contains three genomic islands relative to other complete and incomplete SUP05 genomes. Notably, strain NP1 is the first sequenced representative from the SUP05 clade that does not code for RuBisCO, indicating that it lacks the ability to fix inorganic carbon using the Calvin-Bensen-Bassham (CBB) cycle. Our analysis of the complete genome sequence of *Ca.* T. singularis stain NP1 indicates that there is significant genetic diversity between closely related strains of SUP05 that may infer varying metabolic roles in the global ocean.

### *3.2 Keywords*

Arctic96BD-19, T. singularis, SUP05, genome

### *3.3 Abbreviations*

PS1: Puget Sound 1 NP1: North Pacific 1 EF1: Effingham Inlet 1

### *3.4 Introduction*

Members of the abundant and cosmopolitan SUP05 clade of Gammaproteobacteria play important roles in carbon, nitrogen, and sulfur cycling in the world's oceans (Walsh et al., 2009; Swan et al., 2011; Marshall and Morris, 2013; Murillo et al., 2014; Shah et al., 2016). Phylogenetic analysis of the 16S rRNA gene suggests that the SUP05 clade is comprised of two closely related subclades that vary in their metabolic potential (Walsh et al., 2009). The *sup05* subclade includes both free-living and symbiotic strains and contributes significantly to carbon-fixation fueled by sulfur oxidation, particularly in low-oxygen regions where they

may simultaneously contribute to nitrogen loss (Glaubitz et al., 2013; Mattes et al., 2013; Shah et al., 2016). The sister subclade, *Arctic96BD-19*, has the genomic potential for carbon fixation (autotrophy), a complete TCA cycle and an abundance of organic transporters (organoheterotrophy), proteorhodopsin (phototrophy), sulfur oxidation (lithotrophy), and the oxidation of methylated compounds (methylotrophy), indicating the potential for dual mixotrophy (for both carbon and energy) in the *Arctic96BD-19* subclade (Walsh et al., 2009; Anantharaman et al., 2013; Lesniewski et al., 2012).

Ecotype variation and niche-partitioning within phylogenetically distinct clades of bacteria is common among oligotrophic bacteria such as SAR11 and *Prochlorococcus*, which demonstrate adaptive radiation to gradients in the availability of organic substrates and differing light levels, respectively (Schwalbach et al., 2010; Brown et al., 2012; Malmstrom et al., 2012; Morris et al., 2012). Similarly, culture-independent single-amplified genomes obtained from the North and South Pacific indicate that members of the SUP05 clade may also display genotypic variation in the content of genes encoding different metabolisms (Swan et al., 2011). The ability for these different groups of marine bacteria to diversify, while also maintaining small genomes and cell sizes, allows them to be highly competitive for scarce resources and aids in their success in the open ocean (Giovannoni et al., 2014).

While most SUP05 have eluded cultivation, cultivated isolates provide insight into the complete genomic potential and physiological traits within the clade (Marshall and Morris, 2013; Shah et al., 2016; Marshall and Morris, 2015; Shah et al., 2016). Here, we present the complete genome sequence of the third cultivated free-living representative from the SUP05, *Candidatus* Thioglobus singularis strain NP1 and compare its genomic content to other complete or nearly complete SUP05 genomes to elucidate patterns of genetic diversity that have enabled this clade of bacteria to populate the global ocean.

## 3.5 Organism classification

### 3.5.1 Classification and features

Based on 16S rDNA sequence identity, strain NP1 is a member of the genus *Thioglobus* and species *singularis*, within the *Gammaproteobacteria* (Figure 3.1). The closest cultivated relative to strain NP1 is *Ca.* T. singularis strain PS1 (99.6% rDNA sequence identity) and these two strains cluster within the *Arctic96BD-19* subclade of SUP05. The sister subclade (*sup05*) contains isolate *Ca.* T. autotrophicus strain EF1, which shares 96.9% 16S rDNA sequence identity with *Ca.* T. singularis strain NP1 (Figure 3.1). NP1 cells are coccoid with a diameter of 1.0-1.2 $\mu$m (Figure 3.2), similar to *Ca.* T. singularis strain PS1 (Marshall and Morris, 2013). The genus name, *Thioglobus*, refers to intracellular sulfur ("thio") globules ("globus") in SUP05, yet no sulfur globules were detected under optimal growth conditions when excess reduced sulfur was added to the media (Figure 3.2). Other key features of *Ca.* T. singularis strain NP1 are summarized in Table 3.1.

## 3.6 Genome sequencing information

### 3.6.1 Genome project history

*Ca.* T. singularis, designated strain NP1, represents the third cultured isolate from the SUP05 clade and the second from the *Arctic96BD-19* subclade. As part of a larger project to elucidate the genomic potential of the metabolically diverse SUP05 clade, the genome sequence of strain NP1 was obtained. Strain NP1 was isolated from the oligotrophic open-ocean, thus provides a comparison to strain PS1 isolated from a nutrient-rich coastal environment. The comparison between NP1 and PS1 is useful in understanding adaptations to different nutrient regimes within the *Arctic96BD-19* subclade. Genome sequencing and

assembly of *Ca.* T. singularis strain NP1 were conducted at the University of Washington (Seattle, WA) and the sequence has been deposited in GenBank under accession number SAMN07775767. A summary of the genome sequencing project can be found in Table 3.2.

### *3.6.2 Growth conditions and genomic DNA preparation*

Strain NP1 was isolated from the deep chlorophyll maximum (45 m) in the North Pacific gyre (46° 16'26.04"N, 129° 47'43.8"W), located approximately 600 km from Seattle, WA, USA by high-throughput cultivation, as previously described (Marshall and Morris, 2013). Briefly, whole water samples were diluted in seawater sterilized using a tangential-flow-filtration (TFF) unit with a 30kDa cutoff and aliquoted into 2-mL culture wells. Culture experiments were in acid-washed, sterile Teflon culture plates that were incubated in the dark at 10°C. Cultures were screened for growth using a Guava EasyCyte flow cytometer (MilliporeSigma, Burlington, MA, USA) and identified by amplifying and sequencing the 16S rRNA gene. Strain NP1 was selected for further culturing and genome sequencing.

Optimal growth of strain NP1 was obtained on natural seawater media collected from the surface waters of Puget Sound, WA, USA, filter sterilized, and amended with 1mM thiosulfate and a cellular lysate harvested from the diatom, *Thalassiosira pseudonana* CCMP 1335, under aerobic conditions, as previously described (Marshall and Morris, 2013). Strain NP1 was unable to grow to detectable levels on artificial seawater (ASW) media or on a 50/50 mixture of natural seawater media and ASW. To collect DNA for genome sequencing, strain NP1 was grown in twenty 200-mL cultures on natural seawater media amended with diatom lysate and 1mM thiosulfate to satisfy potential heterotrophic and lithotrophic requirements. Cells were harvested during early stationary phase onto 0.2$\mu$m Supor filters. DNA was extracted from filters and purity was verified by terminal restriction fragment

length polymorphism (TRFLP) analysis following previously described methods (Marshall and Morris, 2013). Remaining DNA was purified on a single PowerClean DNA clean-up column (MO Bio Laboratories, Carlsbad, CA, USA). A total of 2.25 $\mu$g of high-quality, purified DNA were submitted for genome sequencing.

### *3.6.3 Genome sequencing and assembly*

Library preparation, sequencing, and analysis, including genome assembly, were conducted at the University of Washington's Genome Science Department's Pacific Biosciences Services lab (http://pacbio.gs.washington.edu). A PacBio Single-Molecule Real-Time (SMRT) 10kb library was constructed and sequenced using a single SMRT cell on the PacBio RSII platform to obtain 1.3 Gbp of DNA (raw reads). De-novo assembly of the genome from reads was conducted using Hierarchical Genome Assembly Process 2 (HGAP2) (Koren et al., 2013). Through the HGAP2 pipeline, reads were preassembled, assembled using Celera, and polished with Quiver. The SMRT Analysis pipeline produced a single linear contig with overlapping ends and mean coverage of 564X. Genome ends were aligned and trimmed in Geneious v7.1.9 to produce a 1.69Mbp genome.

### *3.6.4 Genome annotation*

The complete, finished genome of strain NP1 was annotated using NCBI's Prokaryotic Annotation Pipeline and checked against RAST annotations (Aziz et al., 2008; Overbeek et al., 2014) and IMG annotations (Markowitz et al., 2012). COG classification of proteins were obtained from the IMG annotation pipeline.

### *3.6.5 Genome properties*

The genome of *Ca.* T. singularis strain NP1 contains a single circular chromosome that is 1,689,404 bp long, with a GC content of 34.48% (Figure 3.3). Genome properties of *Ca.* T. singularis strain NP1 are summarized in Table 3.3 and an overview of COG categories are given in Table 3.4. Coding bases accounted for 94% (1,569,404) of the overall genome. Of the 1,738 genes in the strain NP1 genome, 1,692 are predicted to encode proteins and 46 are ribosomal genes. Over three-quarters of the protein-encoding genes were assigned to COG functional categories. Amino acid transport and metabolism proteins (15.79%) comprised the largest proportion of COG-assigned proteins in the NP1.

## ***3.7 Insights from genome***

Three complete genomes from free-living members of the SUP05 showed high similarity at the 16S level, yet vast genomic diversity and variations in metabolic potential (Figure 3.1, SI Figure C.1, SI Table C.3). The average nucleotide identity (ANI) between the entire genomes of strain NP1 and strain PS1 was 76.69%, calculated from 1,302 best blast hits (BBH) between the two genomes. For comparison, the ANI between strains NP1 and EF1 was 70.37%, calculated from 457 BBH (SI Table C.2). Synteny of protein coding regions was high between the two *Arctic96BD-19* strains (NP1 and PS1) but lower between the *Arctic96BD-19* strains and EF1, reflecting evolutionary divergence between the two subclades (Figure 3.4A). Amino acid transport and metabolism proteins also comprised the largest proportion of COG-assigned proteins in strain PS1 (15.6%), but not in strain EF1 (9.23%) (SI table C.1). In strain EF1, proteins involved in translation and ribosome structure and biogenesis were the most abundant COG category (12.66%).

We identified three genomic islands on the NP1 genome compared to other complete

and incomplete SUP05 genome sequences (Figure 3.3). These long stretches of the genome (10-200kbp) are often characterized by GC-content different from the rest of the genome and are commonly flanked by tRNA genes (Juhas et al., 2009). In strain NP1, all three genomic islands (GEI1, GEI2, and GEI3) are over 10kbp and have decreased GC-content compared to the overall average genome GC-content (Figure 3.3). Two of the three genomic islands (GEI2 and GEI3) are flanked by tRNA sequences and are near integrative host factors (IHF) (Figure 3.3), which may be involved with DNA recombination (Juhas et al., 2009). The majority of genes found within the three genomic islands of strain NP1 are hypothetical proteins without a described function or classification, potentially because they have been recently acquired (Wilhelm et al., 2007; Juhas et al., 2009). While the precise role of genomic islands in marine bacteria is not yet understood, they are known to be regions of the genome where genes are rapidly gained and lost due to horizontal gene transfer and are common in other abundant and cosmopolitan, oligotrophic bacterial clades including *Prochlorococcus* and *Pelagibacter* (Wilhelm et al., 2007; Malmstrom et al., 2012). The presence of the three genomic islands detected here may contribute to diversification within the SUP05 clade.

### *3.7.1 Metabolic potential*

A comparison of the complete genome sequence for *Ca.* T. singularis strain NP1 to previously sequenced isolates *Ca.* T. singularis PS1 and *Ca.* T. autotrophicus strain EF1 revealed that the open-ocean strain NP1 genome lacked several defining genes of the genus *Thioglobus* (SUP05 clade) (SI Figure C.1, SI Table C.3). Strain NP1 is missing the genes that encode the large and small subunits of ribulose-bisphosphate carboxylase (RuBisCO), *cbbLS*, and rubisco accessory genes, *cbbR* and *cbbY*. In the closely related strain PS1, *cbbRLS* are flanked by *argJ*, a protein involved in two steps of arginine biosynthesis, and an ATPase related to

*cbbQ*. In the genome of strain NP1 *argJ* and the ATPase are adjacent to each other while *cbbRLS* are missing (Figure 3.4B). This may suggest that strain NP1 has lost the genetic potential to fix inorganic carbon via the Calvin Benson Bassham (CBB) cycle, which makes it unique among closely related strains, *Ca.* T. singularis PS1 and *Ca.* T. autotrophicus strain EF1 (Swan et al., 2011; Marshall and Morris, 2013; Murillo et al., 2014; Shah et al., 2016; Marshall and Morris, 2015). Experimentally, strain EF1 has been shown to be a strict chemoautotroph (Shah et al., 2016); however, it remains unclear whether strain PS1 utilizes the CBB to fix inorganic carbon as part of its central metabolism.

Strain NP1 has the genomic potential to use organic carbon substrates through glycolysis, a complete TCA cycle, and glyoxylate bypass. This is similar to coastal strain PS1 but different than strain EF1, which lacks 2-oxoglutarate dehydrogenase necessary for a complete TCA cycle (Shah et al., 2016). Like coastal strain PS1, open ocean strain NP1 codes for an abundance of ABC-transporters for organic substrates, suggesting that its primary role in the global ocean is to assimilate and redistribute organic matter or to convert it back to its inorganic constituents. Interestingly, strains NP1 and EF1 lack the genes for the oxidative phase of the pentose phosphate pathway (PPP) (glucose 6-phosphate dehydrogenase, 6-phosphogluconolactonase, and 6-phosphogluconate dehydrogenase) while maintain the genomic potential for the non-oxidative phase, which differs from strain PS1 that contains both phases of the PPP (SI Table C.3, SI Figure C.1). Subpopulations of another oligotrophic marine bacterium, SAR86, lack the oxidative phase of the PPP, which results in a decrease of NADH generation and has been suggested to be an adaptation to utilize different carbon sources (Dupont et al., 2012).

In the oligotrophic open ocean, nitrogen is thought to be a major limiting nutrient to productivity (Zehr and Kudela, 2011); therefore, marine bacteria fulfill their nitrogen re-

quirements through a variety of organic and inorganic sources. Potential nitrogen sources within the SUP05 clade vary between heterotrophic *Ca.* T. singularis strains NP1 and PS1 and strictly chemoautotrophic *Ca.* T. autotrophicus strain EF1. This is likely due to the differences in the central energetic pathways between the two clades. *Ca.* T. autotrophicus strain EF1 uses nitrate as a terminal electron acceptor under anaerobic conditions and consumes ammonium for biosynthetic nitrogen requirements (Shah et al., 2016). *Ca.* T. autotrophicus strain EF1 has genes that encode three different nitrogen regulatory P-II proteins and a single ammonium transporter (SI Table C.3). Conversely, *Ca.* T. singularis strains NP1 and PS1 contain one and two nitrogen regulatory P-II genes, respectively. Both have two additional ammonium transporter genes. Recently, field and lab-based studies have shown that non-methylotrophic heterotrophic marine bacteria use methylated amines to supplement their nitrogen requirements in coastal marine environments (Lidbury et al., 2015; Taubert et al., 2017). Methylated amines are produced as osmoprotectants in marine organisms and are abundant in productive regions of the worlds oceans, including coastal regions and chlorophyll a maxima in the open ocean (Gibb et al., 1999; Gibb and Hatton, 2004; Beale and Airs, 2016). This may constitute a significant fraction of the dissolved organic nitrogen pool available to heterotrophic organisms. Both *Ca.* T. singularis genomes (PS1 and NP1) contain the genes necessary to use methylated amines as organic carbon and nitrogen sources (SI Table C.3), further expanding the suite of nitrogen substrates available to this metabolically diverse subclade of SUP05.

The SUP05 clade is often characterized by the ability of its members to oxidize reduced inorganic sulfur (Anantharaman et al., 2013). *Ca.* T. singularis strain NP1 is missing key sulfur oxidation genes found in other SUP05 (SI Figure C.1, SI Table C.3). These include membrane-bound sulfur oxidation genes, *dsrCKLM*, and sulfite reductase. NP1 has retained

*aprABM*, ATP sulfurylase, and four rhodaneses. For comparison, coastal strain PS1 has a broader set of sulfur oxidation genes (*aprABM*, ATP sulfurylase, *dsrCKLM*, three rhodaneses, and sulfite reductase) giving it the potential for lithotrophic energy production (Marshall and Morris, 2013; Lalish, 2015). The genome of the strictly chemolithoautotrophic strain EF1, contains a complete set of genes to oxidize inorganic sulfur (*fccAB*, *dsrABCH*, *aprABM*, *soxABXYZ*, ATP sulfurylase, and rhodanese sulfurtransferase) (Shah et al., 2016). Evidence of a significantly reduced set of sulfur oxidation genes suggest that NP1 can oxidize reduced sulfur for energy but may be limited in its ability to oxidize or reduce a wider array of sulfur compounds or carry-out reduction of stored intracellular sulfur for assimilation because it is lacking dsr genes and sulfite reductase (SI Figure C.1, SI Table C.3). While no sulfur globules were found using cryo-EM (Figure 3.2), further research into the genetic pathways for sulfur storage are necessary to determine whether strain NP1 maintains the ability to form intracellular sulfur globules.

### *3.7.2 Type II Secretion*

*Ca.* T. singularis strain PS1 and *Ca.* T. autotrophicus strain EF1 genomes contain an operon of genes that encode a type II secretion system (*pilTSOPQBC*, pseudo-pilin *pulG*, and fimbrial pilin), yet *Ca.* T. singularis strain NP1 lacks this entire operon (SI Figure C.1). Type II secretion systems are thought to be involved in DNA uptake or protein secretion in other heterotrophic marine bacteria, *Pelagibacteria ubique* (Zhao et al., 2017) . Lack of this type II secretion system in strain NP1 suggests that these functions are not essential for heterotrophic SUP05 in the open-ocean. Dissolved DNA concentrations, for example, decrease from coastal to open ocean waters (DeFlaun et al., 1987).

### *3.8 Conclusions*

The complete genome sequence of *Ca.* Thioglobus singularis strain NP1, an open-ocean isolate from the SUP05 clade, lacks the genetic potential for carbon fixation via the CBB yet maintains a complete TCA cycle, organic matter transporters, and organic and inorganic nitrogen assimilation pathways, thus representing the first strictly heterotrophic isolate from the SUP05 clade. Further, strain NP1 has a reduced set of sulfur oxidation genes and lacks a complete operon for a type II secretion system. The metabolic diversity within the SUP05 clade is striking given the high relatedness among members at both the 16S rRNA and whole genome levels. These results suggest that some members of the SUP05 clade, including strain NP1, potentially function strictly as chemoorganotrophs, contrary to previously reports suggesting strict or facultative chemolithoautotrophy throughout the ubiquitous marine clade (Swan et al., 2011; Anantharaman et al., 2013; Glaubitz et al., 2013; Murillo et al., 2014).

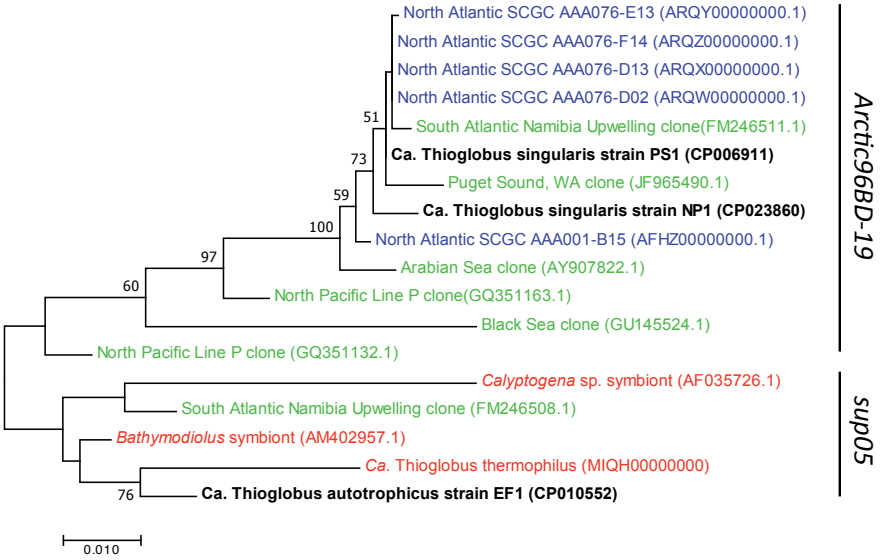


Figure 3.1: Phylogenetic tree showing the position of *Ca.* T. singularis strain NP1 within the SUP05 clade of marine Gammaproteobacteria. Complete genomes from cultured isolates shown in black text, single-cell genomes in blue text, symbionts in red text, and environmental sequences in green text. The 16S rRNA sequences were aligned in ClustalW and trimmed to 1232bp. The phylogenetic tree was constructed using the Maximum Likelihood method based on the Tamura-Nei model in MEGA7. Boostrap values were calculated on 100 interations and values >50 are displayed.

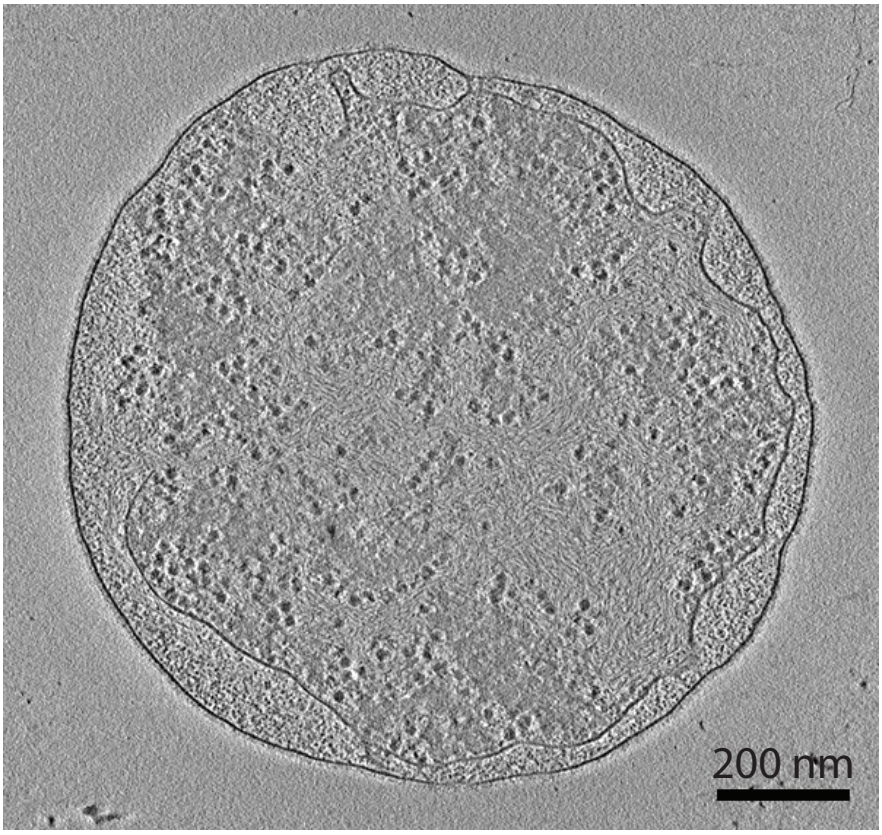


Figure 3.2: Cryo-electron micrography image of *Ca.* Thioglobus singularis strain NP1.

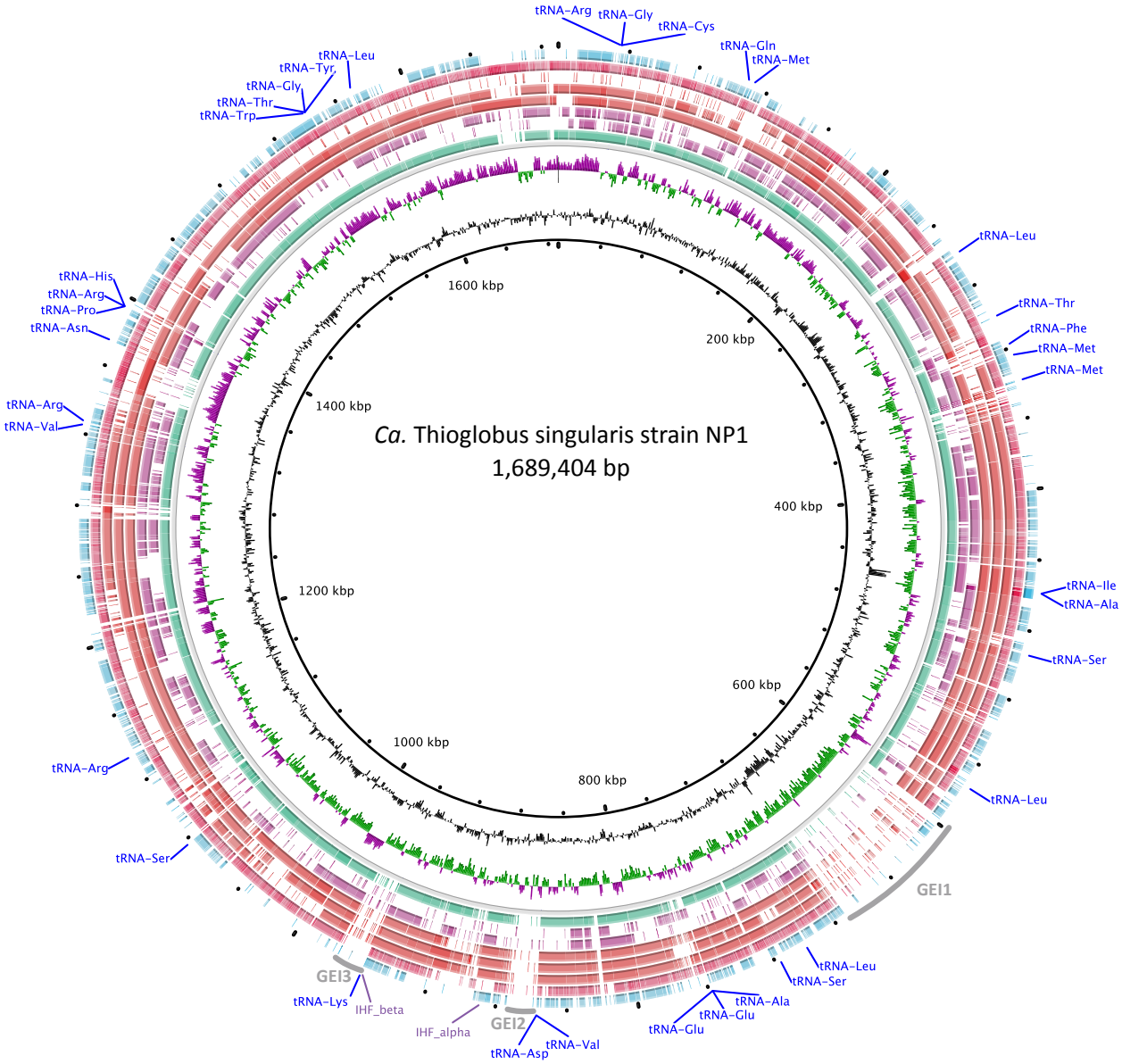


Figure 3.3: Overview of *Ca.* T. singularis strain NP1 genome and genomic comparisons to other *Thioglobus sp.* genomes. Rings from innermost to outermost: strain NP1 GC-content, strain NP1 GC-skew (purple positive skew, green negative skew), *Ca.*T. singularis strain PS1 genome (teal), two single-cell genomes from the deep South Atlantic (purple) (AAA007-B15, AAA007-O20), (red) four single-cell genomes from the surface North Atlantic (AAA076-D13, AAA076-D02, AAA076-F14, AAA076-E13), and *Ca.* T. autotrophica strain EF1 (light blue).

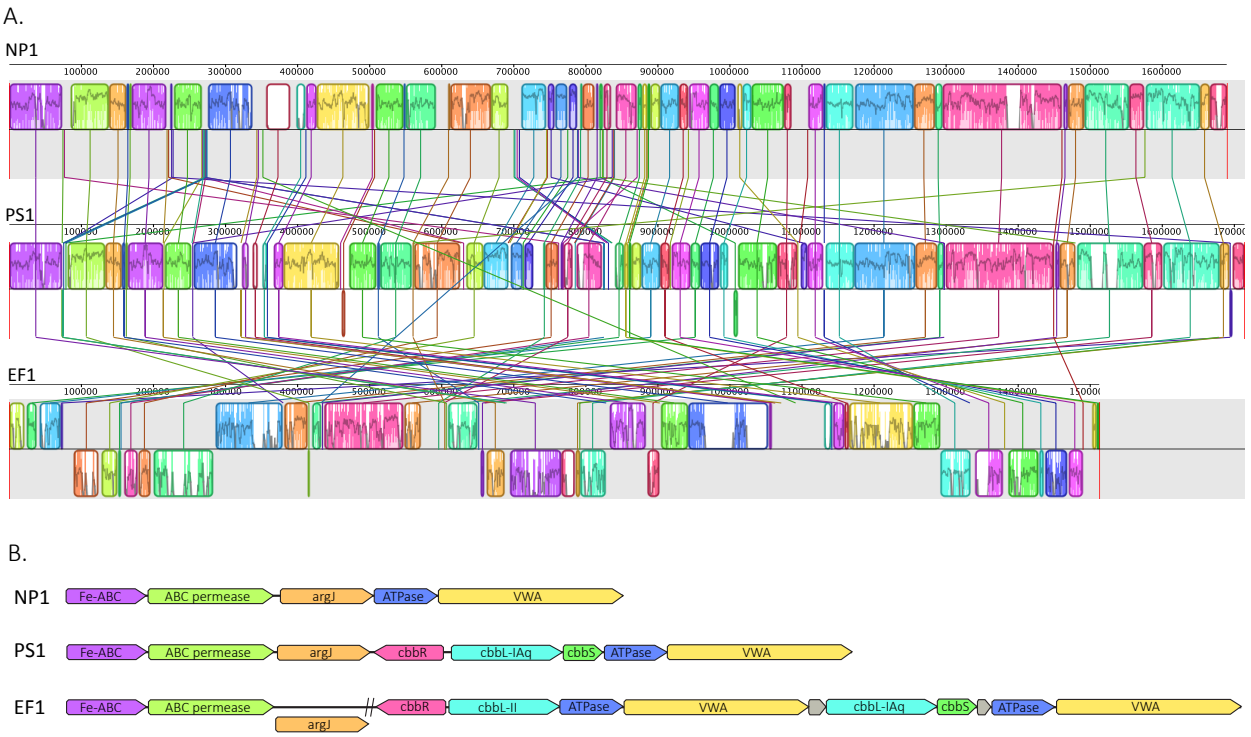


Figure 3.4: Overview of genome synteny and presence of key Calvin-Benson-Basshom (CBB) cycle genes across the complete genome sequences of three free-living isolates of the SUP05 clade. (A) Output from a progressive Mauve alignment implemented in Geneious (v 7.1.9, Newark, NJ). Linear collinear blocks are shown in different colored boxes and were calculated on a minimum of 30,000 matching nucleotides across two or more genomes. (B) CBB regions on the three SUP05 genomes shows that strain NP1 is missing *cbbRLS*. **Fe-ABC**, Fe ABC transporter binding protein; **argJ**, bifunctional arginine biosynthesis protein; **VWA**, von Willebrand factor A.

Table 3.1: Classification and general features of *Ca.* Thioglobus singulariss strain NP1 designation.

| **MIGS ID** | **Property** | **Term** | **Evidence codes**[1] |
|---|---|---|---|
| | Classification | Domain Bacteria | TAS[2] |
| | | Phylum *Proteobacteria* | TAS[3] |
| | | Class *Gammaproteobacteria* | TAS[4] |
| | | Order Unclassified | NAS |
| | | Family Unclassified, S-oxidizing bacteria | NAS |
| | | Genus *Candidatus* Thioglobus | TAS[5] |
| | | Species *Candidatus* Thioglobus singularis | TAS[4] |
| | | (Type) strain NP1 | IDA |
| | Gram stain | Negative | NAS |
| | Cell shape | Coccus | IDA |
| | Motility | Non-motile | IDA |
| | Sporulation | Unknown | |
| | Temp. range | Preference for 10-13°C, psychrophile | IDA |
| | Optimum temp. | Unknown | |
| | pH range; optimum | Unkown, about 8.1 | IDA |
| | Carbon source | Unknown | |
| GS-6 | Habitat | Marine | IDA |
| MIGS-6.3 | Salinity | About 35 PSU | IDA |
| MIGS-22 | Oxygen requirement | Aerobic | IDA |
| MIGS-15 | Biotic relationship | Free-living | TAS[5], IDA |
| MIGS-14 | Pathogenicity | Non-pathogenic | NAS |
| MIGS-4 | Geographic location | North Pacific gyre | TAS[5], IDA |
| MIGS-5 | Sample collection | August 2011 | TAS[5], IDA |
| MIGS-4.1 | Latitude | 46° 16.434'N | TAS[4], IDA |
| MIGS-4.2 | Longitude | 129.° 47.73'W | TAS[4], IDA |
| MIGS-4.4 | Altitude | -45m | TAS[5], IDA |

[1]Evidence codes: TAS, traceable author statement; IDA, inferred from direct assay; NAS, non-traceable author statement

[2]Woese et al. (1990)

[3]Garrity and Bell (2005)

[4]Garrity et al. (2005)

[5]Marshall and Morris (2015)

Table 3.2: Project information.

| **MIGS ID** | **Property** | **Term** |
|---|---|---|
| MIGS-31 | Finishing quality | Complete |
| MIGS-28 | Libraries used | Single PacBio 10kb SMRT cell |
| MIGS-29 | Sequencing platforms | PacBio RSII |
| MIGS-31.2 | Fold coverage | 564x |
| MIGS-30 | Assemblers | HGAP2 |
| MIGS-32 | Gene calling method | GeneMarkS+ (NCBI PGAP); Prodigal (IMG/ER) |
| | Locus tag | CRN91 |
| | Genbank ID | SAMN07775767 |
| | Genbank Date of Release | Upon publication |
| | GOLD ID | Ga0216371 |
| | BIOPROJECT | PRJNA414010 |
| MIGS-13 | Source Material Identifier | Aquatic |
| | Project relevance | Enviromental microbiology, open ocean, free-living, SUP05 |

Table 3.3: Genome statistics.

| **Attribute** | **Value** | **%age** |
|---|---|---|
| Genome size (bp) | 1,689,404 | 100 |
| DNA coding (bp) | 1,596,072 | 94.48 |
| DNA G+C (bp) | 591,027 | 34.98 |
| DNA scaffolds | 1 | 100 |
| Total genes | 1,738 | 100 |
| Protein coding genes | 1,692 | 97.35 |
| RNA genes | 46 | 2.65 |
| Pseudo genes | 0 | 0 |
| Genes in internal clusters | 220 | 12.66 |
| Genes with function prediction | 1,470 | 84.58 |
| Genes assigned to COGs | 1,349 | 77.62 |
| Genes with Pfam domains | 1,516 | 7.23 |
| Genes with signal peptides | 104 | 5.98 |
| Genes with transmembrane helices | 359 | 20.66 |
| CRISPR repeats | 0 | 0 |

Table 3.4: Number of genes associated with general COG functional categories.

| **Code** | **Value** | **%age** | **Descrription** |
|---|---|---|---|
| J | 160 | 10.62 | Translation, ribosomal structure and biogenesis |
| A | 1 | 0.07 | RNA processing and modification |
| K | 42 | 2.79 | Transcription |
| L | 64 | 4.25 | Replication, recombination, and repair |
| B | 1 | 0.07 | Chromatin structure and dynamics |
| D | 15 | 1.00 | Cell cycle control, cell division, chromosome partitioning |
| V | 24 | 1.59 | Defense mechanisms |
| T | 23 | 1.53 | Signal transduction mechanisms |
| M | 89 | 5.91 | Cell wall/membrane biogenesis |
| N | 6 | 0.40 | Cell motility |
| U | 17 | 1.13 | Intracellular trafficking and secretion |
| O | 74 | 4.91 | Posttranslational modification, protein turnover, chaperones |
| C | 111 | 7.37 | Energy production and conversion |
| G | 86 | 5.71 | Carbohydrate transport and metabolism |
| E | 238 | 15.79 | Amino acid transport and metabolism |
| F | 67 | 4.45 | Nucleotide transport and metabolism |
| H | 119 | 7.90 | Coenzyme transport and metabolism |
| I | 85 | 5.64 | Lipid transport and metabolism |
| P | 63 | 4.18 | Inorganic ion transport and metabolism |
| Q | 37 | 2.46 | Secondary metabolites biosynthesis, transport, and catabolism |
| R | 134 | 8.89 | General function prediction only |
| S | 46 | 3.05 | Function unknown |
| - | 389 | 22.38 | Not in COGs |

Chapter 4

# DEEP-SEA VOLCANIC ERUPTIONS CREATE UNIQUE CHEMICAL AND BIOLOGICAL LINKAGES BETWEEN THE SUBSURFACE LITHOSPHERE AND OCEANIC HYDROSPHERE

Rachel Lange Spietz, David Butterfield, Nathaniel J. Buck, Benjamin Larson, William W. Chadwick Jr., Sharon Walker, Deborah Kelley, and Robert Morris [In review]

## *4.1 Abstract*

In April 2015, pressure recorders, seismometers, and hydrophones attached to the Ocean Observatories Initiatives (OOI) Cabled Array on Axial Seamount detected, in real-time, a volcanic eruption predominantly located along the north rift zone (NRZ). Real-time detection enabled a rapid response cruise to augment OOI data with physical, chemical, and biological data about the eruption and the new lava flows along the NRZ, demonstrating the synergistic value of real-time monitoring with rapid response efforts that go beyond the boundaries of a fixed cabled array. We show that the 2015 eruption gave rise to chemically and microbiologically variable hydrothermal plumes over new NRZ lava flows, reflecting unique chemical and biological linkages between the subsurface lithosphere and the oceanic hydrosphere. The warmest and least diluted plume near new lava flows was 0.119°C above background seawater and hosted thermophilic and hyperthermophilic taxa that are typically identified in hydrothermal fluids emanating from the warm subsurface. Cooler and more diluted hydrothermal plumes farther from a hydrothermal fluid source were 0.072-0.078°C

above background seawater and hosted mesophilic and psychrophilic taxa that are typically identified in neutrally buoyant plumes at persistent hydrothermal venting sites. Potentially chemosynthetic microbial lineages, including Epsilonproteobacteria, Gammaproteobacteria, and Methanococcales, were positively correlated with elevated temperature anomalies. These data suggest that hydrothermal fluid flow through new lava flows on the NRZ supported diverse microbial communities for several months following the 2015 eruption and that sub-surface heterogeneity contributed to the structure of unique hydrothermal-plume hosted microbial communities.

## *4.2 Background*

Axial Seamount is the largest and most active volcano on the Juan de Fuca Ridge spreading center. Annual research cruises have collected chemical, geological, and biological data at Axial Seamount for more than two decades, making this one of the longest time-series for a deep-sea volcanically-driven hydrothermal system. In 1998 an eruption occurred on Axial′s south rift zone that was remotely detected by US Navy hydrophones (Dziak et al., 2007; Chadwick et al., 2013). Axial Seamount erupted again in 2011 along the south rift zone resulting in lava flows inside and outside the caldera with numerous snow blower vents that emitted white floc from dense eruption-associated microbial mats (Caress et al., 2012; Meyer et al., 2013; Kelley et al., 2014). In 2014, the Ocean Observatories Initiative (OOI) completed installation of a regional cabled observatory that spans the Juan de Fuca plate with nodes at Axial Seamount, that include over 20 cabled instruments within the caldera (Kelley et al., 2016).

The OOI Cabled Array detected the most recent eruption in real time, which started on April 24, 2015 and lasted for several weeks (Wilcock et al., 2016; Nooner and Chadwick,

2016). This was critical for identifying new lava flows. For the first time, the locations and timing of earthquakes and explosive events were captured as they occurred, which suggested that activity was focused along Axials north rift zone (NRZ). This was confirmed by depth changes detected by bathymetric resurveys and remotely operated vehicles in July and August 2015 (Chadwick et al., 2016; Kelley et al., 2016). The thickest new lava flows were covered by microbial mats, and there were numerous sites of hydrothermal flow (Chadwick et al., 2016; Kelley et al., 2016). The bathymetric and visual observations detected eleven new lava flows from at least 13 new fissures along the NRZ and within Axial caldera (Chadwick et al., 2016), with a combined volume of 1.48 x $10^8$ $m^3$, making the 2015 eruption the largest volume of lava documented since data collection began in the mid-1980s. The morphology and chemical composition of separate lava flows varied from thin sheet flows within the caldera to much thicker and more evolved (lower MgO) pillow lava along the NRZ (Chadwick et al., 2016).

Eruptive events at mid-ocean ridges are responsible for high fluxes in heat, chemicals, and biological matter from the subsurface (Baker et al., 1987; Butterfield et al., 1997; Delaney et al., 1998). Previous studies have found that abrupt changes in the physical and chemical environment of the seafloor and overlying water column follow eruptive events and influence the structure and activity of local microbial communities (Huber et al., 2002, 2003). For example, fresh basalt in areas of hydrothermal flow after eruptive events is rapidly colonized by microbes (Gulmann et al., 2015). These benthic biofilms are dominated by chemolithoautotrophs (Meyer et al., 2013) that support seafloor communities at hydrothermal systems for months to years following eruptions (Tunnicliffe et al., 1997).

Eruptions can produce new venting of hydrothermal fluids due to increased permeability of the crust and new heat sources such as sub-seafloor magma and freshly erupted lava fields

that are still cooling (Baker, 1998; Baker et al., 2004). Neutrally buoyant hydrothermal plumes from long-lived vents are a mixture of entrained background seawater and <0.01% of subsurface-derived hydrothermal fluid (Lupton et al., 1985). A recent high-resolution spatial study of hydrothermal vent plumes along the Eastern Lau Spreading Center suggested that ubiquitous deep-sea microbes populate plume microbial communities, while the contribution of subsurface microbial taxa to plume communities is secondary as a result of extreme dilution within the hydrothermal plume (Sheik et al., 2015). Yet, some subsurface bacteria, such as sulfur-oxidizing Epsilonproteobacteria, demonstrate strong distance-decay relationships within plumes with high abundances near the source of hydrothermal flow and can be considered indicators of hydrothermal activity (Djurhuus et al., 2017). While distance-decay relationships and high rates of entrainment of background deep-sea microbial communities primarily structure plume microbial communities, the geochemical and physical environment of the subsurface source fluids also impose selective pressures on microbial community composition over time (Huber et al., 2007; Opatkiewicz et al., 2009; Anderson et al., 2013).

Newly established links between the seafloor and subsurface magma sources, as the result of a deep-sea eruption, have the potential to create variable sub-seafloor hydrothermal conditions that alter overlying seawater non-uniformly, which may be reflected in the structure of microbial communities in newly formed eruptive hydrothermal plumes. Here, we used physical, chemical, and biological data to characterize hydrothermal plume heterogeneity above new lava flows along the NRZ at Axial Seamount. Our data indicate that differences in subsurface environments associated with new lava flows can create geochemically unique hydrothermal plumes that harbor distinct microbial communities comprised of both deep-ocean and subsurface lineages. This has the potential to significantly impact deep-ocean microbiology, as mid-ocean ridges span 65,000 km of Earth's surface and are responsible for

approximately 70% of the volcanic activity on Earth.

## *4.3 Results and Discussion*

### *4.3.1 Physical and Chemical Characterization of Distinct Post-Eruption Hydrothermal Plumes*

A rapid response cruise was added to the previously planned TN327 expedition to survey the geology, chemistry, and microbiology of new lava flows. CTD tow-yo surveys above the new NRZ lava flows collected continuous measurements for depth, temperature, salinity, and turbidity. Potential temperature and turbidity anomalies verified the presence of near-seafloor hydrothermal plumes 50-100 m thick extending horizontally several kilometers over the northernmost new lava flows (Appendix D, Figure 4.1A, Table 4.1). It is estimated that 92% of the erupted volume was accounted for in the three northernmost lava flows #8-10 (numbering scheme from south to north following Chadwick et al. (2016)). Hydrothermal plumes above these thicker northern flows had higher temperature anomalies than plumes overlying thinner new flows within Axial Caldera or elsewhere along the NRZ. The plumes over new lava flows #9 and 10 (46.08°N and 46.11°N, respectively) were sampled for further characterization of plume chemistry and microbial microbiology. A total of 44 water column samples from the caldera, NRZ and background seawater were analyzed ship board for methane and hydrogen concentrations. From a subset of 12 water-column samples, DNA was extracted and the microbial community was characterized by 16S rRNA gene sequencing (Appendix D).

Temperature and turbidity anomalies, as well as methane and hydrogen concentrations, were not uniform along the NRZ (Figure 4.1B and 4.1C) and showed regions of variability above lava flows #9 and 10 (Table 4.1). The hydrothermal plume above lava flow #9 (samples #6, 8, and 9) had the highest temperature and turbidity anomalies, measured as

ΔNTU (Figure 4.1B, Table 4.1). The plume over lava flow #10 (samples #2, 3, 4, and 5) was characterized by the highest hydrogen and methane concentrations compared to any other plume along the NRZ or within Axial caldera but had lower temperature and turbidity anomalies than in the plume over lava flow #9 (Figure 4.1B and 4.1C, Table 4.1). Two samples (#10 and 12) were collected within the broader plume, not above a new lava flow, to compare lava-associated plumes with the non-lava associated plume (Figure 4.1B). Non-lava temperature anomalies were similar to those over lava flow #10 (average 0.078°C), but turbidity anomalies were lower than in both lava-associated plumes (ΔNTU = 0.024 vs. >0.039). Similarly, methane concentrations in the non-lava plumes were lower than in the lava-associated plumes and hydrogen concentrations were near zero (Table 4.1). One background sample was collected at the NRZ (sample #14) where temperature and turbidity anomalies were not detected. Two additional background samples were collected 24 km NE of Axial Seamount and at depths corresponding to plume samples above lava flow #9 and 10 (1500m and 1700m).

Differences in plume conditions above the two NRZ lava flows indicate variability in the subsurface geology, chemistry, and microbiology. While no obvious differences in lava morphology or composition were detected between flows #9 and 10 (Chadwick et al., 2016), the microbial mats covering lava flows were thicker and more orange in color on lava flow #9 relative to lava flow #10. Additionally, the sharp increase in plume hydrogen and methane concentrations over lava flow #10 indicates differences in the chemistry, and possibly microbiology, over the two flows.

Hydrothermal fluids enriched in reduced chemical species fuel high rates of microbial productivity, which are commonly more productive than photosynthetically-driven parts of the ocean (Lutz et al., 1994; McCollom and Shock, 1997; Shock and Holland, 2004; McCollom,

2000). Geochemical models predict that most chemosynthetic primary production within vent plumes occurs when concentrations of hydrogen and sulfide are highest. This can occur in the early stages of hydrothermal plume development (McCollom, 2000). Microbial characterization of persistent vent plumes has shown that sulfur- and hydrogen-oxidizing autotrophic bacteria dominate vent fluids at their source and are present in diffuse-flow fluids (Sunamura et al., 2004; Anantharaman et al., 2013; Anderson et al., 2013; Mattes et al., 2013). Understanding the heterogeneity in the hydrothermal plumes associated with new lava flows from the 2015 eruptive event is important when considering biogeochemical models of deep-sea eruptions.

### *4.3.2 Microbiological Characterization of Distinct Post-Eruption Hydrothermal Plumes*

To understand how microbial community structure relates to environmental heterogeneity within post-eruptive hydrothermal plumes, we characterized microbial communities in discrete samples collected from the plumes over lava flows #9 and 10, in non-lava associated samples, and in background seawater. Non-metric multidimensional scaling of microbial operational taxonomic units derived from 16S rRNA gene sequence analyses indicate that Bacterial community composition was unique within each region (Figure 4.2A, ANOSIM global $R = 0.8839$, $p < 0.01$). The four regions sampled were also significantly distinct in terms of their Archaeal communities (Figure 4.2B, ANOSIM global $R = 0.4049$, $p < 0.01$). Dispersion ellipses demarcating the standard deviations of sample points within each region indicate that microbial community composition was distinct in each of the four regions sampled.

Distinct Bacterial communities were identified over the four different sampling regions even at a broad taxonomic level (Figure 4.3, SI Figure D.1). The most abundant classes of

bacteria were Alphaproteobacteria, Gammaproteobacteria and Epsilonproteobaceria (Figure 4.3A). Gammaproteobacteria were dominant in plume regions (40.0% to 49.3%). Alphaproteobacteria dominated communities in background seawater, averaging 45.5% of the bacterial 16S rRNA gene sequences recovered from these samples. Many of the top bacterial classes demonstrated strong correlations with temperature and turbidity anomalies, both indicators of hydrothermal input (SI Table D.1). The Gammaproteobacteria and Epsilonproteobacteria had strong normal relationships with temperature (adjusted $R^2$: 0.63 and 0.67, respectively) and turbidity (adjusted $R^2$: 0.57 and 0.96, respectively) anomalies across all twelve samples, while the Alphaproteobacteria had a strong inverse relationship with temperature and turbidity anomalies (adjusted $R^2$: 0.78 and 0.74, respectively) (SI Table D.1).

Shifts in Archaeal communities were also evident, though patterns in distribution were most noticeable among three low abundance orders: Methanococcales, Methanomicrobiales, and Cenarchaeales (Figure 4.3B). Methanococcales were most abundant in the near-seafloor plume over lava flow #9 (average 1.8%), where temperature anomalies were highest and where there were elevated methane and hydrogen concentrations, and over lava flow #10 (average 1.4%) where temperature anomalies were lower, but methane and hydrogen concentrations were highest. Their contribution to Archaeal communities decreased significantly in non-lava samples (average 0.8%) and in background seawater (0.2%). The Methanococcales also had the highest correlations with temperature and turbidity anomalies (adjusted $R^2$: 0.82 and 0.55 respectively) (SI Table D.2). Cultivation-based studies have isolated thermophilic and hyperthermophilic hydrogen-consuming Methanococcales with temperature requirements well above the temperature of the hydrothermal plume fluids from which they were sampled (Holden et al., 1998; Summit and Baross, 1998; Topcuoglu et al., 2016). The implication from these studies is that mesophilic and thermophilic microorganisms in

the heated subsurface environment are released into seawater (Delaney et al., 1998; Holden et al., 1998; Summit and Baross, 1998).

Strictly methanogenic Methanomicrobiales were most abundant in the non-lava associated samples and least abundant in background seawater. Most members of the Methanomicrobiales are pshycrophilic or mesophilic and can use either hydrogen or formate as electron donors for methanogenesis and acetate as a carbon source. It is important to note that while the Methanomicrobiales were most numerous in the non-lava region, their relative abundance compared to other archaeal orders was low (1.82% of total Archaeal community) and that the detection of a 16S rRNA gene sequence in the environment does not imply activity. Therefore, Methanomicrobiales may persist longer than thermophilic and hyperthermophilic methanogens in hydrothermal plumes that cool and become more dilute with entrained seawater, but are likely not contributing as much to methanogenesis within the plume, as indicated by the lower concentrations of methane in mid-plume samples. Methane-oxidizing microbes are known to dominate hydrothermal plumes at Axial Seamount (Mattes et al., 2013). This suggests that they may contribute to the drawdown in methane in the non-lava associated samples. These data suggest that temperature dependent niche-partitioning among sub-seafloor methane producing Methanococcales and Methanomicrobiales is evident in newly formed lava-associated plumes.

Indicator analyses of bacterial taxa detected over new lava flows and in background seawater also suggest that there are detectable differences in species adapted to differences in subsurface temperatures. There were significant increases in the relative abundance of Epsilonproteobacteria in hydrothermal plumes (Figure 4.4). On average, they accounted for 13.2% of the bacterial community in the higher-temperature plume over lava flow #9 and only 0.6% in background seawater. Indicator analyses suggest that thermophilic taxa

from the Epsilonproteobacteria, including *Sulfurimonas sp.*, are significant indicators (SI) in these plumes. This supports the hypothesis proposed by Djurhuus et al. that members of the mesophilic/thermophilic sulfur-oxidizing Epsilonproteobacteria are strong indicators of hydrothermal activity (Djurhuus et al., 2017). We add that individual taxa within the Epsilonproteobacteria demonstrate preferences for different temperature ranges or subsurface geochemistry, which affects the distribution of different taxa across plumes of varying origin (Figures 4.4 and 4.5).

## *4.4 Conclusions*

Real-time data from the OOI Cabled Array at Axial Seamount enabled a rapid response expedition to characterize the geochemistry and microbiology of post-eruptive hydrothermal plumes overlying new lava flows from the 2015 eruption. Our results suggest that differences in the hydrothermal sources associated with distinct lava flows and their subsurface feeder dikes create near-seafloor plume environments that are chemically and microbiologically distinct from upper, non-buoyant hydrothermal plumes. We present a conceptual model that incorporates different hydrothermal venting sites over new lava flows to highlight differences in plume chemistry and microbiology at Axial Seamount (Figure 4.5). In the model, we attribute differences in microbiology to differences in sub-seafloor microbial communities injected into the water column, differences in the degree of mixing due to seawater entrainment, and differences in residence time in the plume. These findings extend our knowledge, indicating that post-eruptive hydrothermal output over new lava flows can continue to influence deep-sea processes for months following an eruptions and that the influence on deep-sea chemistry and microbiology is heterogeneous.

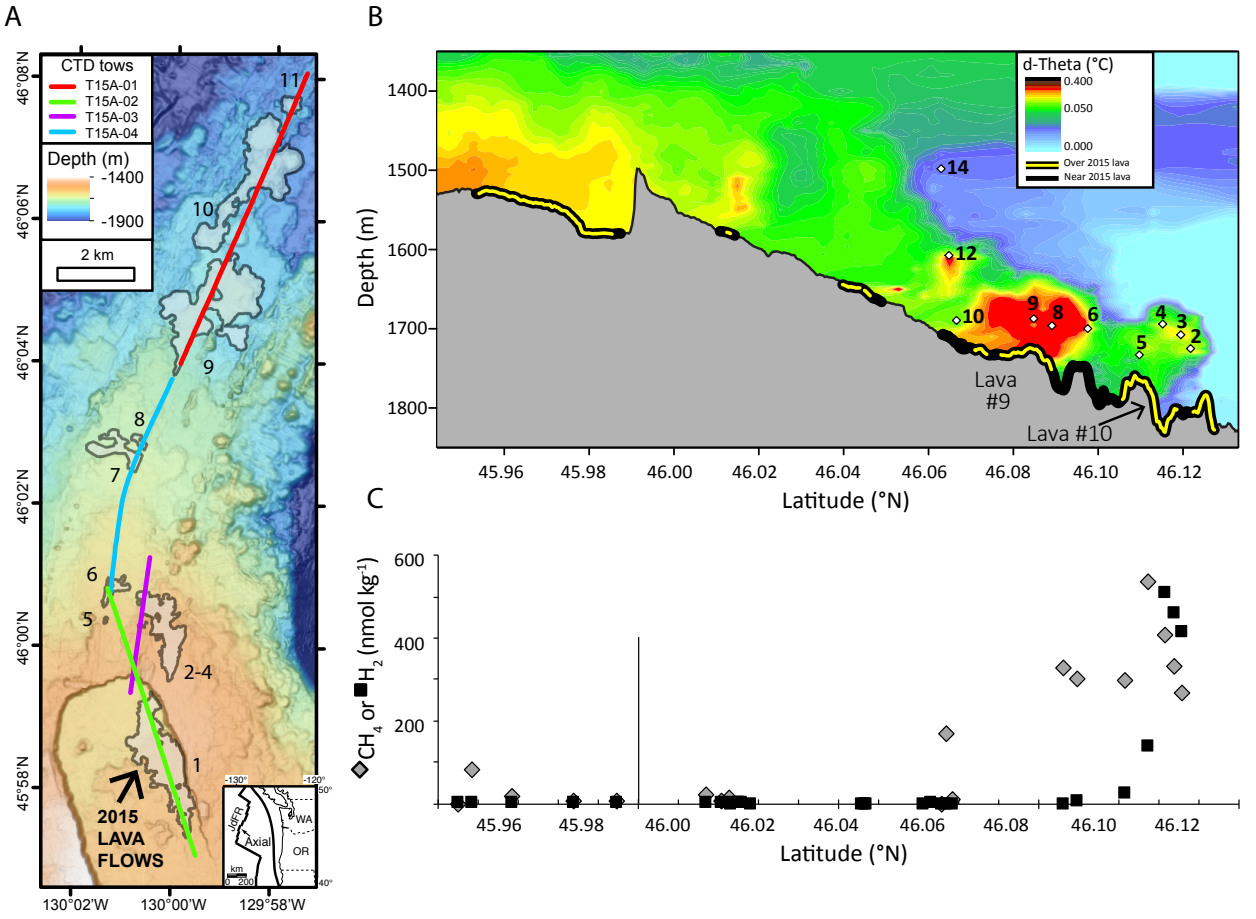


Figure 4.1: Bathymetry, temperature anomaly field, and gas chemistry of hydrothermal plumes over the caldera and north rift zone of Axial Seamount caldera following the 2015 eruption. (A) Map of the 2015 lava flows (black outlines) numbered 1-to-11 from south-to-north, corresponding with the flow numbers in Chadwick et al. (2016). CTD tow-yos during August 2015 are shown by colored lines. Inset shows location of Axial Seamount on the Juan de Fuca Ridge (JdFR) in the NE Pacific, offshore Oregon (OR) and Washington (WA), USA. (B) Temperature anomaly field from CTD tow-yos. Discrete water samples were collected at the locations indicated by open diamonds. Numbered labels refer to sample numbers. (C) Dissolved methane (grey diamonds) and hydrogen (black squares) concentrations in hydrothermal plumes over the caldera and north rift zone. The vertical line indicates the Northern boundary of Axial caldera.

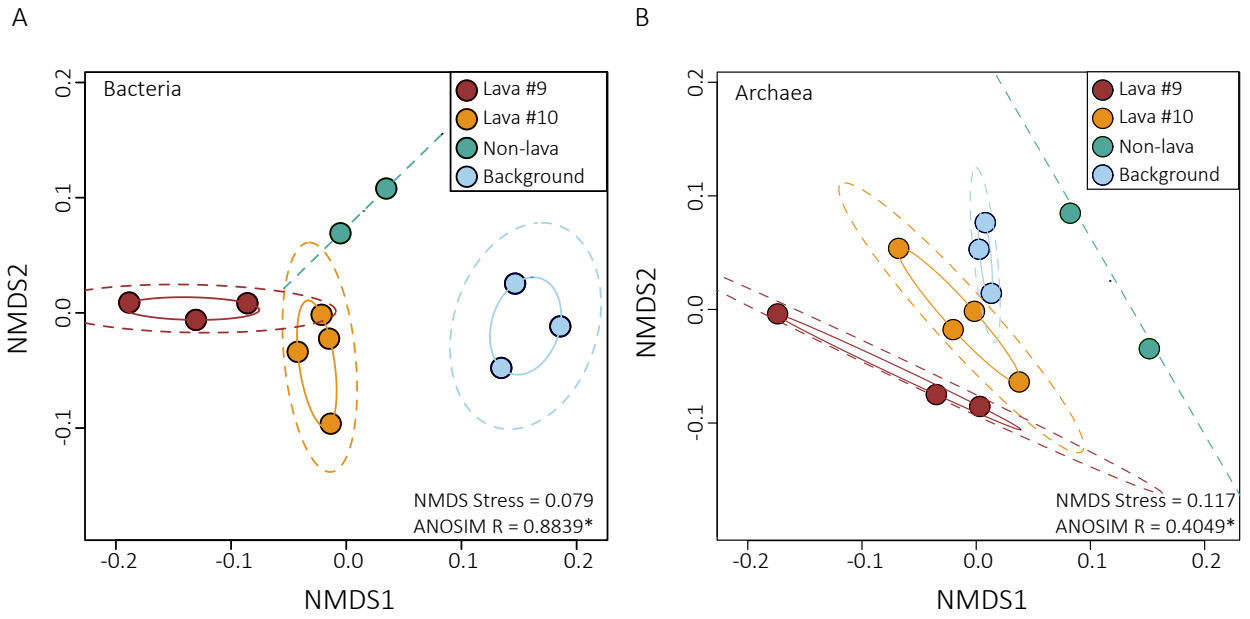


Figure 4.2: Non-metric multidimensional scaling plot showing the differences between microbial communities across sampling regions at Axial Seamount following the 2015 eruption. Microbial communities were assessed for (A) Bacterial and (B) Archaeal composition. Points represent individual samples and the colors of points represent sampling region. Four sampling regions were defined as hydrothermal plumes above lava flow #10 and lava flow #9, non-lava samples, and background seawater not influenced by hydrothermal activity. Solid lines define a convex hull of the set of samples within a group while the dashed-lines show a dispersion ellipse using the standard deviation of sample point scores under a 95% confidence limit. *Indicates p <0.01.

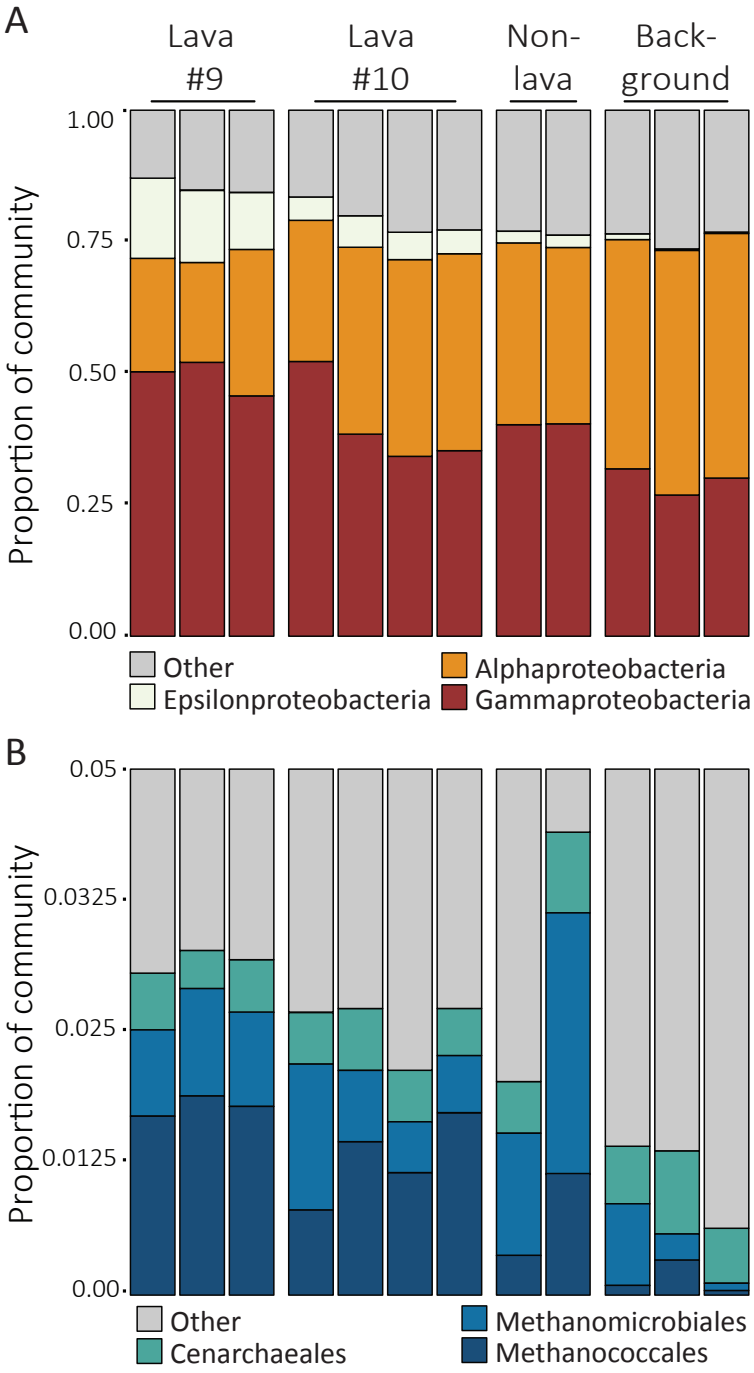


Figure 4.3: Proportions of key (A) Bacterial classes and (B) Archaeal orders detected along the North Rift Zone of Axial Volcano following the 2015 eruption. Samples are ordered by decreasing turbidity anomaly, a proxy for hydrothermal circulation, from left to right. Samples have been classified into four distinct sampling regions: lava flow #9, lava flow #10, non-lava, and background samples not influenced by hydrothermal activity.

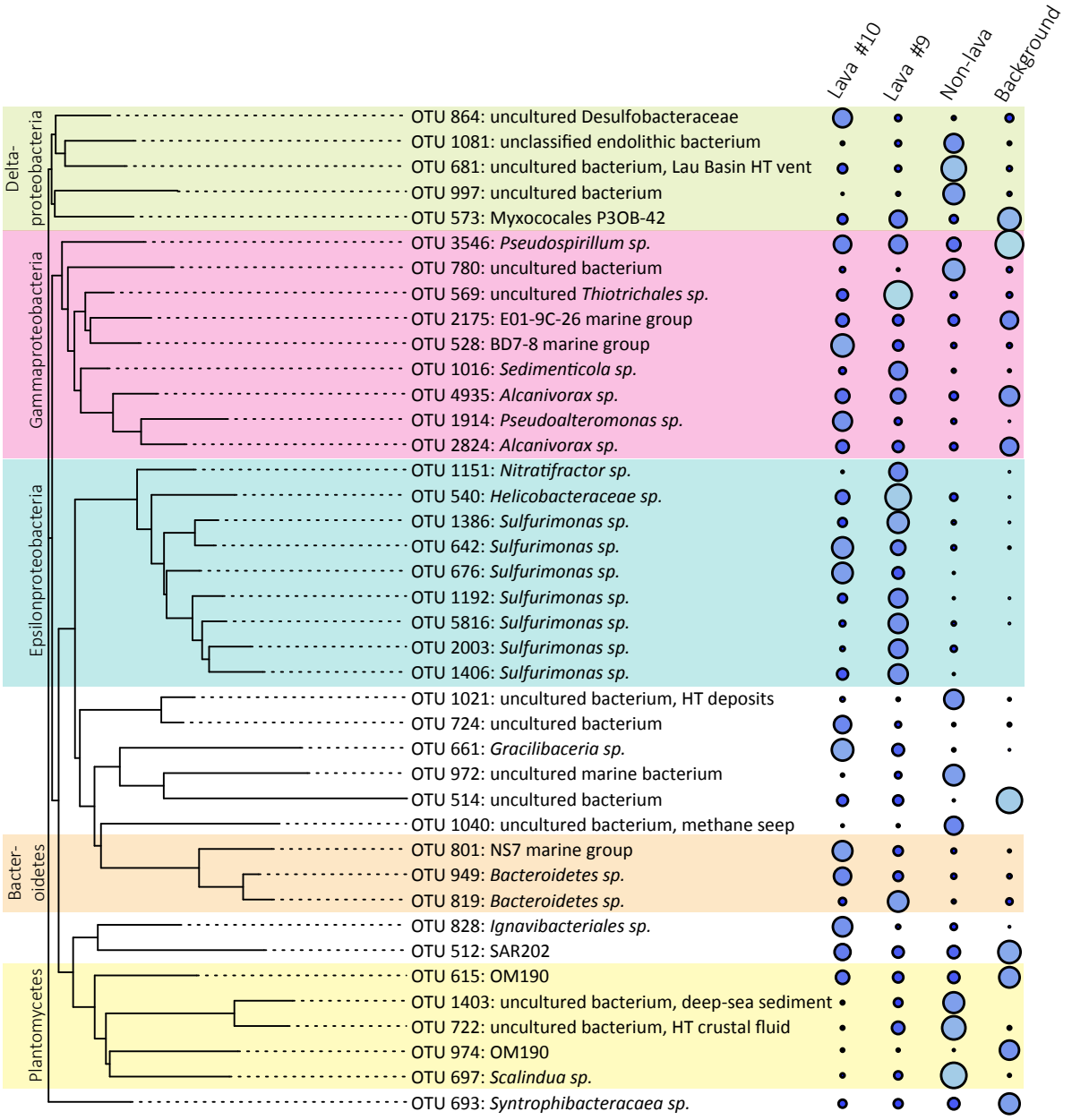


Figure 4.4: Phylogeny and relative abundances of bacterial taxa identified as significant indicators for each region: lava flow #9, lava flow #10, non-lava, and background seawater. All indicator taxa were present in at least three samples and indicator values were significant at alpha $<0.05$. Bubble size is proportional to the log-transformed average abundance of sequences detected within a region.

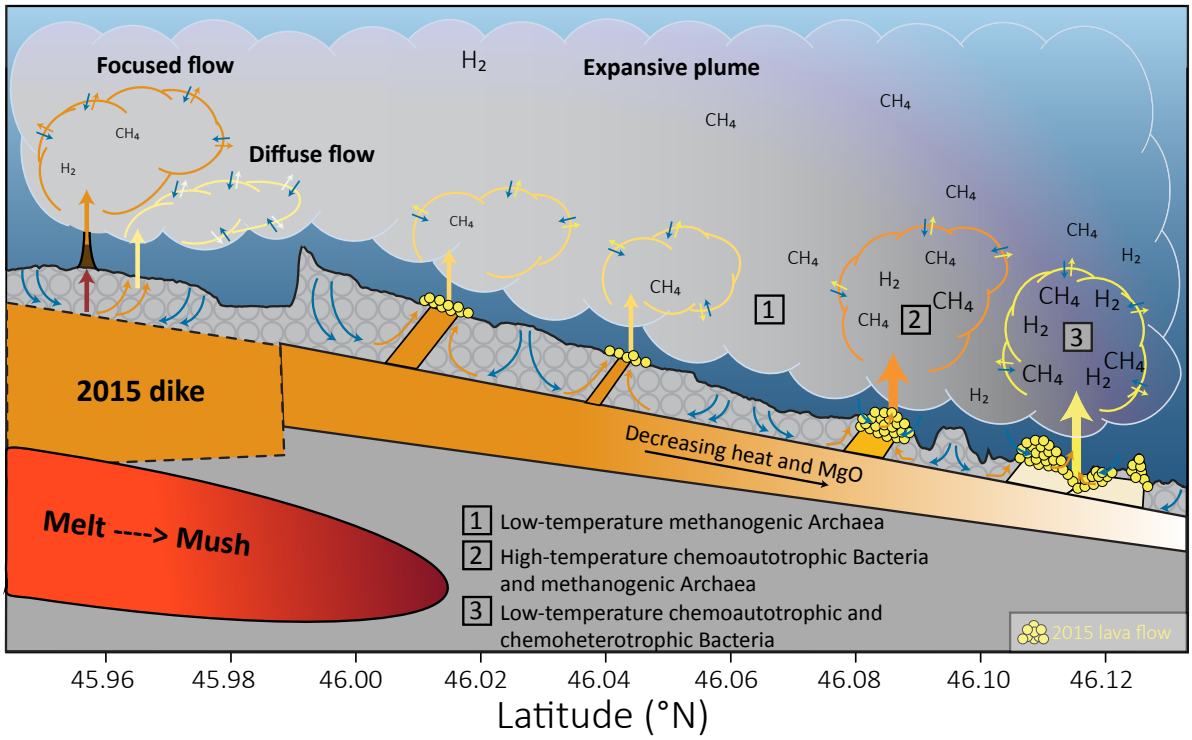


Figure 4.5: Conceptual model showing the different sources of hydrothermal plumes and the implications for overlying seawater microbial communities and chemistry. Interpretation of subsurface processes modified from Chadwick et al. (2016). The weight of arrows corresponds to the estimated relative magnitude of hydrothermal input to plumes. Colors outlining distinct plumes correspond to the relative temperature of the plume, with white being cool and orange being hotter. Dashed lines outlining non-buoyant plumes indicate that the theoretical boundaries of these plumes are not necessarily distinct but rather gradients between regions of hydrothermal input from the subsurface and background seawater. Numbered boxes indicate microbial functional groups strongly associated with each plume region. Note that the vertical axis is exaggerated to show detail.

Table 4.1: Physical and chemical measurements. Discrete samples are categorized by region, also indicated in Figure 1.

| Region | Sample (ID) | Depth (m) | Temperature anomaly (°C) | Turbidity anomaly (ΔNTU) | $CH_4$ (nmol $kg^{-1}$) | $H_2$ (nmol $kg^{-1}$) |
|---|---|---|---|---|---|---|
| Lava flow #10 | 2 | 1725 | 0.070 | 0.057 | 333 | 461 |
| | 3 | 1708 | 0.082 | 0.054 | 406 | 508 |
| | 4 | 1694 | 0.081 | 0.039 | 536 | 140 |
| | 5 | 1733 | 0.054 | 0.035 | 297 | 25 |
| Lava flow #9 | 6 | 1700 | 0.113 | 0.134 | 301 | 4 |
| | 8 | 1696 | 0.121 | 0.202 | NA | NA |
| | 9 | 1687 | 0.122 | 0.241 | NA | NA |
| Non-lava | 10 | 1689 | 0.059 | 0.024 | 8 | 0 |
| | 12 | 1607 | 0.096 | 0.024 | 168 | 0 |
| Back-ground | 14 | 1498 | 0.012 | 0.005 | NA | NA |
| | 1500 | 1500 | -0.012 | 0.001 | BDL | 1 |
| | 1700 | 1700 | 0.000 | 0.001 | 1 | 1 |

NA=not available (data not collected); BDL=below detection limit

# Chapter 5
# CONCLUSION

Microorganisms rule the marine world through cycling key nutrients and energy. Thousands, if not millions, of different types of microbes exist in the oceans, each playing a role in biogeochemical cycling. The majority of microbial taxa exist at low abundances or demonstrate patchy distributions. Yet, few microbial lineages are cosmopolitan in the global ocean. The SUP05 clade of gammaproteobacterial sulfur-oxidizing bacteria are an example of a widely distributed and relatively abundant type of marine bacteria. Thus, the activities of SUP05 bacteria in carbon cycling are important on a global scale.

Members of the SUP05 clade have the potential to drawdown atmospheric $CO_2$ through chemoautotrophic processes while others have the potential to act as a source of $CO_2$ when functioning as organic-carbon consuming heterotrophs. While the distinction between autotrophy and heterotrophy seems clear, many members of the *Arctic96BD-19* subclade of SUP05, including *Thioglobus singularis*, maintain the metabolic potential for mixotrophic growth on several energy sources (light, reduced sulfur, or organic carbon) and inorganic or organic carbon for biomass. The research goal for the first two chapters of this dissertation was to better resolve the functional roles of *T. singularis* in the global marine carbon and energy cycles.

The results from the first two chapters support the hypothesis that *T. singularis* predominantly functions as an organic-carbon consuming heterotroph. Specifically, in chapter one I showed that *T. singularis* has adapted to use extremely dilute organic constituents

for heterotrophic or methylotrophic growth without expressing proteins required for carbon fixation. Further, in chapter two I showed that the second complete genome sequence from *T. singularis* does not contain the genetic potential for carbon fixation. It is clear that *T. singularis* relies on heterotrophic or methylotrophic processes for carbon assimilation and energy acquisition. In chapter one, we began to determine some of the specific carbon compounds that may be important for growth of a single *T. singularis* strain. However, this work has shown that vast microdiversity among closely related ecotypes within the *Artic96BD-19* subclade exists, which could mean that the suite of carbon substrates utilized for growth may also be highly variable throughout the subclade. Further characterization of growth substrates among newly isolated and sequenced ecotypes are necessary to resolve the niche diversity of *Arctic96BD-19*.

The composition and activity of a microbial community is intimately linked with and often limited by the physicochemical conditions of its habitat. Microbial taxa have evolved over millennia to flourish in specific windows of temperature, pH, and oxygen concentrations, to name a few. The availability of chemical substrates for growth further refine the niche-space a taxon occupies. Therefore, in addition to understanding the metabolic potential of individual taxa, we must also aim to further our understanding of how sudden changes in a habitat alters the associated microbial community. These insights could help in modeling and predicting the response of microbial communities under future environmental changes.

In chapter three, I surveyed the effect of a deep-sea eruption on the overlying water-column microbial communities. We found that the effect of the eruption was not uniform across space and that different taxa were indicative of varying levels of hydrothermal influence. The results from this study highlight an important theme in microbial ecology, which is the issue of scale. This was one of the first studies examining the water-column effect

of a deep-sea eruption and the highest-resolution survey in terms of microbial community characterization and sampling space. Because we collected samples in close-proximity and obtained deeply characterized community surveys, we could discern between fine-scale local changes in chemistry and microbiology. In dynamic environments, such as hydrothermal systems, high-resolution sampling across space and time unveil important patchiness and heterogeneity that support diverse microbial communities. Follow-up studies in these hydrothermal plume environments should focus on the active populations and metabolisms on a fine-scale resolution to resolve the extent of hydrothermal impact on deep-sea food webs.

In conclusion, the research presented herein has resolved the roles of marine microbes in carbon and energy cycles and has shown that environmental perturbations have great potential to alter those roles. I suggest that future surveys of the SUP05 clade should be conducted with higher-resolution genetic markers to distinguish between closely related yet functionally-distinct ecotypes. While we have learned much from only three complete isolates of the SUP05 clade, vast diversity awaits exploration.

# BIBLIOGRAPHY

Anantharaman, K., Breier, J. A., Sheik, C. S., and Dick, G. J. (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proceedings of the National Academy of Sciences of the United States*, 110(1):330–335.

Anderson, R. E., Beltrán, M. T., Hallam, S. J., and Baross, J. A. (2013). Microbial community structure across fluid gradients in the Juan de Fuca Ridge hydrothermal system. *FEMS Microbiology Ecology*, 83(2):324–339.

Apprill, A., McNally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology*, 75(2):129–137.

Azam, F., Fenchel, T., Field, J. G., Gray, J. C., Meyer-Reil, L. A., and Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Marine Ecology Progress Series*, 10(3):257–264.

Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008). The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*, 9(1):75.

Baker, E. T. (1998). Patterns of event and chronic hydrothermal venting following a magmatic intrusion: New perspectives from the 1996 Gorda Ridge eruption. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 45(12):2599–2618.

Baker, E. T., German, C. R., and Elderfield, H. (1995). Hydrothermal Plumes Over Spreading-Center Axes : Global Distributions and Geological Inferences. *Geophysical MonographGeophysical Monograph*, 91:47–71.

Baker, E. T., Lowell, R. P., Resing, J. A., Feely, R. A., Embley, R. W., Massoth, G. J., and Walker, S. L. (2004). Decay of hydrothermal output following the 1998 seafloor eruption at Axial Volcano : Observations and models. *Journal of Geophysical Research*, 109:1–14.

Baker, E. T., Massoth, G. J., and Feely, R. A. (1987). Cataclysmic hydrothermal venting on the Juan de Fuca Ridge. *Nature*, 329:149–151.

Barrett, E. L. and Kwan, H. S. (1985). Bacterial reduction of trimethylamine oxide. *Annual review of microbiology*, 39:131–149.

Beale, R. and Airs, R. (2016). Quantification of glycine betaine, choline and trimethylamine N-oxide in seawater particulates: Minimisation of seawater associated ion suppression. *Analytica Chimica Acta*, 938:114–122.

Becker, J. W., Berube, P. M., Follett, C. L., Waterbury, J. B., Chisholm, S. W., Delong, E. F., and Repeta, D. J. (2014). Closely related phytoplankton species produce similar suites of dissolved organic matter. *Frontiers in microbiology*, 5(March):111.

Bracher, A., Sharma, A., Starling-Windhof, A., Hartl, F. U., and Hayer-Hartl, M. (2015). Degradation of potent Rubisco inhibitor by selective sugar phosphatase. *Nature Plants*, 1:1–6.

Brown, M. V., Lauro, F. M., DeMaere, M. Z., Muir, L., Wilkins, D., Thomas, T., Riddle, M. J., Fuhrman, J. A., Andrews-Pfannkoch, C., Hoffman, J. M., McQuaid, J. B., Allen, A., Rintoul, S. R., and Cavicchioli, R. (2012). Global biogeography of SAR11 marine bacteria. *Molecular Systems Biology*, 8(595):1–13.

Butterfield, D. A., Jonasson, I. R., Massoth, G. J., Feely, R. A., Roe, K. K., Embley, R. E., Holden, J. F., McDuff, R. E., Lilley, M. D., and Delaney, J. R. (1997). Seafloor eruptions and evolution of hydrothermal fluid chemistry. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 355(1723):369–386.

Caress, D. W., Clague, D. A., Paduan, J. B., Martin, J. F., Dreyer, B. M., Chadwick, W. W., Denny, A., and Kelley, D. S. (2012). Repeat bathymetric surveys at 1-metre resolution of lava flows erupted at Axial Seamount in April 2011. *Nature Geoscience*, 5(7):483–488.

Carini, P., Steindler, L., Beszteri, S., and Giovannoni, S. J. (2012). Nutrient requirements for growth of the extreme oligotroph Candidatus Pelagibacter ubique' HTCC1062 on a defined medium. *The ISME Journal*, 7(3):592–602.

Chadwick, W. W., Clague, D. A., Embley, R. W., Perfit, M. R., Butterfield, D. A., Caress, D. W., Paduan, J. B., Martin, J. F., Sasnett, P., Merle, S. G., and Bobbitt, A. M. (2013). The 1998 eruption of Axial Seamount: New insights on submarine lava flow emplacement from high-resolution mapping. *Geochemistry, Geophysics, Geosystems*, 14(10):3939–3968.

Chadwick, W. W., Paduan, J. B., Clague, D. A., Dreyer, B. M., and Merle, S. G. (2016). Voluminous eruption from a zoned magma body after an increase in supply rate at Axial Seamount. *Geophysical Research Letters*, 5:63–70.

Chistoserdova, L., Kalyuzhnaya, M. G., and Lidstrom, M. E. (2009). The Expanding World of Methylotrophic Metabolism. *Annual Review of Microbiology*, 63:477–499.

Cho, J. C. and Giovannoni, S. J. (2004). Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. *Applied and Environmental Microbiology*, 70(1):432–440.

Connon, S. A. and Giovannoni, S. J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Applied and Environmental Microbiology*, 68(8):3878–3885.

Dangel, A. W. and Tabita, F. R. (2015). CbbR , the Master Regulator for Microbial Carbon Dioxide Fixation. *Journal of Bacteriology*, 197(22):3488–3498.

DeFlaun, M. F., Paul, J. H., and Jeffrey, W. H. (1987). Distribution and molecular weight of dissolved DNA in subtropical estuarine and oceanic environments. *Marine Ecology Progress Series*, 38:65–73.

Delaney, J. R., Kelley, D. S., Lilley, M. D., Butterfield, D. A., Baross, J. A., D Wilcock, W. S., Embley, R. W., and Summit, M. (1998). The Quantum Event of Oceanic Crustal Accretion: Impacts of Diking at Mid-Ocean Ridges. *Source: Science, New Series*, 281(5374):222–230.

Diaz, M. R., Visscher, P. T., and Taylor, B. F. (1992). Metabolism of dimethylsulfoniopropionate and glycine betaine by a marine bacterium. *FEMS microbiology letters*, 96(1):61–65.

Djurhuus, A., Mikalsen, S.-O., Giebel, H.-A., and Rogers, A. (2017). Cutting through the smoke : Free-living bacterial diversity in deep-sea hydrothermal plumes. *Royal Society Open Science*, 4(160829).

Duarte, C. M., Regaudie-de Gioux, A., Arrieta, J. M., Delgado-Huertas, A., and Agustí, S. (2013). The Oligotrophic Ocean Is Heterotrophic. *Annual Review of Marine Science*, 5(1):551–569.

Ducklow, H. W. and Doney, S. C. (2013). What Is the Metabolic State of the Oligotrophic Ocean? A Debate. *Annual Review of Marine Science*, 5(1):525–533.

Ducklow, H. W. and McCallister, S. L. (2005). The biogeochemistry of carbon dioxide in the coastal oceans. In Robinson, A. R. and Brink, K. H., editors, *The Global Coastal Ocean: Multiscale Interdisciplinary Processes*, chapter 9, pages 269–315. Harvard University Press, Cambridge.

Dupont, C. L., Rusch, D. B., Yooseph, S., Lombardo, M.-J., Richter, R. A., Valas, R., Novotny, M., Yee-Greenbaum, J., Selengut, J. D., Haft, D. H., Halpern, A. L., Lasken, R. S., Nealson, K., Friedman, R., and Venter, J. C. (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME journal*, 6(6):1186–99.

Dziak, R. P., Bohnenstiehl, D. R., Cowen, J. P., Baker, E. T., Rubin, K. H., Haxel, J. H., and Fowler, M. J. (2007). Rapid dike emplacement leads to eruptions and hydrothermal plume release during seafloor spreading events. *Geology*, 35(7):579–582.

Eng, J. K., Hoopmann, M. R., Jahan, T. A., Egertson, J. D., Noble, W. S., and MacCoss, M. J. (2015). A Deeper Look into Comet - Implementation and Features. *Journal of the American Society for Mass Spectrometry*, 26(11):1865–1874.

Eng, J. K., Jahan, T. A., and Hoopmann, M. R. (2013). Comet: An open-source MS/MS sequence database search tool. *Proteomics*, 13(1):22–24.

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Hogberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pederson, T., Rosenthal, Y., Sietzinger, S., Smatacek, V., and Steffen, W. (2000). The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System. *Science*, 290(5490):291–296.

Fuszard, M. A., Wright, P. C., and Biggs, C. A. (2012). Comparative quantitative proteomics of prochlorococcus ecotypes to a decrease in environmental phosphate concentrations. *Aquatic biosystems*, 8(1):7.

Garrity, G. and Bell, J. (2005). Phylum. XIV. Proteobacteria phyl. nov. In Garrity, G., Brenner, D., and Krieg, N., editors, *Bergey's manual of systematic bacteriology*, chapter Vol. 2 Par, page 1. Springer, New York, 2nd edition.

Garrity, G., Bell, K., and Lidburn, T. (2005). Class III. Gammaproteobacteria class. nov. In Garrity, G., Brenner, D., Krieg, N., and Staley, J., editors, *Bergey's manual of systematic bacteriology*, chapter Vol. 2, page 1. Springer, New York, 2nd edition.

Gattuso, J.-P., Frankignoulle, M., and Wollast, R. (1998). Carbon and Carbonate Metabolism in Coastal Aquatic Ecosystems. *Annual Review of Ecology and Systematics*, 29(1):405–434.

Georges, A. A., El-Swais, H., Craig, S. E., Li, W. K., and Walsh, D. A. (2014). Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *The ISME Journal*, 8(10):1301–1313.

Ghosh, W. and Dam, B. (2009). Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. *FEMS Microbiology Reviews*, 33(6):999–1043.

Gibb, S. W. and Hatton, A. D. (2004). The occurrence and distribution of trimethylamine-N-oxide in Antarctic coastal waters. *Marine Chemistry*, 91(1-4):65–75.

Gibb, S. W., Mantoura, R. F. C., Liss, P. S., and Barlow, R. G. (1999). Distributions and biogeochemistries of methylamines and ammonium in the Arabian Sea. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 46(3-4):593–615.

Giovannoni, S. J., Cameron Thrash, J., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *The ISME journal*, (November 2013):1–13.

Glaubitz, S., Kieblich, K., Meeske, C., Labrenz, M., and Jürgens, K. (2013). SUP05 dominates the Gammaproteobacterial sulfur oxidizer assemblages in pelagic redoxclines of the central Baltic and Black Seas. *Applied and environmental microbiology*, 79(8):2767–76.

Gómez-Consarnau, L., Lindh, M. V., Gasol, J. M., and Pinhassi, J. (2012). Structuring of bacterioplankton communities by specific dissolved organic carbon compounds. *Environmental Microbiology*, 14(9):2361–2378.

Gulmann, L. K., Beaulieu, S. E., Shank, T. M., Ding, K., Seyfried, W. E., and Sievert, S. M. (2015). Bacterial diversity and successional patterns during biofilm formation on freshly exposed basalt surfaces at diffuse-flow deep-sea vents. *Frontiers in Microbiology*, 6:1–16.

Halsey, K. H., Giovannoni, S. J., Graus, M., Zhao, Y., Landry, Z., Thrash, J. C., Vergin, K. L., and de Gouw, J. (2017). Biological cycling of volatile organic carbon by phytoplankton and bacterioplankton. *Limnology and Oceanography*, 62:2650–2661.

Hansell, D. A. and Carlson, C. A., editors (2015). *Biogeochemistry of marine dissolved organic matter*. Elsevier, 2nd edition.

Hirsch, P., Bernhard, M., Cohen, S. S., Ensign, J. C., Jannasch, H. W., Koch, A. L., Marshall, K. C., Matin, A., Poindexter, J. S., Rittenberg, S. C., Schmidt, D. C., and Veldkamp, H. (1979). Life under conditions of low nutrient concentrations. In Shilo, M., editor, *Strategies of microbial life in extreme environments*, pages 357–372, Berlin. Springer-Verlag.

Holden, J. F., Summit, M., and Baross, J. A. (1998). Thermophilic and hyperthermophilic microorganisms in 3 - 30C hydrothermal fluids following a deep-sea volcanic eruption. *FEMS Microbiology Ecology*, 25:33–41.

Huber, J. A., Butterfield, D. A., and Baross, J. A. (2002). Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Applied and Environmental Microbiology*, 68(4):1585–1594.

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

Huber, J. a., Mark Welch, D. B., Morrison, H. G., Huse, S. M., Neal, P. R., Butterfield, D. a., and Sogin, M. L. (2007). Microbial population structures in the deep marine biosphere. *Science*, 318(5847):97–100.

Hügler, M. and Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Annual review of marine science*, 3:261–289.

Judd, K. E., Crump, B. C., and Kling, G. W. (2006). Variation in dissolved organic matter controls bacterial production and community composition. *Ecology*, 87(8):2068–79.

Juhas, M., Van Der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., and Crook, D. W. (2009). Genomic islands: Tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, 33(2):376–393.

Kelley, D. S., Delaney, J. R., and Juniper, S. K. (2014). Establishing a new era of submarine volcanic observatories: Cabling Axial Seamount and the Endeavour Segment of the Juan de Fuca Ridge. *Marine Geology*, 352:426–450.

Kelley, D. S., Delaney, J. R., and Team, C. A. (2016). NSF's Cabled Array: A wired tectonic plate and overlying ocean. *OCEANS 2016 MTS/IEEE Monterey, OCE 2016*.

Koren, S., Harhay, G. P., Smith, T. P., Bono, J. L., Harhay, D. M., Mcvey, S. D., Radune, D., Bergman, N. H., and Phillippy, A. M. (2013). Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biology*, 14(9):R101.

Lalish, K. M. (2015). *Elucidating the metabolic activities of SUP05, an abundant group of marine sulfur oxidizing gamma-proteobacteria*. Dissertation, University of Washington.

Lesniewski, R. A., Jain, S., Anantharaman, K., Schloss, P. D., and Dick, G. J. (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *The ISME Journal*, 6(12):2257–2268.

Lidbury, I., Mausz, M. A., Scanlan, D. J., and Chen, Y. (2017). Identification of dimethylamine monooxygenase in marine bacteria reveals a metabolic bottleneck in the methylated amine degradation pathway. *The ISME Journal*, pages 1–10.

Lidbury, I., Murrell, J. C., and Chen, Y. (2014). Trimethylamine N-oxide metabolism by abundant marine heterotrophic bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 111(7):2710–5.

Lidbury, I. D., Murrell, J. C., and Chen, Y. (2015). Trimethylamine and trimethylamine N-oxide are supplementary energy sources for a marine heterotrophic bacterium: implications for marine carbon and nitrogen cycling. *The ISME Journal*, 9(3):760–769.

Louca, S., Hawley, A. K., Katsev, S., Torres-Beltran, M., Bhatia, M. P., Kheirandish, S., Michiels, C. C., Capelle, D., Lavik, G., Doebeli, M., Crowe, S. A., and Hallam, S. J. (2016). Integrating biogeochemistry with multiomic sequence information in a model oxygen minimum zone. *Proceedings of the National Academy of Sciences*, 113(40):E5925–E5933.

Lupton, J., Delaney, J. R., Johnson, H., and Tivey, M. (1985). Entrainment and vertical transport of deep-ocean water by buoyant hydrothermal plumes. *Nature*, 316(6029):621–623.

Lutz, R. A., Shank, T. M., Fornari, D. J., Haymon, R. M., Lilley, M. D., Von Damm, K. L., and Desbruyeres, D. (1994). Rapid growth at deep-sea vents. *Nature*, 371:633–634.

Malmstrom, R. R., Rodrigue, S., Huang, K. H., Kelly, L., Kern, S. E., Thompson, A., Roggensack, S., Berube, P. M., Henn, M. R., and Chisholm, S. W. (2012). Ecology of uncultured Prochlorococcus clades revealed through single-cell genomics and biogeographic analysis. *The ISME journal*, 7:184–198.

Markert, S., Arndt, C., Felbeck, H., Becher, D., Sievert, S. M., Hugler, M., Albrecht, D., Robidart, J., Bench, S., Feldman, R. A., Hecker, M., and Schweder, T. (2007). Physiological Proteomics of the Uncultured Endosymbiont of Riftia pachyptila. *Science*, 315:247–251.

Markowitz, V. M., Chen, I. M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., Ratner, A., Jacob, B., Huang, J., Williams, P., Huntemann, M., Anderson, I., Mavromatis, K., Ivanova, N. N., and Kyrpides, N. C. (2012). IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Research*, 40(D1):115–122.

Marshall, K. T. and Morris, R. M. (2013). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *The ISME Journal*, 7(2):452–455.

Marshall, K. T. and Morris, R. M. (2015). Mixotroph from the SUP05 Clade of Marine Gammaproteobacteria. *Genome Announcements*, 3(5):7–8.

Mattes, T. E., Nunn, B. L., Marshall, K. T., Proskurowski, G., Kelley, D. S., Kawka, O. E., Goodlett, D. R., Hansell, D. A., and Morris, R. M. (2013). Sulfur oxidizers dominate carbon fixation at a biogeochemical hot spot in the dark ocean. *The ISME journal*, 7(12):2349–60.

McCollom, T. M. (2000). Geochemical constraints on primary productivity in submarine hydrothermal vent plumes. *Deep Sea Research Part I: Oceanographic Research Papers*, 47(1):85–101.

McCollom, T. M. and Shock, E. L. (1997). Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochimica et cosmochimica acta*, 61(20):4375–91.

Meyer, J. L., Akerman, N. H., Proskurowski, G., and Huber, J. A. (2013). Microbiological characterization of post-eruption "snowblower" vents at axial seamount, juan de fuca ridge. *Frontiers in Microbiology*, 4:1–13.

Morris, R. M., Frazar, C. D., and Carlson, C. A. (2012). Basin-scale patterns in the abundance of SAR11 subclades, marine Actinobacteria (OM1), members of the Roseobacter clade and OCS116 in the South Atlantic. *Environmental microbiology*, 14(5):1133–44.

Morris, R. M., Nunn, B. L., Frazar, C., Goodlett, D. R., Ting, Y. S., and Rocap, G. (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *The ISME journal*, 4(5):673–85.

Morris, R. M., Rappé, M. S., Connon, S. A., Vergin, K. L., Siebold, W. A., Carlson, C. A., and Giovannoni, S. J. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*, 420(6917):806–810.

Murillo, A. A., Ramírez-Flandes, S., DeLong, E. F., and Ulloa, O. (2014). Enhanced metabolic versatility of planktonic sulfur-oxidizing g-proteobacteria in an oxygen-deficient coastal ecosystem. *Frontiers in Marine Science*, 1(July):1–13.

Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry*, 75(17):4646–4658.

Nesvizhskii, A. I., Vitek, O., and Aebersold, R. (2007). Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods*, 4(10):787–797.

Nooner, S. L. and Chadwick, W. W. (2016). Inflation-predictable behavior and co-eruption deformation at Axial Seamount. *Science*, 354(6318):1399–1403.

Opatkiewicz, A. D., Butterfield, D. A., and Baross, J. A. (2009). Individual hydrothermal vents at Axial Seamount harbor distinct subseafloor microbial communities. *FEMS Microbiology Ecology*, 70(3):413–424.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F., and Stevens, R. (2014). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*, 42(D1):206–214.

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2015). Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environmental Microbiology*, 18(5):1403–1414.

Rappé, M. S., Connon, S. A., Vergin, K. L., and Giovannoni, S. J. (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature*, 418(6898):630–633.

Robinson, C. and Williams, P. J. L. B. (2005). Respiration and its measurement in surface marine waters. In del Giorgio, O. and Williams, P. J. L. B., editors, *Respiration in Aquatic Ecosystems*, chapter 9, pages 148–181. Oxford University Press, Oxford.

Rogge, A., Vogts, A., Voss, M., Jürgens, K., Jost, G., and Labrenz, M. (2017). Success of chemolithoautotrophic SUP05 and Sulfurimonas GD17 cells in pelagic Baltic Sea redox zones is facilitated by their lifestyles as K- and r-strategists. *Environmental Microbiology*, 19(6):2495–2506.

Sarmento, H. and Gasol, J. M. (2012). Use of phytoplankton-derived dissolved organic carbon by different types of bacterioplankton. *Environmental Microbiology*, 14:2348–2360.

Sarmento, H., Romera-Castillo, C., Lindh, M., Pinhassi, J., Sala, M. M., Gasol, J. M., Cèlia, M., and Taylor, G. T. (2013). Phytoplankton species-specific release of dissolved free amino acids and their selective consumption by bacteria. *Limnology and Oceanography*, 58(3):1123–1135.

Schutl, F., Prins, R. A., and Gottschal, J. C. (1997). Oliogotrophy and pelagic marine bacteria: facts and fiction. *Aquatic Microbial Ecology*, 12:177–202.

Schwalbach, M. S., Tripp, H. J., Steindler, L., Smith, D. P., and Giovannoni, S. J. (2010). The presence of the glycolysis operon in SAR11 genomes is positively correlated with ocean productivity. *Environmental Microbiology*, 12:490–500.

Shah, V., Chang, B. X., and Morris, R. M. (2016). Cultivation of a chemoautotroph from the SUP05 clade of marine bacteria that produces nitrite and consumes ammonium. *The ISME Journal*, 11(1):263–271.

Sheik, C. S., Anantharaman, K., Breier, J. A., Sylvan, J. B., Edwards, K. J., and Dick, G. J. (2015). Spatially resolved sampling reveals dynamic microbial communities in rising hydrothermal plumes across a back-arc basin. *The ISME Journal*, 9(6):1434–1445.

Shock, E. L. and Holland, M. E. (2004). Geochemical Energy Sources That Support the Subsurface Biosphere. In Wilcock, W. S., Delong, E. F., Kelley, D. S., Baross, J. A., and Cary, S. C., editors, *The Subseafloor Biosphere at Mid-Ocean Ridges*, pages 153–166. American Geophysical Union, Washington, D. C.

Shteynberg, D., Deutsch, E. W., Lam, H., Eng, J. K., Sun, Z., Tasman, N., Mendoza, L., Moritz, R. L., Aebersold, R., and Nesvizhskii, A. I. (2011). iProphet: Multi-level Integrative Analysis of Shotgun Proteomic Data Improves Peptide and Protein Identification Rates and Error Estimates. *Molecular & Cellular Proteomics*, 10(12):M111.007690.

Sowell, S. M., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Callister, S. J., Smith, R. D., Barofsky, D. F., and Giovannoni, S. J. (2008). Proteomic analysis of stationary phase in the marine bacterium "Candidatus pelagibacter ubique". *Applied and Environmental Microbiology*, 74(13):4091–4100.

Sowell, S. M., Wilhelm, L. J., Norbeck, A. D., Lipton, M. S., Nicora, C. D., Barofsky, D. F., Carlson, C. A., Smith, R. D., and Giovanonni, S. J. (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *The ISME Journal*, 3(1):93–105.

Summit, M. and Baross, J. A. (1998). Thermophilic subseafloor microorganisms from the 1996 North Gorda Ridge eruption. *Deep Sea Research Part II: Topical Studies in Oceanography*, 45(12):2751–2766.

Sun, J., Steindler, L., Thrash, J. C., Halsey, K. H., Smith, D. P., Carter, A. E., Landry, Z. C., and Giovannoni, S. J. (2011). One carbon metabolism in SAR11 pelagic marine bacteria. *PLoS ONE*, 6(8).

Sun, J., Todd, J. D., Thrash, J. C., Qian, Y., Qian, M. C., Guo, J., Fowler, E. K., Aldrich, J. T., Nicora, C. D., Mary, S., Smith, R. D., Leenheer, P. D., Payne, S. H., and Johnston, A. W. B. (2016). The abundant marine bacterium Pelagibacter simultaneously catabolizes dimethylsulfoniopropionate to the gases dimethyl sulfide and methanethiol. *Nature Microbiology*, 11(541):1–24.

Sunamura, M., Higashi, Y., Miyako, C., Ishibashi, J. I., and Maruyama, A. (2004). Two Bacteria Phylotypes Are Predominant in the Suiyo Seamount Hydrothermal Plume. *Applied and Environmental Microbiology*, 70(2):1190–1198.

Swan, B. K., Martinez-Garcia, M., Preston, C. M., Sczyrba, A., Woyke, T., Lamy, D., Reinthaler, T., Poulton, N. J., Masland, E. D. P., Gomez, M. L., Sieracki, M. E., DeLong, E. F., Herndl, G. J., and Stepanauskas, R. (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science*, 333(6047):1296–1300.

Taubert, M., Grob, C., Howat, A. M., Burns, O. J., Pratscher, J., Jehmlich, N., von Bergen, M., Richnow, H. H., Chen, Y., and Colin Murrell, J. (2017). Methylamine as a Nitrogen Source for Microorganisms from a Coastal Marine Environment. *Environmental Microbiology*, 19(6):2246–2257.

Topcuoglu, B. D., Stewart, L. C., Morrison, H. G., Butterfield, D. A., Huber, J. A., and Holden, J. F. (2016). Hydrogen limitation and syntrophic growth among natural assemblages of thermophilic methanogens at deep-sea hydrothermal vents. *Frontiers in Microbiology*, 7:1–12.

Tunnicliffe, V., Embley, R. W., Holden, J. F., Butterfield, D. A., Massoth, G. J., and Juniper, S. K. (1997). Biological colonization of new hydrothermal vents following an eruption on Juan de Fuca Ridge. *Deep-Sea Research Part I: Oceanographic Research Papers*, 44(9-10):1627–1644.

Walsh, D. A., Zaikova, E., Howes, C. G., Song, Y. C., Wright, J. J., Tringe, S. G., Tortell, P. D., and Hallam, S. J. (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science*, 326(5952):578–582.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes : The unseen majority. *Proceedings of the National Academy of Sciences*, 95:6578–6583.

Wilcock, W. S. D., Tolstoy, M., Waldhauser, F., Garcia, C., Tan, Y. J., Bohnenstiehl, D. R., Caplan-Auerback, J., Dziak, R. P., Arnulf, A. F., and Mann, M. E. (2016). Seismic constraints on caldera dynamics from the 2015 Axial Seamount eruption. *Science*, 354(6318):1395–1399.

Wilhelm, L. J., Tripp, H. J., Givan, S. A., Smith, D. P., and Giovannoni, S. J. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct*, 2(1):27.

Williams, T. J., Long, E., Evans, F., Demaere, M. Z., Lauro, F. M., Raftery, M. J., Ducklow, H., Grzymski, J. J., Murray, A. E., and Cavicchioli, R. (2012). A metaproteomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *The ISME journal*, 6(10):1883–900.

Woese, C. R., Kandler, O., and Wheelis, M. L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences*, 87(12):4576–4579.

Wright, J. J., Konwar, K. M., and Hallam, S. J. (2012). Microbial ecology of expanding oxygen minimum zones. *Nature Reviews Microbiology*, 10(6):381–94.

Zehr, J. P. and Kudela, R. M. (2011). Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Annual Review of Marine Science*, 3(1):197–225.

Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., and Mcintosh, J. R. (2017). Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium "Candidatus Pelagibacter ubique". *Applied and environmental microbiology*, 83(3):e02807–16.

# Appendix A

# OPTIMAL TEMPERATURE AND PH RANGE OF *CA.* THIOGLOBUS SINGULARIS STRAIN PS1

Using a defined, artificial seawater media, ASW, (adapted from Carini et al. (2012)), the optimal temperature and pH ranges for *Ca.* Thioglobus singularis strain PS1 were resolved. All cultures were grown on ASW with a dilute diatom cellular lysate (see Chapter 2 Methods) and 1 mM thiosulfate to obtain maximum cell densities. Cultures for the temperature and pH range experiments were grown in 50-mL volumes in 250-mL acid-washed, autoclaved polycarbonate bottles in triplicate. Results from a preliminary experiment found no significant difference in growth rate or maximum cell density when *Ca.* T. singularis PS1 was grown in 24-hour darkness or in a 16hr:8hr light:dark cycle (Figure A.1), so all cultures for the temperature and pH experiment were incubated in the dark. To adjust the pH of ASW media in the pH experiment, 1N KOH or 1N HCl and measured using a pH probe until the desired pH was reached. Cultures for the pH experiment were incubated at 13°C. The pH of culture media was checked at the end of the growth experiments and no significant fluctuations from the initial conditions were detected. Cells were enumerated by staining with SYBR and counted on a Guava EasyFlow Cytometer.

*Ca.* T. singularis strain PS1 reached maximum cell densities when grown at 10°C (Figure A.2) but had the highest growth rate at 13°C (Figure A.3). Growth of *Ca.* T. singularis was not impacted by changing pH within the range of 6.8-8.1, but at pH 6.0 and 8.7 growth was significantly inhibited (Figures A.4 and A.5).

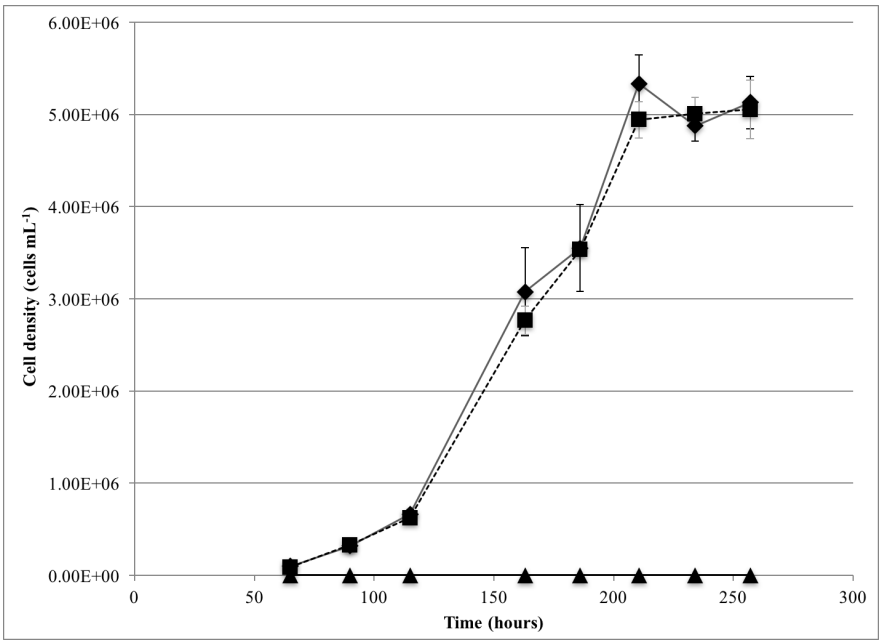


Figure A.1: Effect of light on *Ca.* T. singularis strain PS1. Cultures were grown on ASW with diatom lysate and 1 mM thiosulfate either in light:dark cycles (diamonds) or 24-hr darkness (squares). Negative, no growth control is represented by triangles.

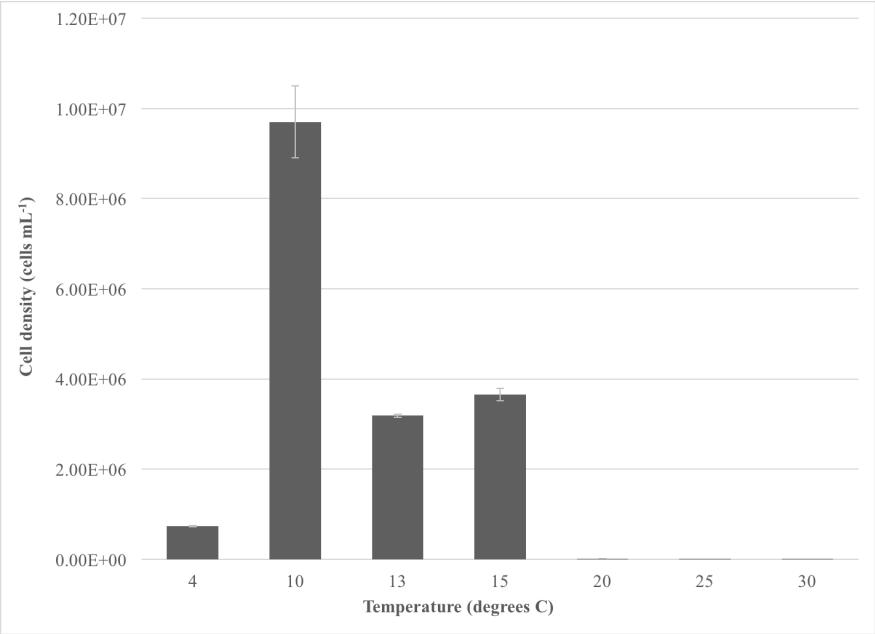


Figure A.2: Maximum cell densities of *Ca.* T. singularis strain PS1 at different temperatures. Error bars show standard deviation among triplicates.

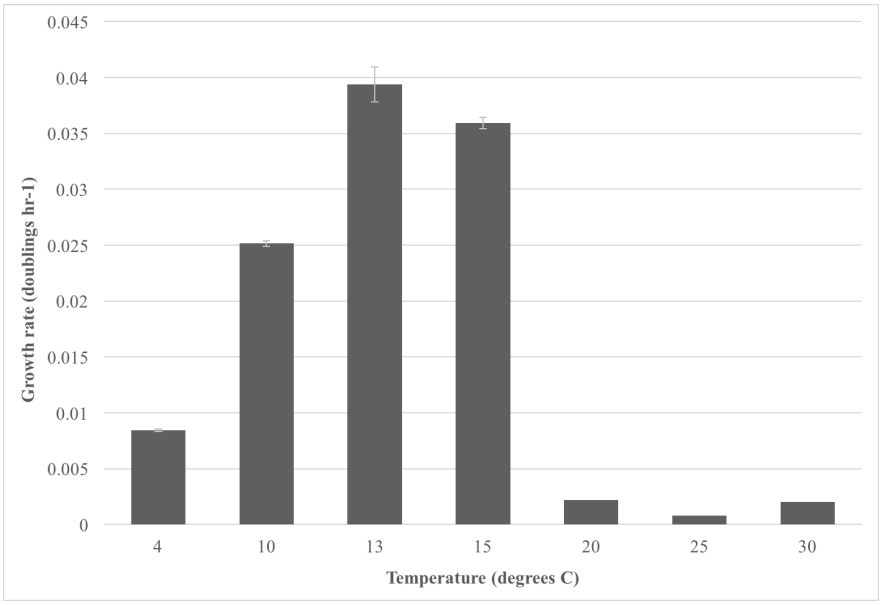


Figure A.3: Growth rates of *Ca.* T. singularis strain PS1 at different temperatures. Error bars show standard deviation among triplicates.

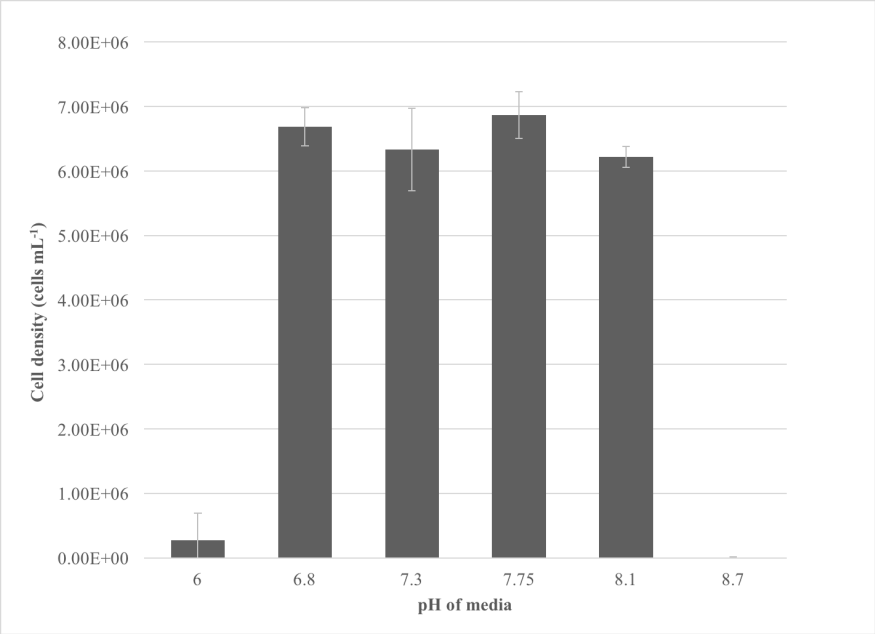


Figure A.4: Maximum cell densities of *Ca.* T. singularis strain PS1 at different pH. Error bars show standard deviation among triplicates.

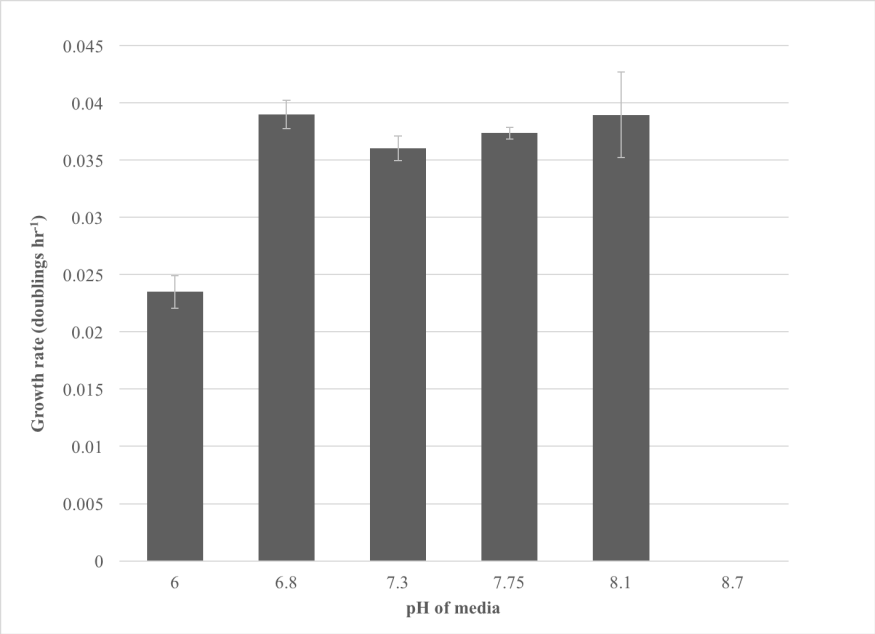


Figure A.5: Growth rates of *Ca.* T. singularis strain PS1 at different pH. Error bars show standard deviation among triplicates.

# Appendix B

# SUPPLEMENTAL INFORMATION TO CHAPTER 2

## *B.1 Supplemental methods*

### *B.1.1 Protein extraction and in-solution proteolytic digestion*

Filters were thawed from -80°C, cut using sterile scissors and each filter was cut in half and resuspended in 0.5 mL of 50 mM ammonium bicarbonate (AMBIC) with 5 mM ethylenediaminetetraacetic acid disodium salt dihydrate (EDTA) buffer (pH 8) in a microcentrifuge tube containing glass beads. Control filters containing only media blanks (i.e., thiosulfate and diatom lysate or only thiosulfate media without PS1) were processed alongside as well as sample tubes containing only the AMBIC/EDTA buffer served as sample processing controls. Each filter suspension was subjected to bead-beating for 30 seconds at 6 m/s, frozen at -20°C for 45 minutes, and thawed. The bead-beating and freeze-thaw cycles were repeated twice followed by a third bead-beating cycle and an overnight freeze at -20°C. Frozen filter resuspensions were thawed, cells further disrupted by vortexing with beads, and then centrifuged at 7,000 rpm (4,600 rcf) for 2 minutes. A 250 $\mu$L aliquot from each tube containing the one-half of its corresponding filter was combined into a new 2 mL protein LoBind eppendorf tube for proteomic processing (see below), resulting in a total of 500 $\mu$L of protein extract for each PS1 treatment.

Estimated bulk protein concentration for lysed cell extracts were measured using the Lowry assay (DC protein assay, Bio-Rad). Lowry assay samples were prepared following

the manufacturer protocol adapted for analysis using a microcuvette (requiring only 5 $\mu$L of sample). Protein concentrations were analyzed using a spectrophotometer (BeckmanCoulter DU800) and determined by comparing sample absorbance units to a bovine serum albumin (BSA) protein concentration calibration curve (prepared only in 50 mM AMBIC, 5 mM EDTA). For downstream proteomics analyses, we also calculated approximate protein concentration from cell density measurements where we assume an 80% extraction efficiency to lyse cells by bead-beating. Thus, we have estimated approximately 1.46 $10^8$ cells ($\sim$ 1.5 $\mu$g) for PS1 grown in only thiosulfate media and 1.49 $10^9$ cells ($\sim$ 15 $\mu$g) for PS1 grown in thiosulfate media with diatom lysate were used for downstream proteomics analyses.

Lysed cells (500 $\mu$L) were subjected to in-solution proteolytic digestion. First, RapiGest SF (Waters), an acid labile surfactant, was added to help facilitate protein solubilization (0.06% w/v). Next, disulfide bonds were reduced with tris(2-carboxyethyl)phosphine (TCEP; to a final concentration of 10 mM) for 1 hr at room temperature in the dark. Reduced samples were then alkylated with iodoacetamide (IAM; to a final concentration of 30 mM) for 1 hr at room temperature in the dark. Excess IAM was quenched by the addition of dithiolthreitol (DTT; to final concentration of 5 mM) for 1 hr at room temperature in the dark. All protein extracts were proteolytically digested with mass spectrometry-grade trypsin/Lys-C mix (Promega) at an estimated substrate-to-enzyme ratio of 25 for 19 hours at 37°C containing a final concentration of 0.05% w/v RapiGest, 50 mM AMBIC and 5 mM EDTA (pH 8). Following trypsin/Lys-C digestion, RapiGest was hydrolyzed by the addition of trifluoroacetic acid (0.5% final v/v, pH <2; concurrently terminating trypsin/Lys-C activity), heated for 45 minutes at 37°C and centrifuged at 15,000 rpm (21,130 rcf) at 4°C for 15 minutes to precipitate the water immiscible decomposition product, dodeca-2-one. The supernatant was removed, transferred to a new protein LoBind eppendorf tube and sam-

ples were dried to near-dryness using a centrivap concentrator (Thermo Savant SPD2010 SpeedVac). Digested samples were desalted using a MacroSpin C18 column (NestGroup) following an established manufacturer protocol. Samples were eluted from the MacroSpin columns using two times 200 $\mu$L of 80% LCMS-grade acetonitrile with 20% LCMS-grade water containing 0.1% LCMS-grade formic acid. Desalted samples were concentrated using the centrivap to near dryness and were immediately resuspended in 95% LCMS-grade water with 5% LCMS-grade acetonitrile containing 0.1% LCMS-grade formic acid. The resuspension solution also contained an internal standard of synthetic peptides (Hi3 Escherichia coli Standard, Waters) at 50 fmol/$\mu$L. The resuspension of 1) PS1 grown in only thiosulfate media was to an approximate 0.15 $\mu$g/$\mu$L (using the smallest possible resuspension volume of 10 $\mu$L) or 1.46 $10^7$ cells/$\mu$L. The resuspension of 2) PS1 grown with thiosulfate and diatom lysate was to an approximate 0.6 g/L protein concentration or 5.97 $10^7$ cells/$\mu$L. All control samples were resuspended with a 25 $\mu$L resuspension volume. Samples were immediately analyzed by liquid chromatography coupled to mass spectrometry (LCMS).

Samples were injected onto the column of a Waters ACQUITY M-class UPLC with an injection volume of 2 $\mu$L in triplicate injections; the PS1 culture with only thiosulfate supplemented media was only injected in duplicates due to limited sample volume. On-column cell count estimates were 1) 2.92 $10^7$ cells (0.3 $\mu$g) and 2) 1.19 $10^8$ cells (1.2 $\mu$g). Peptide separation was performed by reversed-phase chromatography using a nanoACQUITY HSS T3 C18 column (1.8 $\mu$m, 75 $\mu$m x 250 mm; 45°C) with an ACQUITY UPLC M-class Symmetry C18 trapping column (180 $\mu$m x 20 mm). The peptides were trapped at a flow rate of 5 $\mu$L/min at 99% A for 3 minutes. A flow rate of 0.3 $\mu$L/min was used over a gradient between LCMS-grade water (A) and LCMS-grade acetonitrile (B), both modified with 0.1% LCMS-grade formic acid. The total 145 minute gradient method for the separation of

the peptides started at 95% A and ramped to 60% A over the course of 120 minutes. The gradient then switched to 15% A at 122-127 minutes followed by a ramp back to starting conditions 95% A at 128-145 minutes.

The ACQUITY M-class was coupled to a Thermo QExactive HF Orbitrap high-resolution mass spectrometer (HRMS) equipped with a nano-electrospray (NSI) source made in-house following the University of Washington Proteomic Resource (UWPR) design. Using a Micro-Tee (PEEK, 0.025"OD), the commercial analytical column was connected to a commercial emitter (PicoTip, Waters) and the liquid path was applied a high voltage through a platinum wire (adapted from UWPR design). All analyses were carried out in positive mode and a NSI spray voltage of 2.0 kV. Data was collected using data dependent acquisition (DDA) using Xcalibur 4.0 data acquisition software (Thermo Fisher). The MS1 scan range was 400-2000 m/z at 60,000 resolution with a maximum injection time of 30 ms and automated gain control of 1e6. Following each MS1 scan, data-dependent MS2 (dd-MS2) was set to perform on the top 10 ions in a data-dependent manner at 15,000 resolution with a normalized collision energy of 27 eV. Additional selection criteria for dd-MS2 were as follows: maximum injection time of 50 ms with an automated gain control of 5e4, the isolation window was 1.5 Da and dynamic exclusion was set at 20 sec.

### *B.1.2 Data Analysis and Protein Identification*

Data processing was conducted using the software from the trans-proteomic pipeline (TPP v.4.8.1).1 Briefly, raw data was converted to mzML and searched using COMET2,3 against a FASTA protein database consisting of either PS1 (Uniprot 9GAMMA 1,682 proteins accessed July 2016; KEGG estimated protein genes 1,682) or PS1 combined with *Thalassiosira pseudonana* (Uniprot THAPS 12,041 proteins accessed September 2016; KEGG estimated

protein genes 11,672). Each FASTA database also contained the *E.coli* chaperone protein (Hi3 standard; P63284) and was concatenated with a set of randomized sequences. Additional COMET parameters included trypsin enzyme specificity, one allowed non-tryptic termini, up to two missed cleavage sites, carbamidomethylation of cysteine residues as a fixed modification (+57.02146 Da), and oxidation of methionine residues (+15.9949 Da) and clipping of N-terminal methionine as variable modifications. The COMET pep.xml files from each technical triplicate mzML file (or duplicate for PS1 with only thiosulfate treatment) were then grouped and searched in PeptideProphet followed by iProphet4 and ProteinProphet5 within the TPP. The set of randomized sequences were decoy peptides used for the purpose of calculating peptide spectrum matching (PSM) and protein false discovery rates (FDR). Results from each of the combined datasets searched against the PS1 database only, PeptideProphet probabilities greater than 95% for 1) PS1 grown with only thiosulfate, resulted in 1,159 peptides and 3,752 PSM out of a total of 28,601 MS/MS spectra, with a corresponding FDR of 0.2% (PSM level 9/3,752) and for 2) PS1 grown with diatom lysate, resulted in 7,618 peptides and 46,834 PSM out of a total of 120,040 MS/MS spectra, with a corresponding FDR of <0.05% (PSM level 14/46,834).

The list of protein identifications was filtered using a 95% or greater protein probability as calculated from ProteinProphet corresponding to a FDR <0.5%; where 292 and 1,125 proteins were identified in 1) PS1 proteome grown with only thiosulfate and 2) the PS1 proteome grown with diatom lysate, respectively. This represented approximately 17% and 67% of all the predicted proteins encoded in the PS1 genome, respectively.

### *B.1.3 Label-free comparison of relative protein abundances across growth treatments*

Notably, we must consider here the difference in cell biomass injected on-column for the two PS1 growth conditions. PS1 grown with only thiosulfate had much lower biomass on-column and thus many peptides observed were closer to the limit of detection in HRMS. In DDA, the probability of a protein being identified is directly related to a proteins abundance, which is made up of these observed peptide abundances. Thus, because more peptides are near the limit of detection compared to the 2) PS1 grown with thiosulfate and diatom lysate, which had higher biomass on-column, those proteins with higher abundance in the 1) PS1 grown with only thiosulfate are more likely to be observed at >95% protein probability. It is likely that more proteins are expressed in the PS1 grown with only thiosulfate treatment but they were not detected because their peptide abundances exceeded our limit of detection. Overall, this lends to the lower percentage of the proteome being detected in the 1) PS1 grown with only thiosulfate treatment and to the lower percent protein coverage observed. For instance, these analyses resulted in a higher number of proteins identified based on only one or two peptide identifications.

To ensure adequate comparison of relative protein abundances across PS1 growth conditions, we opted to further filter the protein datasets where at least one peptide (or spectral counts) must be observed consistently across individual injections. Raw total spectral counts across triplicates (or duplicates) were then averaged and standard deviations of these averages were calculated for error (for PS1 grown in only thiosulfate averages were calculated based on only two injections). Following this data constraint, 993 total PS1 proteins were observed and used for comparative analyses (Figure 2.2), representing 59% of all the predicted proteins encoded in the PS1 genome. This corresponded to 214 (13%) and 991 (59%) proteins in 1) PS1 proteome grown with only thiosulfate and 2) the PS1 proteome grown

with diatom lysate, respectively (Figure 2.2). Media blank control samples analyzed alongside did not detect any proteins above 95% protein probability or averaged spectral counts greater than the standard deviation, except for the Hi3 internal standard.

Averaged spectral counts of proteins were further normalized to the Hi3 internal standard to account for differences in NSI ionization of peptides due to sample matrix effects (Figure B.4). To compare normalized spectral counts across PS1 growth conditions and account for differences in biomass injected on-column, the relative amount of protein per cell was calculated by dividing by the number of cells injected on-column (see above estimates), resulting in a semi-quantitative estimate of normalized spectral counts per cell (with propagated error). The subcellular localization of each protein is based on predictions from PSORTb (v3.0).

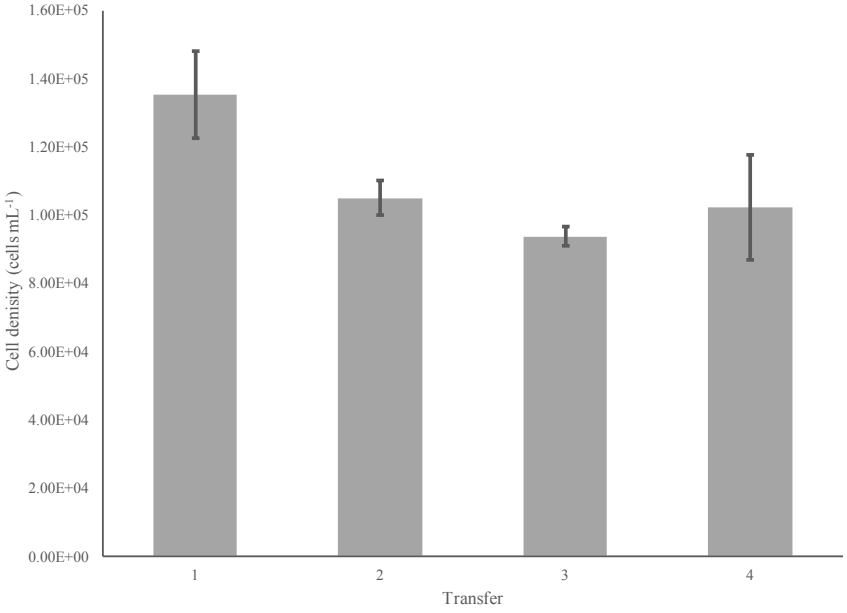


Figure B.1: Final cell densities of subsequent transfers of *Ca.* T. singularis strain PS1 grown in ASW without the addition of nitrogen or vitamins. Error bars indicate standard deviations across triplicate cultures.

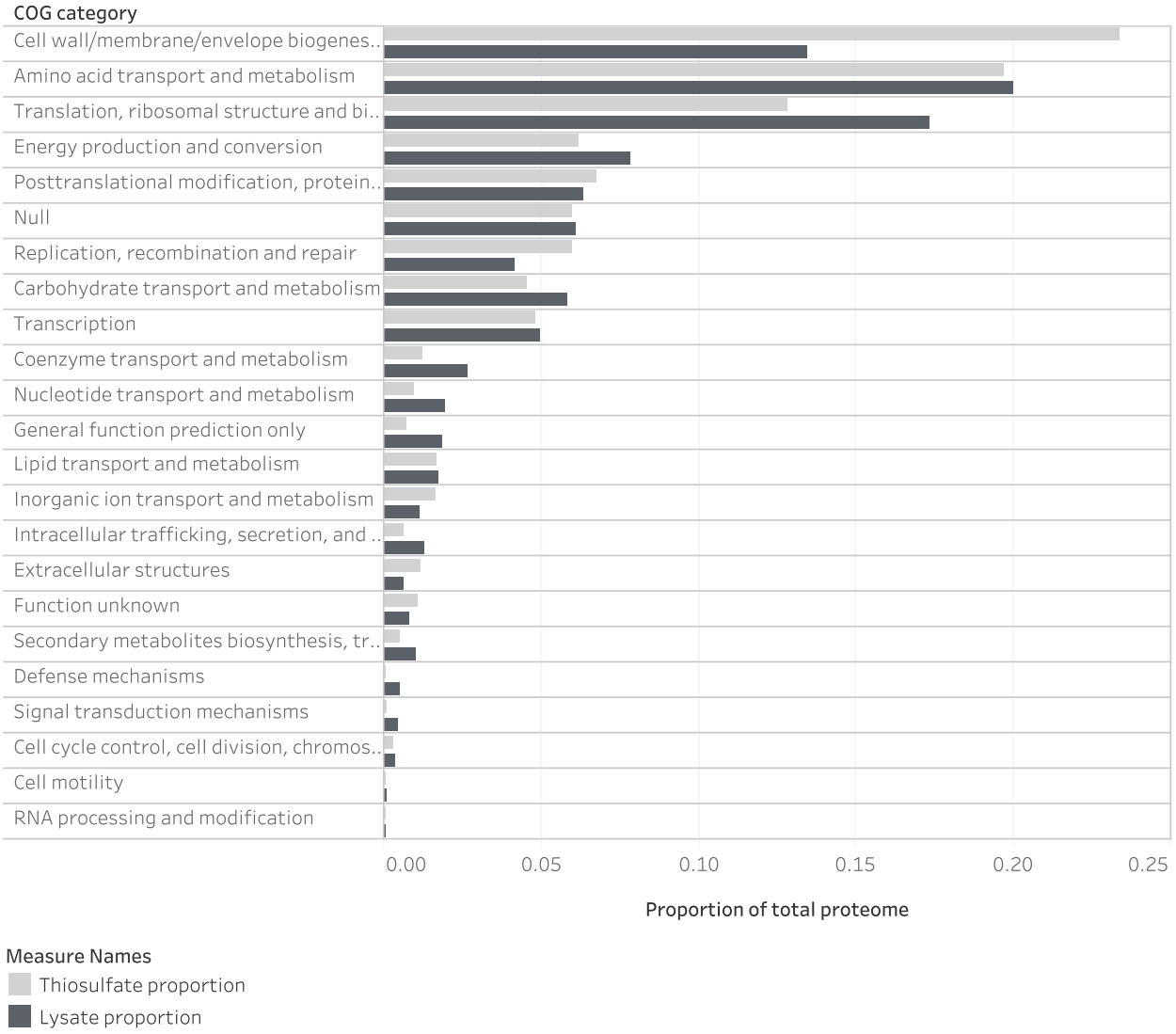


Figure B.2: Complete proteomic profile of *Ca.* T. singularis grown with either thiosulfate only or diatom lysate and thiosulfate categorized into functional annotations by COGs.

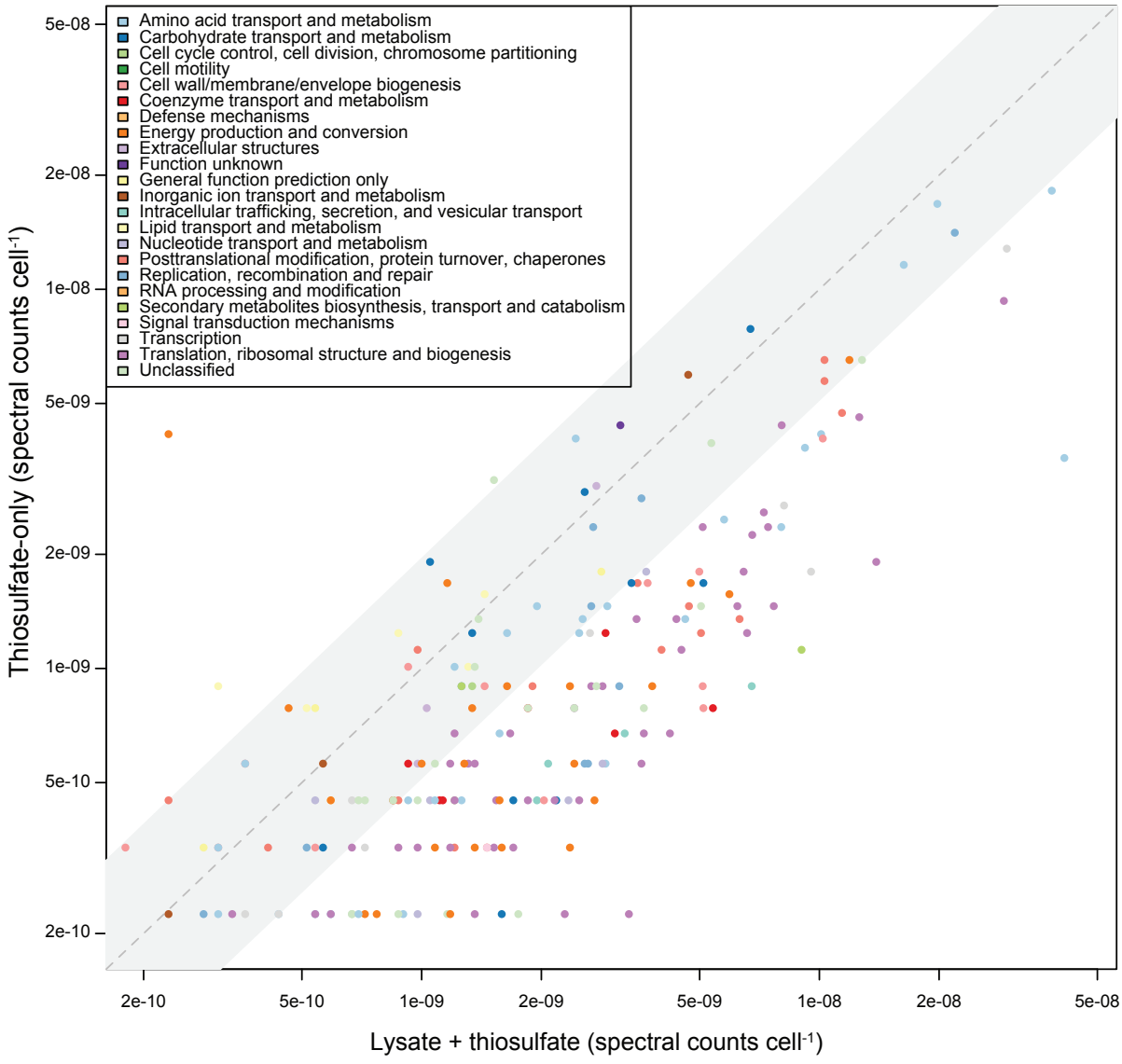


Figure B.3: Differential expression of *Ca.* T. singularis proteins grown with either thiosulfate only or diatom lysate and thiosulfate. The shaded area indicates less than a two-fold change between the two treatments and proteins within that area are considered constitutively expressed. Points are classified by COG functional categories.

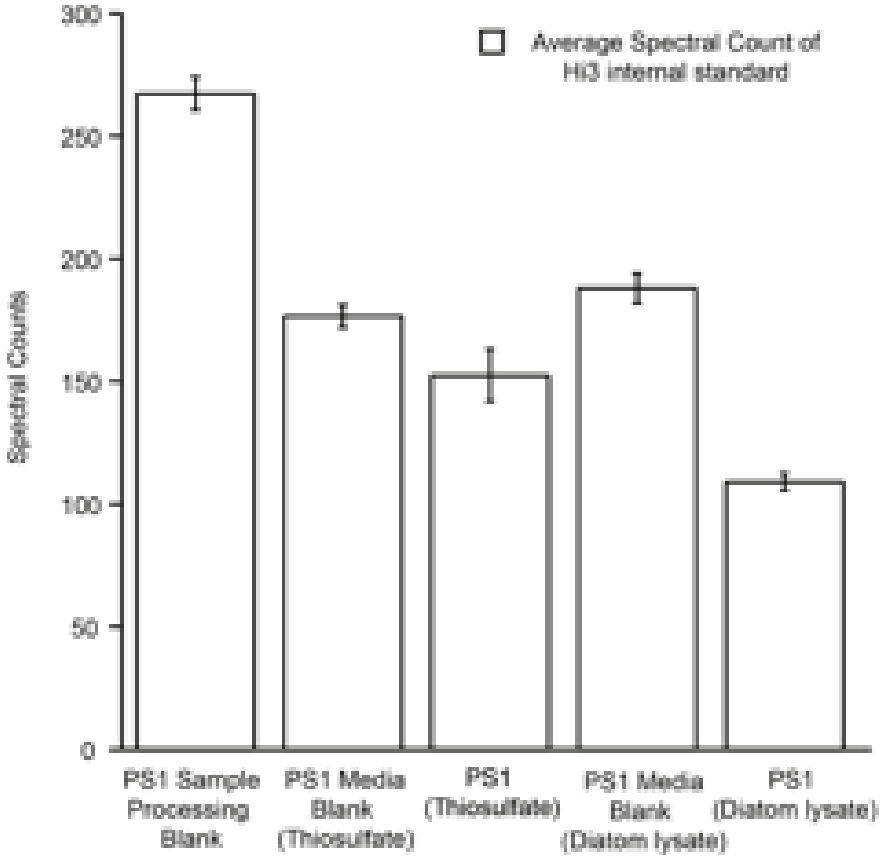


Figure B.4: Matrix effect analysis of Hi3 internal standard peptides as reported in averaged spectral counts (with standard deviation reported as error bars) for multiple injections. Hi3 internal standard was resuspended at an identical concentration for each sample and injected as 100 fmol on-column. Bar graphs represent data from PS1 along with their corresponding sample processing control (AMBIC/EDTA buffer) and media blank controls.

Table B.1: Top 25 most highly expressed proteins in the nutrient-deplete proteome. Expression was quantified as the average normalized spectral counts per cell.

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M3T2F3 | Uncharacterized protein OS | Extracellular | 1.26E-07 | 6.73E-09 | 8.29E-08 | 2.24E-08 | | | |
| A0A0M4LHK8 | Spermidine/putrescine ABC transporter substrate-binding protein OS | Periplasmic | 3.84E-08 | 2.81E-09 | 1.82E-08 | 5.55E-09 | Amino acid transport and metabolism | E | COG0687 |
| A0A0M4L5E0 | Amino acid ABC transporter substrate-binding protein OS | Unknown | 1.98E-08 | 7.06E-10 | 1.68E-08 | 1.73E-09 | Amino acid transport and metabolism | E | COG0834 |
| A0A0M4L3Q3 | Transcriptional regulator OS | Unknown | 2.19E-08 | 2.71E-09 | 1.41E-08 | 2.43E-09 | Replication, recombination and repair | L | COG0776 |
| A0A0M4L4Y3 | Cold-shock protein OS | Cytoplasmic | 2.96E-08 | 1.02E-09 | 1.28E-08 | 3.30E-09 | Transcription | K | COG1278 |
| A0A0M4LGI3 | ABC transporter substrate-binding protein OS | Periplasmic | 1.63E-08 | 1.41E-09 | 1.16E-08 | 1.13E-09 | Amino acid transport and metabolism | E | COG0834 |
| A0A0M4LEH4 | Elongation factor Tu OS | Cytoplasmic | 2.91E-08 | 1.45E-09 | 9.32E-09 | 1.57E-09 | Translation, ribosomal structure and biogenesis | J | COG0050 |
| A0A0M4L584 | Sugar ABC transporter substrate-binding protein OS | Unknown | 6.71E-09 | 3.75E-10 | 7.86E-09 | 1.98E-09 | Carbohydrate transport and metabolism | G | COG1653 |
| A0A0M4M379 | Chaperone protein DnaK OS | Cytoplasmic | 1.03E-08 | 3.61E-10 | 6.51E-09 | 1.35E-09 | Posttranslational modification, protein turnover, chaperones | O | COG0443 |
| A0A0M4LE04 | ATP synthase subunit alpha OS | Cytoplasmic | 1.19E-08 | 5.47E-10 | 6.51E-09 | 1.35E-09 | Energy production and conversion | C | COG0056 |
| A0A0M4LES0 | DNA-directed RNA polymerase subunit beta OS | Cytoplasmic | 1.28E-08 | 4.93E-10 | 6.51E-09 | 1.65E-09 | | | |

Continued on next page.

(SI Table B.1: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M3T1X2 | Taurine ABC transporter substrate-binding protein OS | Periplasmic | 4.68E-09 | 1.90E-10 | 5.95E-09 | 1.49E-09 | Inorganic ion transport and metabolism | P | COG4521 |
| A0A0M4L417 | 10 kDa chaperonin OS | Cytoplasmic | 1.03E-08 | 4.21E-10 | 5.73E-09 | 8.88E-10 | Posttranslational modification, protein turnover, chaperones | O | COG0234 |
| A0A0M3T1Z1 | 60 kDa chaperonin OS | Cytoplasmic | 1.14E-08 | 3.71E-10 | 4.72E-09 | 1.62E-09 | Posttranslational modification, protein turnover, chaperones | O | COG0459 |
| A0A0M4M1X8 | 50S ribosomal protein L7/L12 OS | Unknown | 1.26E-08 | 4.96E-10 | 4.60E-09 | 3.57E-10 | Translation, ribosomal structure and biogenesis | J | COG0222 |
| A0A0M4L3A2 | 50S ribosomal protein L3 OS | Cytoplasmic | 8.04E-09 | 4.03E-10 | 4.38E-09 | 8.50E-10 | Translation, ribosomal structure and biogenesis | J | COG0087 |
| A0A0M4LGE2 | Uncharacterized protein OS | Unknown | 3.16E-09 | 2.52E-10 | 4.38E-09 | 3.43E-10 | Function unknown | S | COG3134 |
| A0A0M4M2R8 | C4-dicarboxylate ABC transporter substrate-binding protein OS | Unknown | 2.31E-10 | 7.74E-11 | 4.15E-09 | 5.57E-10 | Energy production and conversion | C | COG3181 |
| A0A0M4LZX7 | Glutamine synthetase OS | Cytoplasmic | 1.01E-08 | 4.97E-10 | 4.15E-09 | 5.57E-10 | Amino acid transport and metabolism | E | COG0174 |
| A0A0M4M0C3 | DNA-binding protein OS | Unknown | 1.02E-08 | 5.91E-10 | 4.04E-09 | 4.24E-10 | Cell wall, membrane, envelope biogenesis | M | COG1744 |
| A0A0M4M0G8 | Glycine/betaine ABC transporter substrate-binding protein OS | Periplasmic | 2.44E-09 | 2.48E-10 | 4.04E-09 | 1.61E-09 | Amino acid transport and metabolism | E | COG2113 |
| A0A0M4L4P6 | Membrane protein TolA OS | Unknown | 5.35E-09 | 6.46E-10 | 3.93E-09 | 1.46E-09 | | | |
| A0A0M3T1N6 | ABC transporter substrate-binding protein OS | Periplasmic | 9.20E-09 | 4.86E-10 | 3.82E-09 | 9.89E-10 | Amino acid transport and metabolism | E | COG2113 |

Continued on next page.

(SI Table B.1: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M4L4Q0 | Branched-chain amino acid ABC transporter substrate-binding protein OS | Unknown | 4.13E-08 | 2.49E-09 | 3.59E-09 | 2.50E-10 | Amino acid transport and metabolism | E | COG0683 |
| A0A0M4M190 | Uncharacterized protein OS | Membrane | 1.52E-09 | 1.68E-10 | 3.14E-09 | 2.19E-10 | | | |

Table B.2: Top 25 most highly expressed proteins in the nutrient-replete proteome. Expression was quantified as the average normalized spectral counts per cell.

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M3T2F3 | Uncharacterized protein OS | Extracellular | 1.26E-07 | 6.73E-09 | 8.29E-08 | 2.24E-08 | | | |
| A0A0M4L4Q0 | Branched-chain amino acid ABC transporter substrate-binding protein OS | Unknown | 4.13E-08 | 2.49E-09 | 3.59E-09 | 2.50E-10 | Amino acid transport and metabolism | E | COG0683 |
| A0A0M4LHK8 | Spermidine/putrescine ABC transporter substrate-binding protein OS | Periplasmic | 3.84E-08 | 2.81E-09 | 1.82E-08 | 5.55E-09 | Amino acid transport and metabolism | E | COG0687 |
| A0A0M4L4Y3 | Cold-shock protein OS | Cytoplasmic | 2.96E-08 | 1.02E-09 | 1.28E-08 | 3.30E-09 | Transcription | K | COG1278 |
| A0A0M4LEH4 | Elongation factor Tu OS | Cytoplasmic | 2.91E-08 | 1.45E-09 | 9.32E-09 | 1.57E-09 | Translation, ribosomal structure and biogenesis | J | COG0050 |
| A0A0M3T1V7 | ABC transporter substrate-binding protein OS | Periplasmic | 2.66E-08 | 1.31E-09 | | | Carbohydrate transport and metabolism | G | COG1653 |
| A0A0M4L3Q3 | Transcriptional regulator OS | Unknown | 2.19E-08 | 2.71E-09 | 1.41E-08 | 2.43E-09 | Replication, recombination and repair | L | COG0776 |
| A0A0M4L5E0 | Amino acid ABC transporter substrate-binding protein OS | Unknown | 1.98E-08 | 7.06E-10 | 1.68E-08 | 1.73E-09 | Amino acid transport and metabolism | E | COG0834 |
| A0A0M4LGI3 | ABC transporter substrate-binding protein OS | Periplasmic | 1.63E-08 | 1.41E-09 | 1.16E-08 | 1.13E-09 | Amino acid transport and metabolism | E | COG0834 |
| A0A0M4LEI8 | 30S ribosomal protein S4 OS | Cytoplasmic | 1.39E-08 | 1.74E-09 | 1.91E-09 | 2.07E-10 | Translation, ribosomal structure and biogenesis | J | COG0522 |
| A0A0M4LES0 | DNA-directed RNA polymerase subunit beta OS | Cytoplasmic | 1.28E-08 | 4.93E-10 | 6.51E-09 | 1.65E-09 | | | |

Continued on next page.

(SI Table B.2: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M4M1X8 | 50S ribosomal protein L7/L12 OS | Unknown | 1.26E-08 | 4.96E-10 | 4.60E-09 | 3.57E-10 | Translation, ribosomal structure and biogenesis | J | COG0222 |
| A0A0M4LE04 | ATP synthase subunit alpha OS | Cytoplasmic | 1.19E-08 | 5.47E-10 | 6.51E-09 | 1.35E-09 | Energy production and conversion | C | COG0056 |
| A0A0M3T1Z1 | 60 kDa chaperonin OS | Cytoplasmic | 1.14E-08 | 3.71E-10 | 4.72E-09 | 1.62E-09 | Posttranslational modification, protein turnover, chaperones | O | COG0459 |
| A0A0M4M379 | Chaperone protein DnaK OS | Cytoplasmic | 1.03E-08 | 3.61E-10 | 6.51E-09 | 1.35E-09 | Posttranslational modification, protein turnover, chaperones | O | COG0443 |
| A0A0M4L417 | 10 kDa chaperonin OS | Cytoplasmic | 1.03E-08 | 4.21E-10 | 5.73E-09 | 8.88E-10 | Posttranslational modification, protein turnover, chaperones | O | COG0234 |
| A0A0M4M0C3 | DNA-binding protein OS | Unknown | 1.02E-08 | 5.91E-10 | 4.04E-09 | 4.24E-10 | Cell wall, membrane, envelope biogenesis | M | COG1744 |
| A0A0M4LZX7 | Glutamine synthetase OS | Cytoplasmic | 1.01E-08 | 4.97E-10 | 4.15E-09 | 5.57E-10 | Amino acid transport and metabolism | E | COG0174 |
| A0A0M3T1R5 | DNA-directed RNA polymerase subunit beta' OS | Cytoplasmic | 9.53E-09 | 5.33E-10 | 1.80E-09 | 3.41E-10 | Transcription | K | COG0086 |
| A0A0M3T1N6 | ABC transporter substrate-binding protein OS | Periplasmic | 9.20E-09 | 4.86E-10 | 3.82E-09 | 9.89E-10 | Amino acid transport and metabolism | E | COG2113 |
| A0A0M5KRJ8 | C4-dicarboxylate ABC transporter substrate-binding protein OS | Unknown | 9.02E-09 | 3.16E-10 | 1.12E-09 | 7.81E-11 | Secondary metabolites biosynthesis, transport and catabolism | Q | COG4663 |
| A0A0M4LCE7 | DNA-directed RNA polymerase subunit alpha OS | Cytoplasmic | 8.15E-09 | 5.32E-10 | 2.69E-09 | 6.62E-10 | Transcription | K | COG0202 |

Continued on next page.

(SI Table B.2: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|
| A0A0M4L3A2 | 50S ribosomal protein L3 OS | Cytoplasmic | 8.04E-09 | 4.03E-10 | 4.38E-09 | 8.50E-10 | Translation, ribosomal structure and biogenesis | J | COG0087 |
| A0A0M4L6A6 | Dimethylglycine dehydrogenase OS | Cytoplasmic | 8.02E-09 | 6.54E-10 | 2.36E-09 | 5.04E-10 | Amino acid transport and metabolism | E | COG0404 |
| A0A0M4L571 | Polyribonucleotide nucleotidyltransferase OS | Cytoplasmic | 7.68E-09 | 4.03E-10 | 1.46E-09 | 4.87E-10 | Translation, ribosomal structure and biogenesis | J | COG1185 |

Table B.3: Proteins with significantly higher expression under the nutrient-replete treatment compared to the nutrient-deplete treatment. Expression was quantified as the average normalized spectral counts per cell. Fold change was calculated as $\log_2$(Replete expression/Deplete expression). A cutoff $\log_2$-fold change was set at 1.

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4LCD2 | 30S ribosomal protein S10 OS | Cytoplasmic | 3.32E-09 | 3.69E-10 | 2.25E-10 | 1.56E-11 | 3.883186335 | Translation, ribosomal structure and biogenesis | J | COG0051 |
| A0A0M4L4Q0 | Branched-chain amino acid ABC transporter substrate-binding protein OS | Cytoplasmic | 4.13E-08 | 2.49E-09 | 3.59E-09 | 2.50E-10 | 3.524086032 | Amino acid transport and metabolism | E | COG0683 |
| A0A0M4LEI2 | 50S ribosomal protein L14 OS | Extracellular | 2.29E-09 | 1.38E-10 | 2.25E-10 | 1.56E-11 | 3.347350692 | Translation, ribosomal structure and biogenesis | J | COG0093 |
| A0A0M5KRJ8 | C4-dicarboxylate ABC transporter substrate-binding protein OS | Cytoplasmic | 9.02E-09 | 3.16E-10 | 1.12E-09 | 7.81E-11 | 3.009628701 | Secondary metabolites biosynthesis, transport and catabolism | Q | COG4663 |
| A0A0M3T2G2 | Uncharacterized protein OS | Membrane | 1.75E-09 | 1.70E-10 | 2.25E-10 | 1.56E-11 | 2.959358016 | | | |
| A0A0M5KYS1 | Preprotein translocase subunit YajC OS | Extracellular | 6.76E-09 | 8.86E-10 | 8.98E-10 | 3.24E-10 | 2.912235896 | Intracellular trafficking, secretion, and vesicular transport | U | COG1862 |
| A0A0M4LEI8 | 30S ribosomal protein S4 OS | Cytoplasmic | 1.39E-08 | 1.74E-09 | 1.91E-09 | 2.07E-10 | 2.86344034 | Translation, ribosomal structure and biogenesis | J | COG0522 |
| A0A0M3T1W8 | Dihydroxy-acid dehydratase OS | Cytoplasmic | 1.59E-09 | 1.85E-10 | 2.25E-10 | 1.56E-11 | 2.821029859 | Carbohydrate transport and metabolism | G | COG0129 |
| A0A0M5L020 | Phosphoenolpyruvate carboxykinase [ATP] OS | Cytoplasmic | 2.36E-09 | 1.40E-10 | 3.37E-10 | 1.61E-10 | 2.807966363 | Energy production and conversion | C | COG1866 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M3T1V4 | Ketol-acid reductoisomerase OS | Cytoplasmic | 5.40E-09 | 2.17E-10 | 7.86E-10 | 4.80E-10 | 2.78035819 | Coenzyme transport and metabolism | H | COG0059 |
| A0A0M4M2L3 | Outer membrane protein assembly factor BamA OS | Membrane | 5.11E-09 | 3.99E-10 | 7.86E-10 | 1.68E-10 | 2.700722074 | Cell wall, membrane, envelope biogenesis | M | COG4775 |
| A0A0M3T1M5 | 50S ribosomal protein L4 OS | Cytoplasmic | 3.57E-09 | 1.22E-10 | 5.61E-10 | 1.64E-10 | 2.669851398 | Translation, ribosomal structure and biogenesis | J | COG0088 |
| A0A0M4LNJ5 | 50S ribosomal protein L23 OS | Membrane | 4.21E-09 | 1.78E-10 | 6.74E-10 | 4.69E-11 | 2.642999737 | Translation, ribosomal structure and biogenesis | J | COG0089 |
| A0A0M5KY70 | ATP synthase subunit b OS | Cytoplasmic | 2.72E-09 | 1.82E-10 | 4.49E-10 | 3.12E-11 | 2.598819301 | Energy production and conversion | C | COG0711 |
| A0A0M4LG53 | Alanine–tRNA ligase OS | Periplasmic | 1.36E-09 | 1.99E-10 | 2.25E-10 | 1.56E-11 | 2.595609745 | Translation, ribosomal structure and biogenesis | J | COG0013 |
| A0A0M3T2G1 | LPS-assembly protein LptD OS | OuterMembrane | 5.09E-09 | 2.82E-10 | 8.98E-10 | 3.24E-10 | 2.502878306 | Cell wall, membrane, envelope biogenesis | M | COG1452 |
| A0A0M4LE84 | 50S ribosomal protein L21 OS | Cytoplasmic | 2.49E-09 | 1.95E-10 | 4.49E-10 | 3.12E-11 | 2.471358392 | Translation, ribosomal structure and biogenesis | J | COG0261 |
| A0A0M4M1N1 | 30S ribosomal protein S17 OS | Membrane | 3.62E-09 | 2.91E-10 | 6.74E-10 | 4.69E-11 | 2.425169201 | Translation, ribosomal structure and biogenesis | J | COG0186 |
| A0A0M5L010 | 50S ribosomal protein L2 OS | Cytoplasmic | 6.58E-09 | 3.83E-10 | 1.24E-09 | 1.81E-10 | 2.407747463 | Translation, ribosomal structure and biogenesis | J | COG0090 |
| A0A0M3T1R5 | DNA-directed RNA polymerase subunit beta' OS | Cytoplasmic | 9.53E-09 | 5.33E-10 | 1.80E-09 | 3.41E-10 | 2.404479308 | Transcription | K | COG0086 |
| A0A0M4L571 | Polyribonucleotide nucleotidyltransferase OS | Cytoplasmic | 7.68E-09 | 4.03E-10 | 1.46E-09 | 4.87E-10 | 2.395137942 | Translation, ribosomal structure and biogenesis | J | COG1185 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M5KRU1 | 2Fe-2S ferredoxin OS | Cytoplasmic | 1.18E-09 | 9.66E-11 | 2.25E-10 | 1.56E-11 | 2.390789953 | Energy production and conversion | C | COG0633 |
| A0A0M4LFK3 | 5'-nucleotidase OS | Cytoplasmic | 2.34E-09 | 2.81E-10 | 4.49E-10 | 3.19E-10 | 2.38172118 | Nucleotide transport and metabolism | F | COG0737 |
| A0A0M4LN91 | Glycine dehydrogenase (decarboxylating) OS | Cytoplasmic | 2.90E-09 | 1.85E-10 | 5.61E-10 | 1.64E-10 | 2.369980224 | Amino acid transport and metabolism | E | COG1003 |
| A0A0M4LEH9 | Methyltetrahydrofolate:corrinoid methyltransferase OS | Membrane | 1.16E-09 | 1.38E-10 | 2.25E-10 | 1.56E-11 | 2.366127899 | | | |
| A0A0M4M1L3 | Adenylylsulfate reductase subunit beta OS | Unknown | 2.85E-09 | 1.19E-10 | 5.61E-10 | 1.64E-10 | 2.344889243 | Nucleotide transport and metabolism | F | COG1146 |
| A0A0M3T1P7 | 50S ribosomal protein L22 OS | Cytoplasmic | 1.70E-09 | 9.41E-11 | 3.37E-10 | 1.61E-10 | 2.33471425 | Translation, ribosomal structure and biogenesis | J | COG0091 |
| A0A0M3T299 | Glyceraldehyde-3-phosphate dehydrogenase OS | Unknown | 2.18E-09 | 2.06E-10 | 4.49E-10 | 3.12E-11 | 2.279540785 | Carbohydrate transport and metabolism | G | COG0057 |
| A0A0M4M1P0 | 30S ribosomal protein S11 OS | Cytoplasmic | 2.16E-09 | 1.03E-10 | 4.49E-10 | 3.12E-11 | 2.266243962 | Translation, ribosomal structure and biogenesis | J | COG0100 |
| A0A0M3T2E3 | Protein translocase subunit SecA OS | Cytoplasmic | 3.24E-09 | 3.51E-10 | 6.74E-10 | 6.37E-10 | 2.265173317 | Intracellular trafficking, secretion, and vesicular transport | U | COG0653 |
| A0A0M4M1Q1 | Cytochrome C OS | Cytoplasmic | 1.59E-09 | 2.96E-10 | 3.37E-10 | 1.61E-10 | 2.238206269 | Energy production and conversion | C | COG2863 |
| A0A0M5KRH8 | DNA topoisomerase 1 OS | Membrane | 2.62E-09 | 1.14E-10 | 5.61E-10 | 1.64E-10 | 2.223494136 | Replication, recombination and repair | L | COG0550 |
| A0A0M4LZ66 | Trigger factor OS | Cytoplasmic | 6.30E-09 | 3.09E-10 | 1.35E-09 | 9.37E-11 | 2.222392421 | Posttranslational modification, protein turnover, chaperones | O | COG0544 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4M0Y0 | Uncharacterized protein OS | Membrane | 3.62E-09 | 2.34E-10 | 7.86E-10 | 1.68E-10 | 2.20338848 | | | |
| A0A0M3T1Z9 | Single-stranded DNA-binding protein OS | Cytoplasmic | 2.57E-09 | 2.61E-10 | 5.61E-10 | 1.64E-10 | 2.195695683 | Replication, recombination and repair | L | COG0629 |
| A0A0M4L4W1 | S-adenosylmethionine synthase OS | Unknown | 3.06E-09 | 1.88E-10 | 6.74E-10 | 4.69E-11 | 2.182711156 | Coenzyme transport and metabolism | H | COG0192 |
| A0A0M5KZY4 | Membrane protein insertase YidC OS | Cytoplasmic | 2.03E-09 | 2.44E-10 | 4.49E-10 | 3.12E-11 | 2.176692377 | Cell wall, membrane, envelope biogenesis | M | COG0706 |
| A0A0M4LZ29 | 30S ribosomal protein S19 OS | Cytoplasmic | 1.52E-09 | 6.56E-11 | 3.37E-10 | 1.61E-10 | 2.173250827 | Translation, ribosomal structure and biogenesis | J | COG0185 |
| A0A0M4LFF9 | Protein TolB OS | Cytoplasmic | 1.95E-09 | 2.04E-10 | 4.49E-10 | 3.19E-10 | 2.118686774 | Intracellular trafficking, secretion, and vesicular transport | U | COG0823 |
| A0A0M5KRI8 | Adenylate kinase OS | Unknown | 9.77E-10 | 9.43E-11 | 2.25E-10 | 1.56E-11 | 2.118433561 | Nucleotide transport and metabolism | F | COG0563 |
| A0A0M4LN44 | Protein GrpE OS | Cytoplasmic | 1.46E-09 | 1.61E-10 | 3.37E-10 | 1.61E-10 | 2.115147873 | Posttranslational modification, protein turnover, chaperones | O | COG0576 |
| A0A0M4M0K1 | Uncharacterized protein OS | Cytoplasmic | 1.46E-09 | 1.61E-10 | 3.37E-10 | 1.61E-10 | 2.115147873 | Signal transduction mechanisms | T | COG2976 |
| A0A0M4L6M5 | Ubiquinol-cytochrome c reductase iron-sulfur subunit OS | Cytoplasmic | 2.42E-09 | 1.41E-10 | 5.61E-10 | 1.64E-10 | 2.108934372 | Energy production and conversion | C | COG0723 |
| A0A0M5KRF3 | 30S ribosomal protein S2 OS | Unknown | 6.22E-09 | 6.93E-10 | 1.46E-09 | 1.88E-10 | 2.090946211 | Translation, ribosomal structure and biogenesis | J | COG0052 |
| A0A0M4LEM6 | Cytochrome c oxidase subunit 2 OS | Cytoplasmic | 3.80E-09 | 1.69E-10 | 8.98E-10 | 3.24E-10 | 2.081212068 | Energy production and conversion | C | COG1622 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4L548 | 50S ribosomal protein L31 OS | Cytoplasmic | 1.85E-09 | 2.12E-10 | 4.49E-10 | 3.19E-10 | 2.042737921 | Translation, ribosomal structure and biogenesis | J | COG0254 |
| A0A0M4L5N7 | Peptidylprolyl isomerase OS | Cytoplasmic | 5.04E-09 | 3.59E-10 | 1.24E-09 | 1.81E-10 | 2.023083613 | Posttranslational modification, protein turnover, chaperones | O | COG0760 |
| A0A0M4L5G6 | Electron transfer flavoprotein-ubiquinone oxidoreductase OS | Cytoplasmic | 1.36E-09 | 1.99E-10 | 3.37E-10 | 1.61E-10 | 2.012786155 | Energy production and conversion | C | COG0644 |
| A0A0M4LF58 | Elongation factor Ts OS | Cytoplasmic | 4.50E-09 | 1.85E-10 | 1.12E-09 | 6.40E-10 | 2.006426269 | Translation, ribosomal structure and biogenesis | J | COG0264 |
| A0A0M4LZ11 | Argininosuccinate synthase OS | Cytoplasmic | 8.99E-10 | 5.29E-11 | 2.25E-10 | 1.56E-11 | 1.998396114 | Amino acid transport and metabolism | E | COG0137 |
| A0A0M5KS12 | Aspartate kinase OS | Cytoplasmic | 8.74E-10 | 1.21E-10 | 2.25E-10 | 1.56E-11 | 1.957708278 | | | |
| A0A0M3T272 | Short-chain dehydrogenase OS | Unknown | 1.70E-09 | 2.11E-10 | 4.49E-10 | 3.12E-11 | 1.920747396 | Carbohydrate transport and metabolism | G | COG3347 |
| A0A0M4LBY9 | ATP synthase subunit beta OS | Cytoplasmic | 5.94E-09 | 3.36E-10 | 1.57E-09 | 9.59E-10 | 1.919698372 | Energy production and conversion | C | COG0055 |
| A0A0M4LHS2 | Protein translocase subunit SecD OS | Unknown | 2.08E-09 | 2.14E-10 | 5.61E-10 | 1.64E-10 | 1.890510852 | Intracellular trafficking, secretion, and vesicular transport | U | COG0342 |
| A0A0M4L4W7 | Peptidyl-prolyl cis-trans isomerase OS | Cytoplasmic | 1.21E-09 | 2.51E-10 | 3.37E-10 | 1.61E-10 | 1.844186551 | Posttranslational modification, protein turnover, chaperones | O | COG0652 |
| A0A0M4LGM3 | Translation initiation factor IF-2 OS | Cytoplasmic | 6.45E-09 | 3.03E-10 | 1.80E-09 | 6.47E-10 | 1.841302254 | Translation, ribosomal structure and biogenesis | J | COG0532 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4L3Z9 | Chaperone protein HtpG OS | Cytoplasmic | 4.01E-09 | 1.27E-10 | 1.12E-09 | 6.40E-10 | 1.840103504 | Posttranslational modification, protein turnover, chaperones | O | COG0326 |
| A0A0M4M3S5 | Serine–tRNA ligase OS | Cytoplasmic | 1.18E-09 | 1.98E-10 | 3.37E-10 | 1.61E-10 | 1.807966363 | Translation, ribosomal structure and biogenesis | J | COG0172 |
| A0A0M5KS43 | Ubiquinol-cytochrome C reductase OS | Cytoplasmic | 1.57E-09 | 6.68E-11 | 4.49E-10 | 3.12E-11 | 1.805977209 | Energy production and conversion | C | COG2857 |
| A0A0M3T1S7 | Protein RecA OS | Cytoplasmic | 3.14E-09 | 2.56E-10 | 8.98E-10 | 6.25E-11 | 1.805977209 | Replication, recombination and repair | L | COG0468 |
| A0A0M5KS58 | Uncharacterized protein OS | Cytoplasmic | 5.04E-09 | 2.27E-10 | 1.46E-09 | 8.00E-10 | 1.787455365 | | | |
| A0A0M4LF95 | 50S ribosomal protein L32 OS | Cytoplasmic | 1.54E-09 | 2.72E-10 | 4.49E-10 | 3.12E-11 | 1.778143001 | Translation, ribosomal structure and biogenesis | J | COG0333 |
| A0A0M4LFM6 | Ferredoxin–NADP reductase OS | Cytoplasmic | 7.71E-10 | 8.09E-11 | 2.25E-10 | 1.56E-11 | 1.776805859 | Energy production and conversion | C | COG1018 |
| A0A0M4M407 | Peptide ABC transporter substrate-binding protein OS | Cytoplasmic | 4.60E-09 | 2.43E-10 | 1.35E-09 | 9.37E-11 | 1.768674454 | Amino acid transport and metabolism | E | COG0747 |
| A0A0M4L6A6 | Dimethylglycine dehydrogenase OS | Unknown | 8.02E-09 | 6.54E-10 | 2.36E-09 | 5.04E-10 | 1.764815377 | Amino acid transport and metabolism | E | COG0404 |
| A0A0M5KY99 | 30S ribosomal protein S7 OS | Cytoplasmic | 4.37E-09 | 2.73E-10 | 1.35E-09 | 9.37E-11 | 1.694673872 | Translation, ribosomal structure and biogenesis | J | COG0049 |
| A0A0M4LPG9 | Protein HflC OS | Cytoplasmic | 4.70E-09 | 2.15E-10 | 1.46E-09 | 8.00E-10 | 1.686692388 | Posttranslational modification, protein turnover, chaperones | O | COG0330 |
| A0A0M4L2X4 | ATP synthase epsilon chain OS | Periplasmic | 1.08E-09 | 8.44E-11 | 3.37E-10 | 1.61E-10 | 1.680210816 | Energy production and conversion | C | COG0355 |
| A0A0M3T1R3 | Malate dehydrogenase OS | Membrane | 7.20E-10 | 1.20E-10 | 2.25E-10 | 1.56E-11 | 1.678071905 | Energy production and conversion | C | COG0039 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M3T1P5 | 30S ribosomal protein S12 OS | Cytoplasmic | 2.85E-09 | 2.23E-10 | 8.98E-10 | 6.25E-11 | 1.666174569 | Translation, ribosomal structure and biogenesis | J | COG0048 |
| A0A0M3T217 | 30S ribosomal protein S1 OS | Cytoplasmic | 7.43E-09 | 3.98E-10 | 2.36E-09 | 2.28E-10 | 1.654575351 | Translation, ribosomal structure and biogenesis | J | COG0539 |
| A0A0M4LEH4 | Elongation factor Tu OS | Cytoplasmic | 2.91E-08 | 1.45E-09 | 9.32E-09 | 1.57E-09 | 1.642617293 | Translation, ribosomal structure and biogenesis | J | COG0050 |
| A0A0M4M0T4 | Branched-chain amino acid ABC transporter ATPase OS | Unknown | 6.94E-10 | 2.21E-11 | 2.25E-10 | 1.56E-11 | 1.625010661 | Amino acid transport and metabolism | E | COG0411 |
| A0A0M4M2U5 | Peptidyl-prolyl cis-trans isomerase OS | Cytoplasmic | 2.42E-09 | 8.87E-11 | 7.86E-10 | 1.68E-10 | 1.62240583 | | | |
| A0A0M3T1M7 | 50S ribosomal protein L24 OS | Cytoplasmic | 2.42E-09 | 8.87E-11 | 7.86E-10 | 1.68E-10 | 1.62240583 | Translation, ribosomal structure and biogenesis | J | COG0198 |
| A0A0M4M4E0 | Uncharacterized protein OS | Membrane | 2.75E-09 | 2.13E-10 | 8.98E-10 | 6.25E-11 | 1.614644269 | | | |
| A0A0M5KYK2 | ABC transporter substrate-binding protein OS | Unknown | 5.11E-09 | 2.01E-10 | 1.68E-09 | 8.03E-10 | 1.604862058 | Carbohydrate transport and metabolism | G | COG1653 |
| A0A0M4LCE7 | DNA-directed RNA polymerase subunit alpha OS | Unknown | 8.15E-09 | 5.32E-10 | 2.69E-09 | 6.62E-10 | 1.599193887 | Transcription | K | COG0202 |
| A0A0M3T2A8 | 30S ribosomal protein S16 OS | Cytoplasmic | 6.78E-09 | 3.99E-10 | 2.25E-09 | 6.54E-10 | 1.591360272 | Translation, ribosomal structure and biogenesis | J | COG0228 |
| A0A0M4L4R9 | 50S ribosomal protein L11 OS | Cytoplasmic | 2.67E-09 | 3.04E-10 | 8.98E-10 | 6.25E-11 | 1.572052392 | Translation, ribosomal structure and biogenesis | J | COG0080 |
| A0A0M4LI50 | Uncharacterized protein OS | Membrane | 6.68E-10 | 4.93E-11 | 2.25E-10 | 1.56E-11 | 1.569923101 | | | |
| A0A0M4L5B2 | Peptide ABC transporter ATP-binding protein OS | Unknown | 6.68E-10 | 4.93E-11 | 2.25E-10 | 1.56E-11 | 1.569923101 | | | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M5L035 | 50S ribosomal protein L18 OS | Cytoplasmic | 9.77E-10 | 2.38E-10 | 3.37E-10 | 1.61E-10 | 1.535609971 | Translation, ribosomal structure and biogenesis | J | COG0256 |
| A0A0M3T1U8 | Aldehyde dehydrogenase OS | Extracellular | 4.75E-09 | 4.76E-10 | 1.68E-09 | 1.97E-10 | 1.49946628 | Energy production and conversion | C | COG1012 |
| A0A0M4M1M3 | Elongation factor G OS | Membrane | 7.25E-09 | 3.08E-10 | 2.58E-09 | 2.40E-10 | 1.490609929 | Translation, ribosomal structure and biogenesis | J | COG0480 |
| A0A0M4LF66 | O-acetylhomoserine aminocarboxypropyltransferase OS | Cytoplasmic | 1.26E-09 | 5.99E-11 | 4.49E-10 | 3.12E-11 | 1.488636384 | Amino acid transport and metabolism | E | COG2873 |
| A0A0M4M2M9 | Flagellar motor protein MotB OS | Unknown | 4.99E-09 | 3.32E-10 | 1.80E-09 | 3.41E-10 | 1.471042909 | Cell wall, membrane, envelope biogenesis | M | COG2885 |
| A0A0M4M1X8 | 50S ribosomal protein L7/L12 OS | Membrane | 1.26E-08 | 4.96E-10 | 4.60E-09 | 3.57E-10 | 1.453717967 | Translation, ribosomal structure and biogenesis | J | COG0222 |
| A0A0M3T1R7 | Rod shape-determining protein MreB OS | Periplasmic | 1.21E-09 | 1.24E-10 | 4.49E-10 | 3.19E-10 | 1.430219697 | Cell cycle control, cell division, chromosome partitioning | D | COG1077 |
| A0A0M4LZJ5 | Glycine–tRNA ligase beta subunit OS | Cytoplasmic | 1.21E-09 | 1.24E-10 | 4.49E-10 | 3.12E-11 | 1.430219697 | Translation, ribosomal structure and biogenesis | J | COG0751 |
| A0A0M4M063 | Electron transfer flavoprotein subunit beta OS | Unknown | 2.36E-09 | 2.47E-10 | 8.98E-10 | 6.25E-11 | 1.39399951 | Energy production and conversion | C | COG2086 |
| A0A0M4L4E1 | Phospho-2-dehydro-3-deoxyheptonate aldolase OS | Unknown | 5.91E-10 | 4.83E-11 | 2.25E-10 | 1.56E-11 | 1.393233129 | Amino acid transport and metabolism | E | COG0722 |
| A0A0M4LZC6 | 30S ribosomal protein S15 OS | Membrane | 5.91E-10 | 1.95E-10 | 2.25E-10 | 1.56E-11 | 1.393233129 | Translation, ribosomal structure and biogenesis | J | COG0184 |
| A0A0M4LFR5 | Lysine–tRNA ligase OS | Cytoplasmic | 8.74E-10 | 1.21E-10 | 3.37E-10 | 1.61E-10 | 1.374884688 | Translation, ribosomal structure and biogenesis | J | COG1190 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4LG31 | 50S ribosomal protein L1 OS | Cytoplasmic | 3.47E-09 | 2.99E-10 | 1.35E-09 | 6.42E-10 | 1.361976256 | Translation, ribosomal structure and biogenesis | J | COG0081 |
| A0A0M4M0C3 | DNA-binding protein OS | Membrane | 1.02E-08 | 5.91E-10 | 4.04E-09 | 4.24E-10 | 1.336141954 | Cell wall, membrane, envelope biogenesis | M | COG1744 |
| A0A0M4LZI6 | Adenosylhomocysteinase OS | Unknown | 1.13E-09 | 9.60E-11 | 4.49E-10 | 3.12E-11 | 1.331535423 | Coenzyme transport and metabolism | H | COG0499 |
| A0A0M4LFQ5 | 50S ribosomal protein L15 OS | Cytoplasmic | 1.67E-09 | 6.93E-11 | 6.74E-10 | 3.21E-10 | 1.309027606 | Translation, ribosomal structure and biogenesis | J | COG0200 |
| A0A0M4M0S5 | GTP cyclohydrolase FolE2 OS | Membrane | 1.11E-09 | 2.73E-10 | 4.49E-10 | 3.12E-11 | 1.305772326 | Coenzyme transport and metabolism | H | COG1469 |
| A0A0M4LZX7 | Glutamine synthetase OS | Membrane | 1.01E-08 | 4.97E-10 | 4.15E-09 | 5.57E-10 | 1.283172051 | Amino acid transport and metabolism | E | COG0174 |
| A0A0M3T229 | Threonine–tRNA ligase OS | Membrane | 1.36E-09 | 1.99E-10 | 5.61E-10 | 1.64E-10 | 1.277533976 | Translation, ribosomal structure and biogenesis | J | COG0441 |
| A0A0M3T1Z1 | 60 kDa chaperonin OS | Cytoplasmic | 1.14E-08 | 3.71E-10 | 4.72E-09 | 1.62E-09 | 1.27217506 | Posttranslational modification, protein turnover, chaperones | O | COG0459 |
| A0A0M3T1N6 | ABC transporter substrate-binding protein OS | Periplasmic | 9.20E-09 | 4.86E-10 | 3.82E-09 | 9.89E-10 | 1.268061223 | Amino acid transport and metabolism | E | COG2113 |
| A0A0M5KWH5 | 2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase OS | Cytoplasmic | 1.08E-09 | 1.58E-10 | 4.49E-10 | 3.12E-11 | 1.266243962 | Amino acid transport and metabolism | E | COG2171 |
| A0A0M5KW88 | Serine acetyltransferase OS | Cytoplasmic | 5.40E-10 | 7.90E-11 | 2.25E-10 | 1.56E-11 | 1.263034406 | Amino acid transport and metabolism | E | COG1045 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M5KYN3 | Aspartyl/glutamyl-tRNA(Asn/Gln) amidotransferase subunit C OS | Cytoplasmic | 5.40E-10 | 7.90E-11 | 2.25E-10 | 1.56E-11 | 1.263034406 | Translation, ribosomal structure and biogenesis | J | COG0721 |
| A0A0M4LR98 | Uncharacterized protein OS | OuterMembrane | 1.85E-09 | 2.12E-10 | 7.86E-10 | 1.68E-10 | 1.234924053 | | | |
| A0A0M4LF20 | Phosphomethylpyrimidine synthase OS | Unknown | 1.85E-09 | 1.65E-10 | 7.86E-10 | 1.68E-10 | 1.234924053 | Coenzyme transport and metabolism | H | COG0422 |
| A0A0M4LI96 | Uncharacterized protein OS | Cytoplasmic | 2.90E-09 | 2.01E-10 | 1.24E-09 | 1.81E-10 | 1.22571278 | Coenzyme transport and metabolism | H | COG4206 |
| A0A0M3T1K4 | Ribose-phosphate pyrophosphokinase OS | Cytoplasmic | 1.05E-09 | 1.97E-10 | 4.49E-10 | 3.19E-10 | 1.225601978 | Nucleotide transport and metabolism | F | COG0462 |
| A0A0M3T276 | 50S ribosomal protein L27 OS | Cytoplasmic | 1.31E-09 | 1.60E-10 | 5.61E-10 | 1.64E-10 | 1.223494136 | Translation, ribosomal structure and biogenesis | J | COG0211 |
| A0A0M4L3W9 | Amino acid ABC transporter substrate-binding protein OS | Membrane | 5.76E-09 | 3.08E-10 | 2.47E-09 | 3.61E-10 | 1.22155777 | Amino acid transport and metabolism | E | COG0834 |
| A0A0M4LEW9 | Putrescine-binding periplasmic protein OS | Cytoplasmic | 1.57E-09 | 6.68E-11 | 6.74E-10 | 4.69E-11 | 1.219944063 | Amino acid transport and metabolism | E | COG0687 |
| A0A0M4L4Y3 | Cold-shock protein OS | Cytoplasmic | 2.96E-08 | 1.02E-09 | 1.28E-08 | 3.30E-09 | 1.209453366 | Transcription | K | COG1278 |
| A0A0M4L4K3 | Putative K(+)-stimulated pyrophosphate-energized sodium pump OS | Unknown | 1.28E-09 | 9.79E-11 | 5.61E-10 | 1.64E-10 | 1.190071134 | Energy production and conversion | C | COG3808 |
| A0A0M5KWA0 | Membrane protein OS | Membrane | 3.70E-09 | 4.96E-10 | 1.68E-09 | 1.97E-10 | 1.139064038 | Cell wall, membrane, envelope biogenesis | M | COG1744 |
| A0A0M3T2I3 | Uncharacterized protein OS | Cytoplasmic | 9.77E-10 | 5.43E-11 | 4.49E-10 | 3.12E-11 | 1.121643117 | | | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4L4I1 | 50S ribosomal protein L9 OS | Membrane | 5.09E-09 | 2.82E-10 | 2.36E-09 | 2.28E-10 | 1.108878797 | Translation, ribosomal structure and biogenesis | J | COG0359 |
| A0A0M4LPG0 | Transcription termination factor Rho OS | Cytoplasmic | 2.65E-09 | 2.11E-10 | 1.24E-09 | 1.81E-10 | 1.095652239 | Transcription | K | COG1158 |
| A0A0M4LFV6 | DNA-binding protein OS | Cytoplasmic | 7.20E-10 | 1.20E-10 | 3.37E-10 | 1.61E-10 | 1.095248315 | Transcription | K | COG1278 |
| A0A0M5KYA8 | Lon protease OS | Unknown | 1.90E-09 | 1.32E-10 | 8.98E-10 | 6.25E-11 | 1.081212068 | Posttranslational modification, protein turnover, chaperones | O | COG0466 |
| A0A0M4LHK8 | Spermidine/putrescine ABC transporter substrate-binding protein OS | Cytoplasmic | 3.84E-08 | 2.81E-09 | 1.82E-08 | 5.55E-09 | 1.077167861 | Amino acid transport and metabolism | E | COG0687 |
| A0A0M4LFX9 | Ribosome-recycling factor OS | Cytoplasmic | 1.18E-09 | 2.39E-10 | 5.61E-10 | 1.64E-10 | 1.072714184 | Translation, ribosomal structure and biogenesis | J | COG0233 |
| A0A0M4L3Z2 | Membrane protein OS | Cytoplasmic | 3.49E-09 | 3.81E-10 | 1.68E-09 | 4.91E-10 | 1.054765803 | Posttranslational modification, protein turnover, chaperones | O | COG0330 |
| A0A0M4L3W0 | 2-isopropylmalate synthase OS | Cytoplasmic | 9.25E-10 | 1.37E-10 | 4.49E-10 | 3.12E-11 | 1.042737921 | Amino acid transport and metabolism | E | COG0119 |
| A0A0M5KT98 | Formate–tetrahydrofolate ligase OS | Cytoplasmic | 3.67E-09 | 4.05E-10 | 1.80E-09 | 1.25E-10 | 1.027783157 | Nucleotide transport and metabolism | F | COG2759 |
| A0A0M4LCG9 | Glycine/betaine ABC transporter OS | Membrane | 2.49E-09 | 1.79E-10 | 1.24E-09 | 1.81E-10 | 1.005805622 | Amino acid transport and metabolism | E | COG4175 |
| A0A0M3T254 | ABC transporter substrate-binding protein OS | Cytoplasmic | 2.93E-09 | 1.21E-10 | 1.46E-09 | 8.00E-10 | 1.004932296 | Amino acid transport and metabolism | E | COG2113 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4LGH5 | Fructose-bisphosphate aldolase OS | Cytoplasmic | 3.37E-09 | 1.93E-10 | 1.68E-09 | 1.12E-09 | 1.004287358 | Carbohydrate transport and metabolism | G | COG3588 |

Table B.4: Proteins with significantly higher expression under the nutrient-deplete treatment compared to the nutrient-replete treatment. Expression was quantified as the average normalized spectral counts per cell. Fold change was calculated as $\log_2$(Deplete expression/Replete expression). A cutoff $\log_2$-fold change was set at 0.5.

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4M2R8 | C4-dicarboxylate ABC transporter substrate-binding protein OS | Cytoplasmic | 2.31E-10 | 7.74E-11 | 4.15E-09 | 5.57E-10 | 4.1671 | Energy production and conversion | C | COG3181 |
| A0A0M5KYJ2 | Succinate-semialdehyde dehydrogenase OS | Unknown | 3.08E-10 | 9.80E-12 | 8.98E-10 | 3.24E-10 | 1.5438 | Lipid transport and metabolism | I | COG0318 |
| A0A0M4M190 | Uncharacterized protein OS | Cytoplasmic | 1.52E-09 | 1.68E-10 | 3.14E-09 | 2.19E-10 | 1.0467 | Unknown | | |
| A0A0M4LFL2 | Iron-binding protein IscA OS | Unknown | 2.31E-10 | 7.74E-11 | 4.49E-10 | 3.12E-11 | 0.9588 | Posttranslational modification, protein turnover, chaperones | O | COG0316 |
| A0A0M5KRH3 | Lipid-A-disaccharide synthase OS | Cytoplasmic | 1.80E-10 | 4.49E-11 | 3.37E-10 | 1.61E-10 | 0.9048 | Cell wall, membrane, envelope biogenesis | M | COG0763 |
| A0A0M4LFP2 | Sugar ABC transporter substrate-binding protein OS | Unknown | 1.05E-09 | 1.22E-10 | 1.91E-09 | 4.95E-10 | 0.8632 | Carbohydrate transport and metabolism | G | COG1653 |
| A0A0M5KRF8 | Citrate synthase OS | Membrane | 4.63E-10 | 1.47E-11 | 7.86E-10 | 1.68E-10 | 0.7635 | Energy production and conversion | C | COG0372 |
| A0A0M4M0G8 | Glycine/betaine ABC transporter substrate-binding protein OS | Cytoplasmic | 2.44E-09 | 2.48E-10 | 4.04E-09 | 1.61E-09 | 0.7275 | Amino acid transport and metabolism | E | COG2113 |
| A0A0M5KRW2 | GXGXG motif-containing protein OS | Cytoplasmic | 3.60E-10 | 4.60E-11 | 5.61E-10 | 1.64E-10 | 0.6400 | Amino acid transport and metabolism | E | COG0070 |
| A0A0M4LFZ7 | Glucose-1-phosphate thymidylyltransferase OS | Membrane | 3.60E-10 | 4.60E-11 | 5.61E-10 | 1.64E-10 | 0.6400 | Cell wall, membrane, envelope biogenesis | M | COG1209 |

Continued on next page.

(SI Table B.4: Continued from previous page)

| Protein ID | Protein name | Localization | Replete expression | Replete error | Deplete expression | Deplete error | Fold change | COG Category | COG code | COG ID |
|---|---|---|---|---|---|---|---|---|---|---|
| A0A0M4LFR1 | Acetyl-CoA acetyltransferase OS | Unknown | 5.14E-10 | 4.74E-11 | 7.86E-10 | 1.68E-10 | 0.6128 | Lipid transport and metabolism | I | COG0183 |
| A0A0M4LGJ2 | Exodeoxyribonuclease 7 small subunit OS | Membrane | 2.31E-10 | 7.35E-12 | 3.37E-10 | 1.61E-10 | 0.5449 | Replication, recombination and repair | L | COG1722 |
| A0A0M4M166 | Gamma carboxymuconolactone decarboxylase OS | Cytoplasmic | 5.40E-10 | 1.72E-11 | 7.86E-10 | 1.68E-10 | 0.5416 | General function prediction only | R | COG0599 |
| A0A0M4LPJ7 | Isocitrate lyase OS | Unknown | 1.16E-09 | 8.54E-11 | 1.68E-09 | 8.03E-10 | 0.5343 | Energy production and conversion | C | COG2224 |
| A0A0M4L4J9 | Acyl-CoA dehydrogenase OS | Cytoplasmic | 8.74E-10 | 5.25E-11 | 1.24E-09 | 1.81E-10 | 0.5046 | Lipid transport and metabolism | I | COG1960 |

# Appendix C

# SUPPLEMENTAL INFORMATION TO CHAPTER 3

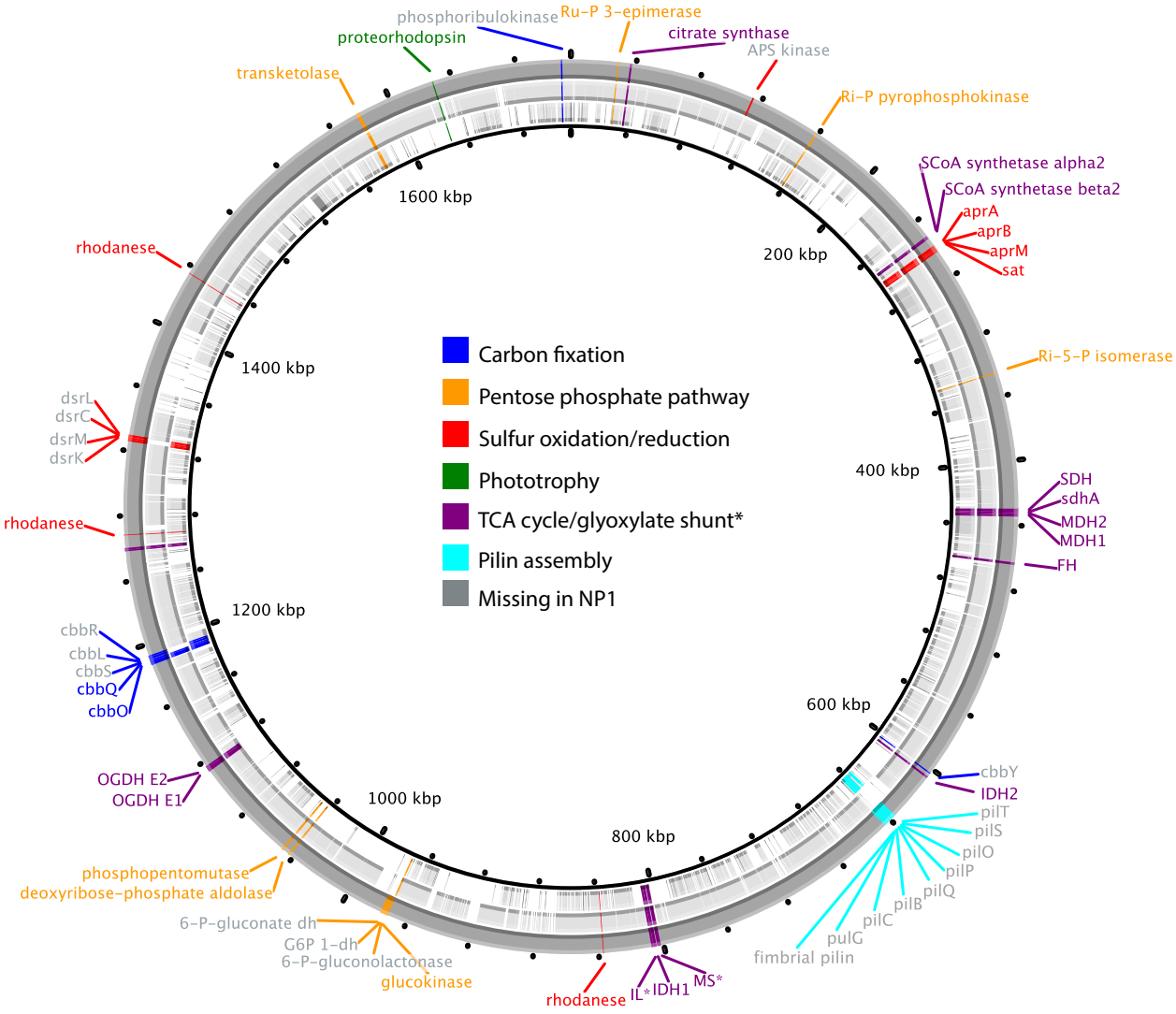


Figure C.1: Genomic comparisons of key metabolic genes in *Thioglobus sp.* Rings from innermost to outermost: *Ca.*T. singularis strain NP1, *Ca.* T. autorophicus strain EF1, and *Ca.* T. singularis strain PS1. Key genes are classified by pathways: carbon fixation (blue), pentose phosphate pathway (orange), sulfur oxidation/reduction (red), phototrophy (green), TCA cycle and gloxylate bypass (purple) and genes unique to glyoxylate bypass denoted with *, pilin assembly (light blue). Genes that were absent from the *Ca.* T. singularis strain NP1 genome are in shown in grey text.

Table C.1: Comparison of COG categories across three complete *Thioglobus sp.* genomes.

| | Strain NP1 | | Strain PS1 | | Strain EF1 | | |
|---|---|---|---|---|---|---|---|
| Code | Value | %age | Value | %age | Value | %age | Description |
| J | 160 | 10.62 | 165 | 10.64 | 166 | 12.66 | Translation, ribosomal structure and biogenesis |
| A | 1 | 0.07 | 1 | 0.06 | 1 | 0.08 | RNA processing and modification |
| K | 42 | 2.79 | 51 | 3.29 | 40 | 3.05 | Transcription |
| L | 64 | 4.25 | 65 | 4.19 | 72 | 5.49 | Replication, recombination and repair |
| B | 1 | 0.07 | 1 | 0.06 | 0 | 0 | Chromatin structure and dynamics |
| D | 15 | 1 | 15 | 0.97 | 17 | 1.3 | Cell cycle control and division, chromosome partitioning |
| V | 24 | 1.59 | 20 | 1.29 | 23 | 1.75 | Defense mechanisms |
| T | 23 | 1.53 | 27 | 1.74 | 37 | 2.82 | Signal transduction mechanisms |
| M | 89 | 5.91 | 87 | 5.61 | 109 | 8.31 | Cell wall/membrane biogenesis |
| N | 6 | 0.4 | 14 | 0.9 | 12 | 0.92 | Cell motility |
| U | 17 | 1.13 | 20 | 1.29 | 19 | 1.45 | Intracellular trafficking and secretion |
| O | 74 | 4.91 | 75 | 4.84 | 92 | 7.02 | Posttranslational modification, protein turnover, chaperones |
| C | 111 | 7.37 | 116 | 7.48 | 114 | 8.7 | Energy production and conversion |
| G | 86 | 5.71 | 100 | 6.45 | 41 | 3.13 | Carbohydrate transport and metabolism |
| E | 238 | 15.79 | 242 | 15.6 | 121 | 9.23 | Amino acid transport and metabolism |
| F | 67 | 4.45 | 64 | 4.13 | 47 | 3.59 | Nucleotide transport and metabolism |
| H | 119 | 7.9 | 121 | 7.8 | 110 | 8.39 | Coenzyme transport and metabolism |
| I | 85 | 5.64 | 84 | 5.42 | 46 | 3.51 | Lipid transport and metabolism |
| P | 63 | 4.18 | 58 | 3.74 | 72 | 5.49 | Inorganic ion transport and metabolism |
| Q | 37 | 2.46 | 39 | 2.51 | 15 | 1.14 | Secondary metabolites biosynthesis, transport, catabolism |
| R | 134 | 8.89 | 122 | 7.87 | 81 | 6.18 | General function prediction only |
| S | 46 | 3.05 | 52 | 3.35 | 48 | 3.66 | Function unknown |
| - | 389 | 22.38 | 379 | 21.45 | 458 | 27.59 | Not in COGs |

Table C.2: Comparison of genome statistics across three complete *Thioglobus sp.* genomes.

| | Strain NP1 | | Strain PS1 | | Strain EF1 | |
|---|---|---|---|---|---|---|
| Attribute | Value | %age | Value | %age | Value | %age |
| Genome size (bp) | 1,689,404 | 100 | 1,714,148 | 100 | 1,512,449 | 100 |
| DNA coding (bp) | 1,596,072 | 94 | 1,621,248 | 95 | 1,420,141 | 94 |
| DNA G+C (bp) | 591,027 | 35 | 641,748 | 37 | 592,030 | 39 |
| DNA scaffolds | 1 | 100 | 1 | 100 | 1 | 100 |
| Total genes | 1738 | 100 | 1767 | 100 | 1660 | 100 |
| Protein coding genes | 1692 | 97 | 1719 | 97 | 1617 | 97 |
| RNA genes | 46 | 3 | 48 | 3 | 43 | 3 |
| Pseudo genes | 0 | 0 | 0 | 0 | 0 | 0 |
| Genes in internal clusters | 220 | 13 | 246 | 14 | 187 | 11 |
| Genes with function prediction | 1470 | 85 | 1492 | 84 | 1318 | 79 |
| Genes assigned to COGs | 1349 | 78 | 1388 | 79 | 1202 | 72 |
| Genes with Pfam domains | 1516 | 87 | 1564 | 89 | 1387 | 84 |
| Genes with signal peptides | 104 | 6 | NA | NA | 147 | 9 |
| Genes with transmembrane helices | 359 | 21 | NA | NA | 317 | 19 |
| CRISPR repeats | 0 | 0 | 0 | 0 | 0 | 0 |

Table C.3: Genomic content of key genes across three complete (strains PS1, NP1, and EF1) and six incomplete *Thioglobus sp.* genomes.

| | | | *Arctic96BD-19* | | | | | | *Sup05* |
|---|---|---|---|---|---|---|---|---|---|
| | Surface | | N. Atlantic (Boothbay) Surface | | | | S. Atlantic 800m | | Oxycline |
| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
| ***Carbon-fixation*** | | | | | | | | | |
| RuBisCo, small subunit(cbbS) | W908_06140 | | X | X | X | | X | X | SP60_06730 |
| RuBisCo, large subunit (cbbL) | W908_06145 | | X | X | X | X | X | X | SP60_06725 |
| | | | | | | | | | SP60_06705 |
| Rubisco operon transcriptional regulator (cbbR) | W908_06150 | | X | X | X | X | X | X | SP60_06700 |
| cbbQ | W908_06135 | CRN91_04215 | X | X | X | X | X | X | SP60_06710 |
| cbbO | W908_06130 | CRN91_04220 | X | X | X | X | X | X | SP60_06715 |
| cbbY | W908_03090 | | X | X | | X | | | SP60_00755 |
| Carbonic anhydrase | | | | | | | | | |
| phosphoglycerate kinase (cbbK) | W908_08030 | CRN91_02110 | X | X | X | X | X | X | SP60_04470 |
| glyceraldehyde-3-phosphate dehydrogenase (cbbG) | W908_08035 | CRN91_02105 | X | X | X | X | X | X | SP60_04460 |
| phosphoglycolate phosphatase | W908_00810 | CRN91_00580 | | X | | X | X | X | SP60_04780 |
| phosphoribulokinase (cbbP) | W908_08755 | | X | X | | X | X | X | SP60_04035 |
| ***Glycolysis and Gluconeogenesis*** | | | | | | | | | |
| glucokinase | W908_05045 | CRN91_05435 | X | X | X | X | | X | |
| glucose-6-phosphate isomerase | | | | | | | | | SP60_01990 |
| 6-phosphofructokinase | W908_00670 | CRN91_00725 | | X | X | X | X | X | SP60_04625 |
| fructose-bisphosphatase | | | | | | | | | |
| fructose-bisphosphate aldolase | W908_01035 | CRN91_02120 | X | X | X | X | X | X | |
| | W908_08020 | | | | | | | | |
| triosephosphate isomerase | W908_00020 | CRN91_01370 | X | X | | X | X | X | SP60_04005 |
| glyceraldehyde-3-phosphate dehydrogenase | W908_08035 | CRN91_02105 | X | X | X | X | X | X | SP60_04460 |
| phosphoglycerate kinase | W908_08030 | CRN91_02110 | X | X | X | X | X | X | SP60_04470 |
| phosphoglyceromutase | W908_00140 | CRN91_01245 | X | X | | X | X | X | SP60_03905 |
| phosphoglucomutase | W908_03710 | CRN91_06625 | X | X | X | X | X | X | SP60_00450 |
| enolase | W908_06245 | CRN91_04115 | X | X | X | X | | X | SP60_01225 |
| pyruvate kinase | W908_08025 | CRN91_02115 | X | X | X | X | X | X | SP60_04475 |
| phosphoenolpyruvate synthase | W908_08045 | CRN91_02095 | X | X | X | X | X | X | SP60_05945 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| phosphoenolpyruvate carboxykinase | W908_01120 | CRN91_00420 | X | X | | X | | | |
| aldose 1-epimerase | W908_00480 | | X | X | X | X | | | |
| Glucose-6-phosphate 1-dehydrogenase | W908_05060 | | | | | | | | |
| ***TCA Cycle*** | | | | | | | | | |
| citrate synthase | W908_00175 | CRN91_01210 | X | X | | X | | X | SP60_03870 |
| 2-oxoglutarate dehydrogenase E1 | W908_05700 | CRN91_04700 | *X | | *X | | | | |
| 2-oxoglutarate dehydrogenase E2 (EC 2.3.1.61) | W908_05705 | CRN91_04695 | X | X | X | X | | X | |
| succinyl-CoA synthetase subunit alpha | W908_01220 W908_06420 | CRN91_00330 | X | X | X | X | X | X | SP60_05115 |
| succinyl-CoA synthetase subunit beta | W908_01225 W908_06425 | CRN91_00325 | X | X | X | X | X | X | SP60_05120 |
| succinate dehydrogenase | W908_02195 | CRN91_07915 | | | | | | | |
| succinate dehydrogenase sdhA | W908_02190 | CRN91_07920 | X | X | | X | | | SP60_05960 |
| fumarate hydratase | W908_02330 | CRN91_07780 | X | X | | X | | | SP60_05965 |
| malate dehydrogenase | W908_02200 | CRN91_07910 | X | X | | X | | | SP60_05970 |
| | W908_02205 | CRN91_07905 | | | | | | | |
| isocitrate dehydrogenase | W908_03105 | CRN91_07050 | X | X | | X | | X | SP60_00740 |
| | W908_04115 | CRN91_06190 | | | | | | | SP60_01900 |
| dihydrolipoamide dehydrogenase | W908_07500 | CRN91_02690 | X | X | X | X | | X | SP60_03105 |
| ***Glyoxylate shunt*** | | | | | | | | | |
| malate synthase | W908_04105 | CRN91_06200 | | | | | | | SP60_01890 |
| isocitrate lyase | W908_04120 | CRN91_06185 | | | | | | | SP60_01905 |
| ***Pentose phosphate pathway*** | | | | | | | | | |
| glucose-6-phosphate 1-dehydrogenase | W908_05060 | | X | | X | X | | | |
| 6-phosphogluconolactonase | W908_05055 | | X | | X | X | | | |
| 6-phosphogluconate dehydrogenase | W908_05065 | | X | | X | X | X | | |
| ribose 5-phosphate isomerase | W908_01760 | CRN91_08335 | X | X | | X | | | SP60_05560 |
| ribulose-phosphate 3-epimerase | W908_00135 | CRN91_01250 | X | X | | X | | X | SP60_03910 |
| transketolase | W908_08040 | CRN91_02100 | X | X | X | X | X | X | SP60_04455 |
| ribose-phosphate pyrophosphokinase | W908_00765 | CRN91_00625 | X | X | * | X | X | X | SP60_04735 |
| phosphopentomutase | W908_05390 | CRN91_05090 | X | X | X | X | X | X | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| deoxyribose-phosphate aldolase | W908_05375 | CRN91_05105 | X | X | X | X | | | |
| gluconate kinase (glucokinase) | W908_05045 | CRN91_05435 | | | | | X | | |
| ***THF-linked oxidation*** | | | | | | | | | |
| formate dehydrogenase, alpha unit (fdhA) | W908_08055 | CRN91_02085 | X | X | X | X | X | X | |
| NAD-dependent formate dehydrogenase, beta subunit (fdhB)? | W908_08060 | CRN91_02080 | X | X | X | X | X | X | |
| formate dehydrogenase, delta subunit (fdhD) | W908_08065 | CRN91_02075 | X | X | X | X | X | X | |
| molybdenum cofactor guanylyltransferase (mobA) | W908_06080 | CRN91_04260 | | | | | | | SP60_06820 |
| molybdopterin molybdochelase (moeA) | W908_06085 | CRN91_04255 | X | | X | X | | X | SP60_06815 |
| formate tetrahydrofolate ligase (fhs) | W908_08070 | CRN91_02070 | X | X | X | X | X | X | |
| methylenetetrahydrofolate reductase (metF) | W908_08095 | CRN91_02045 | X | X | X | X | XX | XX | SP60_01770 |
| methylenetetrahydrofolate dehydrogenase (folD) | W908_06730 | CRN91_03655 | X | XX | XX | XX | X | | SP60_01710 |
| ***Methylamine oxidation*** | | | | | | | | | |
| glutamate-putrescine ligase | W908_02485 | CRN91_07640 | X | X | | X | | X | |
| trimethylamine monooxygenase (Tmm) | W908_02175 | CRN91_07935 | X | X | | | | | |
| trimethylamine N-oxide demethylase (Tdm) | W908_05660 | CRN91_04740 | X | | X | | | | |
| dimethylamine monooxygenase (Dmm) | | | | | | | | | |
| gamma-glutamylmethylamide synthetase (GmaS) | W908_01950 | CRN91_08145 | X | X | | X | | X | |
| N-methylglutamate synthase (MgsA) or glutamate synthase, GltB1 subunit (glxB) | W908_01945 | CRN91_08150 | X | X | | X | | | |
| N-methylglutamate synthase (MgsB) or glutamate synthase, GltB3 subunit (glxC) | W908_01940 | CRN91_08155 | X | X | | X | | | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| N-methylglutamate synthase (MgsC) or glutamate synthase, GltB2 subunit (glxD) | W908_01935 | CRN91_08160 | X | X | | X | | | |
| N-methylglutamate dehydrogenase (MgdA) or sarcosine oxidase, beta subunit (soxB2) | W908_05620 | CRN91_04775 | X | X | X | X | | | |
| N-methylglutamate dehydrogenase (MgdB) or sarcosine oxidase, delta subunit (soxD2) | W908_05615 | CRN91_04780 | X | X | X | X | | | |
| N-methylglutamate dehydrogenase (MgdC) or sarcosine oxidase, alpha subunit (soxA2) | W908_05610 | CRN91_04785 | X | X | X | X | | | |
| N-methylglutamate dehydrogenase (MgdD) or sarcosine oxidase, gamma subunit (soxG2) | W908_05605 | CRN91_04790 | X | X | X | X | | | |
| ***Choline/betaine/glycine transporters*** | | | | | | | | | |
| glycine betaine/l-proline transport ATP-binding protein (opuAA) | X | | X | X | X | X | | | |
| glycine betaine/L-proline transport system permease (opuAB) | X | | ? | ? | | X | | | |
| substrate-binding region of ABC-type glycine betaine transport system (opuAC) | X | | X | X | X | X | | X | |
| ATP-binding cassette domain of proline/glycine betaine uptake system (opuBA) | X | | X | X | X | X | | X | |
| glycine betaine/L-proline transport ATP-binding protein (proV) | XX | | XX | XX | | XX | X | XX | |
| glycine betaine/L-proline transport system permease protein (proW) | XX | | XX | XX | | XX | | XX | |
| ABC-type proline/glycine betaine transport systems (proX) | XX | | XX | XX | | XX | X | XX | |
| ***Choline/betaine/glycine metabolism*** | | | | | | | | | |
| choline dehydrogenase (betA) | XXXX | | XXXX | XXXX | X | XXXX | | XX | |
| betaine aldehyde dehydrogenase (betB or dhaS) | X | | X | X | | X | | | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| transcriptional regulator (betI) | | | | | | | | | |
| betaine-homocysteine S-methyltransferase (bhmT) | | | | | | | | | |
| dimethylglycine dehydrogenase (dmgdh) | X | | X | X | | X | X | X | |
| sarcosine dehydrogenase (sardh) | XX | | XXX | XXX | XX | XXX | XX | XX | |
| sarcosine oxidase beta subunit (soxB1) | X | | X | X | X | X | X | X | |
| Sarcosine oxidase delta subunit (soxD1) | X | | X | X | X | X | X | X | |
| Sarcosine oxidase alpha subunit (soxA1) | X | | X | XX | XX | X | X | X | |
| Sarcosine oxidase gamma subunit (soxG1) | X | | X | X | X | X | | | |
| Monomeric sarcosine oxidase | X | | X | X | X | X | | | |
| glcyine system cleavage protein T2 (gcvT) | XXXXX | | XXXX | XXXXX | XXX | XXXXX | XX | XX | X |
| glycine cleavage H protein (gcvH) | X | | X | X | X | X | X | X | X |
| glycine cleavage system protein P2, glycine dehydrogenase (gcvP) | X | | | X | X | X | | | |
| glycine cleavage system transcriptional repressor (gcvR) | | | | | | | | | |
| glycine cleavage system transcriptional activator | | | | | | | | | |
| glycine cleavage operon activator | | | | | | | | | |
| glycine cleavage system regulatory protein? | X | | | X | X | X | X | X | |
| ***Sulfur oxidation*** | | | | | | | | | incomplete |
| aprM | W908_01280 | CRN91_00265 | X | X | | X | X | X | SP60_04580 |
| aprA | W908_01270 | CRN91_00275 | X | X | | X | X | X | SP60_04590 |
| aprB | W908_01275 | CRN91_00270 | X | X | | X | X | X | SP60_04585 |
| sat | W908_01285 | CRN91_00260 | X | X | | X | X | X | SP60_04575 |
| APS kinase (assimilatory sulfate reduction) | W908_00550 | | | | | | | | |
| dsrA | | | | | | | | | SP60_01870 |
| dsrB | | | | | | | | | SP60_01865 |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| dsrC | W908_06825 | | X | X | X | X | | | SP60_03695 |
| | | | | | | | | | SP60_07285 |
| | | | | | | | | | SP60_01845 |
| dsrE | | | | | | | | | SP60_01855 |
| dsrF | | | | | | | | | SP60_01860 |
| dsrH | | | | | | | | | SP60_01850 |
| dsrK | W908_06815 | | X | X | X | X | | | SP60_01835 |
| dsrL/glutamate synthase | W908_06830 | | X | X | X | X | | | SP60_01830 |
| dsrM/narG | W908_06820 | | X | X | X | X | | | SP60_01840 |
| dsrR | | | | | | | | | SP60_01805 |
| rhodanese | W908_04280 | CRN91_02650 | X | | X | | X | | SP60_01445 |
| rhodanese | W908_07340 | CRN91_02905 | X | | X | X | | | SP60_02965 |
| rhodanese | W908_06490 | CRN91_03880 | X | | X | X | | | SP60_03470 |
| rohdanese | | CRN91_06025 | | | | | | | SP60_03535 |
| soxA | | | | | | | | | SP60_06060 |
| soxB | | | | | | | | | SP60_05195 |
| soxX | | | | | | | | | SP60_06045 |
| soxY | | | | | | | | | SP60_06050 |
| soxZ | | | | | | | | | SP60_06055 |
| sulfite reductase | W908_01190 | | X | X | | X | X | | |
| ***Denitrification*** | | | | | | | | | |
| narG | | | | | | | | | SP60_06775 |
| narH | | | | | | | | | SP60_06780 |
| narJ | | | | | | | | | SP60_06785 |
| narI | | | | | | | | | SP60_06790 |
| narQ | | | | | | | | | ? |
| narL | | | | | | | | | SP60_03540 |
| napB | | | | | | | | | ? |
| napA | | | | | | | | | SP60_06890 |
| napH | | | | | | | | | SP60_06885 |
| napG | | | | | | | | | ? |
| napD | | | | | | | | | ? |
| norD | | | | | | | | | SP60_04225 |
| norBC | | | | | | | | | SP60_04200 |
| ***Phototrophy*** | | | | | | | | | |
| proteorhodopsin | W908_08270 | CRN91_01865 | X | | X | X | | | |
| ***Oxidative phosphorylation*** | | | | | | | | | |

Continued on next page.

(SI Table B.3: Continued from previous page)

| Gene name | Strain PS1 | Strain NP1 | AAA076-D13 | AAA076-D02 | AAA076-E13 | AAA076-F14 | AAA001-B15 | AAA007-O20 | Strain EF1 |
|---|---|---|---|---|---|---|---|---|---|
| ubiquinol-cytochrome C reductase | W908_08725 | CRN91_01425 | X | X | | X | | X | SP60_04065 |
| cytochrome B | W908_08720 | CRN91_01430 | X | X | | X | | X | SP60_04070 |
| ubiquinol-cytochrome C reductase | W908_08715 | CRN91_01435 | X | X | | X | | X | SP60_04075 |
| cytochrome oxidase | | | | | | | | | SP60_04175 |
| cytochrome oxidase subunit II (coxB or B559 subunit alpha) | W908_08615 | CRN91_01535 | X | X | | X | | | SP60_04180 |
| cytochrome C | W908_01610 | CRN91_08550 | X | X | | X | | X | SP60_04195 |
| cytochrome C oxidase | | | | | | | | | SP60_04205 |
| cytochrome C (c4) | W908_01615 | CRN91_08545 | X | X | | X | | X | SP60_05430 |
| cytochrome C (c5) | W908_01795 | CRN91_08295 | X | X | | X | | | SP60_05575 |
| cytochrome cbb3 | W908_05785 | CRN91_04670 | X | X | X | | | | |
| cytochrome soxA | | | | | | | | | SP60_06060 |
| sdhA | W908_02190 | CRN91_07920 | X | X | | X | | | SP60_05960 |
| cytochrome C4 | W908_06015 | CRN91_04320 | X | X | X | X | | | |
| cytochrome oxidase subunit I (ctaD) | W908_08610 | CRN91_01540 | X | X | | X | | | |
| cytochrome C oxidase subunit III (MFS transporter?) | W908_08595 | CRN91_01550 | X | X | | X | X | | |
| cytochrome B | | CRN91_04610 | X | X | | X | | | |
| ***Nitrogen assimilation/regulation*** | | | | | | | | | |
| Nitrogen regulatory protein P-II (1) | | | | | | | | | SP60_05495 |
| Nitrogen regulatory protein P-II (2) | W908_07670 | | | | X | X | | X | SP60_05800 |
| Nitrogen regulatory protein P-II (3) | W908_02110 | CRN91_05415 | X | X | | X | | X | SP60_05830 |
| Ammonium transporter (1) | W908_05715 | CRN91_04685 | X | | X | | | | |
| Ammonium transporter (2) | W908_02115 | CRN91_05420 | X | X | X | X | | X | SP60_05835 |

# Appendix D

# SUPPLEMENTAL INFORMATION TO CHAPTER 4

## *D.1 Supplemental methods*

### *D.1.1 Sample collection*

Hydrothermal plumes within the caldera of Axial volcano and over the new Apprill et al. (2015) North Rift Zone lava flows were mapped during the rapid response cruise (TN-327, August 14-29, 2015) using a Seabird 9plus CTD with integrated turbidity (optical backscatter) and oxidation-reduction potential (ORP) sensors. During tow-yo's, the CTD package (which also included 24 10-L Niskin bottles) was cycled between 10-200 m above the seafloor while towing along the trackline at speeds of 1-2 knots Baker et al. (1995). Turbidity is reported here as non-dimensional Nephelometric Turbidity Units (NTU; American Public Health Association, 1985), and NTU is the turbidity anomaly above the local non-plume background value. The ORP sensor responds quickly and sensitively to reduced chemical species (e.g., $Fe^{2+}$, $HS^{-}$, $H_2$) common in hydrothermal plumes. Real-time monitoring of the optical backscatter and ORP sensors aided in acquiring both plume and non-plume samples. The hydrothermal temperature anomaly ($\triangle\theta$) was calculated as

$$\triangle\theta = \theta - [(m_2 * \sigma_\theta^2) + (m_1 * \sigma_\theta) + m_0]$$

where $\theta$ is potential temperature, $\sigma_\theta$ is potential density, and $m_0$, $m_1$ and $m_2$ are constants

from a polynomial regression between $\theta$ and $\sigma_\theta$ in hydrothermally unaffected water at, or just above, plume depths Lupton et al. (1985). Water samples for DNA analyses for this study were obtained during CTD tow T15A-01 (August 18, 2015).

Once on deck, water samples were processed for physicochemical measurements (pH, gases, trace metals, DOM, and nutrients) and biological samples (cells for DNA). A total of 44 water samples from plumes over the caldera, NRZ, and background were analyzed on board ship for methane and hydrogen concentration by gas chromatography (SRI GC with flame ionization and pulse discharge detectors). Water for DNA was either processed immediately once shipboard or stored at 4°C in the dark for up to 8 hours until processing. To collect cells for DNA, 2L of water were filtered onto 0.2$\mu$m Supor polyethersulfone filters, flash frozen in liquid nitrogen, and stored at -80°C until further processing.

### *D.1.2 SSU rRNA gene sequencing*

DNA was extracted from Supor filters following previously described methods Marshall and Morris (2013), quantified using a Qubit fluorometer (Invitrogen, CA), concentrated using a SpeedVac (Thermo Scientific) and reconstituted in nuclease-free water (Qiagen).

The DNA was submitted to MRDNA (Stillwater, TX) for tag-sequencing. At MRDNA, the V4 hypervariable region of the 16S rRNA gene was amplified using 515F and 806R primers Apprill et al. (2015); Parada et al. (2015) with minor revisions to better amplify marine bacteria and archaea using PCR conditions of an initial denaturation at 94°C for 3 min. followed by 28 cycles of 94°C for 30 seconds, 53°C for 40 seconds, and 72°C for 1 minute, and then a final elongation step at 72°C for 5 minutes. PCR products were checked in 2% agarose gel to determine amplification success and relative intensity of amplification bands. All samples were pooled together in equal proportions, then pooled samples were

purified and calibrated using AMpure XP beads. The pooled and purified PCR products were used to prepare the DNA library at MRDNA on an Illumina MiSeq using manufacturers guidelines (Illumina, CA).

### *D.1.3 Community analyses*

Sequence data was processed by the MRDNA service facility. Their analysis pipeline has been developed to join paired ends, delete barcodes, remove sequences <150bp or with ambiguous bases, denoise sequences, generate OTUs (based on 97% identity), remove chimeras, and finally assign taxonomy using BLASTn against a curated database derived from RDPII and NCBI (http://rdp.cme.msu.edu, www.ncbi.nl.nih.gov). Patterns in bacterial and archaeal communities were analyzed using alpha and beta diversity measures. To minimize the effect of sampling effort, samples were rarefied to the size of the smallest sample. Alpha diversity was measured calculating the observed richness, or the number of OTUs present in each sample. Non-metric multidimensional scaling (NMDS) was used to compare the community composition across samples for either the bacterial or archaeal taxa. A Bray-Curtis distance matrix was calculated by randomly subsampling each sample, calculating Bray-Curtis dissimilarity, and finding the average over 1000 iterations. Analysis of similarity (ANOSIM) tests were employed to determine whether the differences in community composition were significantly larger between samples than within samples, over 999 random permutations of the dataset. Lastly, taxa significantly correlated with changes in community composition across samples were identified as indicator taxa by using a similarity percentage (SIMPER) analysis. For the indicator analysis, taxa that had an average abundance of 100 sequences or more were not included because they were present, and abundant, in all 12 samples. Rather, taxa that had an average abundance between 10 and 100 sequences were considered

most informative because they are not present in all samples, thus demonstrated informative distributional patterns. All community analyses were conducted in R (R Core Team, 2014).

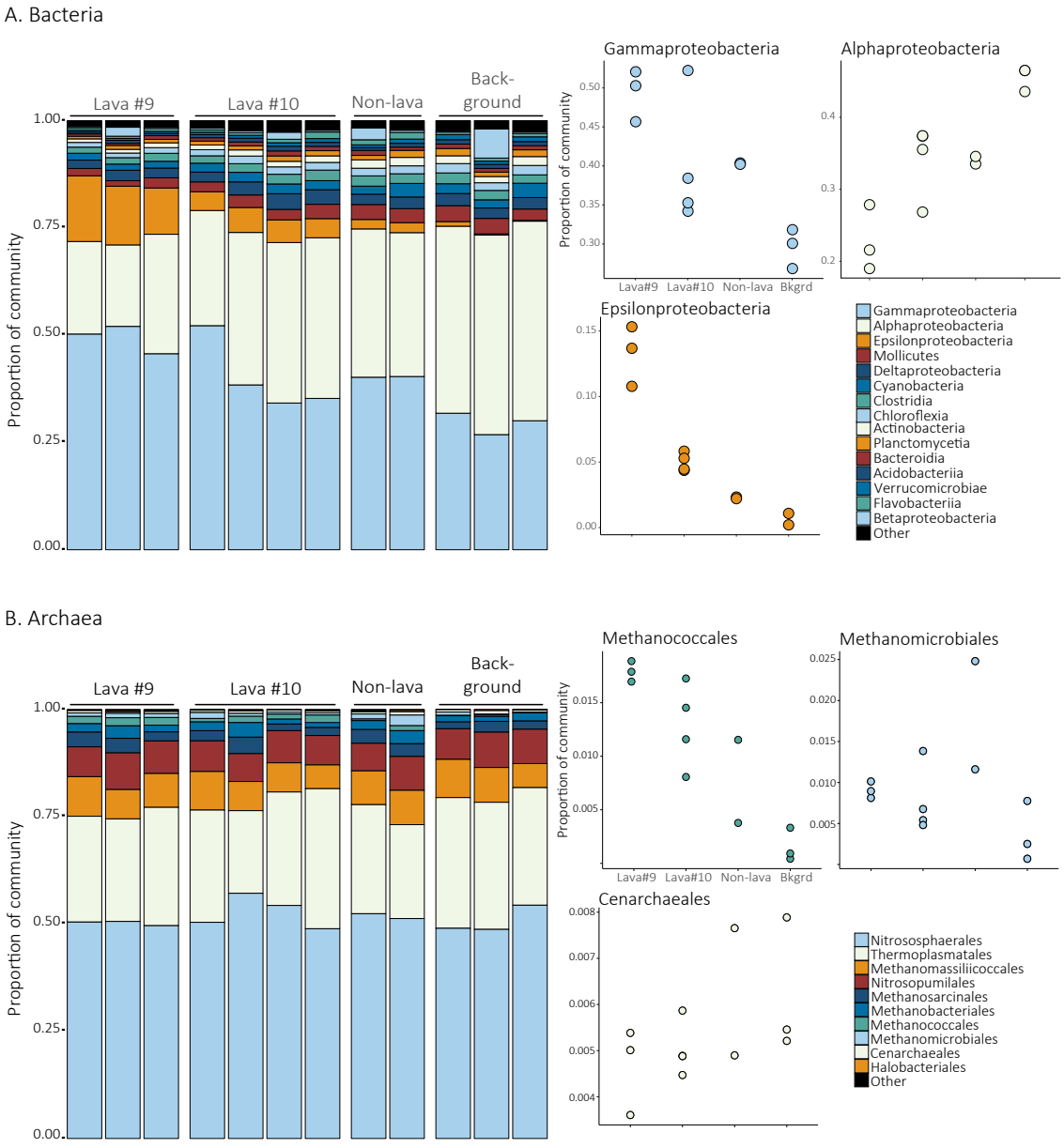


Figure D.1: Proportions of most abundant (A) Bacterial classes and (B) Archaeal orders detected along the North Rift Zone of Axial Volcano following the 2015 eruption. Taxonomic groupings that represented less than 1% or 0.1% of the total community, respectively, were combined into the other category. Samples are ordered by decreasing turbidity anomaly, a proxy for hydrothermal circulation, from left to right. Samples have been classified into four distinct sampling regions: lava flow #9, lava flow #10, non-lava, and background samples not influenced by hydrothermal activity.

Table D.1: Correlation coefficients, adjusted $R^2$, between bacterial class proportions and physicochemical parameters. Positive values indicate strong correlations while negative values indicate no correlation.

| Class | $\Delta$NTU | $\Delta\theta$ | $CH_4$ | $H_2$ |
|---|---|---|---|---|
| Acidobacteriia | 0.5508 | 0.4499 | -0.1414 | -0.1428 |
| Actinobacteria | 0.6284 | 0.3794 | 0.3910 | 0.1374 |
| Alphaproteobacteria | 0.7402 | 0.7838 | 0.1296 | 0.0332 |
| Bacteroidia | 0.5387 | 0.0611 | 0.0160 | -0.0835 |
| Betaproteobacteria | -0.0493 | -0.0304 | 0.0066 | -0.0433 |
| Chloroflexia | 0.7930 | 0.6637 | 0.1600 | -0.0355 |
| Clostridia | 0.6528 | 0.1965 | -0.1338 | 0.1546 |
| Cyanobacteria | 0.3612 | 0.3691 | 0.1003 | -0.1214 |
| Deltaproteobacteria | 0.1770 | -0.0346 | 0.1859 | -0.1362 |
| Epsilonproteobacteria | 0.9594 | 0.6732 | 0.3879 | -0.0730 |
| Flavobacteriia | 0.0821 | -0.0848 | -0.0357 | 0.2003 |
| Gammaproteobacteria | 0.5717 | 0.6325 | 0.0127 | 0.1058 |
| Mollicutes | 0.6784 | 0.3461 | 0.1814 | 0.0539 |
| Other | 0.4872 | 0.3750 | -0.1329 | -0.1411 |
| Planctomycetia | 0.4575 | 0.4545 | 0.1345 | 0.0702 |
| Verrucomicrobiae | 0.5486 | 0.6565 | 0.1226 | 0.0277 |

$\Delta$NTU= turbidity anomaly in nephelometric turbidity units; $\Delta\theta$= temperature anomaly

Table D.2: Correlation coefficients, adjusted $R^2$, between archaeal order proportions and physicochemical parameters. Positive values indicate strong correlations while negative values indicate no correlation.

| Order | $\Delta$NTU | $\Delta\theta$ | $CH_4$ | $H_2$ |
|---|---|---|---|---|
| Cenarchaeales | 0.1454 | -0.0152 | -0.1397 | 0.0575 |
| Halobacteriales | -0.0933 | -0.0727 | -0.0257 | -0.0847 |
| Methanobacteriales | -0.0890 | 0.0565 | -0.0909 | 0.1421 |
| Methanococcales | 0.5515 | 0.8154 | 0.6264 | 0.1378 |
| Methanomassiliicoccales | -0.0977 | -0.0922 | -0.1190 | 0.2300 |
| Methanomicrobiales | -0.0964 | 0.1097 | -0.1351 | 0.0183 |
| Methanosarcinales | -0.0501 | 0.1486 | -0.0456 | 0.0749 |
| Nitrosopumilales | -0.0372 | -0.0986 | -0.0349 | -0.1384 |
| Nitrososphaerales | -0.0421 | -0.1000 | 0.0032 | -0.1348 |
| Other (less than 0.1%) | 0.2628 | 0.5083 | -0.0925 | -0.1243 |
| Thermoplasmatales | -0.0836 | 0.1441 | 0.0157 | -0.0089 |

$\Delta$NTU= turbidity anomaly in nephelometric turbidity units; $\Delta\theta$= temperature anomaly