GEOMICROBIOLOGY OF HYDROTHERMAL PLUMES: ELUCIDATING THE ROLE OF MICROORGANISMS IN DEEP OCEAN CARBON AND SULFUR BIOGEOCHEMICAL CYCLES

by

Karthik Anantharaman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Earth and Environmental Sciences)
in the University of Michigan
2014

Doctoral Committee:

Assistant Professor Gregory James Dick, Chair
Research Professor Jeffrey C. Alt
Professor Joel D. Blum
Professor George W. Kling
Associate Professor Patrick D. Schloss

To my family: My parents, Jayanthi and V.J. Anantharaman, my brother, Prashant, and my wife, Pavana.

ACKNOWLEDGEMENTS

This dissertation is truly a product of many wonderful interactions, collaborations and support from a number of sources. Firstly, I would like to thank my advisor, Gregory Dick, for providing valuable advice and guidance, and showing immense patience during the course of my graduate studies. He afforded me the freedom to pursue my scientific ideas and research interests and his knowledge and perspective of how my research fits into the big picture have allowed me to mature as a scientist, for which I shall forever be grateful. I have also been privileged to have an extremely helpful committee: Joel Blum, Jeffrey Alt, George Kling and Patrick Schloss. Each committee member brought a different perspective that has enhanced my research and this dissertation.

I would like to thank my co-authors and collaborators: John "Chip" Breier, Cody Sheik, Brandy Toner, Kathleen Wendt, and Melissa Duhaime. Specifically, I would like to thank Chip who has been a co-author on all my papers and his expertise with sampling instrumentation and thermodynamic modeling has been vital to my research. Melissa's patience, help, advice and expertise were important to the successful completion of Chapter IV. I am also extremely grateful to Brandy for her assistance with mineralogical analyses in Chapter IV. My collaborations with all of you have helped me break ground and provided me the opportunity to truly do interdisciplinary research.

I would like to thank all the members of the Greg Dick lab at Michigan, past and present for their assistance. Specifically I am grateful to our bioformatics gurus; Sunit Jain and Daniel Marcus for their assistance with analyses. I would also like to thank Ryan Lesniewski, Alex Voorhies, Meng Li, Daniel Reed, Brett Baker, Paul Den Uyl, Michael Lerner and Cody Sheik, whom I have had the pleasure of interacting and collaborating. My time in Ann Arbor has truly been memorable thanks to many shared experiences and would not be the same without all of you.

My research at ELSC would not have been possible without the support of a number of individuals. I would like to thank the chief scientists, Anna-Louise Reysenback and Meg Tivey for providing me the cruise opportunity and providing us valuable sampling time. I thank the captain and crew of R/V *Thomas G. Thompson* and *ROV Jason II* for logistical support.

I am grateful to funding agencies for supporting my dissertation research. The Guaymas Basin project was supported by Gordon and Betty Moore Foundation Grant GBMF2609, the National Science Foundation Grant OCE 1029242 and the University of Michigan Rackham Graduate School Faculty Research Fellowship Program. The ELSC project was funded in part by the Gordon and Betty Moore Foundation Grant GBMF2609 and National Science Foundation Grant OCE1038006. Funding was also provided by a Scott Turner Graduate Student Research Award from the Department of Earth and Environmental Sciences, University of Michigan and the Rackham Graduate School at the University of Michigan.

Thanks to the graduate students of Earth and Environmental Sciences and all of my friends in Ann Arbor for making this fun and enjoyable. Thanks to the Office Staff and Support Staff at the Department of Earth and Environmental Sciences, University of Michigan for administrative assistance. A special thanks to the University of Michigan Cricket Club and the Troy Cricket Club for providing me the opportunity to play cricket in Michigan for some truly memorable experiences.

Last but not the least; I would like to thank my family. I am grateful to my parents, Jayanthi and Anantharaman and my kid brother, Prashant for their unwavering love, support and encouragement that makes me who I am. Finally, I would like to thank my wife, Pavana, for her constant support and love. None of this would be possible without all of you.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

ABSTRACT x

CHAPTER I INTRODUCTION 1

1.1 Hydrothermal plume microbiology 2

1.2 Chemosynthesis in hydrothermal plumes 3

1.3 Hydrogen oxidation in the deep sea 4

1.4 Sulfur biogeochemistry 5

1.5 The 'SUP05' clade of uncultured *Gammaproteobacteria* 7

1.6 Viruses in the deep sea 8

1.7 Field Sites 10

1.8 Structure of the dissertation 11

1.9 References 12

CHAPTER II EVIDENCE FOR HYDROGEN OXIDATION AND METABOLIC PLASTICITY IN WIDESPREAD DEEP-SEA SULFUR OXIDIZING BACTERIA 18

2.1 Abstract 18

2.2 Introduction 19

2.3 Materials and Methods 21

Sampling. 21

Extraction of nucleic acids, metagenomic and metatranscriptomic sequencing. 21

Assembly and Annotation. 22

cDNA mapping. 22

SSU rRNA gene amplicon pyrosequencing. 22

2.4 Results and Discussion 23
Diversity and distribution of SUP05 at Guaymas Basin. 24
Recovery and comparative analysis of SUP05 genomes. 24
Metabolic plasticity in SUP05 – genes for H2 oxidation and O2 respiration. 28
Carbon and nitrogen metabolism. 31
Population-specific metatranscriptomic mapping. 32
Dynamic expression of Ni-Fe hydrogenase genes. 33
Thermodynamic model for estimation of plume chemistry and bioenergetics. 34
2.5 Conclusions 35
2.6 Appendix A 38
CHAPTER II Supplementary Information 38
2.7 References 67
CHAPTER III METAGENOMIC RESOLUTION OF MICROBIAL FUNCTIONS IN DEEP-SEA HYDROTHERMAL PLUMES ACROSS THE EASTERN LAU SPREADING CENTER 75
3.1 Abstract 75
3.2 Introduction 76
3.3 Materials and Methods 79
Sample Collection 79
Extraction of nucleic acids and DNA sequencing. 80
De novo genomic assembly and annotation. 80
Read mapping. 81
Binning and conserved gene analysis. 81
Functional gene analysis. 82
Sequence alignment and phylogeny. 82
2-D Physical/Bioenergetic and thermodynamic modeling. 83
3.4 Results 85
Metagenomic sequencing, de novo assembly and binning. 86
Functional resolution of metagenomic bins. 91
Bioenergetic modeling of potential electron donors. 96
Overall distribution of lithotrophic metabolisms across the ELSC. 100

3.5 Discussion 101
3.6 Conclusions 105
3.7 Appendix B 107
CHAPTER III Supplementary Information 107
3.8 References 132
CHAPTER IV SULFUR OXIDATION GENES IN DIVERSE DEEP-SEA VIRUSES 140
4.1 Abstract 140
4.2 Introduction 141
4.3 Materials and Methods 141
Sample Collection 141
Extraction of nucleic acids and multiple displacement amplification of DNA. 142
DNA sequencing and pre-assembly data processing. 142
De novo genomic assembly. 143
Annotations. 143
Binning. 144
Comparative genomics. 144
Sequence alignment and phylogeny. 144
Thermodynamic modeling. 145
Micro-probe X-ray Diffraction (μXRD). 148
4.4 Results and Discussion 148
4.5 Conclusions 156
4.6 Appendix C 158
CHAPTER IV Supplementary Information 158
4.7 References 210
CHAPTER V CONCLUSIONS AND FUTURE DIRECTIONS 217
5.1 Introduction 217
5.2 Hydrogen oxidation in the deep ocean 218
5.3 Complexity of microbial communities inhabiting hydrothermal plumes 219
5.4 Viruses of chemolithotrophic microorganisms. 222
5.5 References 224

LIST OF FIGURES

Figure 1.1 Map of pathways for oxidation of reduced sulfur species 6

Figure 2.1 Content and transcript abundance of genes from Guaymas Basin SUP05 populations and comparison to genomes of other sequenced SUP05. 26

Figure 2.2 Map of pathways for sulfur oxidation by GB SUP05. 27

Figure 2.3. A, B. Organization and transcript abundance of GB-1 and 2 (A) and putative SUP05 (B) hydrogenase genes and comparison to closely related sequences from Genbank. 29

Figure 2.4 Phylogeny of group 1 membrane bound Ni-Fe hydrogenase large subunit inferred with maximum likelihood. 30

Figure 3.1 Assignment of assembled contigs from ELSC to specific bacterial, archaeal and eukaryotic populations using ESOM implemented with tetranucleotide frequencies. 87

Figure 3.2 Assignment of assembled contigs from ELSC to specific bins containing extrachromosomal elements using ESOM implemented with tetranucleotide frequencies. . 88

Figure 3.3 Identification of eukaryotic bin "Lau8" 90

Figure 3.4 Normalized abundance, energy and carbon metabolism of the 50 most abundant bins identified in ELSC hydrothermal plumes. 92

Figure 3.5 Phylogeny of group 1 membrane-bound Ni, Fe hydrogenase large subunit inferred with maximum likelihood. 94

Figure 3.6 Comparison of gene abundance and thermodynamic-bioenergetic estimates of available free energy associated with electron donors for lithotrophy in rising ELSC hydrothermal plumes. 99

Figure 3.7 2-D Physical model of the ABE-A1 hydrothermal plume coupled to bioenergetic model of elemental sulfur oxidation and normalized gene abundance 100

Figure 3.8 Genomic abundance of organisms using six abundant electron donors in ELSC hydrothermal plumes. 101

Figure 4.1 Gene content of 15 phage genomes from 3 viral families retrieved from Lau and Guaymas basins. 150

Figure 4.2 Phylogenetic tree of rdsrA genes inferred by Maximum Likelihood and detailed view of the SUP05 rdsrA clade. 152

Figure 4.3 Modeled free energies of catabolic reactions as a percentage of total available free energy at 2.5 and 5^{o}C and identification of elemental sulfur in the Abe hydrothermal plume 156

ABSTRACT

Deep-sea hydrothermal vents are distributed globally across mid-ocean ridges and back-arc basins and represent an important interface at which elements and energy are transferred between the lithosphere and the oceans. Deep-sea hydrothermal plumes occur where hot fluids rise from hydrothermal vents on the ocean floor, enriched with chemically reduced elements and compounds such as hydrogen sulfide, hydrogen, methane, iron, manganese and ammonia that serve as energy sources for microbial growth. Microbial activity in plumes influences the speciation of hydrothermal and oceanic elements and nutrients, with broad implications for marine biogeochemistry. Microbial chemosynthesis (fixation of carbon linked to oxidation of inorganic compounds) in plumes also contributes significantly to organic carbon in the deep oceans. Recent work suggests that microbial chemosynthesis is also surprisingly pervasive throughout the dark oceans, serving as a significant CO_2 sink even at sites far-removed from vents. Although ammonia and sulfur have been identified as potential electron donors for such chemosynthesis, they do not fully account for measured rates of carbon fixation in the dark oceans. Thus, there is a need to identify potential electron donors for chemosynthesis in the dark oceans and refine rates of carbon fixation to better resolve marine carbon budgets.

In order to address these research needs, we used DNA and cDNA sequencing of samples from Guaymas Basin (GB) (Gulf of California) and the Eastern Lau Spreading Center (ELSC) (Western Pacific Ocean) to elucidate the genetic potential and activity of microorganisms in

hydrothermal plumes and the surrounding deep oceans and to advance our understanding of the relationship between biogeochemistry and microbial diversity. First, we characterized the gene content and expression of SUP05, a globally widespread group of uncultured sulfur-oxidizing bacteria, in the GB hydrothermal plume. GB SUP05 contains and highly expresses genes for H_2 oxidation; this is the first H_2-oxidizing primary producer to be identified in the pelagic deep oceans. We also provide evidence for metabolic versatility associated with electron donors (H_2 and sulfur) and acceptors (oxygen, nitrate, and nitrite) in SUP05. These results indicate a capacity to influence and link the global cycles of sulfur, nitrogen, and carbon. Second, we show that hydrothermal plumes at ELSC host a complex and diverse microbial community comprising archaea, bacteria, eukarya and viruses. At ELSC, prominent differences in the geochemistry of hydrothermal vents did not manifest in the composition of microbial communities of hydrothermal plumes, which were dominated by sulfur-based chemolithotrophic energy metabolism at all vent sites. Finally, we identified five distinct viruses that infect SUP05 bacteria at GB and ELSC and showed that they contain SUP05-derived genes for sulfur oxidation, thereby providing the first evidence of viral genes involved in lithotrophy. We suggest that the SUP05 viruses use a novel ecological strategy to access abundant elemental sulfur in the environment by supplementing sulfur oxidation metabolism in their hosts in order to support viral infection and replication. This work implicates viruses as an important component of the global biogeochemical cycle of sulfur.

CHAPTER I

INTRODUCTION

Deep-sea hydrothermal vents represent an important interface at which elements and energy are transferred between the lithosphere and the oceans. The magnitudes of these transfers are such that hydrothermal vents exert a substantial influence on the chemistry of the global oceans (Elderfield and Schultz 1996). Deep-sea hydrothermal plumes occur where hot fluids arise from hydrothermal vents on the ocean floor, enriched with chemically reduced elements and compounds like hydrogen sulfide, hydrogen, methane, iron, manganese and ammonia that serve as potential microbial energy sources. Hydrothermal vents are distributed globally across mid-ocean ridges and back-arc basins and continue to be discovered at a rapid pace (Baker and German 2013, Baker et al 2013). Two distinct types of hydrothermal plumes can be defined based on physical and chemical properties: (1) "Rising plumes" are buoyant fluids that form by rapid mixing of hot hydrothermal vent fluids (~350^{o}C) with cold oxic sea water (2-4^{o}C). They potentially rise hundreds of meters off the seafloor before reaching neutral buoyancy. (2) "Neutrally-buoyant plumes" form when rising plumes are progressively diluted by seawater (finally containing ~0.001% hydrothermal vent fluid), dispersing hundreds of kilometers away from the source.

Hydrothermal vents are a significant source of iron and manganese to the world's oceans (Tagliabue et al 2010). Abiotic and biotic transformations of iron and manganese in hydrothermal plumes form oxide minerals that impact ocean geochemistry by controlling the reactivity and fate of these metals (Cowen et al 1986, Dick et al 2009). The resultant iron and manganese oxides are highly reactive and scavenge rare earth elements, potassium, vanadium, arsenic, chromium, uranium, and phosphorus from plumes and deposit them on the seafloor in the form of metalliferrous sediments (Goldberg 1954, Kadko 1993). In addition, processes within hydrothermal plumes may also help disperse iron and manganese through stabilizing associations with organic complexes (Breier et al 2012), sulfide nanoparticles (Yücel et al 2011), or through microbial cellular uptake (Li et al 2014). Thus, hydrothermal plumes affect ocean geochemistry on global scales and elemental transformations taking place within them are critical to understanding global elemental cycles.

1.1 Hydrothermal plume microbiology

Early investigations into the microbiology of hydrothermal plumes indicated that plume microbes are likely sourced from hydrothermal vent-chimney and near-vent associated microbial communities (Winn et al 1986). Near-vent environments are heavily influenced by ocean bottom water whose microbial community composition is dictated by low-temperature hydrothermal fluids emanating from the seafloor, which host microbial communities distinct from the ocean water column (Orcutt et al 2011, Santelli et al 2008). Recent studies have leveraged molecular tools such as analyses of the small subunit (SSU) ribosomal RNA gene to elucidate the composition of microbial communities inhabiting hydrothermal plumes. Sunamura et al (2004)

showed that just two groups of bacteria dominate the hydrothermal plumes at Suiyo Seamount, while Dick et al (2010) showed that the microbial communities inhabiting neutrally buoyant plumes at Guaymas Basin are similar to those in the pelagic ocean water column. German et al (2010) showed that microorganisms from the seafloor can be detected in rising hydrothermal plumes at the Mid-Cayman Rise. However these studies provide no insights into the metabolic potential and activity of microorganisms in hydrothermal plumes, the study of which can be made possible through the use of tools such as community DNA (metagenomics) and cDNA sequencing (metatranscriptomics).

1.2 Chemosynthesis in hydrothermal plumes

Microbial communities in hydrothermal-vent environments are driven by chemolithoautotrophic microorganisms (primary producers that fix carbon and gain energy from oxidation of inorganic compounds) (Jannasch and Mottl 1985, Nakagawa and Takai 2008). The presence of reduced elements, gases, and compounds in hydrothermal plumes allows microbes to derive energy through oxidation reactions. Chemolithotrophic microorganisms primarily exploit the redox disequilibrium of compounds involving carbon, hydrogen, sulfur, iron and manganese whose abiotic oxidation reactions are thermodynamically inhibited and thus proceed slowly without microbial catalysis (McCollom 2000). The primary electron donors for microbial metabolism in hydrothermal plumes are reduced sulfur compounds, hydrogen, methane, iron and manganese. Their potential oxidation reactions with oxygen are summarized as follows:

Hydrogen oxidation: $H_2 + 1/2O_2 \rightarrow H_2O$

Methane oxidation: $CH_4 + 2O_2 \rightarrow HCO_3^- + H^+ + H_2O$

Ammonia oxidation: $NH_3 + O_2 \rightarrow NO_2^- + 3H^+$

Sulfide oxidation: $HS^- + 2O_2 \rightarrow SO_4^{2-} + H^+$

Sulfur oxidation: $S^0 + 3/2\ O_2 + H_2O \rightarrow SO_4^{2-} + 2\ H^+$

Iron oxidation: $Fe^{2+} + 1/4O_2 + 2.5H_2O \rightarrow Fe\ (OH)_3 + 2H^+$

Manganese oxidation: $Mn^{2+} + 1/2O_2 + H_2O \rightarrow MnO_2 + 2H^+$

Estimates show that such hydrothermally derived chemolithotrophic metabolisms can account for up to 25% of the deep ocean organic carbon inventory (Maruyama et al 1998). Recent studies also indicate that chemosynthesis is predominant in the dark oceans at sites far away from hydrothermal vents (Reinthaler et al 2010, Swan et al 2011). However, our current knowledge of the nature and source of electron donors and specific microbial metabolisms that underpin carbon fixation in hydrothermal plumes and the broader deep oceans is insufficient and major discrepancies in the deep ocean carbon budget remain unresolved (Aristegui et al 2009, Burd et al 2010).

1.3 Hydrogen oxidation in the deep sea

Hydrogen oxidation has the potential to provide the greatest amount of metabolic energy to pelagic lithotrophs amongst all dissolved compounds in hydrothermal plumes through the 'Knallgas' reaction (McCollom 2000). Hydrogen concentrations in hydrothermal vent fluids range between 0.033 to 19 mM (Shock and Canovas 2010), orders of magnitude higher than background deep-sea concentrations of about 0.4 nM. In seafloor and near vent environments dominated by thermophilic microorganisms, hydrogen is hypothesized to be amongst the largest

sources of energy for microbial metabolism (Amend and Shock 2001). Recent studies have shown that hydrogen oxidation is a significant source of energy for bacteria in symbioses with hydrothermal-vent associated mussels (Petersen et al 2011). In addition, hydrogen oxidation genes have been identified in anaerobic and facultatively aerobic *Epsilonproteobacteria* from low-temperature hydrothermal vent fluids and animal associations (Nakagawa and Takai 2008)., In the pelagic ocean water column, chemolithotrophic hydrogen oxidation was first hypothesized to be associated with sinking particles (Karl et al 1984), yet hydrogen oxidation genes have only been identified in heterotrophic deep-sea clades of *Alteromonas Gammaproteobacteria* (Ivars-Martinez et al 2008). Thus, hydrogen oxidizing primary producers remain to be identified both in hydrothermal plumes and the dark oceans.

1.4 Sulfur biogeochemistry

Sulfur plays a key role in the biogeochemistry of deep-sea hydrothermal ecosystems. Sulfur can exist in a broad range of oxidation states ranging from -2 to +6, thereby allowing sulfur compounds to act as both microbial electron donors and acceptors depending on environmental conditions. The most abundant reduced sulfur compounds that serve as potential microbial energy sources in hydrothermal plumes include hydrogen sulfide (H_2S), elemental sulfur (S_0), thiosulfate ($S_2O_3^{2-}$), and sulfite (SO_3^{2-}). Recent studies also show that polysulfides (S_6, S_8) and metal sulfides (Fe, Cu, Zn, Mn) are also abundant in hydrothermal plumes (Breier et al 2012). The genetic mechanisms for oxidation of reduced sulfur species include flavocytochrome c sulfide dehydrogenase (*fcc*) and sulfide quinone oxidoreductase (*sqr*), mediating the oxidation of sulfide to elemental sulfur , the Sox enzyme complex (*soxABXYZ*) for

oxidation of thiosulfate to elemental sulfur, rhodanese sulfurtranferase for oxidation of thiosulfate to sulfite, reverse-acting dissimilatory sulfite reductase complex (*rdsrAB*) for oxidation of elemental sufur to sulfite, and adenosine 5′-phosphosulfate reductase (*apr*) and sulfate adenylyltransferase (*sat*) for oxidation of sulfite to sulfate (Fig. 1.1).

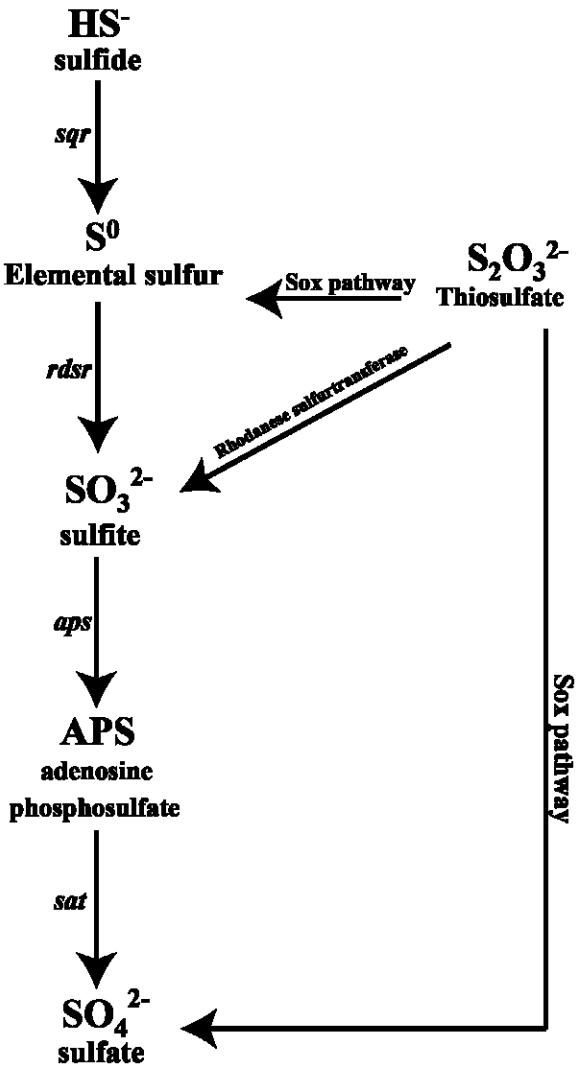


Fig 1.1 Map of pathways for oxidation of reduced sulfur species.

Hydrothermal vent fluids are rich in the most reduced form of sulfur, hydrogen sulfide, which has been measured in vent fluids at concentrations as high as 20 mM (Shock and Canovas 2010). Hydrogen sulfide is the primary electron donor for chemosynthesis in symbiotic bacteria residing within animals such as hydrothermal vent-associated tube-worms, mussels, clams and snails (Cavanaugh et al 1981, Duperron et al 2006, Felbeck 1981, Peek et al 1998, Suzuki et al 2006). Despite the abundance of hydrogen sulfide in vent fluids, thermodynamic models indicate that oxidation of elemental sulfur constitutes the single largest source of energy for microbial metabolism in hydrothermal plumes (McCollom 2000). The high concentrations of elemental sulfur in hydrothermal plumes have three potential causes: (1) abiotic oxidation of hydrogen sulfide to form elemental sulfur, polysulfides and metal sulfides (Breier et al 2012); (2) active microbial metabolic pathways that transform hydrogen sulfide and thiosulfate to extracellular elemental sulfur; (3) sulfur-oxidizing microbes that store elemental sulfur intracellularly in the form of globules that can be used in times of starvation (Hensen et al 2006). This sequestration of elemental sulfur by microorganisms, coupled with its slow kinetics of transformations in comparison to other sulfur species, results in the formation of a 'bottleneck' in the sulfur biogeochemical cycle (Ghosh and Dam 2009).

1.5 The 'SUP05' clade of uncultured *Gammaproteobacteria*

"SUP05"is an uncultivated clade of ubiquitous chemolithotrophic *Gammaproteobacterial* sulfur oxidizers (GSOs) that were first observed to be dominant members of the hydrothermal plume at Suiyo Seamount in the Izu-Bonin-Mariana Arc, located in the Western Pacific Ocean (Sunamura et al 2004). Recent studies have shown that SUP05 are abundant in a variety of

sulfur-rich marine habitats, including hydrothermal diffuse flow (Huber et al 2003), hydrotheraml plumes (Dick and Tebo 2010, Mattes et al 2013), and globally distributed oxygen minimum zones (OMZs) (Glaubitz et al 2013, Lavik et al 2009, Walsh et al 2009). SUP05 bacteria were first identified as important primary producers in OMZs, where they use hydrogen sulfide as an electron donor with nitrate as the electron acceptor, and thereby detoxify the oceans of highly poisonous dissolved sulfide (Lavik et al 2009). In this process, they produce a potent greenhouse gas in the form of nitrous oxide (N_2O) (Walsh et al 2009). Subsequent studies also show that SUP05 bacteria participate in a cryptic cycle of sulfur in OMZs (Canfield et al 2010).

SUP05 bacteria also dominate microbial assemblages associated with the gills of deep-sea hydrothermal vent-associated animals such as clams and mussels in a symbiotic association where they oxidize reduced sulfur species and hydrogen with oxygen as the primary electron acceptor (Kuwahara et al 2007, Newton et al 2007, Petersen et al 2011). SUP05 bacteria often occur in close association with another group of GSOs, the ARCTIC96BD-19 clade (Bano and Hollibaugh 2002). This recently cultivated clade of obligately aerobic bacteria is also ubiquitous in the pelagic oceans and has the genetic potential to fix carbon (Marshall and Morris 2012, Swan et al 2011). Considering the wide metabolic repertoire of the SUP05 group and their ubiquity in deep-sea hydrothermal environments (Dick and Tebo 2010, Huber et al 2003, Mattes et al 2013), it is imperative to understand their ecological role in hydrothermal plumes and the deep-oceans.

1.6 Viruses in the deep sea

Viruses are an important control on marine microorganisms that mediate biogeochemical cycles such as those described above (Fuhrman 1999). Although viral abundance in the oceans is highest in the euphotic zones, viruses are also ubiquitous in the dark oceans (Hara et al 1996). By lysing and turning over host populations, viruses release dissolved organic matter, stimulate nutrient cycling, and influence global biogeochemical cycles (Breitbart 2012, Fuhrman 1999, Suttle 2007). Marine viruses also influence host metabolism and evolution via virally-encoded, host-derived auxiliary metabolic genes (Breitbart 2007) (AMGs). The best studied example of an AMG is the viral cyanobacteria-derived *psbA* gene involved in photosynthesis (photosystem II) that are expressed during infection of cyanobacteria (Lindell et al 2005). Cyanobacterial viruses carry *psbA* genes to ensure their own success (Sullivan et al 2006) while impacting bacterial genome evolution (Avrani et al 2011) through processes such as horizontal gene transfer (Lindell et al 2004). Recent studies also show that AMGs are not limited to photosystem II genes in cyanobacterial viruses, but involve a wide variety of key genes in diverse viruses that encode for photosynthetic electron transport, stress response, pigment biosynthesis, phosphate metabolism, purine biosynthesis, pyrimidine biosynthesis, ribonucleotide reduction and carbon metabolism (Breitbart 2012, Bryan et al 2008, Hurwitz et al 2013). Considering the ubiquity of viral AMGs in natural systems, there is a developing paradigm that these genes serve to relieve metabolic bottlenecks during viral infection by prolonging host fitness long enough to ensure viral propagation (Breitbart 2012). Nearly all of this wealth of knowledge on marine viral ecology comes from studies of the surface oceans, hence, little is known about viruses of deep-sea chemolithotrophic bacteria. Although hints of viral impacts on chemolithotrophic communities exist in the form of observations of rampant horizontal gene transfer (Klein et al 2001), viral AMGs involved in chemolithotrophy remain to be identified.

1.7 Field Sites

The natural settings for the research described in this dissertation are the hydrothermal systems of Guaymas Basin (GB) in the Gulf of California and the Eastern Lau Spreading Center (ELSC) in the Western Pacific Ocean. Sampling these contrasting hydrothermal systems allows us to compare the plume microbiology of a sediment-hosted hydrothermal system to a back-arc basin, study the effect of different availabilities of electron donors and acceptors on plume microorganisms, and investigate the contribution of geographically distinct water masses to plume microbiology.

Guaymas Basin is a deep-sea hydrothermal system that is located at a water depth of about 2000m and that forms the northern-most segment of the East Pacific Rise. Unlike most mid-ocean ridge hydrothermal systems, it is located in a semi-enclosed basin close to the coast and sits underneath highly productive surface waters. The high rates of productivity result in the formation of a large OMZ at a depth of ~400m to ~1000m. Due to high rates of sedimentation from the surface waters, the ridge axis is covered by a thick layer of organic-rich sediment that exerts a significant influence on the geochemistry of hydrothermal vent fluids (Bazylinski et al 1988). In particular, the vent fluids that emerge from the seafloor are enriched in ammonia, methane and hydrocarbons while also possessing a high Mn/Fe ratio due to interactions between hydrothermal fluids and sediments (Von Damm et al 1985). Hydrothermal vent fluids at Guaymas Basin have a temperature of 315^{o}C, pH of 5.9, H_2 concentrations of 3.4 mM and H_2S concentrations of 6 mM. In addition, Guaymas Basin plumes exhibit other unusual characteristics such as low oxygen and high carbon dioxide concentrations due to restricted

mixing of bottom waters in the basin with oxygenated deep-sea water. GB is the primary field site to study the microbiology of neutrally buoyant plumes.

The **Eastern Lau Spreading Center** (ELSC) is a deep-sea hydrothermal system located in a back-arc basin in the Western Pacific Ocean. Although back-arc basins generally exhibit considerable variance in their tectonic characteristics and hydrothermal activity in comparison to mid-ocean ridge systems, ELSC displays the most pronounced and remarkable gradients in the chemistry of underlying rocks, spreading rates and hydrothermal vent geochemistry (Martinez et al 2006, Mottl et al 2011), and these properties are manifest in the geochemistry of the hydrothermal plumes. The five different hydrothermal vent fields sampled at ELSC, Kilo Moana, Tahi Moana, Abe, Mariner and Tui Malila, are located along a 245 km long north-south transect. Along this north-south axis, the underlying rock substrates change from basaltic to andesitic and water depth decreases from 2640m to 1877m (Ferrini et al 2008). The five vent fields display significant inter-field variability in the geochemistry of hydrothermal vent fluids (Mottl et al 2011). The vent fluids at ELSC exhibit H_2S concentrations in the range 2.8-19 mM and H_2 concentrations of 0.033-0.498 mM (Flores et al 2012). In contrast to GB, ELSC does not exhibit any sedimentary influence resulting in little to no output of methane, ammonia and hydrocarbons. Also in contrast to GB, the background deep waters are well-oxygenated. ELSC is the primary field site to study the microbiology of rising plumes and the influence of inter-field geochemical variability on microbial community structure.

1.8 Structure of the dissertation

The research I present in my dissertation is motivated by fundamental questions such as: Which microbes inhabit the dark oceans? What are their metabolic capabilities? What are the biogeochemical impacts of microbes on elemental cycles in the deep-oceans? I utilize the natural settings of deep-sea hydrothermal plumes as an *in situ* laboratory to study deep-sea microbes and elucidate their functional ecology. In **Chapter II**, I ask the question: What is the energy metabolism of ubiquitous SUP05 bacteria that inhabit hydrothermal plumes? I reconstructed genomes of two SUP05 bacteria from GB and described them as the first known hydrogen (and sulfur) oxidizing free-living chemolithotrophs in the deep water column. In **Chapter III**, I ask the question: What is the role of hydrothermal geochemistry shaping the distribution, diversity and energy metabolism of microbes in hydrothermal plumes? I use ~900 distinct reconstructed genomes of diverse bacteria, archaea, eukarya and viruses to investigate how energy metabolisms, especially electron donors for lithotrophy and autotrophy, are distributed in plumes along the ELSC geochemical gradient. In **Chapter IV**, I ask the question: Can we identify viruses putatively infecting SUP05 bacteria? I use reconstructed genomes of five diverse viruses inferred to infect SUP05 bacteria and describe their novel ecological strategies, including the first description of chemolithotrophic sulfur oxidation genes in viruses.

1.9 References

Amend JP, Shock EL (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiology Reviews* **25:** 175-243.

Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54:** 1501-1529.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011). Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474:** 604-608.

Baker ET, German CR (2013). On the Global Distribution of Hydrothermal Vent Fields. *Mid-Ocean Ridges*. American Geophysical Union. pp 245-266.

Baker ET, German CR, Elderfield H (2013). Hydrothermal Plumes Over Spreading-Center Axes: Global Distributions and Geological Inferences. *Seafloor Hydrothermal Systems: Physical, Chemical, Biological, and Geological Interactions*. American Geophysical Union. pp 47-71.

Bano N, Hollibaugh JT (2002). Phylogenetic Composition of Bacterioplankton Assemblages from the Arctic Ocean. *Applied and Environmental Microbiology* **68:** 505-518.

Bazylinski DA, Farrington JW, Jannasch HW (1988). Hydrocarbons in surface sediments from a Guaymas Basin hydrothermal vent site. *Organic Geochemistry* **12:** 547-558.

Breier JA, Toner BM, Fakra SC, Marcus MA, White SN, Thurnherr AM *et al* (2012). Sulfur, sulfides, oxides and organic matter aggregated in submarine hydrothermal plumes at 9°50′N East Pacific Rise. *Geochimica et Cosmochimica Acta* **88:** 216-236.

Breitbart M (2012). Marine Viruses: Truth or Dare. *Annual Review of Marine Science* **4:** 425-448.

Breitbart M, L.R. Thompson, C.A. Suttle, and M.B. Sullivan (2007). Exploring the vast diversity of marine viruses. *Oceanography* **20:** 135-139.

Bryan MJ, Burroughs NJ, Spence EM, Clokie MRJ, Mann NH, Bryan SJ (2008). Evidence for the Intense Exchange of MazG in Marine Cyanophages by Horizontal Gene Transfer. *PLoS ONE* **3:** e2048.

Burd AB, Hansell DA, Steinberg DK, Anderson TR, Arístegui J, Baltar F *et al* (2010). Assessing the apparent imbalance between geochemical and biochemical indicators of meso- and bathypelagic biological activity: What the @$♯! is wrong with present calculations of carbon budgets? *Deep Sea Research Part II: Topical Studies in Oceanography* **57:** 1557-1571.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330:** 1375-1378.

Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, Waterbury JB (1981). Prokaryotic Cells in the Hydrothermal Vent Tube Worm Riftia pachyptila Jones: Possible Chemoautotrophic Symbionts. *Science* **213:** 340-342.

Cowen JP, Massoth GJ, Baker ET (1986). Bacterial scavenging of Mn and Fe in a mid- to far-field hydrothermal particle plume. *Nature* **322:** 169-171.

Dick GJ, Clement BG, Webb SM, Fodrie FJ, Bargar JR, Tebo BM (2009). Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guaymas Basin deep-sea hydrothermal plume. *Geochimica et Cosmochimica Acta* **73:** 6517-6530.

Dick GJ, Tebo BM (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12:** 1334-1347.

Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP *et al* (2006). A dual symbiosis shared by two mussel species, Bathymodiolus azoricus and Bathymodiolus puteoserpentis (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environmental Microbiology* **8:** 1441-1447.

Elderfield H, Schultz A (1996). Mid-Ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annual Review of Earth and Planetary Sciences* **24:** 191-224.

Felbeck H (1981). Chemoautotrophic Potential of the Hydrothermal Vent Tube Worm, Riftia pachyptila Jones (Vestimentifera). *Science* **213:** 336-338.

Ferrini VL, Tivey MK, Carbotte SM, Martinez F, Roman C (2008). Variable morphologic expression of volcanic, tectonic, and hydrothermal processes at six hydrothermal vent fields in the Lau back-arc basin. *Geochemistry, Geophysics, Geosystems* **9:** Q07022.

Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Geoff Wheat C *et al* (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* **10:** 333-346.

Fuhrman JA (1999). Marine viruses and their biogeochemical and ecological effects. *Nature* **399:** 541-548.

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

Glaubitz S, Kießlich K, Meeske C, Labrenz M, 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:** 2767-2776.

Goldberg ED (1954). Marine Geochemistry 1. Chemical Scavengers of the Sea. *The Journal of Geology* **62:** 249-265.

Hara S, Koike I, Terauchi K, Kamiya H, Tanoue E (1996). Abundance of viruses in deep oceanic waters. *Marine Ecology Progress Series* **145:** 269-277.

Hensen D, Sperling D, Trüper HG, Brune DC, Dahl C (2006). Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum. *Molecular Microbiology* **62:** 794-810.

Huber JA, Butterfield DA, Baross JA (2003). Bacterial diversity in a subseafloor habitat following a deep-sea volcanic eruption. *FEMS Microbiol Ecol* **43:** 393-409.

Hurwitz BL, Hallam SJ, Sullivan MB (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14:** R123.

Ivars-Martinez E, Martin-Cuadrado AB, D'Auria G, Mira A, Ferriera S, Johnson J *et al* (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* **2:** 1194-1212.

Jannasch HW, Mottl MJ (1985). Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science* **229:** 717-725.

Kadko D (1993). An assessment of the effect of chemical scavenging within submarine hydrothermal plumes upon ocean geochemistry. *Earth and Planetary Science Letters* **120:** 361-374.

Karl DM, Knauer GA, Martin JH, Ward BB (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* **309:** 54-56.

Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, Abicht H *et al* (2001). Multiple Lateral Transfers of Dissimilatory Sulfite Reductase Genes between Major Lineages of Sulfate-Reducing Prokaryotes. *Journal of Bacteriology* **183:** 6028-6035.

Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M *et al* (2007). Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, Calyptogena okutanii. *Curr Biol* **17:** 881-886.

Lavik G, Stuhrmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P *et al* (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* **457:** 581-584.

Li M, Toner BM, Baker BJ, Breier JA, Sheik CS, Dick GJ (2014). Microbial iron uptake as a mechanism for dispersing iron from deep-sea hydrothermal vents. *Nat Commun* **5:** 3192.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 11013-11018.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86-89.

Marshall KT, Morris RM (2012). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J*.

Martinez F, Taylor B, Baker ET, Resing JA, Walker SL (2006). Opposing trends in crustal thickness and spreading rate along the back-arc Eastern Lau Spreading Center: Implications for controls on ridge morphology, faulting, and hydrothermal activity. *Earth and Planetary Science Letters* **245:** 655-672.

Maruyama A, Urabe T, Ishibashi J, Feely R, Baker ET (1998). Global hydrothermal primary production rate estimated from the southern East Pacific Rise. *Cahiers de Biologie Marine* **39:** 249-252.

Mattes TE, Nunn BL, Marshall KT, Proskurowski G, Kelley DS, Kawka OE *et al* (2013). Sulfur oxidizers dominate carbon fixation at a biogeochemical hot spot in the dark ocean. *ISME J* **7:** 2349-2360.

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

Mottl MJ, Seewald JS, Wheat CG, Tivey MK, Michael PJ, Proskurowski G *et al* (2011). Chemistry of hot springs along the Eastern Lau Spreading Center. *Geochimica et Cosmochimica Acta* **75:** 1013-1038.

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

Newton IL, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC *et al* (2007). The Calyptogena magnifica chemoautotrophic symbiont genome. *Science* **315:** 998-1000.

Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ (2011). Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. *Microbiology and Molecular Biology Reviews* **75:** 361-422.

Peek AS, Feldman RA, Lutz RA, Vrijenhoek RC (1998). Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proceedings of the National Academy of Sciences of the United States of America* **95:** 9962-9966.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476:** 176-180.

Reinthaler T, van Aken HM, Herndl GJ (2010). Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Research Part II: Topical Studies in Oceanography* **57:** 1572-1580.

Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML *et al* (2008). Abundance and diversity of microbial life in ocean crust. *Nature* **453:** 653-656.

Shock E, Canovas P (2010). The potential for abiotic organic synthesis and biosynthesis at seafloor hydrothermal systems. *Geofluids* **10:** 161-192.

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW (2006). Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts. *PLoS Biol* **4:** e234.

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

Suttle CA (2007). Marine viruses — major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Suzuki Y, Kojima S, Sasaki T, Suzuki M, Utsumi T, Watanabe H *et al* (2006). Host-Symbiont Relationships in Hydrothermal Vent Gastropods of the Genus Alviniconcha from the Southwest Pacific. *Applied and Environmental Microbiology* **72:** 1388-1393.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333:** 1296-1300.

Tagliabue A, Bopp L, Dutay J-C, Bowie A, Chever F, Jean-Baptiste P *et al* (2010). Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geoscience* **3:** 252-256.

Von Damm KL, Edmond JM, Measures CI, Grant B (1985). Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochimica Et Cosmochimica Acta* **49:** 2221-2237.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* **326:** 578-582.

Winn CD, Karl DM, Massoth GJ (1986). Microorganisms in deep-sea hydrothermal plumes. *Nature* **320:** 744-746.

Yücel M, Gartman A, Chan CS, George W. Luther I (2011). Hydrothermal vents as a kinetically stable source of iron-sulphide-bearing nanoparticles to the ocean. *Nature Geoscience* **4:** 367-371.

CHAPTER II

EVIDENCE FOR HYDROGEN OXIDATION AND METABOLIC PLASTICITY IN WIDESPREAD DEEP-SEA SULFUR OXIDIZING BACTERIA

Karthik Anantharaman[a], John A. Breier[b], Cody S. Sheik[a], and Gregory J. Dick[a,c]

a. Department of Earth and Environmental Sciences , University of Michigan, Ann Arbor, MI 48109
b. Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA 02543
c. Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109

Published in *Proceedings of the National Academy of Sciences* 110: 330-335.
doi: 10.1073/pnas.1215340110
Article first published online: December 20, 2012

2.1 Abstract

Hydrothermal vents are a well-known source of energy that powers chemosynthesis in the deep sea. Recent work suggests that microbial chemosynthesis is also surprisingly pervasive throughout the dark oceans, serving as a significant CO_2 sink even at sites far-removed from vents. Ammonia and sulfur have been identified as potential electron donors for this chemosynthesis, but they do not fully account for measured rates of dark primary production in the pelagic water column. Here we use metagenomic and metatranscriptomic analyses to show that deep-sea populations of the SUP05 group of uncultured sulfur-oxidizing *Gammaproteobacteria,* which are abundant in widespread and diverse marine environments,

contain and highly express genes encoding group 1 Ni-Fe hydrogenase enzymes for H_2 oxidation. Reconstruction of near-complete genomes of two co-occurring SUP05 populations in hydrothermal plumes and deep waters of the Gulf of California enabled detailed population-specific metatranscriptomic analyses, revealing dynamic patterns of gene content and transcript abundance. SUP05 transcripts for genes involved in H_2 and sulfur oxidation are most abundant in hydrothermal plumes where these electron donors are enriched. In contrast, a second hydrogenase has more abundant transcripts in background deep sea samples. Coupled with results from a bioenergetic model that suggest that H_2 oxidation can contribute significantly to the SUP05 energy budget, these findings reveal the potential importance of H_2 as a key energy source in the deep ocean. This study also highlights the genomic plasticity of SUP05, which enables this widely distributed group to optimize its energy metabolism (electron donor and acceptor) to local geochemical conditions.

2.2 Introduction

Deep-sea hydrothermal vent ecosystems depend on microorganisms that utilize reduced chemicals such as sulfur, methane, ammonium, and hydrogen (H_2) as electron donors for chemosynthesis (de Angelis et al 1993, Distel et al 1988, Jannasch and Mottl 1985, Lam et al 2004, Petersen et al 2011). Recent work suggests that microbial chemosynthesis is also far more prevalent in the broader deep oceans than previously recognized, extending throughout the water column of the dark open ocean, where it serves as a significant source of organic carbon (Aristegui et al 2009, Reinthaler et al 2010). The fuels for this pelagic primary production remain unknown, but recent studies show that ammonium (Lam et al 2004) and sulfur (Swan et al 2011,

Walsh et al 2009) are potential electron donors in the water column. Hydrogen (H_2), long known as an energy source for free-living bacteria in seafloor hydrothermal systems, was also recently identified as an electron donor in hydrothermal vent animal symbioses (Petersen et al 2011). Although microbial communities at seafloor hydrothermal vent sites have attracted much attention, hydrothermal vent plumes remain poorly characterized despite their importance as habitats for free-living chemolithoautotrophs (Winn et al 1986). These plume microorganisms mediate the hydrothermal transfer of elements from the lithosphere to the oceans (Dick et al 2009b, Toner et al 2009) and contribute significantly to organic carbon in the deep oceans via carbon fixation (de Angelis et al 1993, Dick and Tebo 2010, Lilley 1995, McCollom 2000a).

We investigated hydrothermal vent plumes in Guaymas Basin (GB) where hydrothermal enhancement of microbial activity is evident through increased total RNA concentrations (Lesniewski et al 2012*)* and rapid microbially-catalyzed Mn oxidation rates (Dick et al 2009b) in comparison to background waters of the deep Gulf of California. Among the most active and abundant microorganisms in GB plumes are sulfur-oxidizing bacteria of the SUP05 group of *Gammaproteobacteria* (Dick and Tebo 2010, Lesniewski et al 2012). SUP05 are dominant members of microbial communities in diverse marine environments such as hydrothermal vent plumes, symbiotic associations with hydrothermal vent clams and mussels, and oxygen minimum zones (OMZ) across the world`s oceans (Canfield et al 2010, Kuwahara et al 2007, Lavik et al 2009, Newton et al 2007, Sunamura et al 2004, Walsh et al 2009, Wright et al 2012).

In the present study, we use a combination of DNA, cDNA, SSU rRNA amplicon sequencing, and thermodynamic/bioenergetic modeling to elucidate the genetic potential, transcriptional activity and distribution of two uncultivated lineages of SUP05 bacteria in hydrothermal plumes and surrounding deep-sea waters. We report evidence for H_2 oxidation as

an important source of electrons for microbial growth in the deep oceanic water column and suggest that the SUP05 group displays metabolic plasticity that underlies the phylogenetic diversity of these widespread bacteria.

2.3 Materials and Methods

Sampling. Samples were collected on three cruises aboard *R/V New Horizon* in 2004 and 2005 as described previously (Dick et al 2009b, Dick and Tebo 2010). Metadata and chemical/physical characteristics of samples used for shotgun DNA and cDNA sequencing are presented in detail in Lesniewski et al 2012 (Lesniewski et al 2012), while summaries of these samples along with those used for SSU rRNA gene amplicon sequencing are described in Supplementary Table 1.

Extraction of nucleic acids, metagenomic and metatranscriptomic sequencing. DNA and RNA extraction were done as described previously (Dick and Tebo 2010, Lesniewski et al 2012). Purified DNA was used to prepare DNA libraries for sequencing using standard protocols (454 Life Sciences). An overall summary of DNA sequencing obtained using 454 GS FLX Titanium is presented in Lesniewski et al 2012 (Lesniewski et al 2012). cDNA synthesis was performed as described previously (Frias-Lopez et al 2008). cDNA sequencing produced a total of 1,558,905 reads from the plume (664,240 from Plume-3 (Cast 21-6#2) and 894,665 from Plume-4 (Cast 12-27a#1)) and 1,008,693 reads from the background deep sea (514,607 from Background-1 (Cast 12-8#12) and 504,086 from Background-2 (Cast 34-2#7)) using 454 GS FLX Titanium. A plume and background cDNA sample each were prepared for resequencing (for the purpose of comparison with 454) using standard protocols (Illumina) and a total of 103,078,758 reads from

the plume (Cast 21-6#2, Plume-3) and 122,259,588 reads from the background deep sea (Cast 12-8#12, Background-1) were obtained using Illumina HiSeq2000.

Assembly and Annotation. *De novo* metagenomic assembly was performed as described previously (Lesniewski et al 2012) using MIRA (Chevreux 2005) with parameters as follows: (-job=denovo, genome, accurate, 454 -notraceinfo -CL:pec=no -GE:not=8 -AS:urd=no -SK:bph=12:pr=80 454_SETTINGS -AS:mrl=50 -CO:mrpg=3 -AL:mrs=80). Gene annotations of assembled contigs was done through Integrated Microbial Genomes & Metagenomics (IMG/M) system (Markowitz et al 2008) as described previously (Lesniewski et al 2012). See *SI Appendix* for information on binning, identification and separation of the SUP05 contigs.

cDNA mapping. Transcript reads were mapped to predicted proteins using BLASTN *(bitscore≥50, E≤ 1 x E^{-5}, percent identity≥95%)*. Numbers of hits per gene were normalized by dividing the total cDNA hits by gene length, multiplying by 1000 and adjusting for the total size of the data set to enable comparison across the multiple data sets in the background deep sea and hydrothermal plume. Trends in normalized transcript abundances were similar across both 454 and Illumina data sets.

SSU rRNA gene amplicon pyrosequencing. DNA was extracted from a ¼ filter with the MoBio PowerSoil DNA isolation kit (Carlsbad, CA, USA). In addition to bead beating, filters were incubated at 65^{o}C for 20 min to facilitate cellular lysis. Bead beating was performed using the MP-Bio FastPrep-24 (Santa Ana, CA, USA) for 45 seconds at setting 6.5. The 16S rRNA gene was amplified in triplicate 25 µL reactions containing the following (final concentration): 12.5 µL 5 Prime HotMasterMix (Gaithersburg, MD, USA), 2 µL (15 µM) each forward and reverse primers, 1 µL community DNA. Previously described 16S rRNA gene primers targeting

the V4 region (515F/806R) (Bates et al 2010) were used and the reverse primers contained a 12-base barcode (Fierer et al 2008). PCR thermocycler conditions were as follows: initial denaturation 95^{o}C -4 min followed by 30 rounds of 95^{o}C for 30 sec, 50^{o}C for 1 min, 72^{o}C for 1 min and final elongation 72^{o}C for 10 min. Triplicate PCRs were combined and cleaned using a MoBio UltraClean PCR Clean-up kit (Carlsbad, CA, USA). DNA concentration was quantified using PicoGreen (Invitrogen, Carlsbad, CA, USA). Individual barcoded samples were combined into a single sample at equivalent concentrations then sent to Engencore (http://engencore.sc.edu) for pyrosequencing using 454 Titanium chemistry. Amplicon reads were corrected with Pyronoise *(Quince et al 2009)* implemented in MOTHUR (v. 1.26.0)(Schloss et al 2009). Operational taxonomic units (OTUs) were binned at 99% similarity and chimera checked using the OTUpipe (http://drive5.com/otupipe) command within Qiime (ver 1.4.0) (Caporaso et al 2010). Default parameters were used with the exceptions of initial clustering at 100% similarity and low abundance OTUs being kept for downstream analysis of rare phylotypes. OTUs were taxonomically classified with BLASTn (ver 2.2.22, e-values cutoff 10^{-8}) using Greengenes taxonomy and fasta files (available at http://qiime.wordpress.com), which were customized to include SUP05 16S rRNA sequences recovered from Guaymas Basin metagenomic libraries. Binning of OTUs at 99% was necessary in order to distinguish the two SUP05 phylotypes (GB-1 and GB-2). Using the full-length 16S rRNA genes recovered from the metagenomic libraries, we determined that for the V4 region used in the pyrosequencing study, an OTU cutoff of >98.5% would be necessary to distinguish the GB-1 & 2 phylotypes.

2.4 Results and Discussion

Diversity and distribution of SUP05 at Guaymas Basin. Phylogenetic analysis of SUP05 small subunit (SSU) rRNA gene sequences from Guaymas Basin hydrothermal plumes revealed the presence of two distinct SUP05 populations (Supplementary Figure 1) (hereby referred to as GB-1 and GB-2) that share 96.7% SSU rRNA nucleotide sequence identity (Lesniewski et al 2012). Our analyses also show that these two SUP05 lineages cluster closely with all previously identified SUP05 populations and fall into two co-occurring distinct sub-clades, similar to sequences retrieved from the African Shelf Namibian Upwelling zone and from the Saanich Inlet oxygen minimum zone (OMZ). The closest relatives of the GB SUP05 are the SUP05 SI-1 lineage (GB-1) (Walsh et al 2009) and symbionts of *Bathymodiolus* mussels from hydrothermal vents (GB-2) (Duperron et al 2006, Petersen et al 2012). High-throughput sequencing of the SSU rRNA gene amplicons from the Guaymas Basin water column indicate that GB-1 & 2 dominate the deep waters of the GB (>1700m), comprising up to 30% of the microbial community (Supplementary Figure 2). The abundance of SUP05 is tightly coupled to hydrothermal signals and also shows a minor increase in the oxygen minimum zone of the upper GB water column (Supplementary Figure 2).

Recovery and comparative analysis of SUP05 genomes. High-throughput sequencing of community genomic DNA and cDNA was used to reconstruct the metagenomes and metatranscriptomes of GB-1 and GB-2 in hydrothermal plumes and surrounding waters of the deep Gulf of California. *De novo* metagenomic assembly and binning by tetranucleotide signatures (Supplementary Figure 3) and BLAST (Supplementary Methods) yielded draft genomes of GB-1 & 2 that span 1.24 and 1.26 million base pairs (Mbp) of consensus sequence respectively, with an average coverage of ~13x for both genomes (Supplementary Table 2). To

confirm that they represented near-complete genomes, we identified a complete set of universally conserved genes present in each SUP05 genome (Supplementary Table 3).

GB-1 & 2 shared 83% of predicted genes with each other, and 60% of predicted genes with SUP05 populations from the Saanich Inlet OMZ (Walsh et al 2009) and the clam symbionts, *Candidatus* Ruthia Magnifica (Newton et al 2007) and *Candidatus* Vesicomyosocius okutanii (Kuwahara et al 2007) (Fig. 2.1). Like other SUP05 populations sequenced to date, GB-1 & 2 possess the complete repertoire of genes for carbon fixation and oxidation of reduced sulfur compounds, consistent with a common sulfur-based chemolithoautotrophic metabolism within the SUP05 group. These genes encode enzymes for the oxidation of reduced sulfur compounds (H_2S, $S_2O_3^{2-}$, S^0, SO_3^{2-}) including sulfide quinone oxidoreductase (*sqr*), mediating the oxidation of sulfide (HS^-) to elemental sulfur (S^o), the Sox enzyme complex (*soxABXYZ*) for oxidation of thiosulfate ($S_2O_3^{2-}$) to elemental sulfur, rhodanese sulfurtranferase for oxidation of thiosulfate to sulfite, reverse dissimilatory sulfite reductase complex (*dsrAB*) for oxidation of elemental sufur to sulfite (SO_3^{2-}), adenosine 5'-phosphosulfate reductase (*aps*) and sulfate adenylyltransferase (*sat*) for oxidation of sulfite to sulfate (SO_4^{2-}) (Fig. 2.2). Absence of *soxCD* genes in SUP05 populations may result in storage of elemental sulfur and provisioning of SUP05 with an electron donor (Hensen et al 2006), similar to the recently cultivated heterotrophic ARCTIC96BD-19 clade bacterium (Marshall and Morris 2012).

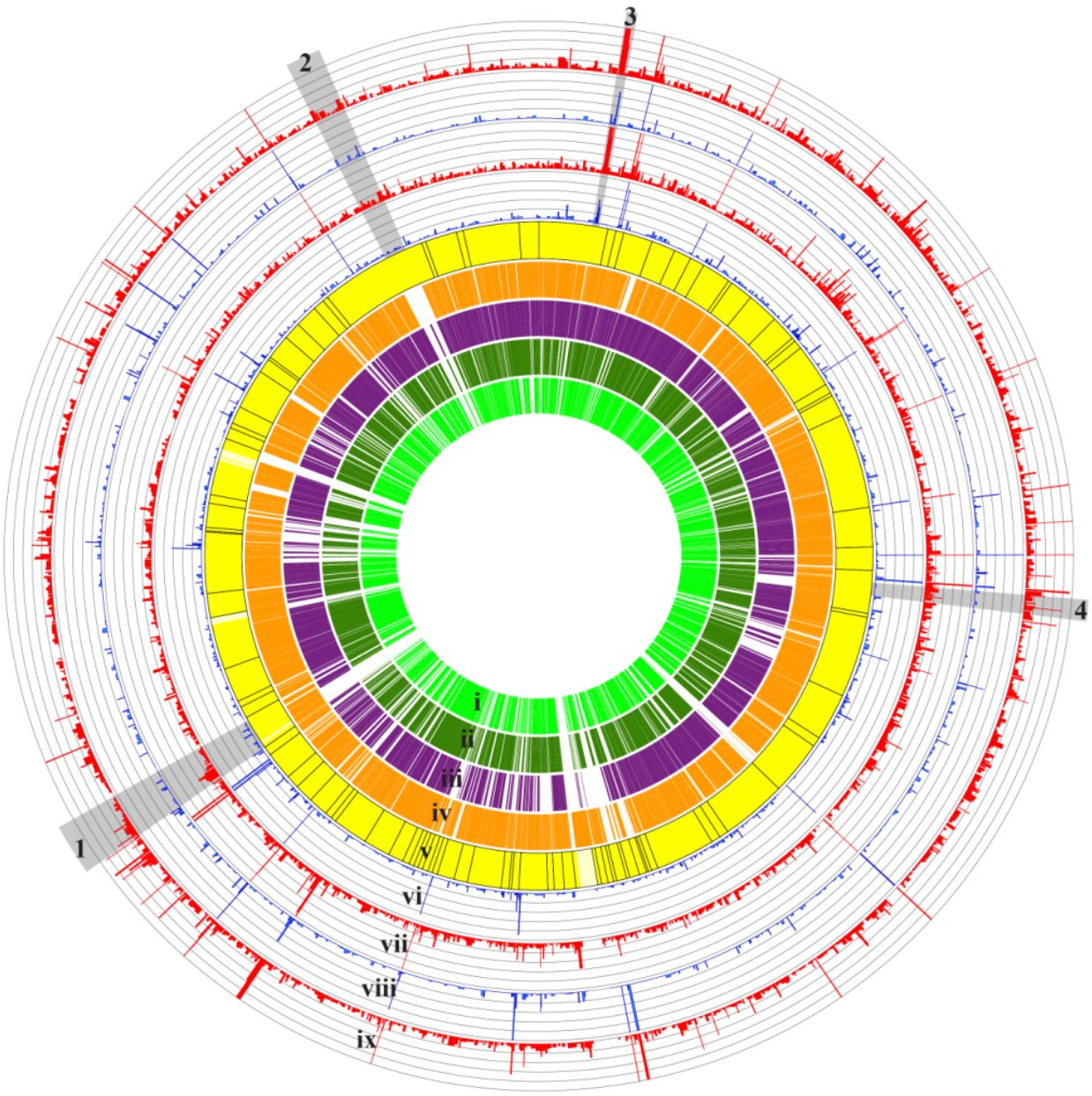


Figure 2.1 Content and transcript abundance of genes from Guaymas Basin SUP05 populations and comparison to genomes of other sequenced SUP05. Nested circles from innermost to outermost represent: (i) – (v) gene content with reference to GB-1 – (i) *Candidatus* Vesicomyosocius okutanii; (ii) *Candidatus* Ruthia magnifica; (iii) Saanich Inlet OMZ SUP05; (iv) GB-2; (v) GB-1. Gaps indicate the absence of genes in comparison to other SUP05 genomes. Black lines on GB-1 denote the separation of contigs that comprise the metagenome. (vi) – (ix) normalized abundance of 454 transcripts: (vi) GB-2 transcripts in background (blue); (vii) GB-2 transcripts in plume (red); (viii) GB-1 transcripts in background (blue); (ix) GB-1 transcripts in plume (red). Grey highlights on outermost circles indicate genes of interest: 1 - hydrogenase operon; 2 – urease operon; 3 – *sox* operon; 4 – cytochrome c oxidase complex.

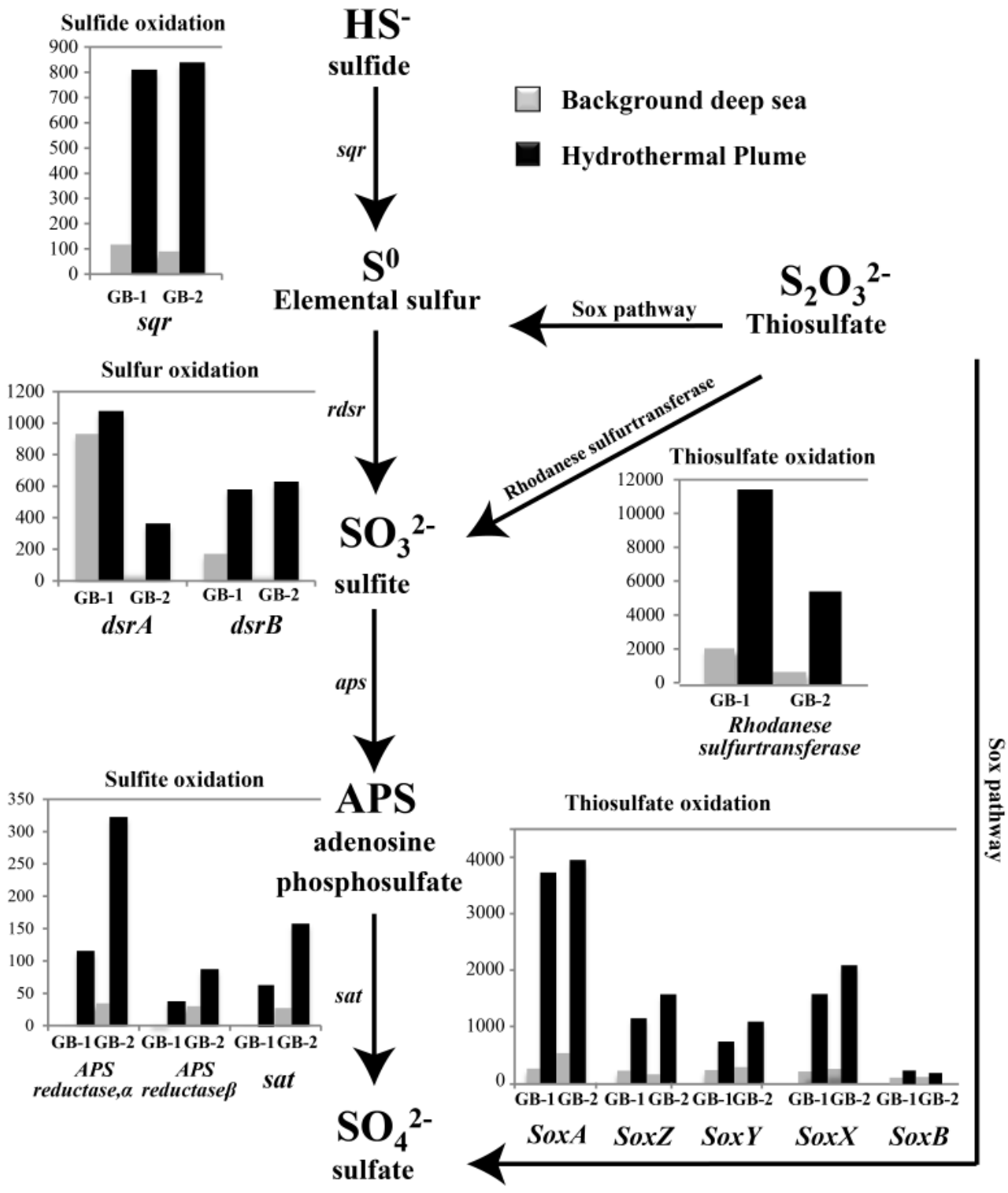


Figure 2.2 Map of pathways for sulfur oxidation by GB SUP05. Inset histograms depict the gene transcript abundance for individual genes in GB-1 and GB-2. Transcript abundance is normalized for gene length and total number of reads per dataset.

Metabolic plasticity in SUP05 – genes for H2 oxidation and O2 respiration. The Guaymas Basin SUP05 populations also harbor genes that set them apart from their Saanich Inlet and clam symbiont counterparts (Fig. 2.1). Key among these unique genes is a membrane-bound group 1 Ni-Fe hydrogenase for H_2 oxidation (Vignais and Billoud 2007). This enzyme and its associated maturation factors are encoded in both SUP05 populations by a set of 18 genes, 17 of which are adjacent on contigs (Fig. 2.3) confidently assigned to SUP05 by tetranucleotide frequency and by the fact that genes flanking the hydrogenase operon share synteny and high sequence similarity with other SUP05 genomes. Although this GB-SUP05 hydrogenase is not present in the Saanich Inlet OMZ SUP05 (Walsh et al 2009*)* or the clam symbionts *Candidatus* Ruthia Magnifica (Newton et al 2007) and *Candidatus* Vesicomyosocius okutanii (Kuwahara et al 2007), it is phylogenetically affiliated with other hydrothermal vent-derived hydrogenases (Fig. 2.4), including those from recently discovered H_2-oxidizing symbionts of *Bathymodiolus* mussels that are the first known H_2-powered chemosynthetic symbiosis at deep-sea hydrothermal vents (Petersen et al 2011). Genes in the SUP05 hydrogenase operons display synteny and high sequence identity (92 and 94% for HupS & HupL) with genes from the *Bathymodiolus* symbionts for structural assembly, synthesis, hydrogen uptake and oxidation, suggesting a similar role in H_2 oxidation for the purpose of energy production (Fig. 2.3A).

The SUP05 genomic bin also contains a contig (AJXC01001965) with genes encoding a second group I Ni-Fe hydrogenase that displays distinct operon structure (Fig. 2.3B) and phylogeny (Fig. 2.4) to the first. This putative SUP05 hydrogenase clusters with Ni-Fe hydrogenases from epipelagic *Gammaproteobacteria* (Huggett and Rappé 2012), *Flavobacteria* (Cho and Giovannoni 2004, Woyke et al 2009), and *Deltaproteobacteria* (Chitsaz et al 2011), possibly indicating a different evolutionary origin and/or physiological role. Because the second

hydrogenase-containing contig cannot be scaffolded onto other SUP05 genomes, and in view of its complex phylogeny, we cannot conclusively determine the taxonomic origin at this time.

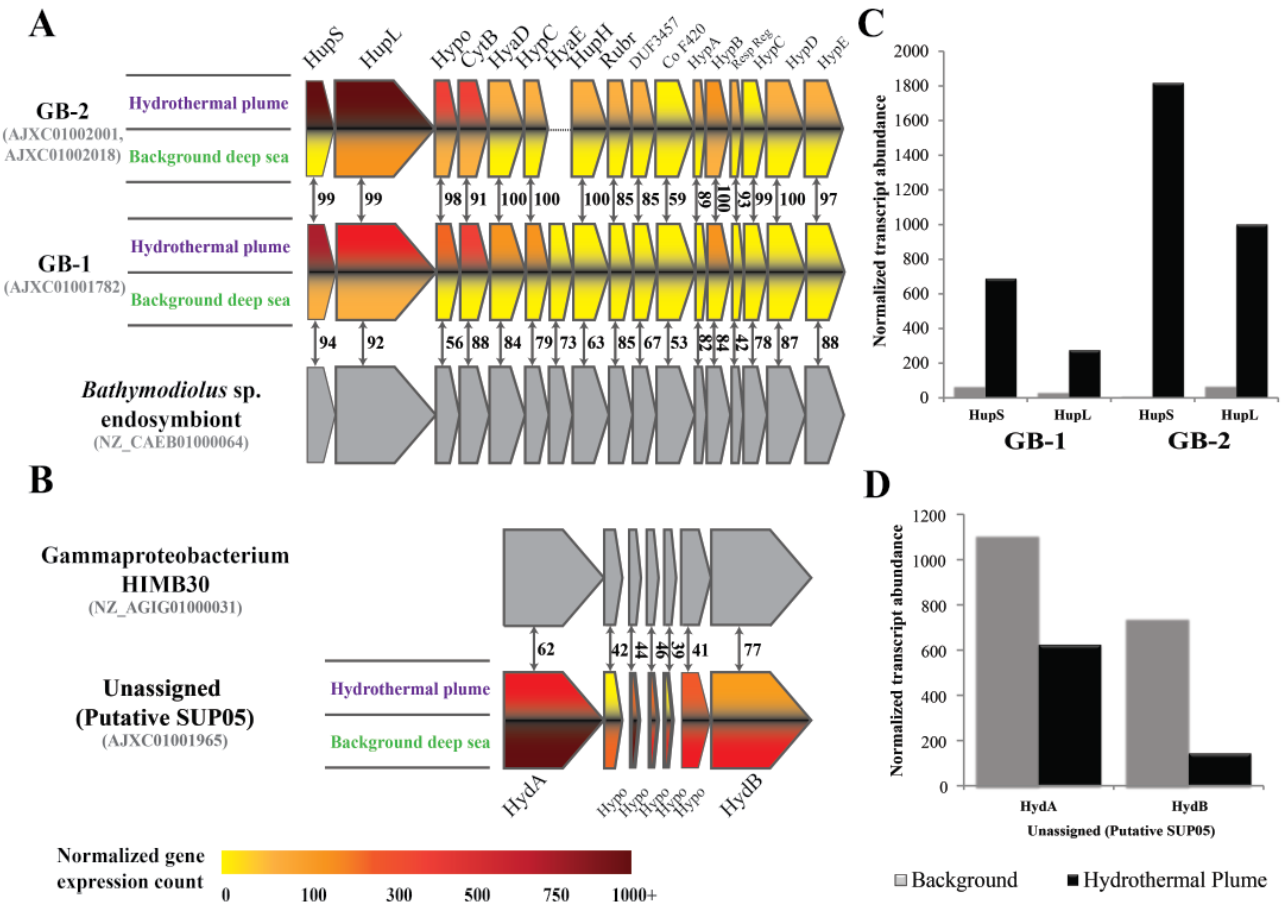


Figure 2.3. A, B. Organization and transcript abundance of GB-1 & 2 (A) and putative SUP05 (B) hydrogenase genes and comparison to closely related sequences from Genbank. Genes are colored according to normalized transcript abundance in plume and background. Arrows indicate shared genes and percent amino acid identity between predicted proteins. Dotted line in GB-2 indicates separation of contigs. **C, D**. Normalized transcript abundance for genes encoding small (HydA, HupS) and large subunits (HydB, HupL) of GB-1 & 2 (C) and putative SUP05 (D) hydrogenases.

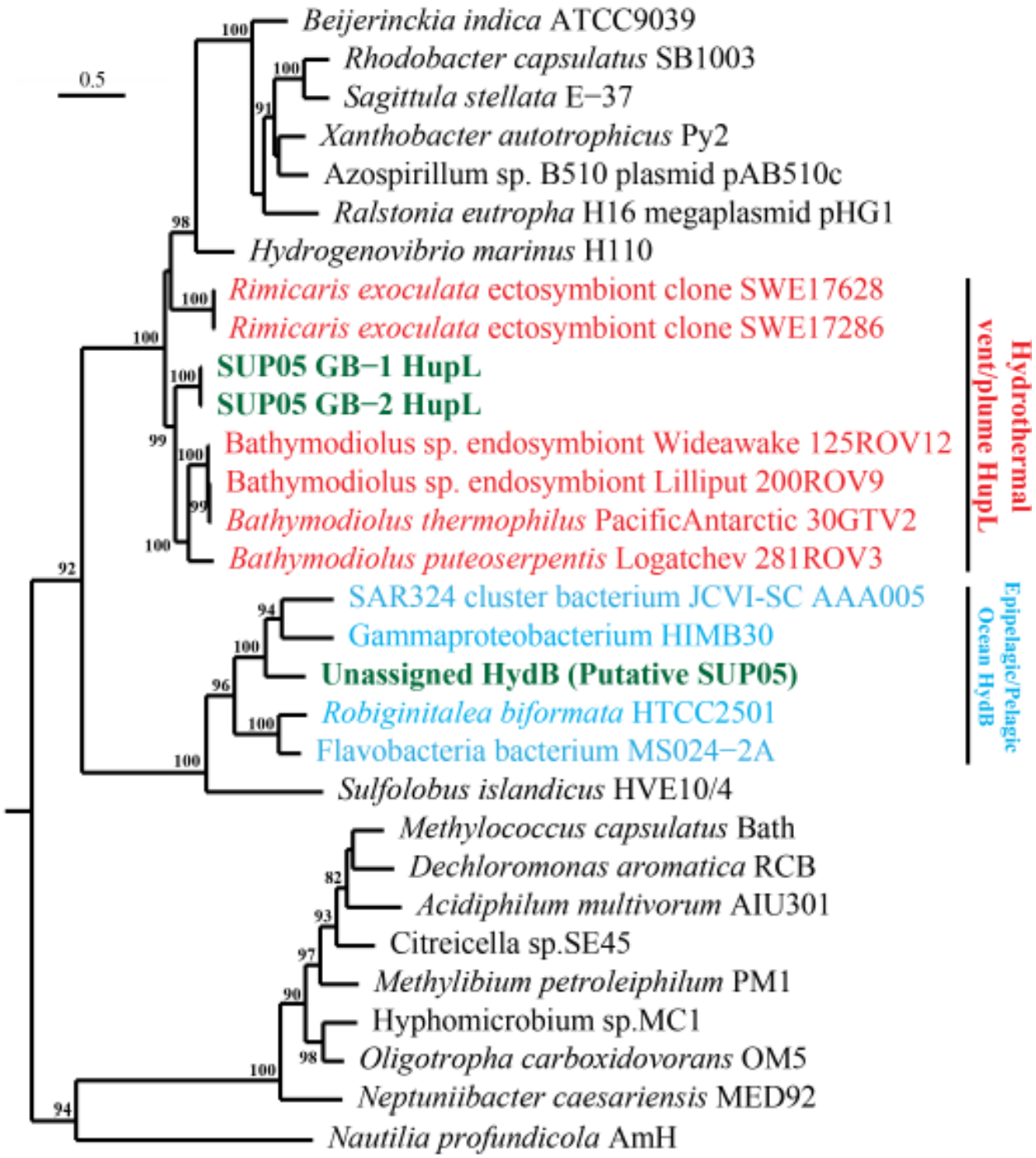


Figure 2.4 Phylogeny of group 1 membrane bound Ni-Fe hydrogenase large subunit inferred with maximum likelihood. Bootstrap values greater than 80 are shown. Sequences in green are from Guaymas Basin, sequences in red are hydrothermal vent derived and sequences in blue are from the epipelagic ocean.

The Guaymas Basin SUP05 genomes also display genomic and metabolic diversity in terms of electron acceptors for energy metabolism. GB-1 & 2 genomes encode for both a cytochrome c oxidase and a cbb_3-type terminal cytochrome c oxidase. Both these cytochrome c oxidase complexes are shared by the SUP05 clam symbionts (Kuwahara et al 2007, Newton et al 2007) but are absent in the free-living OMZ SUP05 (Walsh et al 2009). The presence of these genes enables the use of oxygen as a terminal electron acceptor in both oxic and microoxic environments that exist in the stratified water column of Guaymas Basin. The set of genes encoding for dissimilatory nitrate/nitrite reduction to N_2O, which are present in OMZ SUP05, are absent in GB-1 & 2 (Supporting Information) except for a single dissimilatory nitrite reductase (NO-forming *nirK*) in GB-2, hinting at a possible role in either a partial dissimilatory denitrification pathway or in nitrite detoxification rather than the full denitrification pathway of Saanich Inlet SUP05 (Walsh et al 2009). Evidence consistent with such partial denitrification has been found in the Eastern Tropical South Pacific OMZ, where SUP05 are abundant, and sulfide-dependent reduction of nitrate produces NO_2^- and N_2O as well as N_2 (Canfield et al 2010). Genes for dissimilatory nitrate and nitrite reductases and associated cofactors were identified on short contigs with low genomic coverage (Supplementary Table 4) suggesting that they stem from minor genome variants within the community. Genes for reduction of nitric oxide (NO) to nitrous oxide (N_2O) (*norB*, *norC*) were absent from the metagenome.

Carbon and nitrogen metabolism. The GB-1 & 2 genomes contain genes encoding the Calvin-Benson-Bassham (CBB) cycle including a single form II ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) for the purpose of carbon fixation. This form II RuBisCO is also present in the Saanich Inlet OMZ SUP05 (Walsh et al 2009) and the clam symbionts, *Candidatus* Ruthia Magnifica (Newton et al 2007) and *Candidatus* Vesicomyosocius okutanii (Kuwahara et al

2007). In contrast, the H_2-oxidizing symbionts of *Bathymodiolus* mussels (Petersen et al 2011) possess genes for a form I RuBisCO, which is optimized for higher O_2 and lower CO_2 concentration (Badger and Bek 2008). The presence of genes in GB SUP05 encoding for form II RuBisCO enzymes typically adapted to low O_2 and high CO_2 concentrations is consistent with the low O_2 conditions of the deep Guaymas Basin. Genes for gluconeogenesis and the non-oxidative branch of the pentose phosphate pathway were also identified, along with all components of the tricarboxylic acid cycle (TCA) except for those encoding the α-ketoglutarate dehydrogenase enzyme, consistent with GB SUP05 being primarily autotrophs (Walsh et al 2009). This is also evidenced by the lack of known transporters for organic carbon except the two noted below.

The GB SUP05 genomes possess two ABC-type transporters (HAAT and PAAT family) annotated as amino acid transporters. GB-2 also contains a single putative di/tri carboxylate transporter. The presence of these transporters is intriguing because it may suggest an alternative source of carbon and nitrogen and hint at a mixotrophic lifestyle with the ability to utilize organic carbon as in the recently cultivated and closely related ARCTIC96BD-19 clade bacterium (Marshall and Morris 2012, Swan et al 2011). For the purpose of nitrogen assimilation and metabolism, GB-1 & 2 genomes have multiple copies of genes for ammonium transport and a full complement of assimilatory nitrite reduction genes for reduction of nitrite (NO_3^-) to ammonia (NH_3). Also present are genes for breakdown of urea and amides by an amidohydrolase (GB-1 & 2) and a urease (GB-1 only) that are absent in the SUP05 clam symbionts and the Saanich Inlet OMZ SUP05 genomes.

Population-specific metatranscriptomic mapping. In order to examine the transcriptional activity of GB-1 & 2, we used their assembled genomes as a framework to map

metatranscriptomic reads. Population-specific mapping of Illumina cDNA reads to both SUP05 genomes assigned a total of 104,075 transcripts to GB-1 and 136,524 transcripts to GB-2. Both of these SUP05 genomes recruited more total transcripts in the hydrothermal plume than background by a ratio of approximately three, indicating that they are stimulated in hydrothermal plumes. Amongst the most abundant transcripts in the metatranscriptome were those mapping to genes involved in chemolithoautotrophy (Supplementary Figure 4), including H_2 oxidation (Fig 2.3C, 2.3D), O_2 respiration (Supplementary Figure 5), oxidation of reduced sulfur species (Fig. 2.2), and carbon fixation (Supplementary Figure 6). Both GB-1 & 2 preferentially expressed genes for oxidation of multiple reduced sulfur species (H_2S, $S_2O_3^{2-}$, S^0, SO_3^{2-}) in the plume compared to the background, indicating that reduced sulfur species are important electron donors in the plume. High transcript abundances of the RuBisCO genes in both GB-1 & 2 metatranscriptomes implicates the deep-sea SUP05 populations in carbon fixation and underscores their importance as key autotrophs in the deep sea. All SUP05 genes for nitrogen metabolism were recovered in the metatranscriptome, with genes encoding ammonium and amino acid/amide uptake having high transcript abundances in both the hydrothermal plume and background deep-sea, again suggesting that GB SUP05 actively obtain amino acids from the environment (Supplementary Figure 7). These trends in transcript abundance for SUP05 genes were similar across both 454 and Illumina-based metatranscriptomes.

Dynamic expression of Ni-Fe hydrogenase genes. A major difference is evident in patterns of hydrogenase transcript abundance between plume and background. The hydrothermal vent-related hydrogenases are highly enriched in metatranscriptomes of plumes (Fig. 2.3C). Conversely, the epipelagic-related hydrogenase (putative SUP05) is enriched in the background metatranscriptome relative to the plume (Fig. 2.3D). Because SUP05 abundance in the

metagenome is similar between plume and background *(13, 16),* the dynamic patterns of transcript abundance we observe for the Ni-Fe hydrogenase genes suggests that their expression is regulated rather than constitutive (Petersen et al 2011). Based on the increased H_2 concentrations expected in plumes versus background, we suggest that H_2 concentration is the likely regulator of this observed differential expression. We speculate that the hydrothermal vent-related and the epipelagic-related hydrogenases are distinguished in their affinity for H_2, the former being adapted to higher H_2 concentrations in environments such as hydrothermal plumes, and the latter to low H_2 concentrations typically available in the background deep ocean.

As abundant members of both hydrothermal plume and background deep ocean communities, SUP05 populations likely take advantage of H_2 derived not only from hydrothermal fluids but also from mineral precipitation reactions in the plume (McCollom 2000a) and possibly anaerobic decomposition on sinking particles, a source of H_2 posited long ago (Karl et al 1984). Further, high levels of expression of Ni-Fe hydrogenases in the background deep ocean, far from the hydrothermal plumes, may also indicate the presence of a significant but currently unrecognized source of H_2.

Thermodynamic model for estimation of plume chemistry and bioenergetics. H_2 concentrations of up to 3 mM measured in GB end-member hydrothermal fluids are the result of the reaction of seawater with mantle-derived basalt in the oceanic crust at high temperature and pressure (Lilley et al 1982, McCollom 2008, Welhan and Craig 1979). Unfortunately, to our knowledge, no direct measurements of H_2 concentration have been made on GB plumes. Thus, we used equilibrium thermodynamic reaction path modeling to estimate the concentrations of H_2 and other potential electron donors in the GB plume (Supplementary Information). Results predict that H_2 concentrations range from 0.5 to 50 nM in plumes sampled here (2.93-2.97^{o}C),

which are up to ~100 times greater than typical H_2 concentrations of 0.4 nM in the background deep sea.

To assess the relative importance of H_2 and sulfur as energy sources for SUP05, we compared the free energy yields for a number of metabolisms including those using H_2, S^0, H_2S, $S_2O_3^{2-}$ and particulate metal sulfides as electron donors (Supplementary Table 6). Our model estimated the free energy available from H_2 oxidation in the hydrothermal plume to be 0.04 J/kg of plume fluid at a temperature of 2.95°C, representing 17% of the energy budget for SUP05 (Supplementary Figure 8). Further, these results indicate that H_2 oxidation can account for up to 22% of the energy budget of SUP05 in warmer fluids of rising hydrothermal plumes (3.0-5.9°C), which have not yet been studied from a microbiological perspective. This prominent role for H_2 oxidation is consistent with previous studies that have modeled available energy in hydrothermal plumes (McCollom 2000a); H_2 oxidation is expected to play an even more important role in ultramafic-hosted hydrothermal systems (Amend et al 2011). Amongst sulfur species, we found S^0 oxidation with both oxygen and nitrate to be thermodynamically favored relative to H_2S, thiosulfate and particulate metal sulfides. Although there is uncertainty with regard to sulfur speciation in the plume and the actual form of sulfur utilized by SUP05 is unknown, these results suggest that Guaymas Basin SUP05 populations utilize environmentally-supplied sulfur species other than dissolved H_2S. Overall, the modeling results presented here indicate that oxidation of H_2 and reduced sulfur species are both potentially significant sources of free energy for growth of SUP05 populations in Guaymas Basin hydrothermal plumes.

2.5 Conclusions

This study advances our understanding of the chemolithotrophic metabolism of a widespread group of marine bacteria, providing insight into potential genetic and physiological underpinnings and biogeochemical implications of microbial diversity observed within the SUP05 group. As abundant microorganisms in the pelagic realm of the dark ocean, SUP05 have the capacity to influence and link the global cycles of sulfur, nitrogen, and carbon in an environment that holds the largest reservoir of reactive dissolved inorganic carbon on the Earth's surface. Recognition of H_2 as a significant electron donor for microbial growth in the pelagic water column may shed light on discrepancies in current oceanic carbon budgets (Aristegui et al 2009, Reinthaler et al 2010). Additional molecular studies are needed to determine the prevalence of SUP05 hydrogen oxidation genes beyond the Gulf of California (Supporting Information), and geochemical measurements of H_2 oxidation rates are required to directly and quantitatively evaluate the contribution of H_2 to chemosynthesis in the deep sea. Although these experiments are challenging due to the low H_2 concentrations (nM) and remote nature of the deep sea, the molecular evidence presented here provides the impetus to develop such methods. The genetic and metabolic plasticity of electron donors (H_2 and reduced sulfur species) and acceptors (oxygen, nitrate, and nitrite) across the SUP05 group revealed here underscores the importance of taking fine-scale microbial functional diversity into account when tracking microbial biogeochemistry. Given the central role of SUP05 in the biogeochemistry of globally expanding OMZs and associated feedbacks on cycling of carbon, nitrogen, sulfur, and greenhouse gases (Wright et al 2012), such resolution will be critical to understanding and predicting marine ecosystem dynamics in the context of environmental change.

Accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AJXC00000000. The version described in this paper is the first version, AJXC01000000.

Acknowledgements. This project is funded in part by the Gordon and Betty Moore Foundation and the National Science Foundation (OCE 1029242). We also thank the University of Michigan Rackham Graduate School Faculty Research Fellowship Program for their support and Brett Baker and Sunit Jain for their insightful assistance. 454 DNA sequencing was conducted by Lynn Tomsho in the laboratory of Stephan Schuster at Penn State University. Illumina sequencing was conducted at the University of Michigan DNA Sequencing Core.

Foot notes

Correspondence and request for materials should be addressed to G.J.D. (gdick@umich.edu).

Author Contributions: G.J.D. collected the samples. K.A. and G.J.D. designed the study. G.J.D. and K.A did the DNA and cDNA sequencing. K.A. did the data analyses. J.A.B. did the thermodynamic modeling. C.S.S. did the SSU rRNA amplicon sequencing. K.A. and G.J.D. wrote the manuscript. K.A., J.A.B and C.S.S wrote the supplementary information.

The authors declare no conflict of interest.

2.6 Appendix A

CHAPTER II Supplementary Information

Contents

1. **Supplementary Materials and Methods**
2. **Supplementary Figure 1**
3. **Supplementary Figure 2**
4. **Supplementary Figure 3**
5. **Supplementary Figure 4**
6. **Supplementary Figure 5**
7. **Supplementary Figure 6**
8. **Supplementary Figure 7**
9. **Supplementary Figure 8**
10. **Supplementary Table 1**
11. **Supplementary Table 2**
12. **Supplementary Table 3**
13. **Supplementary Table 4**
14. **Supplementary Table 5**
15. **Supplementary Table 6**
16. **Supplementary Table 7**

Supplementary Materials and Methods

Identification of GB SUP05 contigs

Binning of the assembled contigs in the GB metagenome was performed using tetra-nucleotide frequencies signatures and emergent self-organizing maps (ESOM) (Dick et al 2009a) with a contig length minimum cutoff of 2.5 Kb. 17 reference genomes were used in the ESOM mapping to achieve a greater resolution in the binning process: *Alteromonas macleodii* str.'Deep ecotype' (Ivars-Martinez et al 2008), *Methylococcus capsulatu*s str.bath (Ward et al 2004), *Nitrosopumilus maritimus* str.SCM1 (Walker et al 2010), *Pelagibacter ubique* str.HTCC 1062(Giovannoni et al 2005), Candidatus *Ruthia magnifica* str.Cm (Newton et al 2007), Saanich Inlet OMZ SUP05 (Walsh et al 2009), Candidatus *Vesicomyosocius okutanii* str.HA (Kuwahara et al 2007), *Acidimicrobium ferrooxidans* str.ICP (Clum et al 2009), *Marinobacter aquaeolei* str.VT8 (Singer et al 2011), *Planctomyces limnophilus* str.Mu290 (Labutti et al 2010), *Pseudoalteromonas atlantica* str.T6c, *Thermofilum pendens* str.Hrk5 (Anderson et al 2008), Marine *gammaproteobacterium* HTCC 2080 (Thrash et al 2010a), *Endoriftia persephone* str.Hot96 (Robidart et al 2008), *Gammaproteobacterium* HTCC 5015 (Thrash et al 2010b), *Deltaproteobacterium* MLMS-1, *Gammaproteobacterium* NOR51-B. The GB SUP05 cluster closely with the Saanich Inlet OMZ SUP05 and are highlighted in the ESOM map shown in Supplementary Figure 3. Further, reciprocal BLASTX (Altschul et al 1990) searches were performed against a database containing Saanich Inlet OMZ SUP05, *Candidatus* Ruthia magnifica and *Candidatus* Vesicomyosocius okutanii using the following cutoffs ($E \leq 1 \times E^{-5}$, *percent identity≥60%*) were done and contigs were recruited to the GB SUP05 bin if all genes on the contig hit at least two of the three genomes in the database.

Separation of GB-1 and GB-2 contigs

Contigs identified as SUP05 through ESOM were separated into GB-1 and GB-2 populations using a combination of BLASTN, comparison of gene order on contigs, and contig scaffolding by reciprocal blast searches with Saanich Inlet OMZ SUP05, Candidatus Ruthia magnifica, and Candidatus Vesicomyosocius okutanii. Scaffolding of GB-1 & 2 contigs was performed manually using the following approaches: (i) Comparison of gene orders with the above mentioned SUP05 genomes; (ii) BLASTN on overlapping contigs; (iii) BLASTX on overlapping contigs/individual genes on contigs. Comparison of gene orders was performed starting from the contigs containing the 16s rRNA gene to enable identification of the individual genomes. BLASTN and BLASTX were utilized to resolve the contigs where separation on the basis of gene order was difficult. Average amino acid identity between the two GB SUP05 genomes was 83%, enabling separation using BLASTX where other methods failed. Average coverage data for both genomes was virtually identical (~13x) and was hence not useful for separation of contigs. Thus, manual curation produced two distinct, almost complete genomes as outlined by the presence of distinct rRNA genes and conserved single copy genes outlined in Supplementary Table 3.

Absence of genes involved in nitrate reduction

The absence of genes for nitrate reduction in the genomes of GB-1 and GB-2 is supported by the following lines of evidence. First, whereas the overall GB SUP05 genomes exhibit a high degree of synteny with other SUP05 genomes, the nitrate reduction genes are absent from the GB SUP05 genomic locus in which they reside in the Saanich Inlet SUP05. Gene order and synteny

of flanking regions are conserved between these two genomes. Manual examination of the genomic assembly at this locus confirmed that there are no apparent chimeras or fluctuations in read coverage that might indicate assembly error. Second, unassembled reads and non-SUP05 contigs were searched exhaustively for nitrate reduction genes. Although some were recovered, their average read coverage was far lower than the read coverage of other parts of the SUP05 genome (Supplementary Figure 5). The contigs containing these genes are too small (< 2 Kb) to be conclusively binned and assigned as SUP05 contigs by tetranucleotide frequency. A number of these genes are similar to those from OMZ SUP05 (Supplementary Table 4), and we hypothesize that they belong to minor populations of SUP05 in the GB microbial community. Indeed, there is evidence of numerous low-abundance SUP05 sub-groups. Finally, metatrancriptomes generated using two different sequencing technologies (454 Titanium and Illumina) failed to detect significant numbers of transcripts for nitrate reduction, and no transcripts were detected for *norB* and *norC*, which are involved in reduction of NO to N_2O.

SSU rRNA gene analysis and phylogenetic tree

Small Subunit Ribosomal RNA (SSU rRNA) gene libraries constructed previously (Dick and Tebo 2010), GB-1 and GB-2 SSU rRNA gene sequences described here, and 33 other representative SSU rRNA gene sequences belonging to the groups SUP05 and ARCTIC96BD-19 were aligned with MOTHUR (Schloss et al 2009) using the SILVA reference alignment (Pruesse et al 2007). The aligned sequences were imported into ARB (Ludwig et al 2004) and the GB-1 and GB-2 group clusters were identified based on their position in the SUP05 group within the SILVA reference tree. The phylogenetic tree (Supplementary Figure 2) was constructed using PHYML (Guindon and Gascuel 2003) using the Hasegawa, Kishino and Yano (HKY)

(Hasegawa et al 1985) model of evolution and estimated values for the proportion of invariable sites, the transition/transversion ratio, α parameter of the γ distribution, and 4 substitution categories. A consensus tree was built from 1000 bootstrap replicates to determine the confidence of each node.

Ni, Fe Hydrogenase gene analysis and phylogeny

Representative group 1 Ni, Fe hydrogenase large subunit sequences identified in Vignais et al (2007) were aligned with GB-1 and GB-2 hydrogenases and the unassigned putative SUP05 hydrogenase by MUSCLE (Edgar 2004). The phylogenetic tree (Fig. 2.4) was constructed with PHYML using the HKY model of evolution and estimated values for the proportion of invariable sites, the transition/transversion ratio, α parameter of the γ distribution, and 4 substitution rate categories. 1000 bootstrap replicates were used to determine the consensus tree and analyze the confidence in each node. The two distinct hydrogenase forms, the 'Hydrothermal vent HupL' and the 'Epipelagic/Pelagic Ocean HydB' were identified based on their position in the hydrogenase tree and their sample source.

Accession numbers for sequences in tree: Citreicella sp.SE45 (NZ_GG704601), *Sagittula stellata* E37 (NZ_AAYA01000022), *Oligotropha carboxidovorans* OM5 (NC_005873), Oceanospirillum sp.MED92 (NZ_AAOW01000009), Azospirillum sp.B510 plasmid pAB510c (288913807), Beijerinckia indica ATCC9039 (CP001016), *Methylibium petroleiphilum* PM1 (CP000555), *Ralstonia eutropha* H16 megaplasmid pHG1 (AY305378), *Xanthobacter autotrophicus* Py2 (CP000781), *Rhodobacter capsulatus* SB1003 (CP001312), *Methylococcus capsulatus* str.Bath (AE017282), *Dechloromonas aromatica* RCB(CP000089), *Acidiphilium multivorum* AIU301 (325049009), *Bathymodiolus puteoserpentis* Endosymbiont clone

Logatchev 281ROV3 (FR851255), *Bathymodiolus aff.thermophilus* Endosymbiont clone PacificAntarctic_30GTV2 (FR851256), Bathymodiolus sp. Endosymbiont clone Lilliput 200ROV9 (FR851257), Bathymodiolus sp. Endosymbiont clone Wideawake 125ROV12 (FR851258), *Rimicaris exoculata* Ectosymbiont clone SWE17286 (FR851274), *Rimicaris exoculata* Ectosymbiont clone SWE17628 (FR851269), (ROOT) *Candidatus* Kuenenia stuttgartiensis (CT573072), *Robiginitalea biformata* HTCC2501 (CP001712), *Nautilia profundicola* AmH (CP001279), Flavobacteria bacterium MS024-2A (ABVV01000001), *Hydrogenovibrio marinus* H110 (AB070719), Hyphomicrobium sp.MC1 (338736863), *Sulfolobus islandicus* HVE10/4 (CP002426), SAR324 cluster bacterium JCVI-SC AAA05 (AGAU01000544), Gamma proteobacterium HIMB30 (NZ_AGIG01000031).

To assess the global distribution of hydrogenases described here, we searched publicly available metagenomic datasets from environments where SUP05 may be present (NCBI/Genbank – Nucleotide collection (nr/nt), whole-genome shotgun contigs (wgs), transcriptome shotgun assembly (TSA), high throughput genomic sequences (HTGS), expressed sequence tags (est), non-redundant protein sequences (nr), metagenomic proteins (env_nr), Joint Genome Institute-Integrated Microbial Genomes(JGI-IMG) and CAMERA). Several SUP05-like hydrogenases were detected in a number of marine environments (Supplementary Table 7). Many of these hits were to single reads in Sequence Read Archives and hence cannot be definitely assigned to SUP05. The only matches to assembled contigs were in a set of 110 unpublished metagenomic datasets from oxygen minimum zones of Saanich Inlet and the Eastern North Pacific Subarctic oceans (IMG Project ID:1785). However, because these genes are significantly different in terms of sequence and operon structure to the genes we report here, they cannot be confidently

assigned to SUP05. We note that metagenomic datasets in which SUP05 are abundant are currently few in number and limited in terms of sequence coverage, so the results presented here provide only a preliminary glimpse into the geographic distribution of SUP05-like hydrogenase genes. Overall, these preliminary results are consistent with a variable distribution of hydrogenase genes in SUP05, with presence and type of hydrogenase genes linked to environmental availability of H_2.

Thermodynamic model for plume concentrations

Equilibrium thermodynamic reaction path modeling was used to predict mineral precipitation, chemical concentrations, and activity coefficients resulting from the mixing of seawater with Guaymas Basin endmember vent fluid (Supplementary Table 5). Our approach follows those of previous studies (Bowers et al 1985, Janecky and Seyfried 1984, McCollom 2000b). The geochemical predictions of our specific implementation have compared well with observed plume mineralogy from 9°N East Pacific Rise (Breier et al Submitted). The plume reaction path is modeled through a mixing path that ends at a vent fluid to seawater dilution of 1 part in 10,000, representing the dilution achieved at the non-buoyant plume heights sampled in this study. Upper limits on the available chemical energy for the chemosynthetic metabolisms listed in Supplementary Table 6 were estimated from predicted plume chemistry. To achieve an upper limit constraint on available energy, abiotic dissolved phase redox reactions were not permitted. The only abiotic redox reactions permitted during the plume reaction path were the precipitation of elemental S, bornite, pyrite, chalcopyrite, and covellite using H_2 producing reactions (McCollom 2000b).

Vent fluid chemical composition was based on measurements made in 1982 and 2000 (Von Damm et al 1985, Von Damm et al 2005). *In situ* pH was calculated from measurements of pH at 25° C using an equilibrium reaction path model that increased the temperature of the measured fluid to the original vent fluid temperature. Similarly, the endmember vent fluid $S_2O_3^-$ concentration, a known electron donor in seafloor hydrothermal communities, was estimated as the concentration that results from the thermodynamic equilibrium speciation of the measured H_2S. The concentration of vent fluid NH_3 is based on the measurements of Von Damm et al *(28)*. Thermodynamic equilibrium modeling predicts vent fluid dissolved inorganic N to exist predominantly as NH_3; the predicted concentrations of vent fluid NO_3^- and NO_2^- are vanishingly small and assumed to be zero in this case. Vent fluid N_2 concentrations are assumed to be 83% of background seawater concentrations as a result of a 17% conversion of seawater N_2 to NH_3 during hydrothermal circulation (Brandes et al 1998). Note, these estimates assume thermodynamic equilibrium of S and N species in endmember vent fluid, which may not be valid. Abiotic redox reactions were permitted when estimating thermodynamic equilibrium speciation in the endmember vent fluid with the exception of N_2 to NO_3^- equilibration, which was suppressed in order to apply the experimentally based constraint of Brandes et al (Brandes et al 1998). Background seawater dissolved O_2 concentration was based on previous measurements reported for Guaymas Basin hydrothermal plumes (Campbell and Gieskes 1984). Background seawater dissolved NO_3^- and NO_2^- concentrations are based on measurements from WOCE section P18 *(32)*. Background seawater dissolved H_2 and N_2 concentrations are assumed to be controlled by atmospheric equilibrium; this is consistent with previous findings for H_2 in the Atlantic Ocean (Conrad and Seiler 1988) and N_2 in the Pacific Ocean (Weiss and Craig 1973).

Note, this available data pre-dates this study; actual vent chemistry at the time of this study may have been different.

Reaction path modeling was performed with REACT, part of the Geochemist's Workbench package (Bethke 2007). Following assumptions used in previous models, conductive cooling was neglected and mixture temperatures were a strict function of conservative end-member mixing. Precipitated minerals were allowed to dissolve and their constituents to re-precipitate based on thermodynamic equilibrium constraints. Gibbs free energies of reaction for the metabolisms in Supplementary Table 6 were predicted using

$$\Delta G = \Delta G^{\circ} + RT\ln(Q), \quad (1)$$

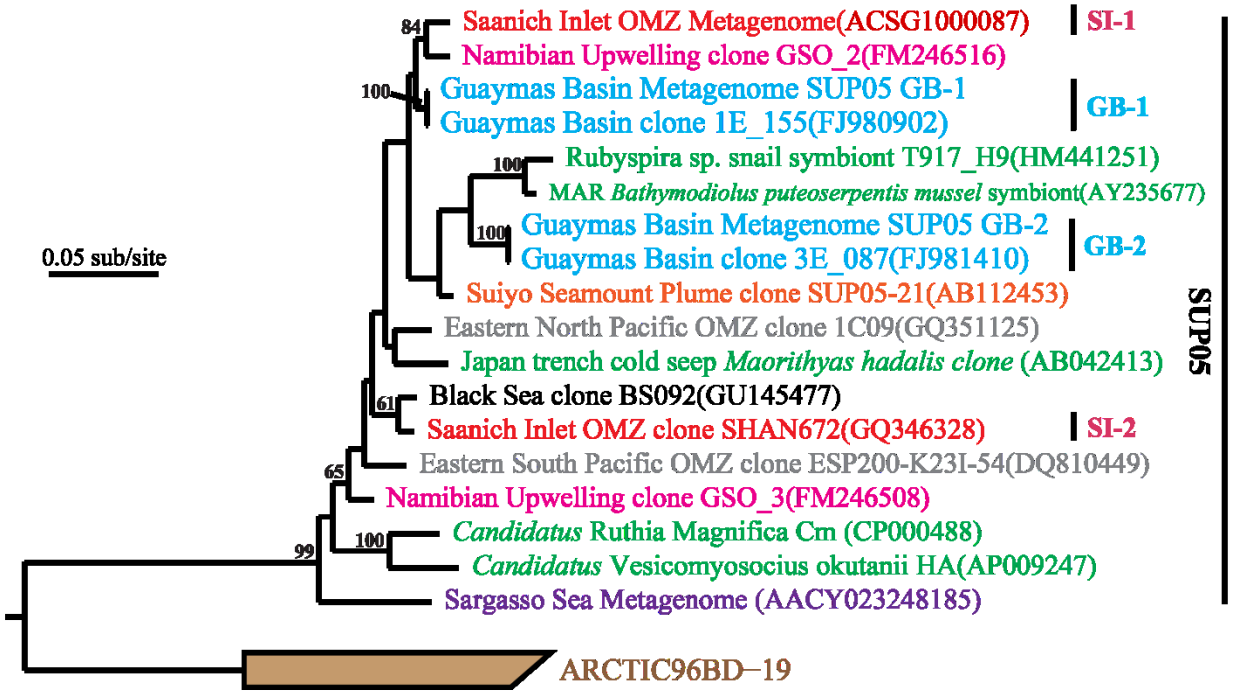
where ΔG is the Gibbs free energy of reaction, ΔG° is the standard Gibbs free energy of reaction, R is the universal gas constant, T is the absolute temperature, and Q is the reaction quotient. Using the generic reaction $aA + bB \leftrightarrow cC + dD$ as an example, Q is defined to be

$$Q = \frac{\gamma_c[C]^c \gamma_d[D]^d}{\gamma_a[A]^a \gamma_b[B]^b}, \quad (2)$$

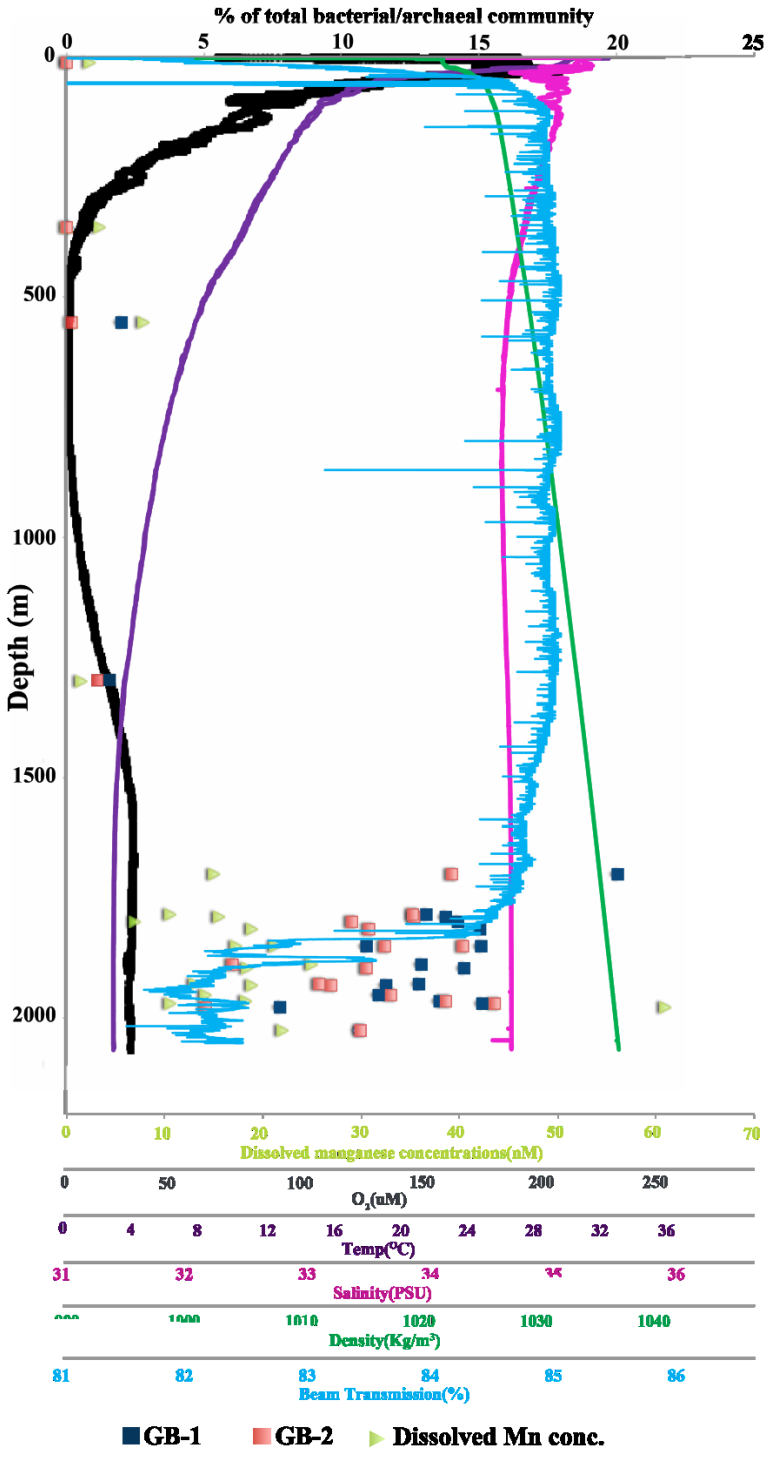
where γ_c, γ_d, γ_a, and γ_b are the activity coefficients of chemical species A, B, D, and C; [A], [B], [C], and [D] are their molar concentrations, and a, b, c, and d are their stoichiometric coefficients. The available energy per kilogram plume fluid was estimated by calculating ΔG for each metabolic reaction using the results of the reaction path model, and multiplying ΔG by the concentration of the most limiting reactant (McCollom 2000b). The most limiting reactant was determined for each metabolism from among those reactants listed in Supplementary Table 6 with the exception of H^+, which was assumed to never be limiting due to the self-ionization of water. At each point in the reaction path, metabolic activity was allowed to modify the

availability of products and reactants with the exception of H^+. This was done by (*i*) calculating the reactant consumption and product yield for each metabolism in the sequential order from greatest to least energy yield in terms of ΔG per electron transferred, and (*ii*) iterating through this process until the total energy at that point the reaction changed by <1%. Gibbs free energies are reported on a per kg plume fluid basis and on a per kg vent fluid basis by multiplying the former by the mass of the plume solution into which the 1 kg vent fluid was mixed.

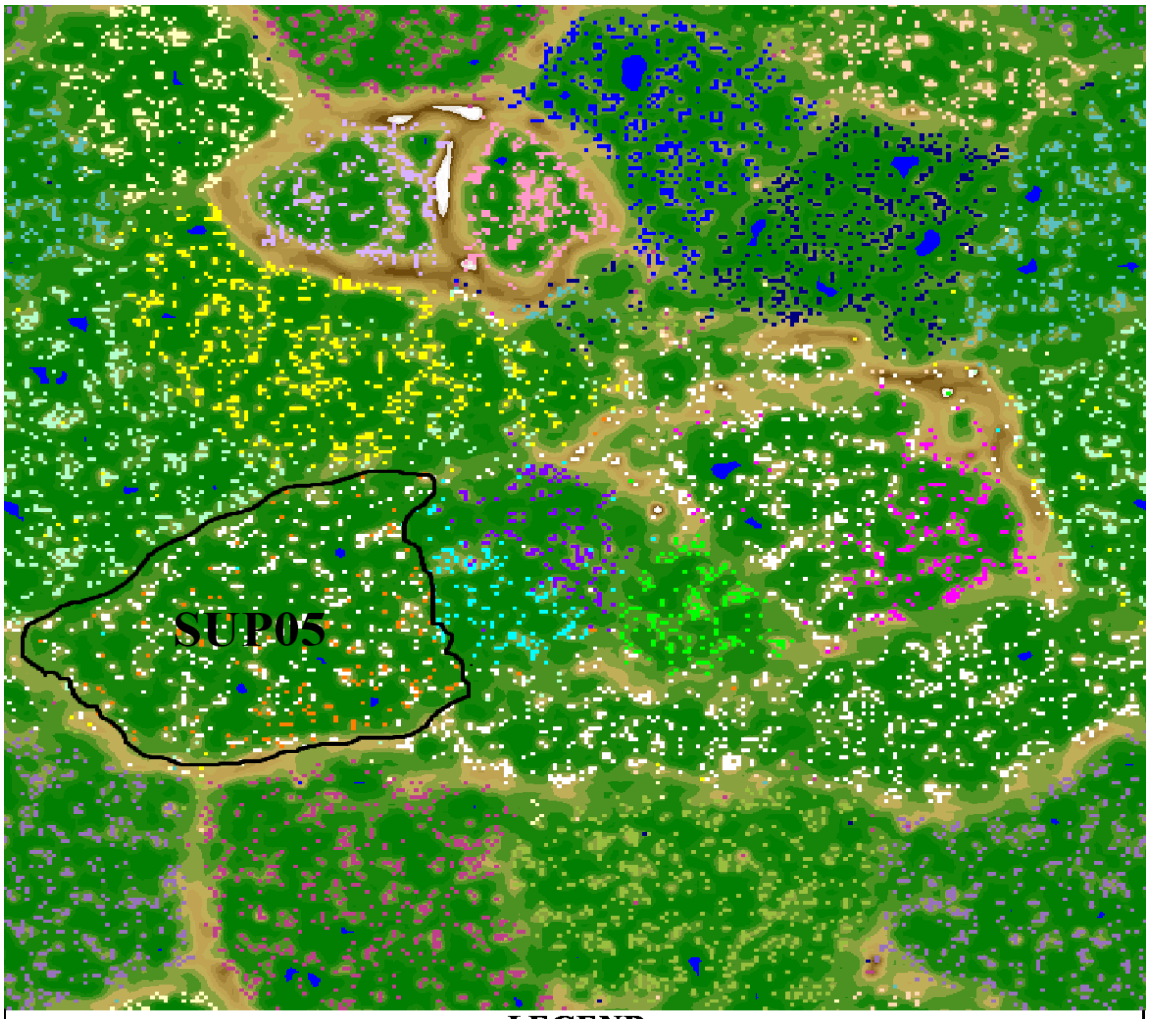
Thermodynamic data was predicted by SUPCRT95 (Johnson et al 1992) for the temperature range of 1-425° C (specifically 1, 25, 60, 100, 225, 290, 350, and 425° C) and a pressure of 500 bar, a pressure and temperature range that is representative of all known deep sea vents. Standard Gibbs free energy of reactions for the metabolisms in Supplementary Table 6 were similarly predicted by SUPCRT95 at 1, 25, and 100° C and linearly interpolated as needed up to a temperature of 121° C. SUPCRT95 uses previously published thermodynamic data for minerals, gases, and aqueous species (Helgeson et al 1978, McCollom and Shock 1997, Saccocia and Seyfried Jr 1994, Shock and Helgeson 1988, Shock et al 1989, Shock et al 1997, Sverjensky et al 1997). Thermodynamic data for pyrolusite, bixbyite, hausmannite, marcasite, and $Fe(OH)_3$ were added for our study (Robie et al 1979, Wagman et al 1982). The B-dot activity model was used (Helgeson 1969, Helgeson and Kirkham 1974). Temperature dependent activity coefficients were used for aqueous CO_2 and water in a NaCl solution (Bethke 2007, Cleverley and Bastrakov 2005, Drummond 1981).



Supplementary Figure 1. Phylogenetic tree of SUP05 group SSU rRNA genes inferred by maximum likelihood. Sequences are colored by origin; blue – Guaymas Basin; red – Saanich Inlet; green – mussel, snail and clam symbionts; pink – Namibian upwelling zone; black – Black Sea, grey – Eastern Pacific, purple – Sargasso sea; orange – Suiyo Seamount.

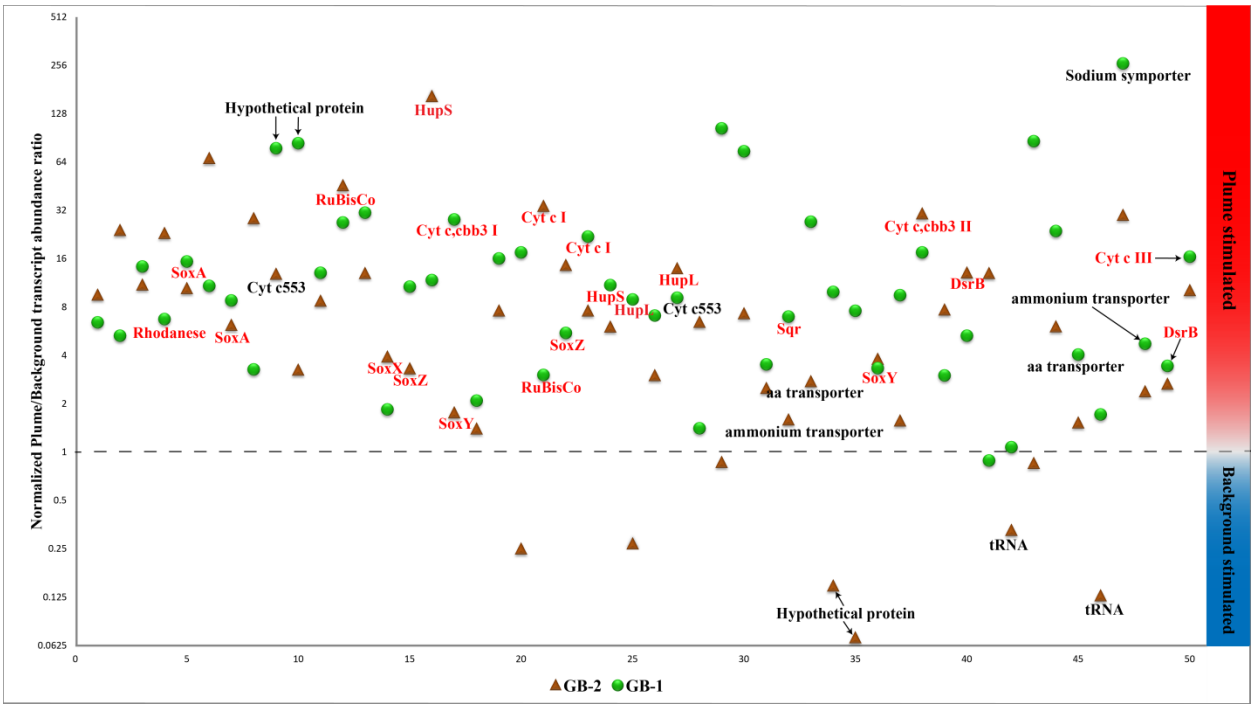


Supplementary Figure 2. Abundance of SUP05 GB-1 and GB-2 as a percentage of the total bacterial/archaeal community (squares) through the water column at Guaymas Basin based SSU rRNA gene amplicon sequencing. Dissolved Mn concentrations are used as a proxy for plume strength. Note that Mn and SUP05 data come from several different CTD casts whereas other water column parameters are from one representative cast.

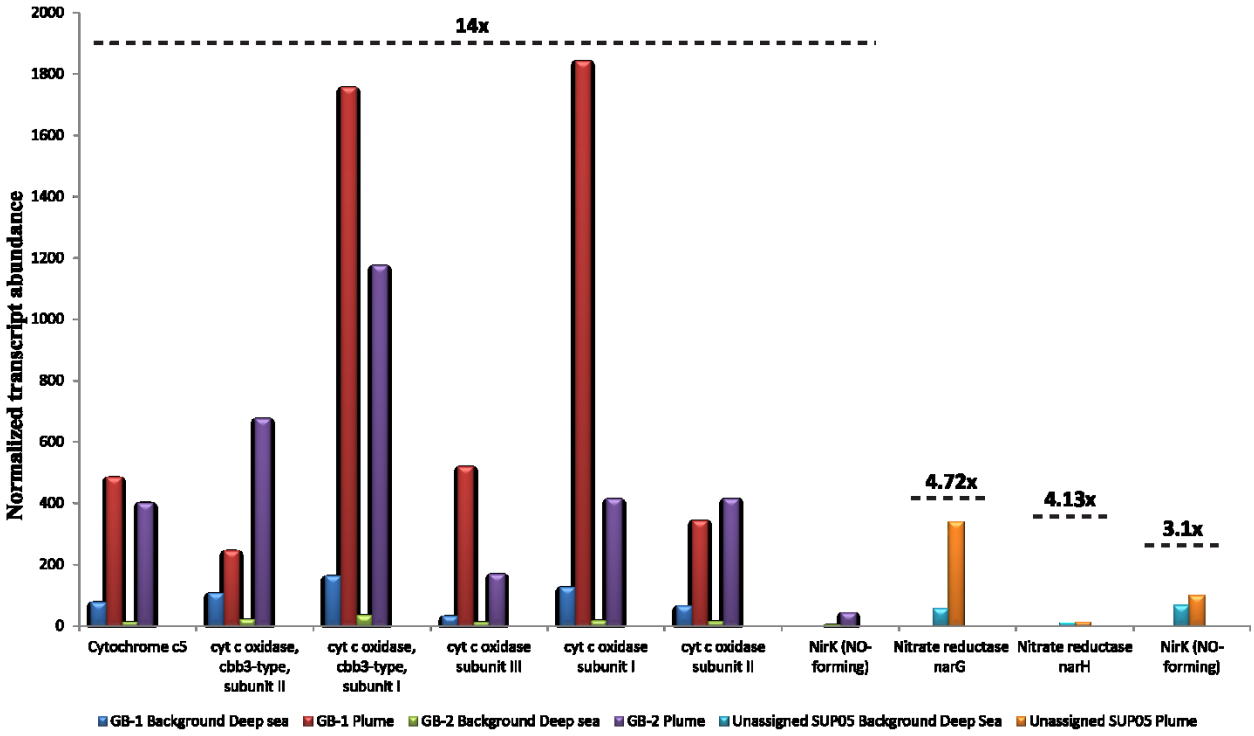


LEGEND			
Colour	**Organism**	**Colour**	**Organism**
	Guaymas Basin metagenome		Marinobacter aquaeolei VT8
	Alteromonas macleodii 'Deep ecotype'		Planctomyces limnophilus Mu290
	Methylococcus capsulatus bath		Pseodualteromonas atlantica T6c
	Nitrosopumilus maritimus SCM1		Thermofilum pendens Hrk5
	Pelagibacter ubique HTCC1062		Marine gammaproteobacterium HTCC 2080
	Candidatus Ruthia magnifica Cm		Endoriftia persephone Hot96
	Saanich Inlet SUP05		Gammaproteobacterium HTCC 5015
	Candidatus Vesicomyosocius okutanii HA		Deltaproteobacterium MLMS-1
	Acidomicrobium ferrooxidans ICP		Gammaproteobacterium NOR51-B

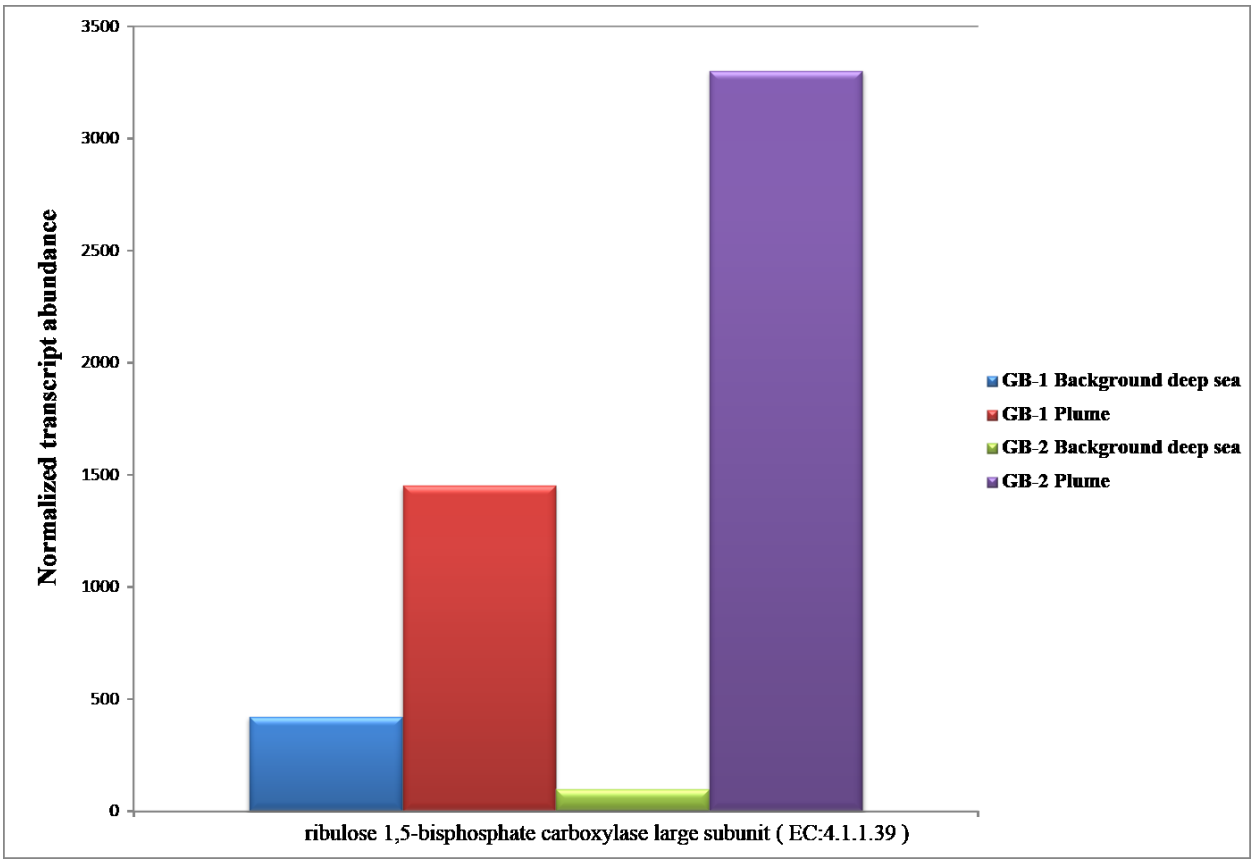
Supplementary Figure 3. Assignment of assembled contigs to specific populations using ESOM implemented with tetranucleotide frequencies. Each colored point represents a contig from a different source population. Background color is dictated by tetranucleotide frequency variance between data points with green/blue indicating similarity and brown ridges demarcating different genomes. **NOTE**: The bound region marked 'SUP05' was defined by the Saanich Inlet SUP05 genome (contigs in orange) that cluster along with the Guaymas Basin SUP05 (contigs in white).



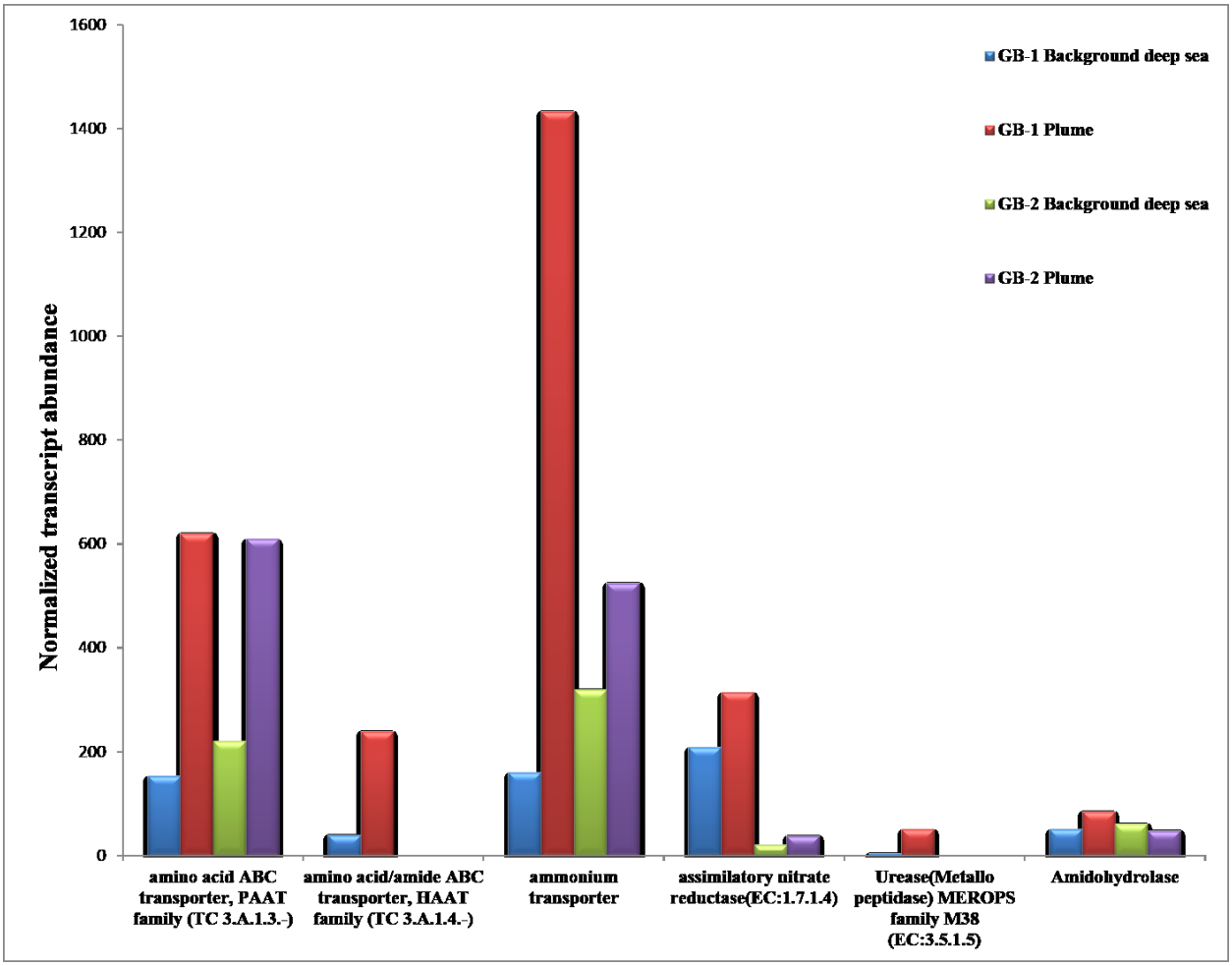
Supplementary Figure 4. Transcript abundance ratio in plume versus background (P/B) for top 50 genes in GB-1 and GB-2, in order of total transcript abundance from left to right. Names of important genes are indicated below symbols. Genes above the dotted line have transcripts that are enriched in the plume, and genes below have transcripts that are enriched in background. Transcript abundance is normalized for gene length and total number of reads per dataset. Genes involved in chemolithoautotrophy are shown in red.



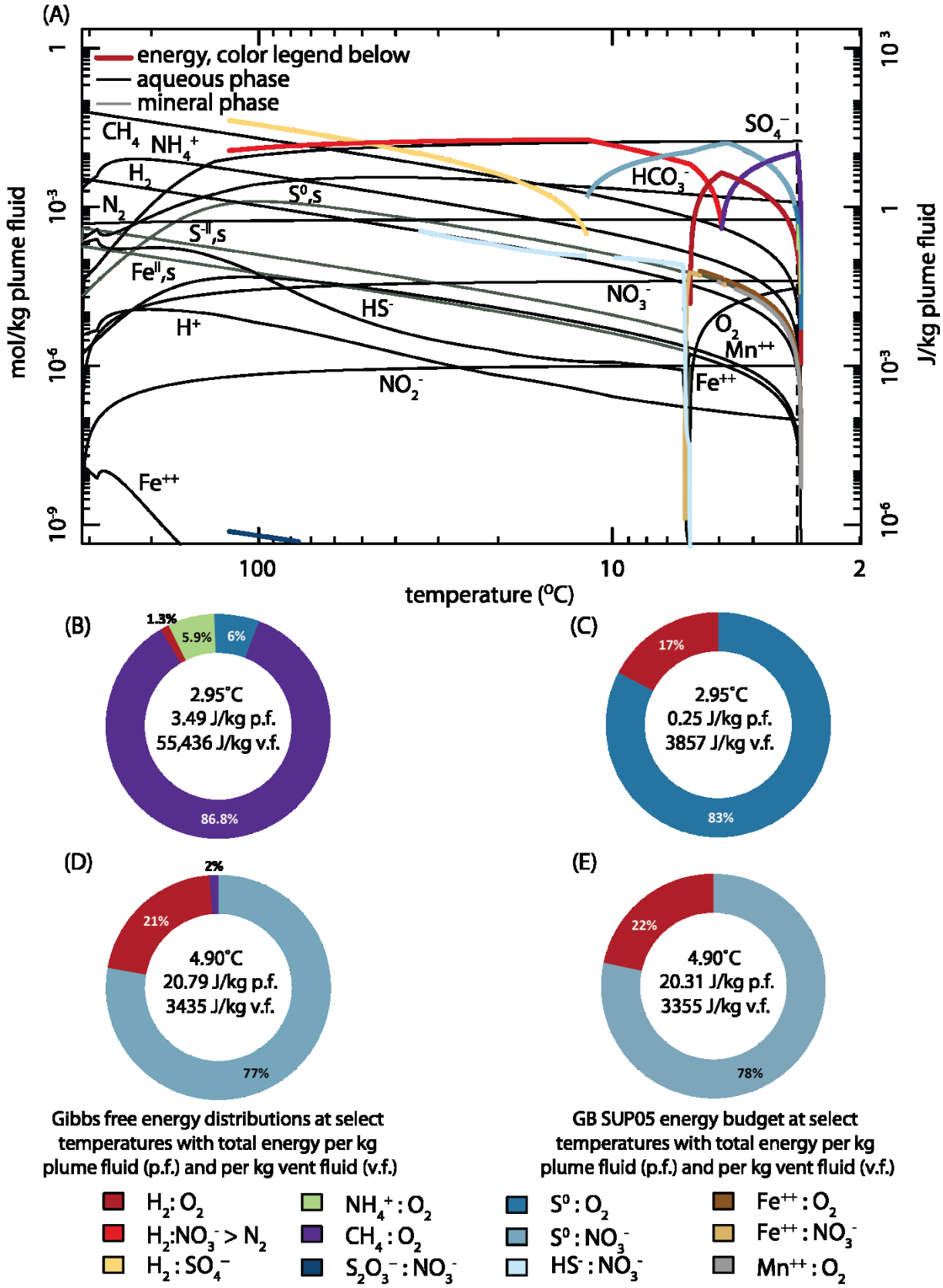
Supplementary Figure 5. Histograms depict the transcript abundance for genes involved in reduction of electron acceptors. Note that GB-1 does not contain a NirK. Dotted line indicates a contiguous genomic fragment with the average coverage listed above. Transcript abundance is normalized for gene length and total number of reads per dataset.



Supplementary Figure 6. Abundance of transcripts for RuBisCo gene involved in carbon fixation. Note: Only a partial RuBisCo gene was recovered for GB-1. Transcript abundance is normalized for gene length and total number of reads per dataset.



Supplementary Figure 7. Abundance of transcripts for key genes involved in Nitrogen assimilation. Note that GB-2 does not possess genes for amino acid/amide ABC transporter (HAAT family) or urease. Transcript abundance is normalized for gene length and total number of reads per dataset.


(A)
energy, color legend below
aqueous phase
mineral phase
CH4
NH4+
H2
N2
S0,s
SII,s
FeII,s
H+
HS-
NO2-
Fe++
SO4--
HCO3-
NO3-
O2
Mn++
Fe++
mol/kg plume fluid
J/kg plume fluid
1
10-3
10-6
10-9
10 3
1
10-3
10-6
100
10
2
temperature (°C)
(B)
1.3%
5.9%
6%
2.95°C
3.49 J/kg p.f.
55,436 J/kg v.f.
86.8%
(C)
17%
2.95°C
0.25 J/kg p.f.
3857 J/kg v.f.
83%
(D)
2%
21%
4.90°C
20.79 J/kg p.f.
3435 J/kg v.f.
77%
(E)
22%
4.90°C
20.31 J/kg p.f.
3355 J/kg v.f.
78%
Gibbs free energy distributions at select temperatures with total energy per kg plume fluid (p.f.) and per kg vent fluid (v.f.)
GB SUP05 energy budget at select temperatures with total energy per kg plume fluid (p.f.) and per kg vent fluid (v.f.)
H2 : O2
H2 :NO3- > N2
H2 : SO4--
NH4+ : O2
CH4 : O2
S2O3-- : NO3-
S0 : O2
S0 : NO3-
HS- : NO3-
Fe++ : O2
Fe++ : NO3-
Mn++ : O2

Supplementary Figure 8. (A) Plume fluid chemical concentrations of hydrothermal vent end members and available free energies of catabolic reactions. The free energies are normalized to kg plume fluid. Dotted lines show energy budget. **(B) and (C)** Free energies of catabolic reactions in the hydrothermal plume at 2.95^{o}C, the typical temperature of plumes in this study. **(B)** Energy of catabolic reactions as a percentage of total available free energy; **(C)** energy of catabolic reactions as a percentage of SUP05-specific metabolisms. **(D) and (E)** Free energies of catabolic reactions in the hydrothermal plume at 4.90^{o}C, a representative temperature for the rising hydrothermal plume **(D)** Energy of catabolic reactions as a percentage of total available free energy; **(E)** energy of catabolic reactions as a percentage of SUP05-specific metabolisms. Total available free energy in the plume is normalized per kg plume fluid and per kg vent fluid as shown in inset.

Supplementary Table 1. Guaymas Basin Sample summary

Cast	Date	Latitude Longitude	Depth (m)	tMn(nM)	dMn(n M)	O_2(µM)	Temp (C)	No. of sequencing reads (SSU rRNA amplicon unless noted)
31-12#4	8/7/2005	27° 0.0311 N 111° 0.438 W	1901	24.57	20.93	24.66	2.91	133
11-2#14	7/18/2004	27°0.823 N 111°24.654 W	1996	356	145	27.4	2.96	576,187 (DNA)
11-1#2	7/11/2004	27° 0.95 N 111° 25.5 W	1993	266	78	29.13	2.91	1785
31-12#2	8/7/2005	27° 0.0311 N 111° 0.438 W	1979	61	15	25.23	2.91	930
11-2#8	7/18/2004	27° 1.09 N 111° 25.06 W	1969	217	40	27.07	2.93	1180
21-6#2	5/2/2005	27°2.240 N 111°21.790 W	1963	315	62	26.1	2.93	664,240 (cDNA)
12-8#4	7/14/2004	27° 28.55 N 111° 22.72 W	1954	219	19	27.74	2.92	1069
12-27a#1	7/16/2004	27°30.360 N 111°20.818 W	1950	288	40	27.7	2.93	894,665 (cDNA), 103,078,758 (cDNA)
11-1#4	7/11/2004	27° 0.95 N 111° 25.5 W	1942	175	28	28.41	2.92	1498
11-1#14	7/11/2004	27° 0.95 N 111° 25.5 W	1910	287	55	27.57	2.93	1010
11-1#6	7/11/2004	27° 0.95 N 111° 25.5 W	1907	218	22	27.05	2.92	2079
34-2#7	8/4/2005	26°22.755 N 110°43.433 W	1900	21	7	46.5	2.59	406,533(DNA), 504,086(cDNA)
11-1#10	7/11/2004	27° 0.95 N 111° 25.5 W	1875	222	21.84	27.12	2.92	1581
11-2#4	7/18/2004	27° 1.09 N 111° 25.06 W	1853	235	25	26.62	2.93	1337
12-8#6	7/14/2004	27° 28.55 N 111° 22.72 W	1852	186	16	27.87	2.92	1211
31-12#20	8/7/2005	27° 0.0311 N 111° 0.438 W	1799	7.28	11.83	22.02	2.92	1369
11-1#12	7/11/2004	27° 0.95 N 111° 25.5 W	1797	85	13	26.44	2.93	1136
11-2#2	7/18/2004	27° 1.09 N 111° 25.06 W	1791	183	46	27.13	2.93	1005
11-2#10	7/18/2004	27° 1.09 N 111° 25.06 W	1785	110	9	27.57	2.93	994
11-1#8	7/11/2004	27°1.852 N 111°24.000 W	1775	257	59	27	2.95	563,818 (DNA)
31-12#22	8/7/2005	27° 0.0311 N 111° 0.438 W	1700	15.47	12.74	21.82	2.94	983
12-8#12	7/13/2004	27°29.174 N 111°21.844 W	1600	55	5	28.5	2.97	358,335(DNA), 514,607(cDNA), 122,259,588 (cDNA)
12-8#14	7/14/2004	27° 28.55 N 111° 22.72 W	1296	7	7	20.1	3.57	1199
11-1#18	7/11/2004	27° 0.95 N 111° 25.5 W	550	6	5	1.88	7.82	1359
11-1#20	7/11/2004	27° 0.95 N 111° 25.5 W	353	10	11	6.98	10.83	693
11-1#24	7/11/2004	27° 0.95 N 111° 25.5 W	12.2	14	6	195.22	28.07	712

Supplementary Table 2. Guaymas Basin SUP05 metagenome features and comparison to Saanich Inlet OMZ SUP05 metagenome (*13*).

Feature	SUP05 GB-1	SUP05 GB-2	OMZ SUP05
No. of Contigs	100	121	90
Total Length (Mb)	1.24	1.26	1.16
Average Coverage	13.07	12.59	7.25
Average G+C Content	40	40	40
No. of ORFs	1401	1387	1333
tRNAs	34	32	29
16s-5s-23s Operon	1	1	1
16s rRNA % identity between GB-1 & GB-2	96.7%		

Supplementary Table 3. Identification of conserved genes in Guaymas Basin SUP05 metagenome

COG Family	Conserved Gene	GB-1 Metagenome gene tag	GB-2 Metagenome gene tag	SI SUP05 Accession number
	Large subunit ribosomal proteins			
COG0080	L11	2062243515	2062122756	EEZ80670
COG0081	L1	2062243516	2062402548	EEZ80669
COG0087	L3	2062249875	2062215754	EEZ79789
		2062249876		
COG0091	L22	2062251116	2062215759	EEZ79794
COG0093	L14	2062251111	2062215764	EEZ79798
COG0094	L5	2062251109	2062215766	EEZ79800
COG0097	L6P/L9E	2062251106	2062215769	EEZ80748
COG0102	L13	2062222949	2062308230	EEZ80067
COG0197	L16/L10E	2062251114	2062215761	EEZ79796
COG0200	L15	2062251102	2062215773	EEZ80752
COG0256	L18	2062251105	2062215770	EEZ80749
	Small subunit ribosomal proteins			
COG0048	S12	2062315196	2062122763	EEZ80549
COG0049	S7	2062315195	2062122762	EEZ80550
COG0052	S2	2062136934	2062107639	EEZ80632
COG0092	S3	2062251115	2062215760	EEZ79795
COG0096	S8	2062251107	2062215768	EEZ80747
COG0098	S5	2062251104	2062215771	EEZ80750
COG0099	S13	2062251100	2062215775	EEZ80754
COG0100	S11	2062251099	2062215776	EEZ80755
COG0103	S9	2062222950	2062308231	EEZ80066
COG0184	S15P/S13E	2062418166	2062215738	EEZ80586
COG0186	S17	2062251112	2062215763	EEZ79797
COG0522	S4	2062251098	2062215777	EEZ80756
	tRNA synthetases			
COG0016	Phenylalanyl-tRNA synthethase alpha subunit	2062411615	2062413646	EEZ79997
COG0018	Arginyl-tRNA synthetase	2062414344	2062294163	EEZ79861
COG0060	Isoleucyl-tRNA synthetase	2062410385	2062169988	EEZ79547
		2062410384	2062169987	
COG0124	Histidyl-tRNA synthetase	2062411648	2062172528	EEZ80161
		2062411649		
COG0143	Methionyl-tRNA synthetase	2062420751	2062406464	EEZ80334
COG0172	Seryl-tRNA synthetase	2062165452	2062412346	EEZ80031
COG0201	Preprotein translocase subunit SecY	2062251101	2062215774	EEZ80753
COG0495	Leucyl-tRNA synthetase	2062136955	2062107661	EEZ79547
COG0525	Valyl-tRNA synthetase	2062365274	2062394691	EEZ79599

			2062394690	
	RNA polymerase subunits			
COG0202	DNA-directed RNA polymerase, alpha subunit/40 kD subunit	2062251097	2062215778	N/A
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	2062134380	2062279941 2062279940	EEZ80666
COG0012	Predicted GTPase	2062136931	2062107636	EEZ80197
COG0533	Metal-dependent protease	2062165482	2062121465 2062121466	ACX30598

NOTE: Genes listed are the 36 universally conserved genes including the 32 universally conserved single copy genes (Ciccarelli et al 2006). Gene tags highlighted in red are potential split genes

Supplementary Table 4. GB SUP05 genes for denitrification

Gene	Number of copies	Average Coverage	Reaction catalyzed	Best BLAST hit in Saanich Inlet SUP05	% amino acid identity
periplasmic nitrate reductase subunit NapA (EC:1.7.99.4)	1	1.95	$NO_3^- \rightarrow NO_2^-$	EEZ79720	68.9
respiratory nitrate reductase alpha subunit narG (EC:1.7.99.4)	10	4.72	$NO_3^- \rightarrow NO_2^-$	EEZ79693	70.3
respiratory nitrate reductase beta subunit narH (EC:1.7.99.4)	4	4.13	$NO_3^- \rightarrow NO_2^-$	EEZ79694	54.1
dissimilatory nitrite reductase (NO-forming) nirK (EC:1.7.2.1)	3	3.10	$NO2^- \rightarrow NO$	EEZ80799	58.9
molybdopterin biosynthesis protein MoeA	4	2.97	Cofactor	EEZ80457	42.4
molybdopterin biosynthesis protein MoeB	5	3.83	Cofactor	EEZ80005	67.8

Supplementary Table 5. Modeling endmembers.

	Guaymas vents[a]	Seawater
T (°C)	315	2.93
pH[b]	5.9	8
O_2, aqueous	0	0.03[c]
NH_4^+ [d]	13.6	0
N_2, aqueous	0.48[e]	0.58[e]
NO_3^-	0	0.04
NO_2^-	0	0.001
H_2, aqueous	3.4[f]	0.0000004
SO_4^{2-}	0	28.0
H_2S, aqueous	5.98	0
$\sum CO_2$, aqueous	61.1[f]	1.8
CH_4, aqueous	63.4[f]	0
Cl^-	637	540
Na^+	513	464
Ca^{2+}	41.5	10.2
Mg^{2+}	0	52.2
K^+	48.5	10.1
SiO_2, aqueous	13.8	0.17
Fe	0.18	0
Mn^{2+}	0.24	0
Cu^+	0.001	0
Zn^{2+}	0.040	0
Ba^{2+}	0.054	0

All concentrations in mmol/kg vent fluid.

(a) Reported Guaymas vent chemistry (Von Damm et al 1985).

(b) *In situ* pH based on 25 °C measurement.

(c) Guaymas background dissolved O_2 (Campbell and Gieskes 1984).

(d) Predicted to exist as NH_3 in vent fluid.

(e) Seawater dissolved N_2 from (Weiss and Craig 1973); vent fluid dissolved N_2 assumed to be 83% of seawater concentration (Brandes et al 1998).

(f) Reported dissolved gases (Von Damm et al 2005).

Supplementary Table 6. Metabolic reactions and standard Gibbs free energies at 1, 25, and 100°C.

Metabolism	Reaction	e^{-}[b]	$\Delta G°$[a] (kJ/mol)		
			1°C	25°C	100°C
H_2 oxidation	$H_2 + 0.5O_2 \rightarrow H_2O$	2	-265	-264	-260
H_2–NO_3^-:H_2O–N_2	$NO_3^- + 2.5H_2 + H^+ \rightarrow 0.5N_2 + 3H_2O$	5	-637	-637	-635
Methanotrophy	$CH_4 + 2O_2 \rightarrow HCO_3^- + H^+ + H_2O$	8	-828	-825	-810
HS^- oxidation[c]	$HS^- + 2O_2 \rightarrow SO_4^{--} + H^+$	8	-798	-793	-768
$S_2O_3^{2-}$ oxidation	$S_2O_3^{2-} + 2O_2 + H_2O \rightarrow 2SO_4^{--} + 2H^+$	8	-773	-766	-736
$S_2O_3^{2-}$–NO_3^-: SO_4^{--}–N_2	$S_2O_3^{2-} + 1.6NO_3^- + 0.2H_2O \rightarrow 2SO_4^{2-} + 0.8N_2 + 0.4H^+$	8	-733	-728	-713
S^0 oxidation	$S^0 + 1.5O_2 + H_2O \rightarrow SO_4^{--} + 2H^+$	6	-540	-535	-512
H_2–NO_3^-:H_2O–NO_2^-	$NO_3^- + H_2 \rightarrow NO_2^- + H_2O$	2	-177	-177	-175
S^0–NO_3^-: SO_4^{2-}–N_2	$S^0 + 1.2NO_3^- + 0.4H_2O \rightarrow SO_4^{2-} + 0.6N_2 + 0.8H^+$	6	-510	-507	-494
HS^-–NO_3^-:H_2O–NO_2^-	$NO_3^- + HS^- + H^+ \rightarrow NO_2^- + H_2O + S^0$	2	-170	-170	-172
Fe reduction	$H_2 + 2Fe^{3+} \rightarrow 2\ Fe^{2+} + 2H^+$	2	-89.6	-92.5	100
H_2–N_2: NH_4^+	0.5N2 + 1.5H2 + H+ -> NH4+	3	-217	-216	-207
NH_4^+ oxidation	NH4+ + 1.5O2 -> 2H+ + NO2- + H2O	6	-261	-264	-271
Sulfate reduction	$SO_4^{2-} + 4H_2 + H^+ \rightarrow HS^- + 4H_2O$	8	-261	-264	-271
Methanogenesis	$4H_2 + HCO_3^-\ H^+ \rightarrow CH_4 + 3H_2O$	8	-232	-232	-229
Fe oxidation[c]	$Fe^{2+} + 0.25O_2 + 2.5H_2O \rightarrow Fe(OH)_3,s + 2H^+$	1	-16.0	-16.0	-18.8
Fe^{2+}–NO_3^-: Fe_s^{3+}–N_2	$Fe^{2+} + 0.2NO_3^- + 2.4H_2O \rightarrow Fe(OH)_3,s + 0.1N_2 + 1.8H^+$	1	-11.0	-11.3	-15.8
Mn oxidation	$Mn^{2+} + 0.5O_2 + H_2O \rightarrow MnO_2,s + 2H^+$	2	-5.65	-5.71	-4.62

(a) Standard Gibbs free energies of reaction predicted by SUPCRT95 (Johnson et al 1992).

(b) Number of electrons transferred during the reaction.

(c) Reactions using particulate and aqueous phase e^- donors were predicted individually.

Supplementary Table 7. Best Hits to SUP05 Hydrogenases in other databases

		Hydrothermal vent/plume HupL				Epipelagic/Pelagic Ocean HydB			
Environment/Metagenome	**Database (CAMERA/ NCBI/IMG)**	**Best Hit**	**E value**	**Max identity**	**Bit score**	**Best Hit**	**E value**	**Max identity**	**Bit score**
Saanich Inlet OMZ SUP05	nr (NCBI)	x	x	x	x	x	x	x	x
SUP05 Mussel symbionts	nr (NCBI)	ZP_09785661	0	92%	1171	x	x	x	x
SUP05 Clam symbionts	nr (NCBI)	x	x	x	x	x	x	x	x
Tevnia jerichonana (vent Tica) symbiont Metagenome	nr (NCBI)	ZP_08817562	0	51%	603	x	x	x	x
Riftia pachyptila symbiont Metagenome	nr (NCBI)	ZP_08829259	0	51%	602	x	x	x	x
Eastern Subarctic North Pacific Ocean OMZ Metagenome	(Unpublished) Project ID:1785 (IMG)	SI53jan11_150mDRAFT_100013415	1.00E-175	52%	607	P4_A09_1300m_0511.00003880	0	63%	782
Eastern Tropical South Pacific OMZ Metagenome	SRA023632 (NCBI)	SRR064450.40007.2	1.00E-34	64%	149	SRR304656.269476.2	1.00E-39	80%	166
Eastern Tropical South Pacific OMZ Metatranscriptome	SRA023632 (NCBI)	SRR064451.300887.2	1.00E-31	64%	138	SRR064449.263720.2	3.00E-23	62%	109
Eastern Tropical South Pacific OMZ Metagenome	SRA025088 (NCBI)	SRR070084.786489.2	1.00E-76	64%	292	SRR304680.924115.2	1.00E-80	74%	305
Global Ocean Survey	Metagenomic proteins (NCBI)	EDJ53379	0	69%	870	EDF78626	0	77%	627
Oregon Coast OMZ Metagenome	(Unpublished) CAM_P_0000692 (CAMERA)	CAM_READ_0417768141	5.00E-47	78%	193	CAM_READ_0442341325	5.00E-104	79%	385
Hawaii Ocean Time Series(ALOHA) Metagenome	CAM_PROJ_HOT (CAMERA)	x	x	x	x	HF_READ_05420203	3.00E-81	81%	283
Lost City Chimney Biofilm Metagenome	CAM_PROJ_HydrothermalVent (CAMERA)	HydrothermalVent_READ_00015031	0	79%	626	x	x	x	x
Bermuda Ocean Metagenome	CAM_P_0000712 (CAMERA)	x	x	x	x	CAM_READ_0354829177	7.00E-94	84%	221
Western Channel, UK Metagenome	CAM_PROJ_WesternChannelOMM (CAMERA)	WesternChannelOMM_READ_01009764	7.00E-59	67%	231	WesternChannelOMM_READ_02610180	9.00E-83	73%	310
Whalefall(Pacific Ocean & Antarctic Shelf)	CAM_PROJ_WhaleFall (CAMERA)	NCBI_READ_1594446127	9.00E-78	58%	226	x	x	x	x
Antarctica Aquatic Microbial Metagenome	CAM_PROJ_AntarcticaAquatic (CAMERA)	NCBI_READ_1112328230670	6.00E-91	55%	333	CAM_READ_0117659869	4.00E-59	72%	236

2.7 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Amend JP, McCollom TM, Hentscher M, Bach W (2011). Catabolic and anabolic energy for chemolithoautotrophs in deep-sea hydrothermal systems hosted in different rock types. *Geochimica et Cosmochimica Acta* 75: 5736-5748.

Anderson I, Rodriguez J, Susanti D, Porat I, Reich C, Ulrich LE *et al* (2008). Genome Sequence of Thermofilum pendens Reveals an Exceptional Loss of Biosynthetic Pathways without Genome Reduction. *Journal of Bacteriology* 190: 2957-2965.

Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* 54: 1501-1529.

Badger MR, Bek EJ (2008). Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO2 acquisition by the CBB cycle. *Journal of Experimental Botany* 59: 1525-1541.

Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N (2010). Examining the global distribution of dominant archaeal populations in soil. *ISME J*.

Bethke CM (2007). *Geochemical and biogeochemical reaction modeling*, Second edn. Cambridge University Press: Cambridge.

Bowers TS, Von Damm KL, Edmond JM (1985). Chemical evolution of mid-ocean ridge hot springs. *Geochimica Et Cosmochimica Acta* 49: 2239-2252.

Brandes JA, Boctor NZ, Cody GD, Cooper BA, Hazen RM, Yoder HS (1998). Abiotic nitrogen reduction on the early Earth. *Nature* 395: 365-367.

Breier JA, Toner BM, Fakra SC, Marcus MA, White SN, Thurnherr AM *et al* (Submitted). Sulfur, sulfides, oxides, and organic matter aggregated in submarine hydrothermal plumes at 9° 50' N East Pacific Rise. *Geochimica Et Cosmochimica Acta*.

Campbell AC, Gieskes JM (1984). Water column anomalies associated with hydrothermal activity in the Guaymas Basin, Gulf of California. *Earth and Planetary Science Letters* 68: 57-72.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* 330: 1375-1378.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Meth* 7: 335-336.

Chevreux B (2005). MIRA: An Automated Genome and EST Assembler Ph.D thesis, German Cancer Research Center Heidelberg, Duisburg.

Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo M-J, Dupont CL, Badger JH *et al* (2011). Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat Biotech* 29: 915-921.

Cho J-C, Giovannoni SJ (2004). Robiginitalea biformata gen. nov., sp. nov., a novel marine bacterium in the family Flavobacteriaceae with a higher G+C content. *International Journal of Systematic and Evolutionary Microbiology* 54: 1101-1106.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science* 311: 1283-1287.

Cleverley JS, Bastrakov EN (2005). K2GWB: Utility for generating thermodynamic data files for The Geochemist's WorkbenchÆ at 0-1000°C and 1-5000 bar from UT2K and the UNITHERM database. *Computers & Geosciences* 31: 756-767.

Clum A, Nolan M, Lang E, Glavina Del Rio T, Tice H, Copeland A *et al* (2009). *Complete genome sequence of Acidimicrobium ferrooxidans type strain (ICP T)*, vol. 1.

Conrad R, Seiler W (1988). Methane and hydrogen in seawater (Atlantic Ocean). *Deep Sea Research Part A Oceanographic Research Papers* 35: 1903-1917.

de Angelis MA, Lilley MD, Baross JA (1993). Methane oxidation in deep-sea hydrothermal plumes of the endeavour segment of the Juan de Fuca Ridge. *Deep Sea Research Part I: Oceanographic Research Papers* 40: 1169-1186.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al* (2009a). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10: R85.

Dick GJ, Clement BG, Webb SM, Fodrie FJ, Bargar JR, Tebo BM (2009b). Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guaymas Basin deep-sea hydrothermal plume. *Geochimica et Cosmochimica Acta* 73: 6517-6530.

Dick GJ, Tebo BM (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* 12: 1334-1347.

Distel DL, Lane DJ, Olsen GJ, Giovannoni SJ, Pace B, Pace NR *et al* (1988). Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *J Bacteriol* 170: 2506-2510.

Drummond SE (1981). Boiling and mixing of hydrothermal fluids: chemical effects on mineral precipitation., Pennsylvania State University.

Duperron S, Bergin C, Zielinski F, Blazejak A, Pernthaler A, McKiness ZP *et al* (2006). A dual symbiosis shared by two mussel species, Bathymodiolus azoricus and Bathymodiolus puteoserpentis (Bivalvia: Mytilidae), from hydrothermal vents along the northern Mid-Atlantic Ridge. *Environmental Microbiology* 8: 1441-1447.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792-1797.

Fierer N, Hamady M, Lauber CL, Knight R (2008). The influence of sex, handedness, and washing on the diversity of hand surface bacteria. *Proceedings of the National Academy of Sciences* 105: 17994-17999.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences* 105: 3805-3810.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al* (2005). Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science* 309: 1242-1245.

Guindon S, Gascuel O (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology* 52: 696-704.

Hasegawa M, Kishino H, Yano T-a (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* 22: 160-174.

Helgeson HC (1969). Thermodynamics of hydrothermal systems at elevated temperatures and pressures. *American Journal of Science* 267: 729-804.

Helgeson HC, Kirkham DH (1974). Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; II, Debye-Huckel parameters for activity coefficients and relative partial molal properties. *American Journal of Science* 274: 1199-1261.

Helgeson HC, Delaney JM, Nesbitt HW, Bird DK (1978). Summary and critique of the thermodynamic properties of rock-forming minerals. *American Journal of Science* 278-A: 1-229.

Hensen D, Sperling D, Trüper HG, Brune DC, Dahl C (2006). Thiosulphate oxidation in the phototrophic sulphur bacterium Allochromatium vinosum. *Molecular Microbiology* 62: 794-810.

Huggett MJ, Rappé MS (2012). Genome Sequence of Strain HIMB30, a Novel Member of the Marine Gammaproteobacteria. *Journal of Bacteriology* 194: 732-733.

Ivars-Martinez E, Martin-Cuadrado A-B, D'Auria G, Mira A, Ferriera S, Johnson J *et al* (2008). Comparative genomics of two ecotypes of the marine planktonic copiotroph Alteromonas macleodii suggests alternative lifestyles associated with different kinds of particulate organic matter. *ISME J* 2: 1194-1212.

Janecky DR, Seyfried WE (1984). Formation of massive sulfide deposits on oceanic ridge crests - incremental reaction models for mixing between hydrothermal solutions and seawater. *Geochimica Et Cosmochimica Acta* 48: 2723-2738.

Jannasch HW, Mottl MJ (1985). Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science* 229: 717-725.

Johnson JW, Oelkers EH, Helgeson HC (1992). SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. *Computers & Geosciences* 18: 899-947.

Karl DM, Knauer GA, Martin JH, Ward BB (1984). Bacterial chemolithotrophy in the ocean is associated with sinking particles. *Nature* 309: 54-56.

Kuwahara H, Yoshida T, Takaki Y, Shimamura S, Nishi S, Harada M *et al* (2007). Reduced genome of the thioautotrophic intracellular symbiont in a deep-sea clam, Calyptogena okutanii. *Curr Biol* 17: 881-886.

Labutti K, Sikorski J, Schneider S, Nolan M, Lucas S, Glavina Del Rio T *et al* (2010). *Complete genome sequence of Planctomyces limnophilus type strain (Mü 290 T)*, vol. 3.

Lam P, Cowen JP, Jones RD (2004). Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiology Ecology* 47: 191-206.

Lavik G, Stuhrmann T, Bruchert V, Van der Plas A, Mohrholz V, Lam P *et al* (2009). Detoxification of sulphidic African shelf waters by blooming chemolithotrophs. *Nature* 457: 581-584.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J*.

Lilley MD, de Angelis MA, Gordon LI (1982). CH4, H2, CO and N2O in submarine hydrothermal vent waters. *Nature* 300: 48-50.

Lilley MD, Feely, R.A., and Trefry, J.H. (ed) (1995) *Chemical and biochemical transformations in hydrothermal plumes*. American Geophysical Union: Washington, DC, USA, 369–391pp.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Research* 32: 1363-1371.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D *et al* (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research* 36: D534-D538.

Marshall KT, Morris RM (2012). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J*.

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

McCollom TM, Shock EL (1997). Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochimica Et Cosmochimica Acta* 61: 4375-4391.

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

McCollom TM (ed) (2008) *Observational, experimental, and theoretical constraints on carbon cycling in mid-ocean ridge hydrothermal systems*. AGU: Washington, D. C., 193-213pp.

Newton IL, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC *et al* (2007). The Calyptogena magnifica chemoautotrophic symbiont genome. *Science* 315: 998-1000.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* 476: 176-180.

Petersen JM, Wentrup C, Verna C, Knittel K, Dubilier N (2012). Origins and Evolutionary Flexibility of Chemosynthetic Symbionts From Deep-Sea Animals. *The Biological Bulletin* 223: 123-137.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* 35: 7188-7196.

Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, Head IM *et al* (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Meth* 6: 639-641.

Reinthaler T, van Aken HM, Herndl GJ (2010). Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Research Part II: Topical Studies in Oceanography* 57: 1572-1580.

Robidart JC, Bench SR, Feldman RA, Novoradovsky A, Podell SB, Gaasterland T *et al* (2008). Metabolic versatility of the Riftia pachyptila endosymbiont revealed through metagenomics. *Environmental Microbiology* 10: 727-737.

Robie RA, Hemingway BS, Fisher JR (1979). *Thermodynamic properties of minerals and related substances at 298.15 K and 1 Bar (10 Pascals) pressure and at higher temperatures. Bulletin 1452*. U.S. Geological Survey: Reston, VA.

Saccocia PJ, Seyfried Jr WE (1994). The solubility of chlorite solid solutions in 3.2 wt% NaCl fluids from 300-400°C, 500 bars. *Geochimica Et Cosmochimica Acta* 58: 567-585.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology* 75: 7537-7541.

Shock EL, Helgeson HC (1988). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C. *Geochimica Et Cosmochimica Acta* 52: 2009-2036.

Shock EL, Helgeson HC, Sverjensky DA (1989). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. *Geochimica Et Cosmochimica Acta* 53: 2157-2183.

Shock EL, Sassani DC, Willis M, Sverjensky DA (1997). Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes. *Geochimica Et Cosmochimica Acta* 61: 907-950.

Singer E, Webb EA, Nelson WC, Heidelberg JF, Ivanova N, Pati A *et al* (2011). Genomic Potential of Marinobacter aquaeolei, a Biogeochemical "Opportunitroph". *Applied and Environmental Microbiology* 77: 2763-2771.

Sunamura M, Higashi Y, Miyako C, Ishibashi J, Maruyama A (2004). Two bacteria phylotypes are predominant in the Suiyo seamount hydrothermal plume. *Appl Environ Microbiol* 70: 1190-1198.

Sverjensky DA, Shock EL, Helgeson HC (1997). Prediction of the thermodynamic properties of aqueous metal complexes to 1000°C and 5 kb. *Geochimica Et Cosmochimica Acta* 61: 1359-1412.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* 333: 1296-1300.

Thrash JC, Cho J-C, Ferriera S, Johnson J, Vergin KL, Giovannoni SJ (2010a). Genome Sequences of Strains HTCC2148 and HTCC2080, Belonging to the OM60/NOR5 Clade of the Gammaproteobacteria. *Journal of Bacteriology* 192: 3842-3843.

Thrash JC, Stingl U, Cho J-C, Ferriera S, Johnson J, Vergin KL *et al* (2010b). Genome Sequence of the Novel Marine Member of the Gammaproteobacteria Strain HTCC5015. *Journal of Bacteriology* 192: 3838-3839.

Toner BM, Fakra SC, Manganini SJ, Santelli CM, Marcus MA, Moffett JW *et al* (2009). Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nature Geosci* 2: 197-201.

Vignais PM, Billoud B (2007). Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chemical Reviews* 107: 4206-4272.

Von Damm KL, Edmond JM, Measures CI, Grant B (1985). Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochimica Et Cosmochimica Acta* 49: 2221-2237.

Von Damm KL, Parker CM, Zierenberg RA, Lilley MD, Olson EJ, Clague DA *et al* (2005). The Escanaba Trough, Gorda Ridge hydrothermal system: Temporal stability and subseafloor complexity. *Geochimica Et Cosmochimica Acta* 69: 4971-4984.

Wagman DD, Evans WH, Parker VB, Schumm RH, Halow I, Bailey SM *et al* (1982). *The NBS tables of chemical thermodynamic properties : selected values for inorganic and C1 and C2 organic substances in SI units*, vol. 11, supplement no. 2. American Chemical Society and the American Institute of Physics for the National Bureau of Standards: Washington, D.C.

Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al* (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences* 107: 8818-8823.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* 326: 578-582.

Ward N, Larsen Ø, Sakwa J, Bruseth L, Khouri H, Durkin AS *et al* (2004). Genomic Insights into Methanotrophy: The Complete Genome Sequence of Methylococcus capsulatus(Bath). *PLoS Biol* 2: e303.

Weiss RF, Craig H (1973). Precise shipboard determination of dissolved nitrogen, oxygen, argon, and total inorganic carbon by gas chromatography. *Deep Sea Research and Oceanographic Abstracts* 20: 291-303.

Welhan JA, Craig H (1979). Methane and Hydrogen in East Pacific Rise Hydrothermal Fluids. *Geophys Res Lett* 6: 829-831.

Winn CD, Karl DM, Massoth GJ (1986). Microorganisms in deep-sea hydrothermal plumes. *Nature* 320: 744-746.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al* (2009). Assembling the Marine Metagenome, One Cell at a Time. *PLoS One* 4: e5299.

Wright JJ, Konwar KM, Hallam SJ (2012). Microbial ecology of expanding oxygen minimum zones. *Nat Rev Micro* 10: 381-394.

CHAPTER III

METAGENOMIC RESOLUTION OF MICROBIAL FUNCTIONS IN DEEP-SEA HYDROTHERMAL PLUMES ACROSS THE EASTERN LAU SPREADING CENTER

Karthik Anantharaman[1], John A. Breier[2] and Gregory J. Dick[1,3,4]

[1]Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA
[2]Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
[3]Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA
[4]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA

3.1 Abstract

Deep-sea hydrothermal plumes affect ocean biogeochemistry on global scales and are critical to understanding global elemental cycles. In comparison to neutrally-buoyant plumes, rising hydrothermal plumes represent an understudied system where microbial metabolism and particle formation processes initiate the transformation of reduced chemicals like hydrogen sulfide, hydrogen, methane, iron, manganese, and ammonia that are abundant in hydrothermal vent fluids. Here we use metagenomics and bioenergetic modeling to describe the abundance and genetic potential of microorganisms in relation to available electron donors in five different hydrothermal plumes and associated background deep-sea waters from the Eastern Lau Spreading Center located in the Western Pacific Ocean. A total of 331 distinct genomic 'bins'

were identified, comprising an estimated 951 genomes of archaea, bacteria, eukarya and viruses. These genomes include a significant proportion of novel microorganisms and thus reveal insights into the energy metabolism of heretofore unknown microbes. Community-wide analyses of genes encoding enzymes that oxidize inorganic energy sources show that use of sulfur constitutes the most abundant and diverse chemolithotrophic microbial metabolism in the community. Genes for sulfur oxidation were commonly present in genomic bins that also contained genes for oxidation of hydrogen and methane, suggesting metabolic versatility in these microbial groups. The diversity and abundance of genes encoding hydrogen oxidation was moderate, whereas that of genes for methane and ammonia oxidation was low. Bioenergetic-thermodynamic modeling supports the metagenomic analyses, showing that oxidation of elemental sulfur with oxygen is the most dominant catabolic reaction in the hydrothermal plumes. We conclude that rising hydrothermal plumes host a complex and diverse microbial community whose functional ecology is dictated by the underlying plume chemistry, with a dominant role for sulfur-based chemolithoautotrophy.

3.2 Introduction

Deep-sea hydrothermal vent plumes occur at mid-ocean ridges and back arc-basins throughout the world's oceans, where chemically reduced hydrothermal vent fluids are mixed with cold, oxic deep ocean water. The enrichment of electron donors such as H_2S, H_2, CH_4, NH_3, Mn^{2+} and Fe^{2+} in plumes fuels chemosynthetic microbial metabolisms (de Angelis et al 1993, Dick et al 2009b, Distel et al 1988, Jannasch and Mottl 1985, Lam et al 2004, Petersen et al 2011). Hydrothermal plumes influence the broader oceans because biogeochemical processes in

plumes control the availability and fate of trace metals and essential micronutrients (Kadko 1993, Tagliabue et al 2010, Toner et al 2009), and primary production in plumes may serve as a significant source of organic carbon to the deep oceans (McCollom 2000). The microbial ecology of hydrothermal plumes shares many similarities with other marine environments where primary production is linked to oxidation of reduced sulfur species, ammonia, and hydrocarbons occurs, including the pelagic ocean water column, oxygen minimum zones, and deep-sea hydrocarbon plumes (Aristegui et al 2009, Canfield et al 2010, DeLong et al 2006, Dick et al 2013, Reinthaler et al 2010, Rivers et al 2013, Swan et al 2011).

Recent studies have begun to elucidate the importance and role of microorganisms and metabolisms that operate within hydrothermal plumes. Surveys of the small subunit ribosomal RNA (SSU) genes using tag pyrosequencing and clone libraries have revealed the composition of plume microbial communities (Dick and Tebo 2010, German et al 2010, Sunamura et al 2004, Sylvan et al 2012). Metagenomic, metatranscriptomic and metaproteomic methods have provided insights into the roles of dominant organisms involved in sulfur, hydrogen, methane, and ammonia oxidation in hydrothermal plumes such as SUP05 *Gammaproteobacteria* (Anantharaman et al 2013), *Methylococcae Gammaproteobacteri*a (Li et al 2013), Marine Group I *Thaumarchaea* (Baker et al 2012), and SAR324 *Deltaproteobacteria* (Sheik et al 2013). New studies also show that rare members of the plume microbial community such as *Alteromonacadae Gammaproteobacteri*a (Li et al 2014) and *Nitrospirae* (Baker et al 2013) are potentially keystone species with roles in iron uptake and nitrite oxidation, respectively. Yet, the metabolic capabilities and ecology of many other populations of bacteria, archaea, eukarya and viruses in the complex plume communities remain unresolved.

The Eastern Lau Spreading Center (ELSC) is a deep-sea hydrothermal system located in the Lau Basin, a back-arc basin in the Western Pacific Ocean. In contrast to mid-ocean ridges (MOR), back-arc basins exhibit greater geologic diversity characterized by asymmetrical seafloor spreading. At ELSC, these characteristics manifest as steep gradients in the chemistry of underlying rocks, tectonic characteristics, and hydrothermal activity along a north-south axis. Spreading rates are fast in the north (~97mm/year) and slower in the south (40mm/year) (Zellmer and Taylor 2001). A deep axial valley (~2640m) characterized by basaltic underlying rock in the north transitions into a shallow axial ridge with andesitic underlying rocks (~1877m) in the south (Martinez et al 2006). Six different vent fields have been identified along the ELSC, and the chemistry of hydrothermal fluids exhibits significant inter-field variability, with properties similar to MORs in the north and highly elevated concentrations of H_2S, Fe and Mn coupled with lower pH towards the south (Ferrini et al 2008, Mottl et al 2011). Recently, Flores et al (2012) surveyed seafloor hydrothermal deposits at ELSC and concluded that they host microbial communities similar to MORs except for the Mariner vent field, where the community was heavily influenced by the vent field's unique geochemistry. Sylvan et al (2013) noted that the microbial diversity of low temperature hydrothermal deposits along ELSC displayed significant heterogeneity that followed host rock composition. Because rising vent plumes reflect the geochemistry of underlying hydrothermal vent fluids, the chemical gradient along the ELSC provides an opportunity to study the impacts of vent geology and geochemistry on the composition of microbial communities in plumes.

We recently analyzed the overall microbial community diversity of hydrothermal plumes across the ELSC by high-throughput tag-sequencing of 92 samples collected from various elevations off the seafloor at five of the six vent fields (Supplementary table 1) (Sheik et al

2014). This study showed that ELSC plumes contain a mixture of vent-associated and pelagic ocean microorganisms, and that the structure of these microbial communities is variable within vent fields, between vent fields, and even within individual plumes as they rise off the seafloor to the point of neutral buoyancy. Here, we follow up the investigations of Sheik et al (2014) by using shotgun metagenomic sequencing to characterize the metabolic functions of microorganisms in hydrothermal plumes at the five different ELSC vent fields and in surrounding ocean background waters. Metagenomic analyses and thermodynamic-bioenergetic models show that despite the microbial and geochemical variability observed previously, the most abundant energy metabolisms and community composition across ELSC plumes are strikingly similar and dominated by diverse populations of sulfur oxidizing chemolithoautotrophs. The ELSC plumes also host heterotrophs and chemolithoautotrophs that use hydrogen, ammonia, methane and nitrite as electron donors. Minor variations between ELSC sites reflect differences in underlying plume chemistry.

3.3 Materials and Methods

Sample Collection. Samples were collected from five different hydrothermal vent fields (Kilo Moana, Tahi Moana, Abe, Tui Malila, and Mariner, from north to south) during cruises TN235 and TN236 aboard the *R/V Thomas G. Thompson* in May-July 2009. Details of samples and the sampling locations are provided in Supplementary table 1. A total of 92 hydrothermal plume and background deep-sea samples were collected. Of these, 78 samples were collected from the rising plume and background deep sea using a Suspended Particle Rosette Sampler (SUPR) (Breier et al 2009) mounted on remotely operated vehicle *ROV Jason II*, while 14 samples were collected with 20 L Niskin bottles via conductivity temperature, and depth (CTD) rosette. Three types of background ocean water samples were collected: (1) near bottom backgrounds were

collected with SUPR away from the hydrothermal vent fields; (2) above plume backgrounds were collected with SUPR continuously pumping at an approximate water depth of 700-1300m; (3) below neutrally buoyant plume backgrounds were collected using the CTD rosette. Water samples collected with SUPR (10-60 l) were filtered *in situ* on to 0.8-µm 37-mm polycarbonate Supor membranes (Pall Corporation, Port Washington, NY, USA) and preserved shipboard in RNAlater (Ambion, Austin, TX, USA). Water samples collected by CTD rosette were pressure filtered with N_2 gas shipboard on to 0.2-µm 47-mm polycarbonate membranes and preserved in RNAlater (Ambion, Austin, TX, USA).

Extraction of nucleic acids and DNA sequencing. DNA was extracted from ¼ filters as described previously (Dick and Tebo 2010). Multiple displacement amplification (MDA) of genomic DNA was performed using the illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare, Piscataway, NJ, USA). Shotgun sequencing of DNA was performed with Illumina HiSeq2000 (Illumina, Inc., San Diego, CA, USA) at the University of Michigan DNA Sequencing Core.

De novo genomic assembly and annotation. Raw shotgun sequencing reads were deprelicated (100% identity over 100% of length) and trimmed using the adaptive read trimmer, Sickle (https://github.com/najoshi/sickle). Samples from the five vent sites (Kilo Moana, Abe, Mariner, Tahi Moana, Tui Malila) were each assembled *de novo* to obtain five separate assemblies. Whole genome *de novo assemblies* were performed using IDBA-UD (Peng et al 2012) with the following parameters: --mink 50, --maxk 92, --step 4, --min_contig 500. rRNA reads were identified using RiboPicker (Schmieder et al 2012) with a custom database (5s+16s+23s rRNA) and assembled separately using IDBA-UD with the following parameters: --mink 50, --maxk 92, --step 4. To check the veracity of the generated contigs, *de novo* whole-genome assemblies were

repeated using Velvet (Zerbino and Birney 2008) in an iterative manner by removing the reads used in formation of contigs at the higher kmer (kmer 91 to 51, steps of 4) using the following parameters: -exp_cov auto –ins_length 214-223 -ins_length_sd 20 –read_trkg yes –min_contig_lgth 2500. This was followed by further refinement with MetaVelvet (Namiki et al 2012) (kmer 91 to 51, steps of 4) using the following parameters: -scaffolding yes, -min_contig_lgth 2500. All major trends were similar across both sets of assemblies. All data presented in this paper are from the assemblies generated by IDBA-UD.

Read mapping. Paired-end sequencing reads were mapped to assembled contigs using the Burrows-Wheeler Aligner (BWA version 0.7.5a) (Li and Durbin 2009). First, individual forward and reverse reads for each sample were mapped to the assembled contigs using the BWA-ALN algorithm implemented using default parameters except for a modified maxDiff parameter (aln, -n 0.02). Second, paired forward and reverse read alignments were generated in the SAM format using the BWA-SAMPE algorithm with default parameters. The mapped read counts were extracted using SAMtools 0.1.17 (Li et al 2009). Referenced contigs were visualized for evenness of mapped read coverage to identify potential chimeric regions using Integrative Genome Viewer (IGV) (Thorvaldsdóttir et al 2013).

Binning and conserved gene analysis. All resulting contigs were assigned into putative taxonomic groups by binning with Emergent Self-Organizing Maps (ESOM) with a combination of tetra-nucleotide frequency (Dick et al 2009a) and coverage of contigs across the five different assemblies (cutoffs: minimum contig size=4 Kb, maximum contig size=8 Kb) as determined by read mapping. All tetramers containing start and stop codons were removed prior to analysis. All resultant bins were manually evaluated for accuracy, completeness, and to estimate number of genome equivalents using the distribution of conserved phylogenetic markers described

previously (Ciccarelli et al 2006). Gene calling and annotations were done through the DOE Joint Genome Institute (JGI) Integrated Microbial Genomes metagenomics expert review (IMG-MER) pipeline (Markowitz et al 2008). All bins were identified to appropriate taxonomic levels in the following order of workflow: (i) identification of SSU rRNA genes on contigs by BLASTN (Altschul et al 1990) to the Silva SSU Database version 111 (Pruesse et al 2007); (ii) ESOM Binning with reference genomes; (iii) taxonomic clustering of annotated genes by protein UBLAST with the IMG database. Bins in which a majority of the predicted open reading frames (ORFs) could not be identified to any taxonomic level or clustered with viruses were annotated as putative extrachromosomal elements (plasmids/phages/prophages). Number of genome equivalents for bacteria and archaea were estimated using the average number of conserved phylogenetic markers (Supplementary table 6, Supplementary table 7, Supplementary 8). Number of genome equivalents for bins containing extrachromosomal elements was estimated by presence of contigs from each of the five rising plumes sampled (e.g., if bin A contained contigs from 2 different rising plumes, then the number of genome equivalents was 2).

Functional gene analysis. Genetic potential of individual bins was investigated by comparison of predicted proteins with COG/Pfam/KEGG families. All results were verified using reciprocal protein blasts with custom databases and the NCBI non-redundant database (nr) (cutoff: evalue<1e-5). Abundance of genes for oxidation of electron donors was calculated by recruitment of individual reads from each of the 12 samples using BLASTX with custom protein databases (cutoffs: minimum alignment length>60 bp, e-value<1e-5, bit score>50).

Sequence alignment and phylogeny. Alignment of Ni-Fe hydrogenase amino acid sequences was performed by MUSCLE (Edgar 2004) using default parameters followed by manual refinement. All predicted Ni-Fe hydrogenase amino acid sequences from the ELSC plumes were

aligned and compared with reference sequences identified previously (Vignais and Billoud 2007). Phylogenetic analysis of Ni-Fe hydrogenase genes was inferred by Maximum Likelihood implemented in RaxML (Stamatakis 2006) using the PROTGAMMAGTR algorithms and bootstrapped 1000 times with the following parameters: -f a –m PROTGAMMAGTR –N 1000 –x 777 –p 333.

2-D Physical, Bioenergetic and thermodynamic modeling. Computational fluid dynamics modeling was used to develop a two-dimensional radially symmetric physical transport model for a high temperature hydrothermal plume emanating from the ABE hydrothermal field. Model parameters are based on measurements of vent temperature and plume physical structure made in the ABE hydrothermal field of the Eastern Lau Spreading Center at the time these samples were collected. The model numerically solves the density weighted unsteady Reynolds-averaged Navier-Stokes and energy equations, with turbulent entrainment and mixing modelled by the realizable k-ε turbulence closure method. This plume model is described in detail in (Jiang 2014). The following is a brief summary of pertinent aspects.

The modeling was performed with ANSYS Fluent (version 13.0.0). The model is axisymmetric with the axis at the vent center line. The radial and vertical spans are 200 and 400 m, respectively. There is a no-slip boundary condition at the seafloor with a constant temperature. There is a pressure-outlet boundary condition at the top of the domain with a constant temperature. The vertical boundary of the domain representing the plume axis is prescribed with an axis boundary condition. The opposing vertical boundary is prescribed with a symmetry boundary condition. The initial background temperature field is a vertically linear varying temperature field. Total integration time is relatively short so Earth's rotation is neglected. Density and specific heat of hydrothermal fluid and seawater were calculated from a

set of nonlinear equations of state for a thermal saline fluid for a temperature range of 0-374°C at fixed salinity (S = 34.65) and pressure (for a depth of 2140 m) (Sun et al 2008). In the absence of data specific to saline fluids, thermal conductivity and viscosity were calculated using equations for pure water according to National Institute of Standards and Technology (NIST) Reference Fluid Thermodynamic and Transport Properties Database (REFPROP), NIST Standard Reference Database 23, Version 9.0.

The model provides a best estimate of steady-state plume flow, in the absence of bottom currents, from a vent (i.e. A1, 20°45'47.5"S 176°11'28.8"W, depth 2140 m) in the northern region of the ABE hydrothermal field along the Eastern Lau Spreading Centre(Ferrini et al 2008, Mottl et al 2011). The effective vent orifice diameter, ~0.14 m, was estimated based on video imagery collected by ROV Jason II (cruise TN235, Dive J2-424). Vent fluid exit temperature used for this model was 309 °C(Mottl et al 2011). Rising plume temperature measurements were made with the temperature probe of the In Situ Electrochemical Analyzer system (Luther III et al 2008). Vertical plume and background seawater structure with respect to temperature, density and optical backscatter was identified by multiple nearby CTD profiles, on a water rosette equipped with a Seapoint turbidity sensor, during R/V Thomas G. Thompson cruises TN235 and TN236. Vent fluid velocity was modelled as 0.2 m s-1. This was estimated on the basis of iteratively varying this parameter to replicate the plume rise height and is consistent with measured hydrothermal vent discharge velocities (Ramondenc et al, 2006).

Model predictions compare well with CTD observations of the neutrally buoyant plumes above this region of the ABE hydrothermal field, and point measurements of temperature at 10 and 40 m elevation within the rising stem of the plume emanating from A1 vent. Comparisons in the near vent region of the plume (<0.5 m elevation) are complicated by the presence of multiple

vents in close proximity in this portion of the ABE field. Specifically, the model predicts a temperature at 0.5 m above A1 vent that is approximately 10 times greater than that measured (Jiang et al 2014). This suggests the parameters used in the model are either not fully representative of mean vent flux at the time of measurement, that the measurement does not reflect effective centreline plume temperature, or even more probable that the aggregate effect of the multiple proximal plumes in this area is not well represented by the specific ABE-A1 vent parameters. Regardless, the model does agree well with the other observations suggesting it is a reasonable representation of the effective transport of the aggregate plume rising from this area of the ABE field. Thus it is useful for understanding the physical structure of hydrothermal plumes in this environment.

In order to map thermodynamic equilibrium reaction path modeling results on to the two-dimensional physical transport model domain, a chemically non-reactive scalar quantity, C, was included in the model. This scalar represents the relative distribution of conservative vent fluid chemical constituents within the plume. The scalar is prescribed to be 1 in vent fluid and 0 in seawater, and is thus numerically equivalent to vent fluid dilution within the plume model.

Equilibrium thermodynamic reaction path modeling was used to predict Fe mineral precipitation, chemical concentrations, and activity coefficients resulting from the mixing of seawater with end member fluid from A1 vent in the ABE hydrothermal field (Mottl et al 2011). Our approach follows those of previous studies (Bowers et al 1985, Janecky and Seyfried 1984, McCollom 2000). This specific plume thermodynamic model implementation is described in Chapter IV.

3.4 Results

We analysed a total of 12 samples, comprising 9 plume and 3 background deep-sea samples from five different vent sites along the ELSC (Kilo Moana, Abe, Mariner, Tahi Moana, Tui Malila) (Supplementary Table 1).

Metagenomic sequencing, de novo assembly and binning. Shotgun sequencing of community genomic DNA on 6 lanes of Illumina HiSeq2000 produced a total of 1,026,438,887 paired-end reads (total reads= 2,052,877,774, average insert-size=221bp). Most reads (90.1%) were above Q30, and the mean quality score was 35.18. *De novo* assemblies of quality filtered reads generated a total of 1,091,773 contigs and 1,511,701,270 bp of consensus sequence (Supplementary Table 2). Prediction of open reading frames resulted in a total of 1,697,862 putative genes (Supplementary Table 3). The individual assembly of sequencing reads from each hydrothermal plume helped to distinguish polymorphisms and variation between near-identical populations and enabled effective comparisons of microbial populations along the ELSC.

Tetranucleotide-based ESOM binning was performed on all contigs of length greater than 4000 bp (62,135 contigs). All five assemblies were binned together to enable identification of similar organisms across the vent sites. Putative assignment of taxonomic groups by SSU rRNA genes, binning with reference genomes, and taxonomic profiling of annotated genes yielded a total of 331 distinct genomic bins comprising 98 bacteria, 7 archaea, 1 eukarya (Fig. 3.1, Supplementary Table 4) and 225 putative extrachromosomal elements (plasmids/phages/prophages) (Fig. 3.2, Supplementary Table 5). In total, these bins represent an estimated total of 951 partial to near-complete genome sequences.

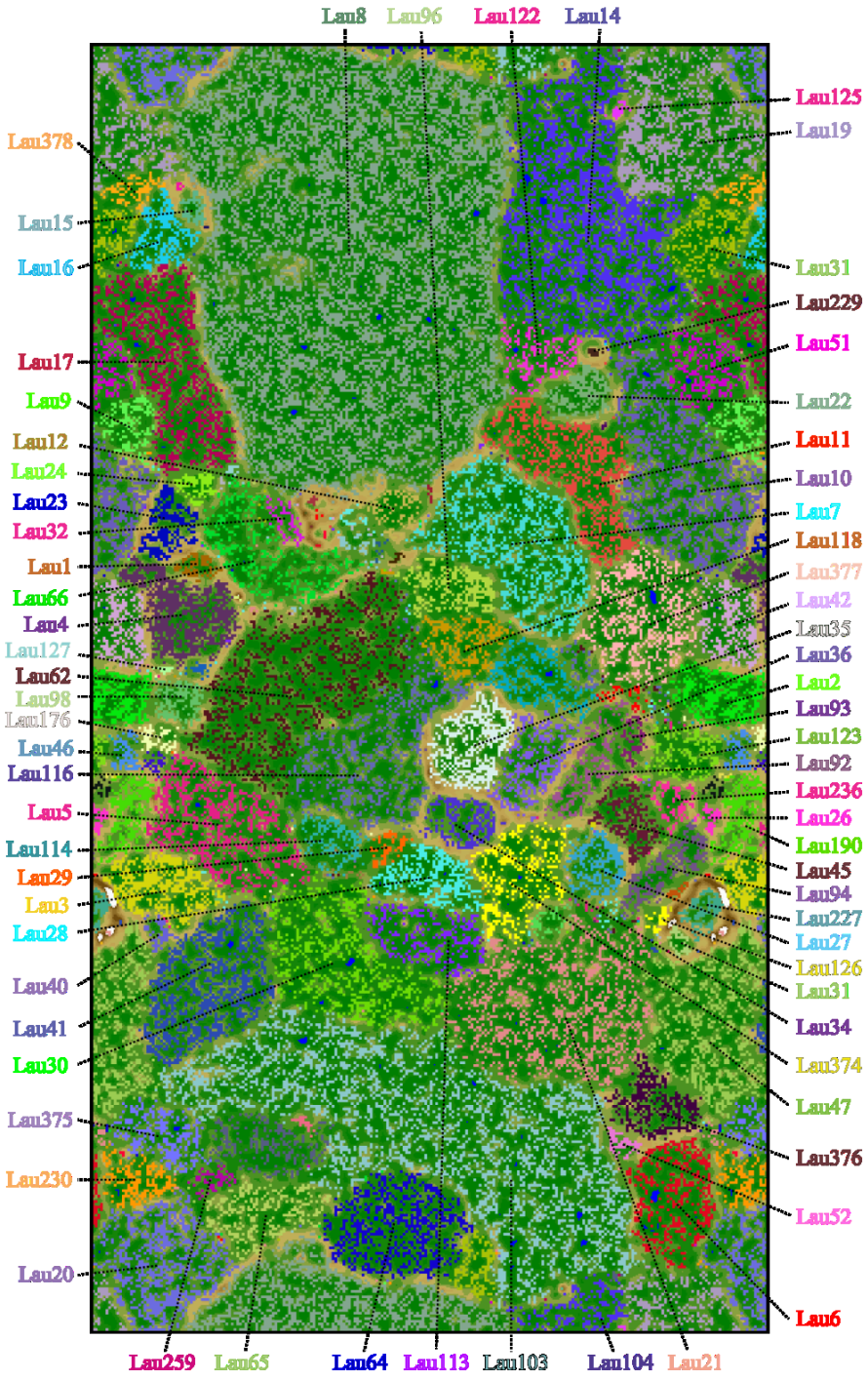


Fig. 3.1 Assignment of assembled contigs from ELSC to specific bacterial, archaeal and eukaryotic populations using ESOM implemented with tetranucleotide frequencies. Each point on the map represents a contig (>4 kb) or contig fragment generated *in silico* (4-8 kb). All identified bins are uniquely color coded as indicated. Background topography color represents euclidean distance of tetranucleotide frequency between data points, with blue indicating highest similarity, followed by green, and brown ridges representing the largest differences. The ESOM map displayed is tiled and torroidal (i.e., continuous from top to bottom and left to right).

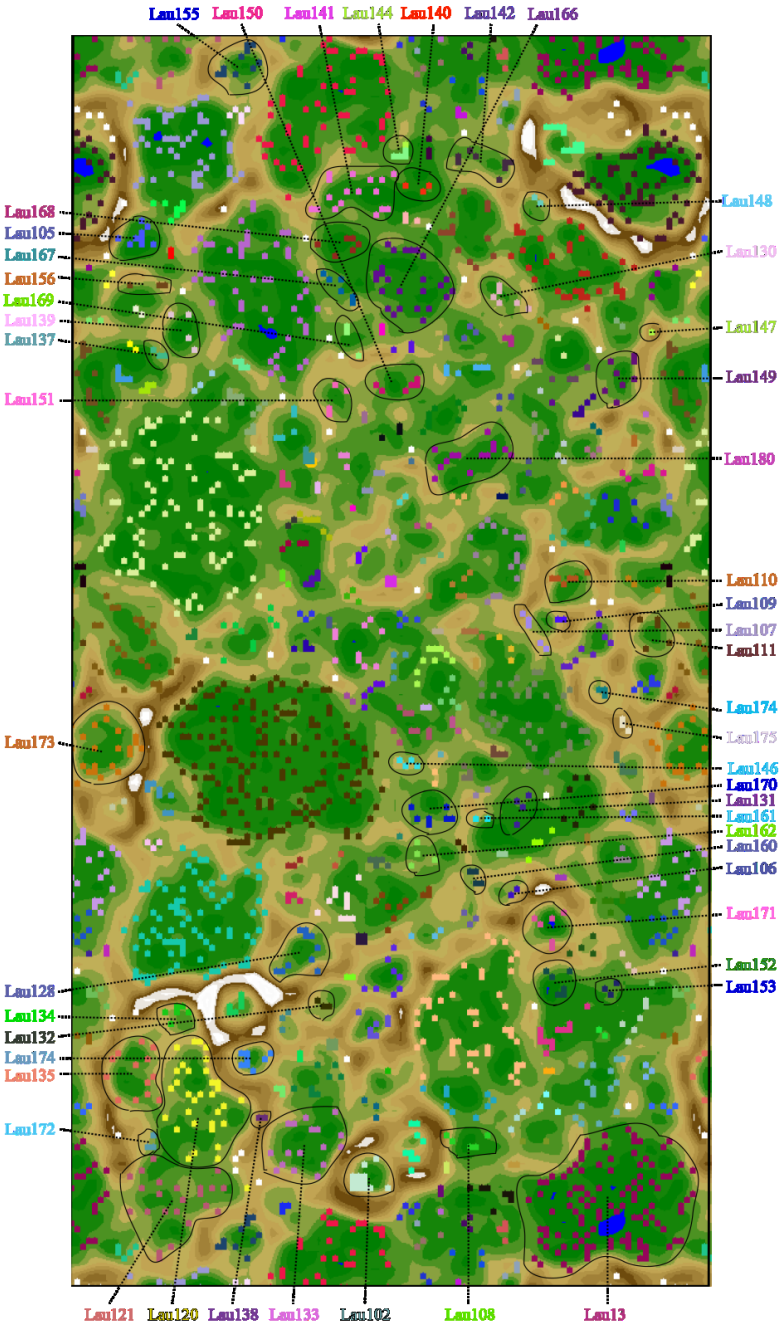


Fig. 3.2 Assignment of assembled contigs from ELSC to specific bins containing extrachromosomal elements using ESOM implemented with tetranucleotide frequencies. Each point on the map represents a contig (>4 kb) or contig fragment generated *in silico* (4-8 kb). All identified bins are indicated by black lines and uniquely color coded as indicated. Background topography color represents euclidean distance of tetranucleotide frequency between data points, with blue indicating highest similarity, followed by green, and brown ridges representing the largest differences. The ESOM map displayed is tiled and torroidal (i.e., continuous from top to bottom and left to right). Only the top 40 bins are shown.

The hydrothermal plumes of ELSC were dominated by microbial populations described previously from the hydrothermal plumes of Guaymas Basin (GB), SUP05 *Gammaproteobacteria*, *Alteromonacadae Gammaproteobacteria*, Marine Group I *Thaumarchaea* and SAR324 *Deltaproteobacteria* (Anantharaman et al 2013, Baker et al 2012, Dick and Tebo 2010, Lesniewski et al 2012, Sheik et al 2013). In addition, we assembled essentially complete, near-complete, and partial genomes for novel organisms previously unknown in the pelagic water column or known only through SSU rRNA genes retrieved from hydrothermal plumes, seafloor or hydrothermal vent chimney communities (Supplementary Table 4). This included *Aquificae* (genomic bins Lau227 and Lau237), *Planctomycetes* (Lau7, Lau35, Lau36, Lau94, Lau96), *Cyanobacteria* (Lau27), *Melainabacteria* (Lau2), Candidate Division TM7 (*Saccharibacteria*) (Lau1), *Chloroflexi* (Lau3 and Lau12), Marine Group A (Lau47, Lau103 and Lau104), *Nitrospinae* (Lau17), *Nitrospirae* (Lau44), *Poribacteria* (Lau21 and Lau377), *Verrucomicrobia* (Lau158, Lau184, Lau190, and Lau191), *Acidobacteria* (Lau40), *Epsilonproteobacteria* (Lau15, Lau129, and Lau229) and *Gemmatimonadetes* (Lau45) (Allers et al 2013, Di Rienzi et al 2013, Flores et al 2012, Swan et al 2011, Sylvan et al 2012, Sylvan et al 2013, Teske et al 2002, Walsh et al 2009, Wright et al 2012, Yamamoto and Takai 2011). Searches for eukaryotic SSU rRNA genes revealed the presence of novel populations of *Archaeplastida*, DH147-EKD23, *Alveolata* (*Ciliophora*, *Dinoflagellata*, OLI11255, *Protalveolata*), *Rhizaria* and *Opisthokonta* (*Fungi* and *Metazoa*). However, we conclusively identified only one distinct bin of Eukarya, the largest bin (Lau8) in the metagenome, putatively identified as a *Bathymodiolinae* mussel with highest SSU rRNA similarity (99.9%) to *Bathymodiolus tangaroa* from the Western Pacific Ocean (Fig. 3.3) (Jones et al 2006). The Lau8 bin contained a total of 54 Mega bases (Mb) of consensus sequence on 9599 contigs, and was

represented most abundantly in the Abe hydrothermal plume metagenome. We also recovered a total of 225 genomic bins containing putative extrachromosomal elements (viruses, plasmids) and comprising a total of ~38 megabases.

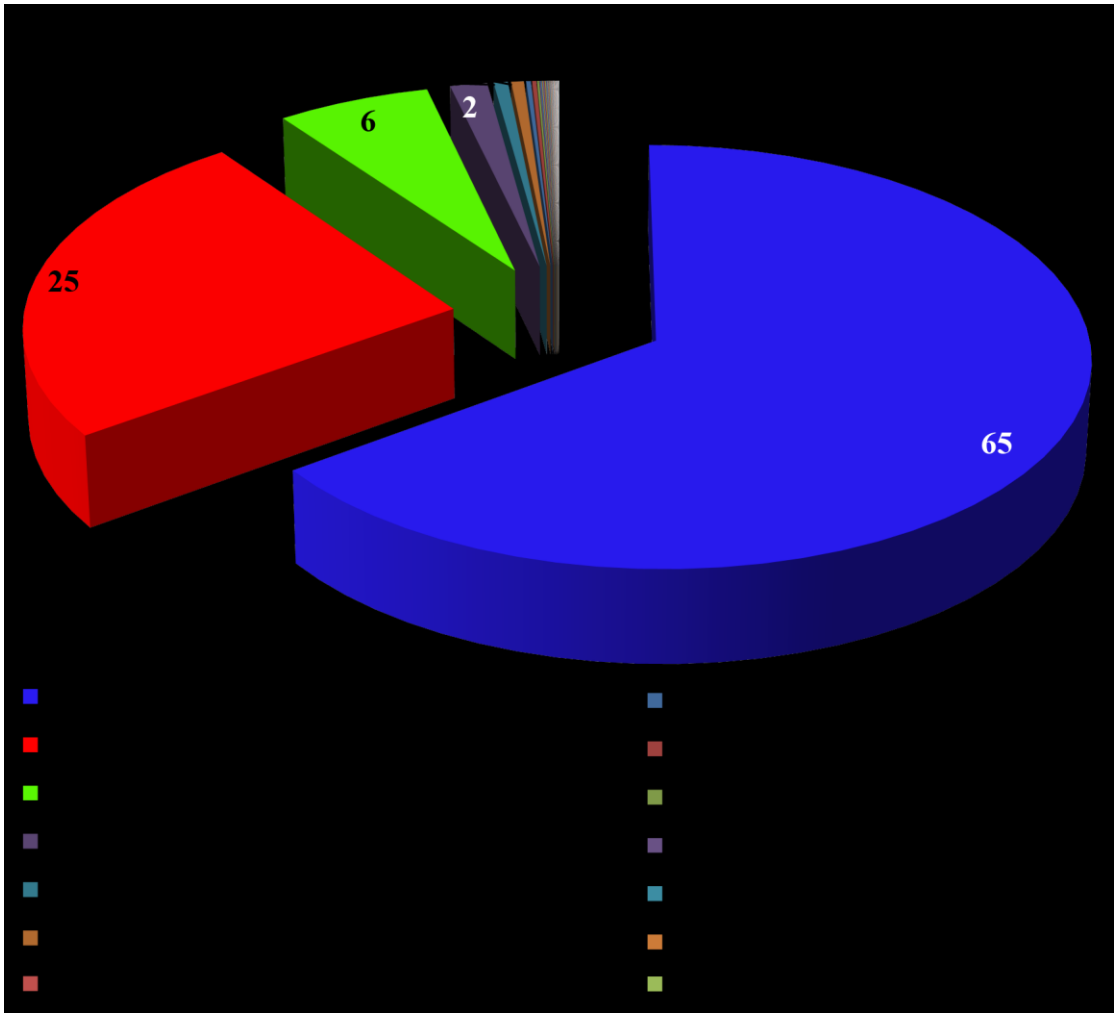


Fig. 3.3 Identification of eukaryotic bin "Lau8". Pie-chart indicates the distribution of blastn hits of all identified ORFs in bin "Lau8" to organisms in NCBI-nt as indicated in the legend. Inset number indicates the percentage (%) of hits.

Functional resolution of metagenomic bins.

We analyzed metagenomic bins across the five hydrothermal plumes at ELSC to determine genomic coverage (a proxy for organism abundance) and identify genes encoding for energy and carbon metabolism, including utilization of electron donors and acceptors and carbon fixation pathways (Fig. 3.4).

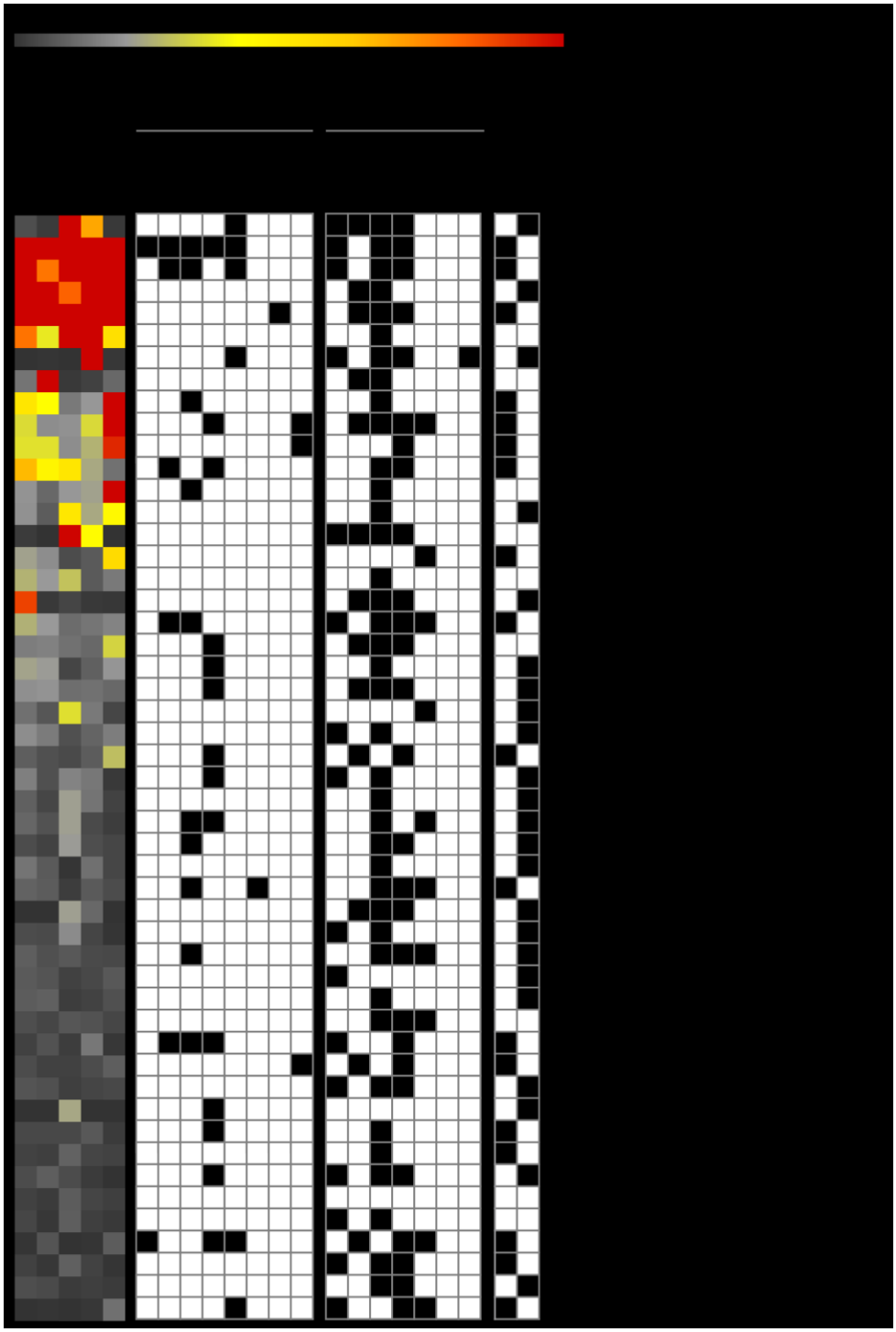


Fig. 3.4 Normalized abundance, energy and carbon metabolism of the 50 most abundant bins identified in ELSC hydrothermal plumes. Heat map on left panel indicates abundance of bins for each sampled hydrothermal plume and is displayed as a percentage of the total community as shown in the legend. Black boxes indicate identified genes encoding carbon metabolism and electron donors and acceptors for energy metabolism.

Oxidation of reduced sulfur. The genetic potential for use of different forms of reduced S species (H_2S, S_0 and $S_2O_3^{2-}$, SO_3^{2-}) as electron donors was determined by searching genomic bins for the following genes: the *sox* pathway (*soxABCDXYZ*) and rhodanese sulfurtransferase for oxidation of $S_2O_3^{2-}$; adenosine phosphosulfate reductase (apr), sulfate adenylyltransferase (*sat*), and sulfite oxidoreductase for oxidation of SO_3^{2-}; sulfur oxygenase reductase (*sor*) and reverse acting-dissimilatory sulfite reductase (*rdsrAB*) for oxidation of S_0; and *flavocytochrome sulfide dehydrogenase* (*fcc*) and sulfide:quinone oxidoreductase (*sqr*) for oxidation of H_2S. Collectively, genes for oxidation of sulfur were the most common genes for lithotrophy identified in genomic bins, being present in 20 of the 50 most abundant organisms including the abundant SUP05 and SAR324 (Fig. 3.4).

***Hydrogen oxidation*.** A total of 24 Nickel-Iron (Ni-Fe) hydrogenase operons were identified in the ELSC metagenomes, comprising three distinct forms of hydrogenases. Organisms possessing the type I membrane-bound Ni-Fe hydrogenases for oxidation of H_2 include *Alteromonacadae* (Lau4), SUP05 (Lau10), *Epsilonproteobacteria* (Lau229), *Poribacteria* (Lau21), SAR324 (Lau20) and *Flavobacteria* (Lau23) (Fig. 3.5). Amongst these, SUP05 and SAR324 possessed the 'hyd' type Ni-Fe hydrogenases previously observed to be transcriptionally active outside of hydrothermal plumes in the deep waters of Carmen Basin and Guaymas Basin in the Gulf of California (Anantharaman et al 2013). Only *Cyanobacteria* (Lau27), *Melainabacteria* (Lau2) and *Flavobacteria* (Lau23) possessed the type IIa cyanobacterial uptake type Ni-Fe hydrogenases. Type III NADP-reducing Ni-Fe hydrogenases were identified in *Melainabacteria (*Lau2) and *Pseudomonadaceae* (Lau 177), while type IV H_2 evolving Ni-Fe hydrogenases were observed in *Aquificae* (Lau227) (Vignais and Billoud 2007). Notably, genes for oxidation of reduced sulfur species were observed in each of the hydrogen oxidizing organisms possessing

the type I Ni-Fe hydrogenases. No other forms of hydrogenases (iron-iron or nickel-iron-selenium) were identified.

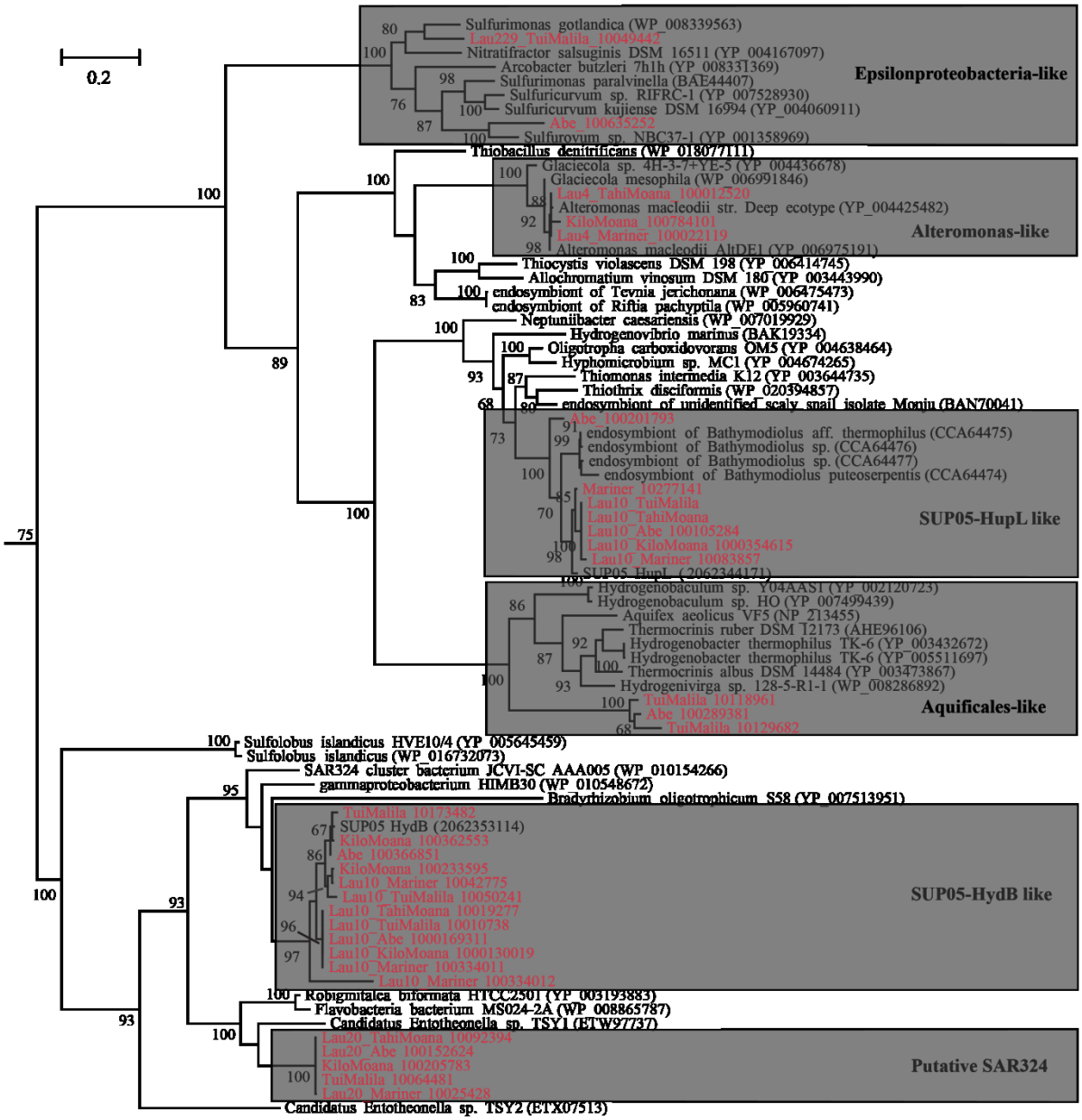


Fig 3.5. Phylogeny of group 1 membrane-bound Ni, Fe hydrogenase large subunit inferred with maximum likelihood. Bootstrap values >60 are shown. Sequences in red are from ELSC. Grey boxes indicate phylogenetic clusters of ELSC sequences.

Ammonia and methane oxidation. In contrast to the diversity of organisms and genes for sulfur and H_2 oxidation, only two genomic bins contained genes encoding the oxidation of NH_3 and CH_4. A single genomic bin (Lau19) of the Marine Group I Archaea (*Thaumarchaea*) containing 10 near-complete and partial genomes had genes for oxidation of ammonia in the form of ammonia monooxygenase (*amo*) (Konneke et al 2005). No ammonia oxidizing bacteria (AOB) were identified in the ELSC plumes. A genomic bin (Lau113) containing 3 near-complete genomes from the OPU-3 group of *Methylococcacae* (*Gammaproteobacteria*) possessed particulate methane monooxygenases (*pmo*) for methane oxidation (Hanson and Hanson 1996) as well as novel *soxXYZA* genes (protein identity of ~55% to hits in NCBI-nr) for oxidation of thiosulfate.

Nitrite oxidation. We identified two bacterial phyla, *Nitrospinae* (Lau17, Lau41) and *Nitrospirae* (Lau44) with the potential for oxidation of nitrite. Nitrite oxidoreductase (*nxr*) genes were identified in the *Nitrospinae* bin but not in the *Nitrospirae* bin (Baker et al 2013, Lücker et al 2010, Luecker et al 2013). However, we recovered nxr genes on short contigs that share homology with both groups, hence their taxonomic assignment cannot be conclusively determined. *Nitrospinae* were observed to be more abundant than *Nitrospirae* at ELSC (Fig 3.4).

Autotrophy. Given the prevalence of primary production by microbes in the dark oceans, both in hydrothermal plumes and in locations far removed from vents (Aristegui et al 2009, Reinthaler et al 2010), we identified genes involved in carbon fixation. Four of the six carbon fixation pathways currently known (Hügler and Sievert 2010) were confidently identified, including the Calvin-Benson-Bassham Cycle (CBB), the Reductive Tricarboxylic Acid Cycle (rTCA), the Reductive Acetyl-CoA pathway (Wood-Ljungdahl pathway), and the 3-Hydroxy propionate/4

Hydroxy-Butyrate Cycle (3HP/4HB). The CBB and R-TCA cycles were the most abundant carbon fixation pathways in the ELSC plumes.

Nitrogen metabolism. Genes for the uptake and use of urea were pervasive in the ELSC metagenomes. Important amongst these are *urea amidohydrolase* (*ureαβγEFH*) and *allophanate hydrolase* similar to previously identified genes in SUP05 and Marine Group I *Thaumarchaea* (Anantharaman et al 2013, Baker et al 2012).

Heterotrophy. Heterotrophic metabolism in the metagenomic bins was inferred based on prior knowledge of specific microbial groups, absence of carbon fixation mechanisms, presence of organic carbon compound transporters, and organic carbon compound degradation mechanisms. Amongst the most abundant heterotrophs observed in the ELSC metagenome were SAR11 (Lau14), Alteromonas (Lau4), Marine Group A/SAR406 (Lau47, Lau104) and Marixanthomonas (Lau23) (Fig.3.4). The most commonly observed genes for uptake of organic carbon compounds were amino acid transporters (annotated as branched chain and polar amino acid transporters) and di and tricarboxylate transporters.

Fermentation. An essentially-complete genome sequence was recovered from of a novel Candidate Division TM7 (*Saccharibacteria*) bacterium with SSU rRNA identity of 84% to the recently identified Candidatus *Saccharimonas aalborgensis* (Albertsen et al 2013). To our knowledge this represents the first description of fermentation in the pelagic ocean, though it was hypothesized long ago (Karl et al 1984).

Bioenergetic modeling of potential electron donors. Because measurements of the concentrations of dissolved chemicals in the samples studied here are not available, we used equilibrium thermodynamic reaction path modeling to estimate the geochemical concentrations

in the hydrothermal plumes at Abe, Kilo Moana and Mariner. Results indicate that similar to the hydrothermal plumes of GB, the ELSC plumes hold little H_2S, with most of the reduced S instead stored in the form of elemental sulfur (S_0). In contrast to Kilo Moana and Abe, plumes at Mariner had high concentrations of particulate Fe^{2+} and Mn^{2+}. In order to assess the relative importance of the different electron donors, we used bioenergetic modeling to estimate the amount of energy in the plume fluids by comparing the free energy yields from different microbial metabolisms involving a combination of diverse electron donors and electron acceptors. In the warmer rising hydrothermal plume fluids with predicted temperatures of 2.5°C-5°C, the total amount of available free energy in the plumes is estimated to range from 3.41-37.94 J/kg of plume fluid (Fig.3.6). Most of the energy is predicted to be derived from the oxidation of elemental S with O_2 (79.8-92.9%), while free energy yield from the oxidation of H_2, CH_4, NH_3, with O_2 was minor in comparison (Fig. 3.6). The free energy yields of Fe^{2+} and Mn^{2+} oxidation ranged from 0.6%- 13.2% and 0.4-4.7% respectively, with the higher values predicted for the Mariner vent site. Oxygen was predicted to be the most commonly used electron acceptor with a minor role for NO_3^- in the fluids of the rising plume (data not shown). Processes involving the use of alternate electron acceptors like NO_3^-, NO_2^-, NO, Fe^{3+}, Mn^{4+} and SO_4^{2-} were predicted to have a greater metabolic role at higher temperatures (>50°C) (data not shown).

Considering the dominant role predicted for aerobic sulfur oxidation, we used a coupled 2-D physical-bioenergetic model at Abe to assess the distribution of energy available from oxidation of elemental sulfur in the hydrothermal vent environment (Fig 3.7A). Our results indicate that the available free energy from elemental sulfur oxidation is highest in the rising plume up to a height of 200m, where the plume attains neutral buoyancy. Comparison of gene abundances associated with reduced sulfur oxidation at Abe indicate that both total abundance of

genes associated with oxidation of reduced sulfur species and those specific for elemental sulfur (*dsr*) correlate strongly with predictions of potential free-energy yield from the bioenergetic model. Genes associated with all reduced sulfur species demonstrated increased abundances in the plume in comparison to the background (Fig 3.7B), while specifically genes associated with elemental sulfur oxidation increased by seven-fold from the lower rising plume (1m) to the upper rising plume (200m).

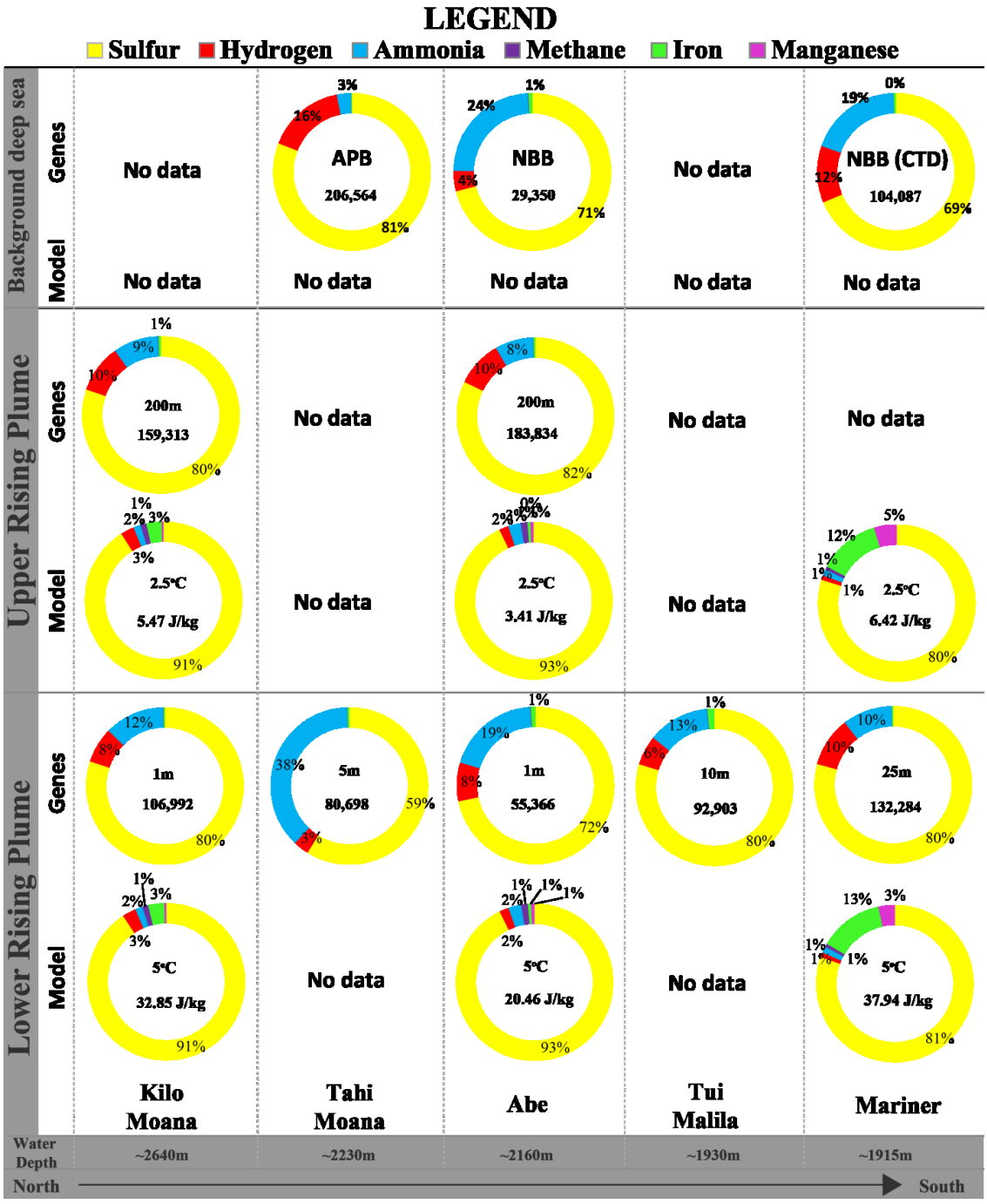


Fig. 3.6 Comparison of gene abundance and thermodynamic-bioenergetic estimates of available free energy associated with electron donors for lithotrophy in rising ELSC hydrothermal plumes. Inset data indicates the following from top to bottom: (1) Height above vent orifice for genes; temperature used for model. (2) Gene abundance associated with electron donor oxidation (normalized for gene length and dataset size); total available free energy per kg of plume fluid for model. APB – above plume background. NBB – near bottom background. CTD denotes sample taken with CTD rosette rather than SUPR. The gene abundance shown for sulfur oxidation is the sum of genes associated with oxidation of all reduced sulfur species.

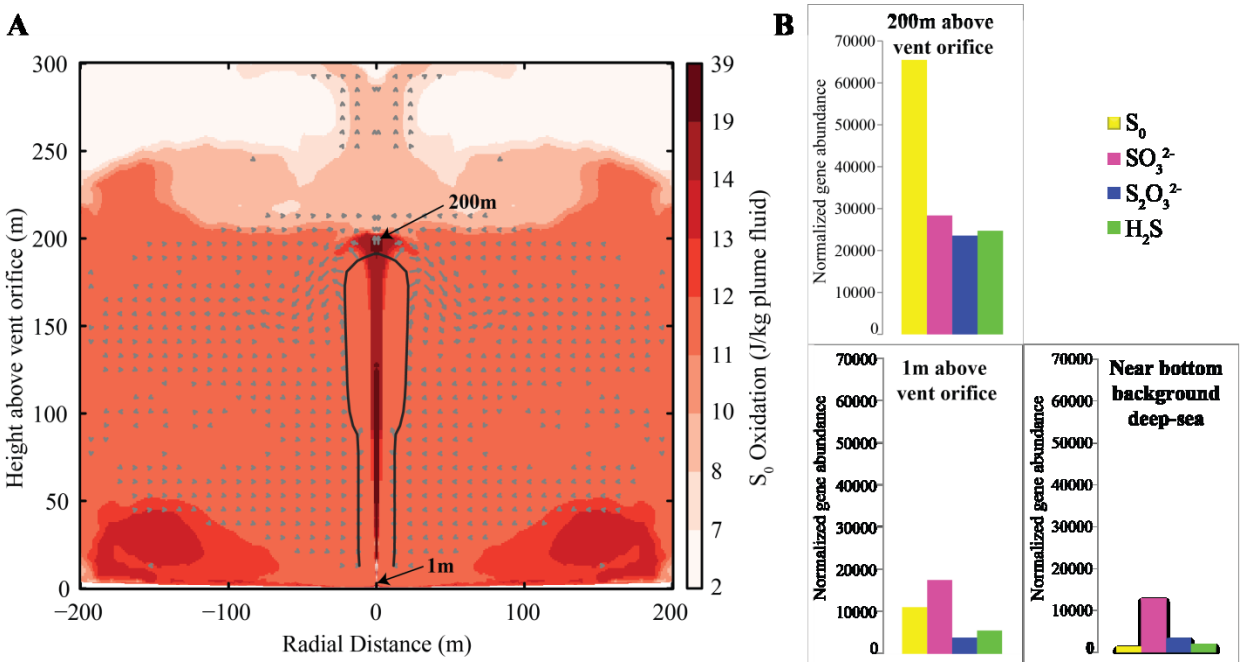


Fig. 3.7 A. 2-D Physical model of the ABE-A1 hydrothermal plume coupled to a bioenergetic model of elemental sulfur oxidation. Arrows illustrate the flow field. Contour line represents a water velocity of 0.0075 m s^{-1}. Arrows representing velocities greater than this value have been omitted for clarity. Available energy in the hydrothermal plume from S_0 oxidation is shown in the form of a heat map colored as indicated in the legend. **B.** Normalized abundance of genes associated with the oxidation of four forms of reduced sulfur in the ABE-A1 hydrothermal plume and background deep-sea (S_0 – *dsrA*; SO_3^{2-} - *aprA*; $S_2O_3^{2-}$ - *soxZ*; H_2S- *fcc*+*sqr*).

Overall distribution of lithotrophic metabolisms across the ELSC. Results from metagenomic analyses indicate that sulfur is the most commonly used electron donor in ELSC communities, consistent with bioenergetic models that indicate that aerobic oxidation of reduced sulfur species dominates the available free energy for chemosynthesis in the hydrothermal plumes of ELSC (Fig. 3.6). Sulfur-oxidizing bacteria dominate both in terms of abundance (i.e., abundance of reads mapped to genomes with sulfur oxidation genes) and diversity (i.e., number of genomes or genomic bins containing sulfur oxidation genes) (Fig. 3.8). Genomes with genes for H_2 oxidation are present at an abundance similar to those for sulfur oxidation but are less diverse, while those with genes for oxidation of nitrogen species and metals account for a

relatively small proportion of the genomic abundance and diversity of the ELSC community (Fig. 3.8).

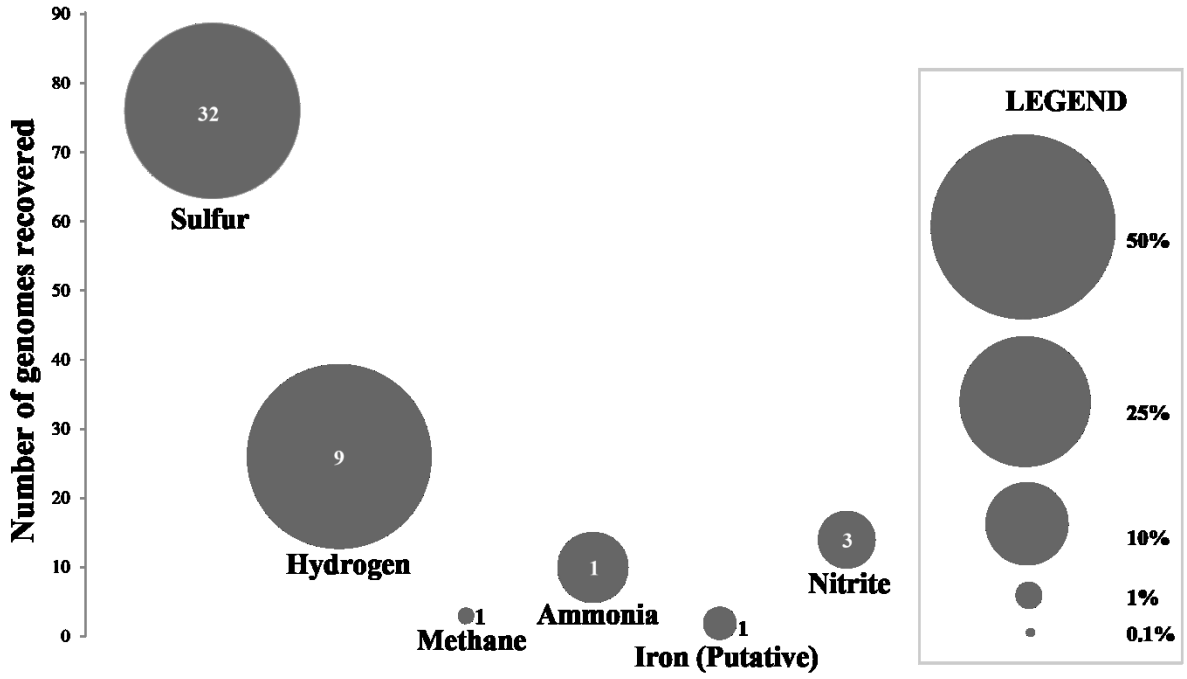


Fig. 3.8 Genomic abundance of organisms using six abundant electron donors in ELSC hydrothermal plumes. Size of the bubble indicates abundance (normalized read coverage) of identified bins as indicated in the legend. Numbers inside the circles indicate the number of distinct genomic bins associated with the oxidation of the respective electron donors.

3.5 Discussion

The Eastern Lau Spreading Center hosts hydrothermal vent fields along a north-south axis where underlying host rock chemistry shapes the hydrothermal vent fluid geochemistry. We used metagenomics to study the impact of these geochemical gradients on the genetic and metabolic potential of ELSC plume microbial communities. Through a combination of *de novo*

genomic assembly and enhanced tetranucleotide binning, distinct genomic bins were identified. Analyses of genomic bins resulted in the resolution of taxonomic groups that provide insights into the energy metabolism and functional biogeochemical roles of microorganisms in ELSC plumes. Our results show that although ELSC plumes are dominated by populations previously reported to be abundant and ubiquitous in the deep ocean, such as SUP05, SAR324, *Thaumarchaea* (Dick and Tebo 2010, Lesniewski et al 2012), and SAR11 (Thrash et al 2014), they also contain novel microbial populations including *Epsilonproteobacteria*, *Aquificales, Chloroflexi* and *Planctomycetes* that are indigenous to hydrothermal environments such as vent chimneys and extinct sulfides (Flores et al 2012, Sylvan et al 2013). The most abundant members of microbial communities of the five different hydrothermal plumes at ELSC are remarkably similar and dominated by sulfur-oxidizing bacteria. This indicates that the high concentration and bioenergetic potential of sulfur in the water column overprints the other geochemical differences in host rocks and end-member hydrothermal fluid chemistry along the ELSC (Mottl et al 2011) that have been found to influence microbial diversity on the seafloor (Flores et al, 2012; Sylvan et al 2013). Some differences in the membership of dominant organisms were observed between vent fields, chiefly the increased abundance of *Alteromonas* and *Marinobacter* populations at Mariner. These microbial groups have been shown to be involved in iron uptake (Li et al 2014) and iron and manganese oxidation (Edwards et al 2004, Singer et al 2011), respectively, and their increased abundance at Mariner correlates with increased iron and manganese concentrations in the plumes. However, determining the abundance of genes associated with iron and manganese oxidation in ELSC plumes is currently not possible as these genes are highly divergent and have yet to be identified conclusively.

Although many of the dominant microbial populations found in GB plumes (Dick and Tebo 2010, Lesniewski et al 2012) are also abundant at ELSC, methanotrophs that were diverse and abundant at GB (Li et al 2013) are only present at low abundance at ELSC (Fig. 3.8). This is likely related to methane concentrations of 5-57 μM in the hydrothermal vent fluids of the ELSC that are three orders of magnitude lower than sediment-hosted hydrothermal vents of GB. Bioenergetic modeling highlights this difference in the two vent systems, indicating that methane oxidation potentially contributes up to 93% of all available energy at GB (Anantharaman et al 2013), but only about 1% of the total available free energy from catabolic reactions at ELSC (Fig 3.6).

We observed the presence of multiple pathways for lithotrophy in many genomes (Fig. 3.4), including potential for oxidation of both sulfur and H_2 in SUP05, SAR324, *Epsilonproteobacteria* and *Aquificales*, and oxidation of methane and sulfur in *Methylococcae*. This suggests that metabolic versatility may be a common metabolic strategy in deep-sea microorganisms. While the ability to grow on multiple electron donors has been reported many times (Anantharaman et al 2013, Nakagawa et al 2005, Petersen et al 2011, Sheik et al 2013), to our knowledge this is the first report of potential for lithotrophy in a methanotroph. This suggests that facultative methanotrophy extends beyond use of organic acids and ethanol (Semrau et al 2011) and includes sulfur as an alternative energy source. Because the data presented here only reveals metabolic potential, we are unable to determine whether these metabolisms are simultaneously active in the same organism as observed previously for H_2 and S oxidation by SUP05 (Anantharaman et al 2013, Petersen et al 2011), or whether they are employed separately in distinct geochemical environments. An important implication of this metabolic plasticity is that the proliferation of some genes (e.g., H_2 oxidation) could be linked to

growth on other substrates (e.g., sulfur oxidation) (Reed et al, 2014). Although further work including cultivation and rate measurements of carbon fixation, methane and reduced sulfur consumption are needed to understand the ecology of these methylotrophic gammaproteobacteria in marine environments (Kessler et al 2011, Li et al 2013, Rivers et al 2013, Tavormina et al 2010), their widespread nature underlines the importance of our study.

Metabolic versatility associated with sulfur oxidation potentially has the ancillary effect of imparting functional redundancy for sulfur oxidation in ELSC plumes (Allison and Martiny 2008). The diversity and versatility of sulfur-oxidizing microorganisms that are abundant both in ELSC plumes as well as in the pelagic ocean could allow the metabolism of sulfur oxidation to be resilient in the dynamic setting of deep-sea hydrothermal plumes, potentially persisting in the face of long periods between reduced sulfur availability. Such extended deficits of sulfur could be brought on by perturbations to ocean currents and hydrothermal activity.

The shotgun sequencing approach employed here yielded sequence information on the whole plume community, including all three domains of life (Bacteria, Archaea, Eukarya) and all ecological roles (primary producers, heterotrophs, grazers, viruses). The prevalence of putative viruses in the ELSC metagenomes indicates that they are abundant in hydrothermal plumes, as has been observed in the broader deep oceans (Hara et al 1996). Indeed, detailed analysis of these viral genomes shows that they often contain sulfur oxidation genes that likely function as auxiliary metabolic genes and serve as an important reservoir of genetic diversity (Chapter IV). In addition, the finding of a *Bathymodiolinae* mussel genome despite no visual signs of animals on the sample filters suggests that either microscopic mussel larvae or tissues were present in the plume, supporting the notion that hydrothermal plumes serve as a dispersal vector for animals over large oceanic distances and regimes (Dick et al 2013, Mullineaux et al 1995). Our results

also indicate that heterotrophy is widespread in hydrothermal plumes; although most hydrothermal plume research has focused on primary production via chemosynthesis, clearly the fate of organic carbon as determined by heterotrophs also warrants attention. Chemosynthetic production in plumes can serve as a regionally significant source of organic carbon to the deep ocean (McCollum, 2000; De Angelis et al, 1993, Lam et al, 2008), providing fuel for heterotrophy in addition to that from sinking particles (Karl et al 1984).

3.6 Conclusions

Overall, our study shows that that ELSC hydrothermal plumes host a complex and diverse microbial community comprising archaea, bacteria, eukarya and viruses. Oxidation of reduced sulfur species constitutes the most abundant chemolithotrophic energy metabolism in ELSC hydrothermal plumes. Sulfur oxidation genes were observed in hydrogen and methane oxidizing organisms, suggesting that dominant primary producers in the ELSC hydrothermal plumes have diverse metabolic strategies. The abundance and diversity of sulfur oxidizing microorganisms, coupled with their metabolic diversity, hints at functional redundancy associated with sulfur oxidation that could allow it to persist in the dynamic settings of hydrothermal plumes. Although additional molecular evidence is necessary to understand the complex interplay between different electron donors utilized by plume microbes, our results provides the impetus to do so. Metagenomic analyses and bioenergetic models also point to a comparatively minor metabolic role for hydrogen, methane and ammonia oxidation in fuelling primary production at ELSC. Despite prominent differences in fluid geochemistry between sites at the ELSC, including highly enriched concentrations of iron at Mariner, the effect on the microbial community appears to be

relatively minor, being masked by the dominance of sulfur metabolism. Finally, the genomic data presented here provides opportunities for future studies with respect to two understudied components of deep-ocean microbial communities, eukarya and viruses.

Acknowledgements This project is funded in part by the Gordon and Betty Moore Foundation and the National Science Foundation Ridge2000 (R2K) program (OCE 1038006). We also thank the University of Michigan Rackham Graduate School Faculty Research Fellowship Program for their support and Sunit Jain for his assistance with bioinformatics.

Author Information The nucleotide sequences are available from DOE JGI-IMG/MER - Taxon Object IDs (Kilo Moana: 3300001680, Abe: 3300001681, Mariner: 3300001678, Tahi Moana: 3300001679, Tui Malila: 3300001676 and Guaymas: 3300001683). The authors declare no competing financial interests. The authors declare no competing financial interests.

3.7 Appendix B

CHAPTER III Supplementary Information

Contents

1. **Supplementary Table 1**
2. **Supplementary Table 2**
3. **Supplementary Table 3**
4. **Supplementary Table 4**
5. **Supplementary Table 5**
6. **Supplementary Table 6**
7. **Supplementary Table 7**
8. **Supplementary Table 8**

Supplementary Table 1. Sampling details

ELSC Site/Vent	Sample	Sample type	Date(DD/MM/YYYY)	Latitude/Longitude	Depth (m)	Filter size (µm)	No. of sequence reads
Mariner/MA3	TN236-J2440-2	Near bottom background	20/06/2009	S 22 10.818293 W 176 36.086423	1915	0.8	x
	TN236-J2440-6	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1915	0.8	x
	TN236-J2440-7	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1915	0.8	x
	TN236-J2440-10	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1910	0.8	x
	TN236-J2440-11	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1910	0.8	x
	TN236-J2440-14	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1900	0.8	x
	TN236-J2440-15	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1900	0.8	x
	TN236-J2440-18	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1890	0.8	185135248
	TN236-J2440-19	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1890	0.8	x
	TN236-J2440-21	Above plume background	20/06/2009	S 22 10.818293 W 176 36.086423	~1300	0.8	x
	TN236-J2440-24	Blank	20/06/2009	x	x	0.8	x
Abe/A1	TN236-J2449-2	Near bottom background	04/07/2009	S 22 45.677706 W 176 11.369574	2155	0.8	169488288
	TN236-J2449-4	Rising Plume	04/07/2009	S 22 45.677706 W 176 11.369574	2155	0.8	x
	TN236-J2449-5	Rising Plume	04/07/2009	S 22 45.677706 W 176 11.369574	2155	0.8	x
	TN236-J2449-10	Rising Plume	04/07/2009	S 22 45.677706 W 176 11.369574	2150	0.8	x
	TN236-J2449-14	Rising Plume	04/07/2009	S 22 45.677706 W 176 11.369574	2150	0.8	x
	TN236-J2449-21	Above plume background	04/07/2009	S 22 45.677706 W 176 11.369574	~1300	0.8	x
	TN236-J2449-24	Blank	04/07/2009	x	x	0.8	x
Abe/A1	TN236-J2435-3	Near bottom background	13/06/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN236-J2435-4	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN236-J2435-5	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN236-J2435-8	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2149	0.8	x
	TN236-J2435-9	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2149	0.8	x
	TN236-J2435-12	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2129	0.8	x
	TN236-J2435-13	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2129	0.8	x
	TN236-J2435-16	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2069	0.8	x
	TN236-J2435-17	Rising Plume	13/06/2009	S 20 45.672883 W 176 11.434418	2069	0.8	x
	TN236-J2435-20	Above plume background	13/06/2009	S 20 45.672883 W 176 11.434418	~1300	0.8	x
	TN236-J2435-24	Blank	13/06/2009	x	x	0.8	x

Tahi Moana/SP2	TN236-J2450-2	Near bottom background	05/07/2009	S 20 40.894100 W 176 10.940463	2235	0.8	x
	TN236-J2450-4	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2235	0.8	x
	TN236-J2450-5	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2235	0.8	x
	TN236-J2450-9	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2229	0.8	186087990
	TN236-J2450-10	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2229	0.8	x
	TN236-J2450-14	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2229	0.8	x
	TN236-J2450-20	Above plume background	05/07/2009	S 20 40.894100 W 176 10.940463	~1300	0.8	x
	TN236-J2450-24	Blank	05/07/2009	x	x	0.8	x
Kilo Moana/KM 1	TN236-J2436-2	Near bottom background	15/06/2009	S 20 3.229502 W 176 8.015363	2665	0.8	x
	TN236-J2436-4	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2665	0.8	x
	TN236-J2436-5	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2665	0.8	x
	TN236-J2436-8	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2655	0.8	x
	TN236-J2436-9	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2655	0.8	x
	TN236-J2436-12	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2635	0.8	x
	TN236-J2436-13	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2635	0.8	x
	TN236-J2436-16	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2605	0.8	174530426
	TN236-J2436-17	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2605	0.8	x
	TN236-J2436-20	Above plume background	15/06/2009	S 20 3.229502 W 176 8.015363	~1300	0.8	x
	TN236-J2436-24	Blank	15/06/2009	x	x	0.8	x
Tui Malila/TM1	TN236-J2447-2	Near bottom background	01/07/2009	S 21 59.401181 W 176 34.124651	1929	0.8	x
	TN236-J2447-4	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1929	0.8	x
	TN236-J2447-5	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1929	0.8	x
	TN236-J2447-9	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1919	0.8	x
	TN236-J2447-10	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1919	0.8	187067650
	TN236-J2447-14	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1899	0.8	x
	TN236-J2447-18	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1899	0.8	x
	TN236-J2447-20	Above plume background	01/07/2009	S 21 59.401181 W 176 34.124651	~1300	0.8	x
	TN236-J2447-24	Blank	01/07/2009	x	x	0.8	x
Tui Malila/TMS 1	TN236-J2442-2	Near bottom background	23/06/2009	S 21 59.274547 W 176 34.060503	1928	0.8	x
	TN236-J2442-5	Rising Plume	23/06/2009	S 21 59.274547 W 176 34.060503	1928	0.8	x

	TN236-J2442-9	Above plume background	23/06/2009	S 21 59.274547 W 176 34.060503	~1300	0.8	x
	TN236-J2442-20	Above plume background	23/06/2009	S 21 59.274547 W 176 34.060503	~1000	0.8	x
	TN236-J2442-24	Blank	23/06/2009	x	x	0.8	x
Tahi Moana/SP1	TN236-J2445-2	Near bottom background	29/06/2009	S 20 40.927843 W 176 11.001806	2230	0.8	x
	TN236-J2445-4	Rising Plume	29/06/2009	S 20 40.927843 W 176 11.001806	2230	0.8	x
	TN236-J2445-10	Rising Plume	29/06/2009	S 20 40.927843 W 176 11.001806	2220	0.8	x
	TN236-J2445-13	Rising Plume	29/06/2009	S 20 40.927843 W 176 11.001806	2220	0.8	x
	TN236-J2445-20.2	Above plume background	29/06/2009	S 20 40.927843 W 176 11.001806	~1300	0.2	181482188
	TN236-J2445-20.8	Above plume background	29/06/2009	S 20 40.927843 W 176 11.001806	~1000	0.8	x
	TN236-J2445-24	Blank	29/06/2009	x	x	0.8	x
Kilo Moana/KM 4	TN235-J2424-7	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2639	0.8	x
	TN235-J2424-8	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2639	0.8	118751402
	TN235-J2424-11	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2629	0.8	x
	TN235-J2424-12	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2629	0.8	x
	TN235-J2424-16	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2599	0.8	x
	TN235-J2424-19	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2599	0.8	x
	TN235-J2424-20	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2439	0.8	157276514
	TN235-J2424-22	Above plume background	22/05/2009	S 20 3.234200 W 176 8.008000	x	0.8	x
Abe/A1	TN235-J2426-7	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	181094744
	TN235-J2426-12	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	2149	0.8	x
	TN235-J2426-16	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	2119	0.8	x
	TN235-J2426-20	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	1959	0.8	168325584
	TN235-J2426-24	Blank	25/05/2009	x	x	0.8	x
Abe/A1	TN235-J2427-2	Near bottom background	27/05/2009	S 20 45.672883 W 176 11.434418	2155	0.8	x
	TN235-J2427-7	Rising Plume	27/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN235-J2427-8	Rising Plume	27/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN235-J2427-19	Rising Plume	27/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
	TN235-J2427-20	Rising Plume	27/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	x
Kilo Moana	TN236-CTD-KM-IP1	Neutrally buoyant plume	13/06/2009	S 20 3.246489 W 176 8.011308	2305	0.2	x
	TN236-CTD-	Neutrally	13/06/2009	S 20 3.246489 W 176	2315	0.2	188964668

	KM-IP2	buoyant plume		8.011308			
	TN236-CTD-KM-BP1	Below Plume background	13/06/2009	S 20 3.246489 W 176 8.011308	2400	0.2	x
	TN236-CTD-KM-BP2	Below Plume background	13/06/2009	S 20 3.246489 W 176 8.011308	2350	0.2	x
Tui Malila	TN236-CTD-Tui-IP1	Neutrally buoyant plume	20/06/2009	S 21.98790 W 176.56768	1675	0.2	x
	TN236-CTD-Tui-IP2	Neutrally buoyant plume	20/06/2009	S 21.98790 W 176.56768	1650	0.2	x
	TN236-CTD-Tui-BP1	Below Plume background	20/06/2009	S 21.98790 W 176.56768	1750	0.2	x
Tahi Moana/1D	TN236-CTD-TM-IP1	Neutrally buoyant plume	05/07/2009	S 20 40.3905 W 176 10.8435	2050	0.2	x
Mariner	TN236-CTD-Mar-IP1	Neutrally buoyant plume	19/06/2009	S 20 10.8035165 W 176 36.074496	1740	0.2	x
	TN236-CTD-Mar-IP2	Neutrally buoyant plume	19/06/2009	S 20 10.8035165 W 176 36.074496	1725	0.2	x
	TN236-CTD-Mar-BP1	Below Plume background	19/06/2009	S 20 10.8035165 W 176 36.074496	1785	0.2	154673072
	TN236-CTD-Mar-BP2	Below Plume background	19/06/2009	S 20 10.8035165 W 176 36.074496	1780	0.2	x
	TN236-CTD-Mar-BP1	Above Plume background	19/06/2009	S 20 10.8035165 W 176 36.074496	1680	0.2	x
	TN236-CTD-Mar-BP2	Above Plume background	19/06/2009	S 20 10.8035165 W 176 36.074496	1600	0.2	x

Supplementary Table 2. Assembly statistics

Velvet+MetaVetvet+Minimus	IDBA-UD
Total Sequences: 640793	Total Sequences: 1091773
Total Bases: 841,239,120	Total Bases: 1,511,701,270
Total Bases in sequences greater than 4000: 143,260,994	Total Bases in sequences greater than 4000: 553,583,049
Longest Sequence Length: 818,219	Longest Sequence Length: 830,465
Shortest Sequence Length: 501	Shortest Sequence Length: 91
Mean Length: 1313 bp	Mean Length: 1385 bp
Mean Length after setting the 4000 base limit: 6886 bp	Mean Length after setting the 4000 base limit: 8909 bp
Number of Sequences with Length greater than 4000 bases: 20804	Number of Sequences with Length greater than 4000 bases: 62135
Nucleotide Distribution after setting the 4000 base limit:	Nucleotide Distribution after setting the 4000 base limit:
A : 27.6824 %	A : 28.9150 %
T : 27.6417 %	T : 28.4189 %
G : 22.2372 %	G : 21.2481 %
C : 22.3155 %	C : 21.4015 %
N : 0.1233 %	N : 0.0165 %
Sequence Length Range: No. of Contigs	Sequence Length Range: No. of Contigs
1-2000: 554653	1-2000: 909021
2001-4000: 65336	2001-4000: 120617
4001-6000: 13164	4001-6000: 31034
6001-8000: 4049	6001-8000: 12248
8001-10000: 1619	8001-10000: 6256
10001-12000: 713	10001-12000: 3417
12001-14000: 367	12001-14000: 2203
14001-16000: 229	14001-16000: 1565
16001-18000: 144	16001-18000: 1055
18001-20000: 95	18001-20000: 773
20001-22000: 63	20001-22000: 581
22001-24000: 45	22001-24000: 440
24001-26000: 51	24001-26000: 355
26001-28000: 29	26001-28000: 284
28001-30000: 24	28001-30000: 216
30001-32000: 16	30001-32000: 212
32001-34000: 19	32001-34000: 182
34001-36000: 13	34001-36000: 130

36001-38000: 13	36001-38000: 127
38001-40000: 14	38001-40000: 90
40001-42000: 16	40001-42000: 105
42001-44000: 6	42001-44000: 86
44001-46000: 13	44001-46000: 60
46001-48000: 11	46001-48000: 52
48001-50000: 3	48001-50000: 60
50001-52000: 6	50001-52000: 47
52001-54000: 4	52001-54000: 38
54001-56000: 9	54001-56000: 38
56001-58000: 7	56001-58000: 38
58001-60000: 3	58001-60000: 25
60001-62000: 6	60001-62000: 23
62001-64000: 2	62001-64000: 30
64001-66000: 5	64001-66000: 19
66001-68000: 4	66001-68000: 16
68001-70000: 3	68001-70000: 19
70001-72000: 1	70001-72000: 16
72001-74000: 1	72001-74000: 13
74001-76000: 1	74001-76000: 20
76001-78000: 1	76001-78000: 14
78001-80000: 2	78001-80000: 13
80001-82000: 7	80001-82000: 9
82001-84000: 1	82001-84000: 13
88001-90000: 3	84001-86000: 9
90001-92000: 1	86001-88000: 9
92001-94000: 1	88001-90000: 14
96001-98000: 1	90001-92000: 7
98001-100000: 2	92001-94000: 7
100001-1000000: 17	94001-96000: 7
	96001-98000: 5
	98001-100000: 5
	100001-1000000: 150

Supplementary Table 3. Details of identified ORFs

Assembly	Protein coding genes	rRNA genes	tRNA genes	COG clusters	Pfam clusters
KiloMoana	401957	294	4218	4456	13850
Abe	664114	188	8039	4443	14587
Mariner	202932	186	2322	4203	12115
TahiMoana	267200	181	2933	4336	13038
TuiMalila	141590	94	1614	3771	10753

Supplementary Table 4. Details of identified archaeal, bacterial and eukarya bins

Bin name	Organism/Group	No. of Contigs	Total size(bp)	%GC	Estimated no. of genomes
Lau1	Bacteria;Candidate Division TM7	5	909,913	43.52	1
Lau2	Bacteria;Cyanobacteria/Melainabacteria;MLE1-12	360	4,104,129	49.49	1
Lau3	Bacteria;Chloroflexi;SAR202 clade	821	6,029,363	56.89	1
Lau4	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae;Alteromonas	661	9,888,618	43.86	2
Lau5	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae;Marinobacter	195	9,438,787	56.12	2
Lau6	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II (Group C)	303	7,737,129	44.58	4
Lau7	Bacteria;Planctomycetes;Phycisphaerae;Phycisphaerales;Phycisphaeraceae;JL-ETNP-F27	1335	11,146,757	45.04	5
Lau8	Eukaryota;Opisthokonta;Metazoa;Mollusca;Bivalvia;Mytiloida;Bathymodiolus tangaroa	9584	54,022,390	33.91	1
Lau9	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;ZD0405 (ARCTIC96BD-19)	291	2,388,907	39.34	1
Lau10	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;SUP05 Clade	2159	14,020,866	38.01	8
Lau11	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Moraxellaceae;Acinetobacter	696	8,573,787	39.02	2
Lau12	Bacteria;Chloroflexi;Anaerolineae;Anaerolineales;Anaerolineaceae	111	682,424	42.15	1
Lau14	Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade	4284	30,724,625	30.06	26
Lau15	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Helicobacteraceae;Sulfurimonas	96	677,211	34.51	1
Lau16	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;SAR86 clade	448	4,266,187	37.37	3
Lau17	Bacteria; Nitrospinae; Nitrospinia; Nitrospinales; Nitrospinaceae	1713	13,893,733	39.39	6
Lau19	Archaea;Thaumarchaeota;Marine Group I	2096	15,151,636	33.46	10
Lau20	Bacteria;Proteobacteria;Deltaproteobacteria;SAR324 clade (Marine Group B)	1435	14,645,818	42.12	8
Lau21	Bacteria;Poribacteria	1247	19,585,039	44.23	2
Lau22	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Mesonia	146	2,464,964	34.89	1
Lau23	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Marixanthomonas	14	3,250,025	40.28	2
Lau24	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae;Marixanthomonas	131	1,056,646	40.83	1
Lau26	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Thalassospira	66	555,626	52.41	1
Lau27	Bacteria;Cyanobacteria;SHA-109	318	2,573,937	54.10	1
Lau28	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae	399	3,845,565	45.21	1
Lau29	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae	87	556,445	50.32	1
Lau30	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales	1134	10,534,609	45.25	2
Lau31	Bacteria;Bacteroidetes;Unclassified	453	4,462,640	38.23	1
Lau32	Bacteria;Bacteroidetes;Unclassified	78	722,333	40.67	1
Lau33	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;Flavobacteriaceae	59	373,613	40.53	1
Lau34	Archaea;Euryarchaeota;Halobacteria;Halobacteriales;Deep Sea Hydrothermal Vent Gp 6(DHVEG-6)	293	2,770,233	52.56	1
Lau35	Bacteria;Planctomycetes;Pla3 lineage	787	6,055,579	57.39	1
Lau36	Bacteria;Planctomycetes;Pla3 lineage	401	4,733,687	60.40	5

Lau40	Bacteria;Acidobacteria;Acidobacteria;BPC102	112	728,843	53.76	1
Lau41	Bacteria; Nitrospinae; Nitrospinia; Nitrospinales; Nitrospinaceae	2066	16,427,353	46.90	5
Lau42	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;NS9 marine group	274	5,807,840	43.39	3
Lau44	Bacteria;Nitrospirae	408	4,397,644	52.21	3
Lau45	Bacteria;Gemmatimonadetes;Gemmatimonadetes;BD2-11 group	378	2,772,136	54.30	1
Lau46	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Hypho monadaceae	111	943,398	55.70	1
Lau47	Bacteria;Deferribacteres;Deferribacteres;Deferribacterales;SAR406 clade (Marine Group A)	688	13,711,201	45.88	7
Lau51	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;SUP 05 Clade	272	3,134,417	36.38	1
Lau52	Bacteria;Bacteroidetes;Unclassified	106	780,364	42.65	1
Lau53	Bacteria;Proteobacteria;Deltaproteobacteria;Unclassified	56	323,536	48.99	1
Lau62	Bacteria;Proteobacteria;Gammaproteobacteria;E01-9C-26 marine group	1126	16,062,766	52.41	5
Lau64	Bacteria;Bacteroidetes;Unclassified	758	9,477,688	43.38	2
Lau65	Unclassified	445	2,679,292	43.14	2
Lau66	Bacteria;Bacteroidetes;Flavobacteria;Unclassified	491	4,091,476	44.27	1
Lau92	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II (Group B)	272	3,084,605	58.48	3
Lau93	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II	84	675,085	55.05	1
Lau94	Bacteria;Planctomycetes;Planctomycetia;Planctomycetales;Planctomy cetaceae	315	2,328,079	53.02	2
Lau95	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	37	223,029	49.01	1
Lau96	Bacteria;Planctomycetes;Planctomycetia;Planctomycetales;Planctomy cetaceae	383	3,436,646	48.68	1
Lau98	Bacteria;Proteobacteria;Deltaproteobacteria;Unclassified	218	1,804,876	52.32	1
Lau101	Bacteria;Proteobacteria;Unclassified	6	41,648	56.77	1
Lau103	Bacteria;Deferribacteres;Deferribacteres;Deferribacterales;SAR406 clade (Marine Group A)	4434	43,187,681	41.00	14
Lau104	Bacteria;Deferribacteres;Deferribacteres;Deferribacterales;SAR406 clade (Marine Group A)	443	3,720,913	38.27	2
Lau112	Bacteria;Unclassified	20	129,093	45.18	1
Lau113	Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales;Hyd2 4-01 (OPU3)	312	5,302,821	42.67	3
Lau114	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	233	2,023,828	51.56	1
Lau115	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	20	149,321	48.72	1
Lau116	Bacteria;Proteobacteria;Gammaproteobacteria;Order Incertae cedis (Tubeworm symbiont)	823	11,148,002	52.75	5
Lau117	Bacteria;Unclassified	3	15,852	49.67	1
Lau118	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales	492	4,917,635	50.47	2
Lau122	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;SUP 05 Clade	302	1,876,788	33.89	1
Lau123	Bacteria;Unclassified	269	1,838,828	55.05	1
Lau125	Archaea;Euryarchaeota;Unclassified	65	415,496	30.15	1
Lau126	Bacteria;Unclassified	40	232,436	49.11	1
Lau127	Bacteria;Proteobacteria;Alphaproteobacteria;Unclassified	32	186,078	44.71	1
Lau129	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Hel icobacteraceae;Sulfurimonas	17	96,262	35.98	1
Lau143	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	6	30,876	34.90	1
Lau158	Bacteria;Verrucomicrobia;Unclassified	35	308,256	38.65	1

Lau159	Bacteria;Proteobacteria;Unclassified	9	80,800	42.17	1
Lau163	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	32	215,013	53.70	1
Lau164	Bacteria;Verrucomicrobia;Unclassified	10	73,599	56.43	1
Lau176	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	59	829,333	54.32	1
Lau177	Bacteria;Proteobacteria;Gammaproteobacteria;Pseudomonadales;Pseudomonadaceae	53	463,345	59.13	1
Lau178	Bacteria;Proteobacteria;Gammaproteobacteria;Methylococcales	40	237,587	53.25	1
Lau179	Bacteria;Proteobacteria;Unclassified	8	107,643	39.59	1
Lau184	Bacteria;Verrucomicrobia;Unclassified	5	23,710	36.52	1
Lau190	Bacteria;Verrucomicrobia;Arctic97B-4 marine group	417	2,886,548	58.15	2
Lau197	Bacteria;Proteobacteria;Deltaproteobacteria;Unclassified	2	20,751	44.98	1
Lau198	Bacteria;Unclassified	4	27,618	46.17	1
Lau203	Bacteria;Proteobacteria;Unclassified	58	438,706	48.30	1
Lau208	Bacteria;Verrucomicrobia;Unclassified	6	30,481	38.88	1
Lau210	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales	5	31,669	39.17	1
Lau227	Bacteria;Aquificae;Aquificae;Aquificales;Aquificaceae	171	1,296,646	48.41	2
Lau229	Bacteria;Proteobacteria;Epsilonproteobacteria	127	1,142,760	36.98	1
Lau230	Bacteria;Proteobacteria;Gammaproteobacteria;Salinisphaerales;Salinisphaeraceae;ZD0417 marine group	376	3,589,670	42.05	3
Lau231	Bacteria;Unclassified	20	190,602	39.64	3
Lau233	Bacteria;Unclassified	35	210,608	37.71	1
Lau236	Bacteria;Unclassified	122	879,107	55.12	1
Lau237	Bacteria;Aquificae;Aquificae;Aquificales;Desulfurobacteriaceae	27	148,555	49.01	1
Lau248	Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Alteromonadaceae;Alteromonas	1	4,388	47.47	1
Lau249	Bacteria;Unclassified	1	63,694	41.61	1
Lau252	Bacteria;Bacteroidetes;Flavobacteria;Flavobacteriales;Cryomorphaceae;Unclassified	21	120,381	38.76	1
Lau259	Bacteria;Unclassified	87	501,503	42.49	1
Lau265	Bacteria;Verrucomicrobia;Unclassified	2	13,130	43.42	1
Lau269	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	12	84,379	46.85	1
Lau331	Archaea;Thaumarchaeota;Marine Group I	12	79,139	40.58	1
Lau371	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	9	45,561	41.34	1
Lau374	Bacteria;Proteobacteria;Alphaproteobacteria;Unclassified	474	3,950,903	49.99	1
Lau375	Bacteria;Proteobacteria;Deltaproteobacteria;SAR324 clade (Marine Group B)	551	3,959,011	42.64	2
Lau376	Bacteria;Proteobacteria;Alphaproteobacteria;Unclassified	598	5,415,484	44.70	2
Lau377	Bacteria;Poribacteria	328	8,726,215	42.97	3
Lau378	Bacteria;Proteobacteria;Gammaproteobacteria;Unclassified	139	1,320,110	38.32	2
Lau379	Bacteria;Unclassified	17	108,542	45.11	1

Supplementary Table 5. Details of viral bins

Bin name	Organism/Group	% of unclassified/viral genes	No. of Contigs	Total size(bp)	%GC	No. of genomes
Lau13	Unclassified/Putative Phage	89.43%	97	906,225	32.81	4
Lau37	Unclassified/Putative Phage	97.17%	24	267,668	39.56	5
Lau38	Unclassified/Putative Phage	97.06%	11	110,943	42.79	4
Lau39	Unclassified/Putative Phage	94.01%	10	218,585	35.75	4
Lau43	Unclassified/Putative Phage	40.91%	161	1,268,671	52.21	5
Lau49	Unclassified/Putative Phage	46.42%	64	665,092	56.63	2
Lau50	Unclassified/Putative Phage	79.11%	39	415,853	35.45	3
Lau54	Unclassified/Putative Phage	90.81%	67	498,303	42.19	4
Lau55	Unclassified/Putative Phage	90.75%	136	958,976	39.62	5
Lau56	Unclassified/Putative Phage	50.00%	2	19,029	52.07	2
Lau57	Unclassified/Putative Phage	98.81%	14	99,492	38.82	3
Lau58	Unclassified/Putative Phage	96.97%	6	40,910	36.58	3
Lau59	Unclassified/Putative Phage	97.87%	13	60,899	53.25	4
Lau61	Unclassified/Putative Phage	95.06%	7	34,875	54.40	3
Lau63	Unclassified/Putative Phage	95.48%	12	65,347	56.15	5
Lau67	Unclassified/Putative Phage	60.00%	12	84,271	45.64	3
Lau68	Unclassified/Putative Phage	86.11%	14	118,281	38.27	4
Lau69	Unclassified/Putative Phage	100.00%	9	55,064	51.12	3
Lau70	Unclassified/Putative Phage	82.86%	4	20,305	53.31	1
Lau72	Unclassified/Putative Phage	100.00%	5	33,263	47.71	3
Lau73	Unclassified/Putative Phage	96.97%	8	51,647	45.93	4
Lau74	Unclassified/Putative Phage	44.44%	6	48,078	41.64	3
Lau75	Unclassified/Putative Phage	86.67%	5	39,146	42.64	3
Lau76	Unclassified/Putative Phage	100.00%	6	42,252	43.52	3
Lau77	Viruses;dsdna viruses (no rna stage);Caudovirales;Myoviridae;T4-like viruses	81.39%	56	823,098	36.81	3
Lau78	Unclassified/Putative Phage	97.02%	12	153,339	45.94	3
Lau79	Unclassified/Putative Phage	100.00%	2	13,172	48.39	1
Lau80	Unclassified/Putative Phage	90.22%	5	54,015	43.61	3
Lau81	Unclassified/Putative Phage	97.59%	14	111,548	46.40	3
Lau82	Unclassified/Putative Phage	100.00%	2	16,979	44.13	2
Lau83	Unclassified/Putative Phage	88.89%	2	19,354	41.04	1
Lau84	Unclassified/Putative Phage	89.83%	9	65,626	42.48	3
Lau85	Viruses;dsdna viruses (no rna stage);Caudovirales;Myoviridae;T4-like viruses	80.85%	59	633,699	36.06	4
Lau86	Unclassified/Putative Phage	70.59%	3	25,238	50.03	3
Lau88	Unclassified/Putative Phage	100.00%	4	31,494	47.11	3
Lau89	Unclassified/Putative Phage	57.23%	12	101,643	46.65	5

Lau90	Unclassified/Putative Phage	34.78%	5	52,667	39.61	2
Lau91	Unclassified/Putative Phage	98.08%	8	60,391	45.55	2
Lau99	Unclassified/Putative Phage	50.00%	5	39,180	47.24	4
Lau100	Unclassified/Putative Phage	98.46%	4	26,653	53.86	3
Lau102	Unclassified/Putative Phage	98.00%	5	156,492	40.85	1
Lau105	Unclassified/Putative Phage	36.84%	9	44,553	30.42	1
Lau106	Unclassified/Putative Phage	25.00%	5	30,862	46.02	3
Lau107	Unclassified/Putative Phage	28.57%	6	28,868	43.20	4
Lau108	Unclassified/Putative Phage	74.07%	11	53,957	38.68	5
Lau109	Unclassified/Putative Phage	100.00%	3	15,772	44.46	2
Lau110	Unclassified/Putative Phage	100.00%	11	86,118	40.52	4
Lau111	Unclassified/Putative Phage	58.33%	11	62,590	42.52	4
Lau120	Unclassified/Putative Phage	88.82%	37	181,802	41.69	2
Lau121	Unclassified/Putative Phage	94.51%	33	163,604	32.91	2
Lau128	Unclassified/Putative Phage	100.00%	15	112,930	56.04	4
Lau130	Unclassified/Putative Phage	85.00%	6	68,113	38.82	3
Lau131	Unclassified/Putative Phage	100.00%	5	36,407	49.18	4
Lau132	Unclassified/Putative Phage	100.00%	7	69,732	53.05	4
Lau133	Unclassified/Putative Phage	100.00%	28	202,724	43.26	3
Lau134	Unclassified/Putative Phage	75.00%	6	34,286	52.18	2
Lau135	Unclassified/Putative Phage	99.15%	19	145,920	43.13	2
Lau136	Unclassified/Putative Phage	93.02%	3	15,424	38.21	1
Lau137	Unclassified/Putative Phage	100.00%	3	13,926	39.18	3
Lau138	Unclassified/Putative Phage	100.00%	2	17,122	44.65	2
Lau139	Unclassified/Putative Phage	82.35%	3	105,599	39.92	3
Lau140	Unclassified/Putative Phage	81.31%	6	65,699	34.84	5
Lau141	Unclassified/Putative Phage	94.74%	20	184,765	36.07	4
Lau142	Unclassified/Putative Phage	95.65%	8	112,031	36.07	4
Lau144	Unclassified/Putative Phage	100.00%	1	4,266	37.79	2
Lau146	Unclassified/Putative Phage	100.00%	4	19,412	48.00	2
Lau147	Unclassified/Putative Phage	57.14%	5	39,643	36.57	3
Lau148	Unclassified/Putative Phage	88.66%	4	39,428	33.40	2
Lau149	Unclassified/Putative Phage	91.40%	12	65,985	37.13	4
Lau150	Unclassified/Putative Phage	97.67%	13	79,549	38.88	3
Lau151	Unclassified/Putative Phage	96.67%	12	116,964	37.17	5
Lau152	Unclassified/Putative Phage	94.12%	12	126,665	41.22	4
Lau153	Unclassified/Putative Phage	97.50%	6	42,312	44.05	3
Lau155	Unclassified/Putative Phage	97.14%	12	67,627	35.44	2
Lau156	Unclassified/Putative Phage	100.00%	4	20,347	35.88	2
Lau160	Unclassified/Putative Phage	64.29%	4	21,657	49.05	2
Lau161	Unclassified/Putative Phage	100.00%	4	24,106	49.63	4

Lau162	Unclassified/Putative Phage	87.65%	6	46,857	51.60	3
Lau166	Unclassified/Putative Phage	91.65%	35	334,702	34.60	4
Lau167	Unclassified/Putative Phage	92.68%	13	81,121	33.29	4
Lau168	Unclassified/Putative Phage	83.33%	10	84,748	34.57	4
Lau169	Unclassified/Putative Phage	60.00%	4	20,769	37.97	4
Lau170	Unclassified/Putative Phage	53.57%	11	54,924	51.80	5
Lau171	Unclassified/Putative Phage	100.00%	12	80,772	45.14	4
Lau172	Unclassified/Putative Phage	84.45%	9	56,636	51.54	1
Lau173	Bacteria;Proteobacteria;Gammaproteobacteria;Oceanospirillales;Halomonadaceae	9.47%	25	155,489	53.94	2
Lau174	Unclassified/Putative Phage	100.00%	5	27,495	47.92	4
Lau175	Unclassified/Putative Phage	85.71%	3	16,053	48.28	2
Lau180	Unclassified/Putative Phage	80.91%	21	216,322	41.30	4
Lau181	Unclassified/Putative Phage	96.55%	5	37,947	41.55	3
Lau182	Unclassified/Putative Phage	94.12%	7	54,907	41.26	4
Lau183	Unclassified/Putative Phage	93.02%	4	57,524	44.51	3
Lau185	Unclassified/Putative Phage	100.00%	6	53,218	42.61	4
Lau186	Unclassified/Putative Phage	54.84%	5	28,437	40.04	5
Lau187	Unclassified/Putative Phage	78.95%	6	32,069	31.03	3
Lau189	Unclassified/Putative Phage	80.69%	56	372,832	25.42	3
Lau192	Unclassified/Putative Phage	99.38%	17	75,955	39.47	4
Lau193	Unclassified/Putative Phage	100.00%	5	24,097	39.59	3
Lau194	Unclassified/Putative Phage	97.96%	7	37,358	39.98	4
Lau195	Unclassified/Putative Phage	81.82%	4	20,184	40.08	3
Lau196	Unclassified/Putative Phage	59.38%	9	44,088	55.73	2
Lau199	Unclassified/Putative Phage	100.00%	5	21,049	45.54	3
Lau200	Unclassified/Putative Phage	100.00%	4	18,435	53.26	3
Lau201	Unclassified/Putative Phage	69.23%	6	58,028	45.03	4
Lau204	Unclassified/Putative Phage	93.33%	7	58,332	60.77	2
Lau205	Unclassified/Putative Phage	100.00%	7	38,521	45.72	3
Lau206	Unclassified/Putative Phage	100.00%	7	41,711	41.51	4
Lau207	Unclassified/Putative Phage	66.67%	5	23,211	40.40	4
Lau209	Unclassified/Putative Phage	66.67%	7	48,035	37.52	4
Lau211	Unclassified/Putative Phage	80.00%	5	26,792	41.41	3
Lau212	Unclassified/Putative Phage	100.00%	4	17,419	40.21	3
Lau213	Unclassified/Putative Phage	100.00%	6	30,420	41.41	3
Lau214	Unclassified/Putative Phage	54.55%	8	40,092	40.06	3
Lau215	Unclassified/Putative Phage	100.00%	6	40,915	40.67	4
Lau216	Unclassified/Putative Phage	100.00%	6	54,858	40.21	3
Lau217	Unclassified/Putative Phage	95.83%	8	62,603	50.88	2
Lau218	Viruses;dsdna viruses (no rna stage);Caudovirales;Podoviridae	73.33%	6	83,236	34.56	3
Lau219	Unclassified/Putative Phage	100.00%	4	19,816	49.43	2

Lau220	Viruses;dsdna viruses (no rna stage);Unclassified	84.91%	32	224,624	41.57	3
Lau221	Unclassified/Putative Phage	66.67%	5	42,040	41.44	3
Lau222	Unclassified/Putative Phage	53.85%	7	54,480	44.04	3
Lau223	Unclassified/Putative Phage	95.00%	5	36,732	43.19	5
Lau224	Unclassified/Putative Phage	73.08%	5	42,140	41.32	3
Lau225	Unclassified/Putative Phage	75.00%	8	60,916	41.35	3
Lau226	Unclassified/Putative Phage	73.17%	4	25,815	43.64	4
Lau228	Unclassified/Putative Phage	64.71%	77	377,454	14.39	3
Lau232	Unclassified/Putative Phage	95.06%	12	65,147	42.50	2
Lau234	Unclassified/Putative Phage	92.50%	16	81,867	19.86	2
Lau235	Unclassified/Putative Phage	85.00%	5	59,588	38.57	3
Lau238	Unclassified/Putative Phage	57.14%	5	31,516	60.58	2
Lau241	Unclassified/Putative Phage	96.43%	6	178,448	41.08	3
Lau242	Unclassified/Putative Phage	98.89%	7	48,694	39.89	3
Lau243	Unclassified/Putative Phage	85.00%	6	34,056	36.57	3
Lau244	Unclassified/Putative Phage	100.00%	3	14,339	43.59	2
Lau245	Unclassified/Putative Phage	71.43%	5	23,944	41.81	5
Lau246	Unclassified/Putative Phage	33.33%	3	39,701	44.89	2
Lau247	Unclassified/Putative Phage	92.86%	10	101,421	51.21	3
Lau254	Unclassified/Putative Phage	76.92%	3	40,485	53.90	2
Lau256	Unclassified/Putative Phage	86.67%	7	50,225	55.35	4
Lau257	Unclassified/Putative Phage	73.68%	11	64,635	49.12	5
Lau260	Unclassified/Putative Phage	50.00%	7	71,726	42.90	5
Lau263	Unclassified/Putative Phage	60.87%	5	24,660	39.09	4
Lau264	Unclassified/Putative Phage	71.43%	6	38,540	40.67	5
Lau267	Unclassified/Putative Phage	75.00%	4	18,746	42.52	3
Lau271	Unclassified/Putative Phage	75.00%	4	28,995	54.83	4
Lau272	Unclassified/Putative Phage	45.83%	26	181,661	53.94	4
Lau273	Unclassified/Putative Phage	40.00%	12	82,859	50.73	3
Lau274	Unclassified/Putative Phage	69.70%	8	71,390	50.94	3
Lau278	Unclassified/Putative Phage	100.00%	4	45,859	50.80	4
Lau279	Unclassified/Putative Phage	93.75%	13	118,039	39.94	3
Lau280	Unclassified/Putative Phage	87.81%	8	54,984	40.97	3
Lau281	Unclassified/Putative Phage	100.00%	2	27,017	35.45	4
Lau282	Unclassified/Putative Phage	78.13%	10	134,212	39.02	4
Lau283	Unclassified/Putative Phage	94.12%	4	26,173	41.54	3
Lau284	Unclassified/Putative Phage	70.83%	6	41,525	41.05	3
Lau285	Unclassified/Putative Phage	86.11%	5	30,187	38.96	3
Lau286	Unclassified/Putative Phage	87.50%	6	36,609	46.08	5
Lau287	Unclassified/Putative Phage	90.91%	7	54,240	41.81	3
Lau288	Unclassified/Putative Phage	75.00%	9	72,763	38.78	5

Lau289	Viruses;dsdna viruses (no rna stage);Caudovirales;Myoviridae;T4-like viruses	91.43%	8	110,857	33.29	5
Lau291	Unclassified/Putative Phage	100.00%	5	32,046	37.11	3
Lau292	Unclassified/Putative Phage	88.89%	7	33,690	39.75	4
Lau293	Unclassified/Putative Phage	83.78%	6	32,332	42.88	2
Lau294	Unclassified/Putative Phage	61.49%	12	102,283	37.90	3
Lau295	Unclassified/Putative Phage	80.00%	5	26,793	39.95	3
Lau296	Unclassified/Putative Phage	100.00%	11	96,278	39.39	4
Lau297	Unclassified/Putative Phage	70.31%	9	62,434	36.25	4
Lau298	Unclassified/Putative Phage	90.24%	5	60,744	36.53	3
Lau300	Unclassified/Putative Phage	95.65%	8	51,386	50.20	4
Lau302	Unclassified/Putative Phage	100.00%	3	14,841	52.10	2
Lau303	Unclassified/Putative Phage	53.33%	5	28,013	43.12	5
Lau304	Unclassified/Putative Phage	50.00%	7	48,782	39.82	5
Lau308	Unclassified/Putative Phage	77.78%	5	25,088	47.84	2
Lau309	Unclassified/Putative Phage	72.73%	6	29,853	42.90	3
Lau311	Unclassified/Putative Phage	99.02%	7	47,886	26.40	3
Lau312	Unclassified/Putative Phage	100.00%	2	12,628	51.04	1
Lau313	Unclassified/Putative Phage	92.86%	4	21,420	35.84	3
Lau314	Unclassified/Putative Phage	100.00%	5	42,723	39.31	3
Lau315	Unclassified/Putative Phage	76.00%	3	26,353	38.61	3
Lau316	Unclassified/Putative Phage	87.50%	7	64,627	39.33	4
Lau317	Unclassified/Putative Phage	70.00%	2	10,969	39.77	2
Lau318	Unclassified/Putative Phage	94.60%	7	39,582	35.83	4
Lau319	Unclassified/Putative Phage	78.95%	5	49,849	40.50	2
Lau320	Unclassified/Putative Phage	80.00%	2	10,834	37.40	1
Lau325	Unclassified/Putative Phage	84.93%	12	81,164	35.11	3
Lau326	Unclassified/Putative Phage	83.33%	5	62,189	36.82	4
Lau327	Unclassified/Putative Phage	72.73%	6	31,947	37.85	4
Lau328	Unclassified/Putative Phage	47.06%	6	36,298	51.14	1
Lau329	Unclassified/Putative Phage	75.00%	2	8,463	37.70	1
Lau330	Unclassified/Putative Phage	76.19%	6	27,500	31.59	4
Lau333	Unclassified/Putative Phage	70.00%	2	10,879	48.00	1
Lau334	Unclassified/Putative Phage	100.00%	4	24,216	49.68	3
Lau336	Unclassified/Putative Phage	100.00%	6	61,907	49.63	4
Lau337	Unclassified/Putative Phage	100.00%	4	23,743	46.45	4
Lau338	Unclassified/Putative Phage	60.00%	3	15,394	39.14	3
Lau340	Unclassified/Putative Phage	96.55%	5	23,862	53.97	2
Lau342	Unclassified/Putative Phage	56.25%	50	308,246	44.44	5
Lau343	Unclassified/Putative Phage	80.00%	7	49,298	47.13	2
Lau344	Unclassified/Putative Phage	83.33%	2	14,406	40.77	2
Lau345	Unclassified/Putative Phage	100.00%	8	68,376	52.23	4

Lau351	Unclassified/Putative Phage	90.91%	5	44,898	44.73	3
Lau352	Unclassified/Putative Phage	100.00%	6	30,618	43.92	3
Lau353	Unclassified/Putative Phage	86.57%	13	119,349	38.58	5
Lau354	Unclassified/Putative Phage	100.00%	5	24,439	35.79	4
Lau355	Unclassified/Putative Phage	75.00%	3	22,836	41.19	2
Lau356	Unclassified/Putative Phage	60.61%	5	37,042	39.92	4
Lau357	Unclassified/Putative Phage	78.38%	7	53,981	40.40	4
Lau358	Unclassified/Putative Phage	91.31%	2	75,100	37.02	2
Lau359	Unclassified/Putative Phage	100.00%	5	31,939	36.89	5
Lau360	Unclassified/Putative Phage	98.31%	7	43,475	36.74	2
Lau367	Unclassified/Putative Phage	90.48%	4	47,563	37.62	3
Lau368	Unclassified/Putative Phage	100.00%	5	34,212	47.17	4
Lau370	Unclassified/Putative Phage	57.14%	8	54,650	40.97	4
Lau372	Unclassified/Putative Phage	100.00%	12	94,976	39.46	4
Lau373	Unclassified/Putative Phage	100.00%	4	59,850	34.21	3
Lau380	Unclassified/Putative Phage	76.00%	2	12,760	36.18	2
Lau381	Unclassified/Putative Phage	100.00%	4	97,482	35.03	4
Lau382	Unclassified/Putative Phage	100.00%	3	16,557	39.47	3
Lau383	Unclassified/Putative Phage	100.00%	2	11,382	51.88	2
Lau384	Unclassified/Putative Phage	100.00%	2	22,564	36.23	2
Lau385	Unclassified/Putative Phage	100.00%	4	28,322	48.67	4
Lau386	Unclassified/Putative Phage	89.47%	1	31,483	34.93	1
Lau387	Unclassified/Putative Phage	42.86%	2	8,740	48.64	2
Lau388	Unclassified/Putative Phage	75.00%	4	23,139	32.52	3
Lau389	Unclassified/Putative Phage	100.00%	3	13,064	37.83	2

Supplementary Table 6. Conserved LSU proteins identified in bacterial and archaeal bins

Bin name	Large subunit ribosomal proteins										
	COG 0080	COG 0081	COG 0087	COG 0091	COG 0093	COG 0094	COG 0097	COG 0102	COG 0197	COG 0200	COG 0256
	L11	L1	L3	L22	L14	L5	L6P/ L9E	L13	L16/ L10E	L15	L18
Lau1	1	1	1	1	1	1	1	1	1	1	1
Lau2	0	0	0	0	0	0	0	1	0	0	0
Lau3	0	0	0	0	1	1	1	1	0	1	1
Lau4	3	3	2	2	2	2	2	1	2	2	2
Lau5	2	2	1	1	1	1	2	2	1	2	2
Lau6	3	3	2	3	3	3	2	4	4	2	2
Lau7	1	1	2	2	2	2	1	9	2	0	1
Lau9	0	0	0	0	0	0	0	1	0	0	0
Lau10	3	4	4	4	5	5	5	4	5	5	5
Lau11	1	1	1	1	1	1	1	3	1	1	1
Lau12	0	0	0	0	0	0	0	0	0	0	0
Lau14	10	9	23	26	27	33	33	19	18	23	21
Lau15	0	0	0	0	0	0	0	0	0	0	0
Lau16	1	1	1	1	2	2	2	3	1	2	2
Lau17	1	1	1	1	1	2	3	5	1	3	3
Lau19	10	11	3	2	2	2	1	8	7	5	6
Lau20	7	8	4	7	7	8	8	5	7	8	8
Lau21	0	0	1	0	0	0	0	3	0	1	1
Lau22	0	0	0	0	0	0	0	0	0	0	0
Lau23	1	1	1	1	1	1	1	1	1	1	1
Lau24	0	0	0	0	0	0	0	0	0	0	0
Lau26	0	0	0	0	0	0	0	0	0	0	0
Lau27	0	0	0	0	0	0	0	0	0	0	0
Lau28	0	0	0	0	0	0	0	0	0	0	0
Lau29	0	0	0	0	0	0	0	0	0	0	0
Lau30	0	0	1	0	0	0	2	3	0	1	1
Lau31	0	0	1	0	0	1	1	1	0	1	1
Lau32	0	0	0	0	0	0	0	0	0	0	0
Lau33	0	0	0	0	0	0	0	0	0	0	0
Lau34	0	0	0	0	0	0	0	0	0	0	0
Lau35	0	0	0	0	0	0	0	1	0	0	0
Lau36	6	7	4	4	5	5	5	0	5	5	5
Lau40	0	0	0	0	0	0	0	0	0	0	0
Lau41	0	0	3	3	2	2	1	5	2	1	1
Lau42	2	2	2	1	1	1	1	3	1	1	2
Lau44	2	2	0	0	2	3	3	3	1	3	3

Lau45	0	0	0	0	0	0	0	0	0	0	0
Lau46	0	0	0	0	0	0	0	0	0	0	0
Lau47	5	5	7	7	7	5	4	7	7	4	4
Lau51	0	0	0	0	0	0	0	0	0	0	0
Lau52	0	0	0	0	0	0	0	0	0	0	0
Lau53	0	0	0	0	0	0	0	0	0	0	0
Lau62	4	4	4	4	4	4	4	4	4	4	3
Lau64	1	1	1	1	1	1	1	2	1	1	1
Lau65	0	3	1	0	0	0	0	5	1	0	0
Lau66	0	0	0	0	0	0	0	0	0	0	0
Lau92	2	2	1	1	1	1	1	1	2	1	1
Lau93	0	0	0	0	0	0	1	0	0	2	2
Lau94	1	1	1	0	1	1	1	1	0	1	1
Lau95	0	0	0	0	0	0	0	0	0	0	0
Lau96	0	0	0	0	1	1	1	1	0	1	1
Lau98	0	0	0	0	0	0	0	0	0	0	0
Lau101	0	0	0	0	0	0	0	0	0	0	0
Lau103	5	5	3	3	6	6	8	14	5	9	8
Lau104	0	0	0	1	2	2	1	3	1	1	1
Lau112	0	0	0	0	0	0	0	0	0	0	0
Lau113	0	1	2	1	2	2	2	2	1	2	2
Lau114	0	0	0	0	0	0	0	1	0	0	0
Lau115	0	0	0	0	0	0	0	0	0	0	0
Lau116	2	2	3	3	4	4	3	1	4	4	3
Lau117	0	0	0	0	0	0	0	0	0	0	0
Lau118	0	0	1	0	0	0	0	2	0	1	0
Lau122	0	0	0	0	0	0	0	0	0	0	0
Lau123	0	0	0	0	0	0	0	1	0	0	0
Lau125	1	1	0	0	0	0	0	0	0	0	0
Lau126	0	0	0	0	0	0	0	0	0	0	0
Lau127	0	0	0	0	0	0	0	0	0	0	0
Lau129	0	0	0	0	0	0	0	0	0	0	0
Lau143	0	0	0	0	0	0	0	0	0	0	0
Lau158	0	0	0	0	0	0	0	0	0	0	0
Lau159	0	0	0	0	0	0	0	0	0	0	0
Lau163	0	0	0	0	0	0	0	0	0	0	0
Lau164	0	0	0	0	0	0	0	0	0	0	0
Lau176	1	1	0	1	1	1	1	1	1	1	1
Lau177	0	0	0	0	0	0	0	1	0	0	0
Lau178	0	0	0	0	0	0	0	0	0	0	0
Lau179	0	0	0	0	0	0	0	0	0	0	0

Lau184	0	0	0	0	0	0	0	0	0	0	0
Lau190	0	0	4	1	2	2	2	0	2	2	2
Lau197	0	0	0	0	0	0	0	0	0	0	0
Lau198	0	0	0	0	0	0	0	0	0	0	0
Lau203	0	0	0	0	0	0	0	0	0	0	0
Lau208	0	0	0	0	0	0	0	0	0	0	0
Lau210	0	0	0	0	0	0	0	0	0	0	0
Lau227	2	2	2	2	2	2	2	2	2	1	2
Lau229	0	0	1	0	0	0	0	1	0	0	0
Lau230	3	3	2	2	2	2	2	1	2	2	2
Lau231	5	5	4	4	4	5	4	0	4	4	4
Lau233	0	0	0	0	0	0	0	0	0	0	0
Lau236	1	1	1	0	0	0	0	0	0	1	1
Lau237	0	0	0	0	0	0	0	0	0	0	0
Lau248	0	0	0	0	0	0	0	0	0	0	0
Lau249	0	0	0	0	0	0	0	0	0	0	0
Lau252	0	0	0	0	0	0	0	0	0	0	0
Lau259	0	0	0	0	0	0	0	0	0	0	0
Lau265	0	0	0	0	0	0	0	0	0	0	0
Lau269	0	0	0	0	0	0	0	0	0	0	0
Lau331	0	0	0	0	0	0	0	0	0	0	0
Lau371	0	0	0	0	0	0	0	0	0	0	0
Lau374	0	0	0	0	0	0	0	0	0	0	0
Lau375	1	1	1	1	1	1	1	0	1	1	1
Lau376	0	0	3	0	0	0	1	2	0	3	2
Lau377	3	3	3	3	3	3	3	1	3	2	2
Lau378	0	0	0	0	0	0	1	3	0	1	1
Lau379	0	0	0	0	0	0	0	0	0	0	0

Supplementary Table 7. Conserved SSU ribosomal proteins in bacterial and archaeal bins

Bin name	Small subunit ribosomal proteins											
	COG 0048	COG 0049	COG 0052	COG 0092	COG 0096	COG 0098	COG 0099	COG 0100	COG 0103	COG 0184	COG 0186	COG 0522
	S12	S7	S2	S3	S8	S5	S13	S11	S9	S15P/ S13E	S17	S4
Lau1	1	1	1	1	1	1	1	1	1	1	1	1
Lau2	0	0	2	0	0	0	0	0	1	0	0	1
Lau3	0	0	3	0	1	1	1	0	1	2	0	1
Lau4	1	1	2	2	2	2	2	2	1	2	3	2
Lau5	1	1	2	1	1	2	2	2	2	2	1	2
Lau6	4	4	4	3	2	2	4	4	4	4	3	4
Lau7	7	8	4	2	1	1	0	0	8	3	2	6
Lau8	0	0	0	0	0	0	0	1	0	0	0	0
Lau9	1	1	0	0	0	0	0	0	1	0	0	0
Lau10	7	5	3	4	5	5	4	4	4	4	5	3
Lau11	1	1	1	1	1	1	1	1	1	4	1	1
Lau12	0	0	0	0	0	0	0	0	0	0	0	0
Lau14	13	15	20	27	34	22	27	25	17	14	27	22
Lau15	0	0	0	0	0	0	0	0	0	0	0	0
Lau16	1	1	1	1	2	2	2	2	3	1	2	2
Lau17	1	1	8	1	2	3	10	10	6	4	1	10
Lau19	12	11	6	2	2	5	8	10	8	10	2	8
Lau20	4	6	8	7	8	8	7	6	2	5	7	5
Lau21	0	0	1	0	0	1	1	1	3	0	0	3
Lau22	1	1	0	0	0	0	0	0	0	0	0	0
Lau23	1	1	1	1	1	1	1	1	1	1	1	1
Lau24	0	0	0	0	0	0	0	0	0	0	0	0
Lau26	0	0	0	0	0	0	0	0	0	0	0	0
Lau27	0	0	0	0	0	0	0	0	0	0	0	0
Lau28	0	0	0	0	0	0	0	0	0	0	0	0
Lau29	0	0	0	0	0	0	0	0	0	0	0	0
Lau30	0	0	2	0	0	1	2	2	3	1	0	4
Lau31	1	1	0	0	1	1	1	1	1	1	0	1
Lau32	0	0	0	0	0	0	0	0	0	0	0	0
Lau33	0	0	0	0	0	0	0	0	0	0	0	0
Lau34	2	2	1	0	0	0	1	1	0	3	0	1
Lau35	0	0	0	0	0	0	0	0	1	3	0	0
Lau36	4	4	4	5	5	5	4	4	0	2	5	4
Lau40	0	0	0	0	0	0	0	0	0	0	0	0
Lau41	3	4	10	2	1	1	1	1	6	8	2	1
Lau42	3	3	3	1	1	1	1	1	3	3	1	1

Lau44	**0**	**0**	**1**	**0**	**3**	**3**	**3**	**2**	**3**	**4**	**1**	**2**
Lau45	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau46	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau47	**7**	**7**	**7**	**7**	**4**	**4**	**4**	**4**	**7**	**1**	**7**	**4**
Lau51	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau52	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau53	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau62	**0**	**5**	**6**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**	**4**
Lau64	**1**	**1**	**2**	**1**	**1**	**1**	**1**	**1**	**2**	**2**	**1**	**1**
Lau65	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**1**	**5**	**1**	**0**	**2**
Lau66	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau92	**8**	**9**	**1**	**1**	**1**	**1**	**2**	**2**	**1**	**1**	**1**	**2**
Lau93	**1**	**1**	**0**	**0**	**1**	**2**	**0**	**0**	**0**	**0**	**0**	**0**
Lau94	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**
Lau95	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau96	**0**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**1**	**0**	**0**	**2**
Lau98	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau101	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau103	**3**	**3**	**16**	**4**	**8**	**9**	**8**	**7**	**15**	**10**	**6**	**6**
Lau104	**0**	**0**	**3**	**1**	**1**	**1**	**2**	**2**	**3**	**0**	**2**	**2**
Lau112	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau113	**0**	**0**	**2**	**1**	**2**	**2**	**2**	**2**	**2**	**0**	**2**	**2**
Lau114	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau115	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau116	**6**	**6**	**5**	**4**	**4**	**4**	**4**	**4**	**1**	**3**	**4**	**4**
Lau117	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau118	**1**	**1**	**0**	**0**	**0**	**0**	**2**	**2**	**3**	**0**	**0**	**3**
Lau122	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau123	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**
Lau125	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau126	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau127	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau129	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau143	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau158	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau159	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau163	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau164	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau176	**0**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**0**	**1**	**1**
Lau177	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
Lau178	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Lau179	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau184	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau190	**3**	**4**	**0**	**2**	**2**	**2**	**1**	**0**	**0**	**0**	**2**	**0**
Lau197	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau198	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau203	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau208	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau210	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau227	**2**	**2**	**1**	**2**	**2**	**2**	**0**	**0**	**2**	**1**	**2**	**0**
Lau229	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**
Lau230	**3**	**3**	**3**	**2**	**2**	**2**	**2**	**2**	**1**	**0**	**2**	**2**
Lau231	**6**	**6**	**0**	**4**	**4**	**4**	**3**	**3**	**0**	**0**	**4**	**3**
Lau233	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau236	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**1**
Lau237	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau248	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau249	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau252	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau259	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau265	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau269	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau331	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau371	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau374	**1**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau375	**7**	**6**	**0**	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**1**	**0**
Lau376	**0**	**0**	**0**	**0**	**1**	**2**	**3**	**3**	**2**	**4**	**0**	**3**
Lau377	**3**	**3**	**3**	**3**	**3**	**2**	**2**	**2**	**1**	**4**	**3**	**1**
Lau378	**0**	**0**	**1**	**0**	**1**	**1**	**1**	**2**	**3**	**3**	**0**	**2**
Lau379	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Supplementary Table 8. Other conserved genes in identified bacterial and archaeal bins

Bin name	tRNA synthetases									RNA polymerase subunits			
	COG0016	COG0018	COG0060	COG0124	COG0143	COG0172	COG0201	COG0495	COG0525	COG0202	COG0085	COG0012	COG0533
	Phenylalanyl-tRNA syntheth ase alpha subunit	Arginyl-tRNA synthe tase	Isoleucyl-tRNA synthe tase	Histidyl-tRNA synthe tase	Methionyl-tRNA synthe tase	Seryl-tRNA synthe tase	Preprotein transl ocase subuni t SecY	Leucyl-tRNA synthe tase	Valyl-tRNA synthe tase	DNA-directed RNA polymer ase, alpha subunit/ 40 kD subunit	DNA-directed RNA polymera se, beta subunit/1 40 kD subunit	Predicted GTPase	Metal-dependent protease
Lau1	0	1	1	1	1	1	1	1	1	1	1	1	1
Lau2	1	1	1	0	0	0	1	0	1	0	0	1	0
Lau3	0	1	2	0	0	2	1	4	2	1	0	0	0
Lau4	1	1	2	2	1	2	2	2	2	2	3	2	1
Lau5	2	2	2	2	2	2	2	2	3	2	2	2	2
Lau6	4	4	2	4	4	4	2	4	4	5	4	4	4
Lau7	1	12	6	12	5	11	0	14	4	9	1	5	7
Lau8	0	0	0	0	0	3	0	0	1	0	0	0	0
Lau9	0	1	1	1	1	2	0	3	3	0	0	5	1
Lau10	9	12	13	10	9	13	4	16	8	6	16	7	3
Lau11	2	3	1	1	1	2	1	2	1	1	1	2	2
Lau12	1	0	0	0	0	1	0	0	1	0	0	1	1
Lau14	18	20	36	14	12	27	27	21	35	24	25	33	35
Lau15	0	0	0	2	0	1	1	0	1	0	0	0	1
Lau16	1	3	4	2	4	3	3	2	6	2	1	2	5
Lau17	6	8	4	7	8	12	10	7	8	10	1	3	5
Lau19	3	10	17	13	17	14	11	11	4	6	14	24	9
Lau20	6	9	9	9	5	6	8	1	7	5	10	7	9
Lau21	4	2	4	0	4	1	1	8	4	2	5	3	1
Lau22	1	1	1	1	1	1	0	1	1	0	0	0	0
Lau23	1	1	2	1	1	1	1	1	2	1	1	1	1
Lau24	0	1	0	0	1	1	0	1	0	0	0	0	0
Lau26	0	0	0	0	0	0	2	0	1	0	0	0	1
Lau27	1	1	0	0	0	0	0	0	2	0	0	0	0
Lau28	0	0	0	0	0	1	0	2	1	0	0	1	2
Lau29	0	0	0	0	1	0	0	1	0	0	0	0	0
Lau30	3	5	3	3	6	5	2	0	4	2	2	4	1
Lau31	1	2	2	2	0	2	1	2	2	0	0	1	2
Lau32	0	0	0	0	0	0	0	0	0	0	0	0	0
Lau33	0	2	0	0	0	0	0	0	0	0	0	0	0
Lau34	1	2	1	0	3	0	0	2	2	1	1	3	2
Lau35	2	0	4	3	2	3	0	3	2	1	0	2	0
Lau36	0	2	3	3	2	1	5	2	2	7	9	0	0

Lau40	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**0**
Lau41	**8**	**7**	**6**	**11**	**4**	**4**	**2**	**15**	**8**	**6**	**4**	**9**	**4**
Lau42	**3**	**3**	**4**	**3**	**3**	**3**	**1**	**3**	**4**	**1**	**1**	**3**	**3**
Lau44	**3**	**1**	**4**	**0**	**1**	**3**	**3**	**2**	**4**	**3**	**3**	**2**	**2**
Lau45	**1**	**3**	**0**	**3**	**2**	**3**	**0**	**0**	**2**	**0**	**0**	**0**	**3**
Lau46	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau47	**7**	**5**	**9**	**6**	**7**	**7**	**4**	**9**	**7**	**4**	**5**	**7**	**8**
Lau51	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**1**	**0**	**0**	**0**	**0**	**0**
Lau52	**0**	**0**	**0**	**0**	**1**	**2**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau53	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau62	**5**	**5**	**4**	**5**	**5**	**6**	**4**	**4**	**5**	**4**	**5**	**4**	**3**
Lau64	**4**	**2**	**3**	**4**	**4**	**2**	**1**	**3**	**5**	**2**	**1**	**3**	**3**
Lau65	**3**	**0**	**4**	**2**	**3**	**1**	**0**	**2**	**3**	**0**	**4**	**1**	**2**
Lau66	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**
Lau92	**1**	**1**	**0**	**0**	**9**	**2**	**2**	**2**	**8**	**2**	**4**	**1**	**8**
Lau93	**0**	**0**	**0**	**0**	**1**	**0**	**2**	**1**	**0**	**0**	**0**	**0**	**1**
Lau94	**1**	**1**	**1**	**1**	**0**	**1**	**1**	**1**	**0**	**4**	**1**	**3**	**2**
Lau95	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau96	**1**	**0**	**0**	**4**	**1**	**2**	**1**	**2**	**2**	**1**	**0**	**2**	**0**
Lau98	**0**	**0**	**1**	**2**	**1**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**1**
Lau101	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau103	**16**	**31**	**30**	**29**	**22**	**16**	**9**	**19**	**24**	**10**	**7**	**27**	**18**
Lau104	**3**	**3**	**2**	**2**	**1**	**4**	**3**	**0**	**2**	**3**	**0**	**5**	**0**
Lau112	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau113	**2**	**2**	**1**	**3**	**3**	**4**	**2**	**4**	**4**	**2**	**2**	**3**	**4**
Lau114	**1**	**1**	**0**	**1**	**2**	**2**	**0**	**3**	**1**	**0**	**0**	**0**	**1**
Lau115	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau116	**5**	**0**	**6**	**3**	**2**	**5**	**4**	**6**	**7**	**3**	**5**	**4**	**2**
Lau117	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau118	**4**	**3**	**1**	**5**	**0**	**4**	**1**	**1**	**2**	**3**	**0**	**3**	**1**
Lau122	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**2**	**0**	**0**
Lau123	**0**	**2**	**0**	**1**	**0**	**0**	**0**	**2**	**2**	**1**	**0**	**1**	**0**
Lau125	**0**	**0**	**1**	**2**	**2**	**3**	**0**	**0**	**1**	**0**	**1**	**0**	**0**
Lau126	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau127	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**
Lau129	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau143	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau158	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau159	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau163	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau164	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**

Lau176	**1**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**
Lau177	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau178	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**
Lau179	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau184	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau190	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**1**	**0**	**1**	**1**	**0**
Lau197	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau198	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau203	**0**	**1**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau208	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau210	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau227	**1**	**0**	**2**	**2**	**2**	**1**	**4**	**3**	**3**	**1**	**2**	**0**	**0**
Lau229	**1**	**1**	**1**	**3**	**1**	**1**	**0**	**1**	**1**	**0**	**0**	**0**	**1**
Lau230	**0**	**4**	**6**	**4**	**3**	**0**	**2**	**0**	**3**	**1**	**3**	**5**	**0**
Lau231	**0**	**0**	**0**	**0**	**0**	**0**	**4**	**0**	**0**	**3**	**6**	**0**	**0**
Lau233	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
Lau236	**0**	**1**	**1**	**0**	**0**	**0**	**1**	**1**	**0**	**1**	**0**	**0**	**0**
Lau237	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau248	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau249	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau252	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau259	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau265	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau269	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau331	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau371	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
Lau374	**0**	**0**	**0**	**3**	**1**	**0**	**0**	**0**	**1**	**1**	**2**	**0**	**0**
Lau375	**0**	**0**	**4**	**1**	**0**	**0**	**0**	**0**	**0**	**0**	**5**	**0**	**7**
Lau376	**3**	**0**	**0**	**0**	**2**	**3**	**3**	**1**	**1**	**2**	**3**	**0**	**0**
Lau377	**0**	**2**	**4**	**3**	**0**	**3**	**2**	**2**	**0**	**2**	**0**	**1**	**3**
Lau378	**3**	**0**	**0**	**0**	**0**	**7**	**1**	**4**	**3**	**2**	**0**	**0**	**4**
Lau379	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**

3.8 References

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31:** 533-538.

Allers E, Wright JJ, Konwar KM, Howes CG, Beneze E, Hallam SJ *et al* (2013). Diversity and population structure of Marine Group A bacteria in the Northeast subarctic Pacific Ocean. *ISME J* **7:** 256-268.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Anantharaman K, Breier JA, Sheik CS, Dick GJ (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proceedings of the National Academy of Sciences* **110:** 330-335.

Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54:** 1501-1529.

Baker BJ, Lesniewski RA, Dick GJ (2012). Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J* **6:** 2269-2279.

Baker BJ, Sheik CS, Taylor CA, Jain S, Bhasi A, Cavalcoli JD *et al* (2013). Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *ISME J* **7:** 1962-1973.

Bowers TS, Von Damm KL, Edmond JM (1985). Chemical evolution of mid-ocean ridge hot springs. *Geochimica Et Cosmochimica Acta* **49:** 2239-2252.

Breier JA, Rauch CG, McCartney K, Toner BM, Fakra SC, White SN *et al* (2009). A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. *Deep Sea Research Part I: Oceanographic Research Papers* **56:** 1579-1589.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330:** 1375-1378.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward Automatic Reconstruction of a Highly Resolved Tree of Life. *Science* **311:** 1283-1287.

de Angelis MA, Lilley MD, Baross JA (1993). Methane oxidation in deep-sea hydrothermal plumes of the endeavour segment of the Juan de Fuca Ridge. *Deep Sea Research Part I: Oceanographic Research Papers* **40:** 1169-1186.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496-503.

Di Rienzi SC, Sharon I, Wrighton KC, Koren O, Hug LA, Thomas BC *et al* (2013). The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *eLife* **2**.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al* (2009a). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.

Dick GJ, Clement BG, Webb SM, Fodrie FJ, Bargar JR, Tebo BM (2009b). Enzymatic microbial Mn(II) oxidation and Mn biooxide production in the Guaymas Basin deep-sea hydrothermal plume. *Geochimica et Cosmochimica Acta* **73:** 6517-6530.

Dick GJ, Tebo BM (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12:** 1334-1347.

Dick GJ, Anantharaman K, Baker BJ, Li M, Reed DC, Sheik CS (2013). The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. *Frontiers in Microbiology* **4**.

Distel DL, Lane DJ, Olsen GJ, Giovannoni SJ, Pace B, Pace NR *et al* (1988). Sulfur-oxidizing bacterial endosymbionts: analysis of phylogeny and specificity by 16S rRNA sequences. *J Bacteriol* **170:** 2506-2510.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792-1797.

Edwards KJ, Bach W, McCollom TM, Rogers DR (2004). Neutrophilic Iron-Oxidizing Bacteria in the Ocean: Their Habitats, Diversity, and Roles in Mineral Deposition, Rock Alteration, and Biomass Production in the Deep-Sea. *Geomicrobiology Journal* **21:** 393-404.

Ferrini VL, Tivey MK, Carbotte SM, Martinez F, Roman C (2008). Variable morphologic expression of volcanic, tectonic, and hydrothermal processes at six hydrothermal vent fields in the Lau back-arc basin. *Geochemistry, Geophysics, Geosystems* **9:** Q07022.

Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Geoff Wheat C *et al* (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* **10:** 333-346.

German CR, Bowen A, Coleman ML, Honig DL, Huber JA, Jakuba MV *et al* (2010). Diverse styles of submarine venting on the ultraslow spreading Mid-Cayman Rise. *Proceedings of the National Academy of Sciences* **107:** 14020-14025.

Hanson RS, Hanson TE (1996). Methanotrophic bacteria. *Microbiological Reviews* **60:** 439-471.

Hara S, Koike I, Terauchi K, Kamiya H, Tanoue E (1996). Abundance of viruses in deep oceanic waters. *Marine Ecology Progress Series* **145:** 269-277.

Hügler M, Sievert SM (2010). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. *Annual Review of Marine Science* **3:** 261-289.

Janecky DR, Seyfried WE (1984). Formation of massive sulfide deposits on oceanic ridge crests - incremental reaction models for mixing between hydrothermal solutions and seawater. *Geochimica Et Cosmochimica Acta* **48:** 2723-2738.

Jannasch HW, Mottl MJ (1985). Geomicrobiology of Deep-Sea Hydrothermal Vents. *Science* **229:** 717-725.

Jiang H, Breier JA, Sylvan JB, Edwards KJ, Madison AS, Luther III GW (2014). Physical controls on mixing and transport within rising submarine hydrothermal plumes: A numerical simulation study. . *Deep Sea Res Part I* **(Submitted)**.

Jiang H, Breier, J.A., Sylvan, J.B., Edwards, K.J., Madison, A.S., Luther III, G.W. (2014). Physical controls on mixing and transport within rising submarine hydrothermal plumes: A numerical simulation study (To be Submitted). *Deep Sea Res Part I*.

Jones WJ, Won YJ, Maas PAY, Smith PJ, Lutz RA, Vrijenhoek RC (2006). Evolution of habitat use by deep-sea mussels. *Marine Biology* **148:** 841-851.

Kadko D (1993). An assessment of the effect of chemical scavenging within submarine hydrothermal plumes upon ocean geochemistry. *Earth and Planetary Science Letters* **120:** 361-374.

Karl DM, Knauer GA, Martin JH, Ward BB (1984). Bacterial chemolithotrophy in the ocean Is associated with sinking particles. *Nature* **309:** 54-56.

Kessler JD, Valentine DL, Redmond MC, Du M, Chan EW, Mendes SD *et al* (2011). A Persistent Oxygen Anomaly Reveals the Fate of Spilled Methane in the Deep Gulf of Mexico. *Science* **331:** 312-315.

Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543-546.

Lam P, Cowen JP, Jones RD (2004). Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiology Ecology* **47:** 191-206.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J* **6:** 2257–2268.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Li M, Jain S, Baker BJ, Taylor C, Dick GJ (2013). Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. *Environmental Microbiology***:** n/a-n/a.

Li M, Toner BM, Baker BJ, Breier JA, Sheik CS, Dick GJ (2014). Microbial iron uptake as a mechanism for dispersing iron from deep-sea hydrothermal vents. *Nat Commun* **5:** 3192.

Lücker S, Wagner M, Maixner F, Pelletier E, Koch H, Vacherie B *et al* (2010). A Nitrospira metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences* **107:** 13479-13484.

Luecker S, Nowka B, Rattei T, Spieck E, Daims H (2013). The genome of Nitrospina gracilis illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Frontiers in Microbiology* **4**.

Luther III GW, Glazer BT, Ma S, Trouwborst RE, Moore TS, Metzger E *et al* (2008). Use of voltammetric solid-state (micro)electrodes for studying biogeochemical processes: Laboratory measurements to real time measurements with an in situ electrochemical analyzer (ISEA). *Marine Chemistry* **108:** 221-235.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D *et al* (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research* **36:** D534-D538.

Martinez F, Taylor B, Baker ET, Resing JA, Walker SL (2006). Opposing trends in crustal thickness and spreading rate along the back-arc Eastern Lau Spreading Center: Implications for controls on ridge morphology, faulting, and hydrothermal activity. *Earth and Planetary Science Letters* **245:** 655-672.

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

Mottl MJ, Seewald JS, Wheat CG, Tivey MK, Michael PJ, Proskurowski G *et al* (2011). Chemistry of hot springs along the Eastern Lau Spreading Center. *Geochimica et Cosmochimica Acta* **75:** 1013-1038.

Mullineaux LS, Wiebe PH, Baker ET (1995). Larvae of benthic invertebrates in hydrothermal vent plumes over Juan de Fuca Ridge. *Marine Biology* **122:** 585-596.

Nakagawa S, Takai K, Inagaki F, Hirayama H, Nunoura T, Horikoshi K *et al* (2005). Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. *Environ Microbiol* **7:** 1619-1632.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40:** e155.

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28:** 1420-1428.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476:** 176-180.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35:** 7188-7196.

Reinthaler T, van Aken HM, Herndl GJ (2010). Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Research Part II: Topical Studies in Oceanography* **57:** 1572-1580.

Rivers AR, Sharma S, Tringe SG, Martin J, Joye SB, Moran MA (2013). Transcriptional response of bathypelagic marine bacterioplankton to the Deepwater Horizon oil spill. *ISME J* **7:** 2315-2329.

Schmieder R, Lim YW, Edwards R (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28:** 433-435.

Semrau JD, DiSpirito AA, Vuilleumier S (2011). Facultative methanotrophy: false leads, true results, and suggestions for future research. *FEMS Microbiology Letters* **323:** 1-12.

Sheik CS, Jain S, Dick GJ (2013). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology***:** n/a-n/a.

Sheik CS, Anantharaman K, Breier JA, Sylvan JB, Dick GJ (2014). Response of deep-ocean, particulate associated microbial communities to buoyant hydrothermal plumes across a back-arc spreading basin. *in prep*.

Singer E, Webb EA, Nelson WC, Heidelberg JF, Ivanova N, Pati A *et al* (2011). Genomic Potential of Marinobacter aquaeolei, a Biogeochemical "Opportunitroph". *Applied and Environmental Microbiology* **77:** 2763-2771.

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

Sun H, Feistel R, Koch M, Markoe A (2008). New equations for density, entropy, heat capacity, and potential temperature of a saline thermal fluid. *Deep Sea Research Part I: Oceanographic Research Papers* **55:** 1304-1310.

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

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333:** 1296-1300.

Sylvan JB, Pyenson BC, Rouxel O, German CR, Edwards KJ (2012). Time-series analysis of two hydrothermal plumes at 9°50′N East Pacific Rise reveals distinct, heterogeneous bacterial populations. *Geobiology* **10:** 178-192.

Sylvan JB, Sia TY, Haddad AG, Briscoe LJ, Toner BM, Girguis PR *et al* (2013). Low temperature geomicrobiology follows host rock composition along a geochemical gradient in Lau Basin. *Frontiers in Microbiology* **4**.

Tagliabue A, Bopp L, Dutay J-C, Bowie A, Chever F, Jean-Baptiste P *et al* (2010). Hydrothermal contribution to the oceanic dissolved iron inventory. *Nature Geoscience* **3:** 252-256.

Tavormina PL, Ussler W, 3rd, Joye SB, Harrison BK, Orphan VJ (2010). Distributions of putative aerobic methanotrophs in diverse pelagic marine environments. *ISME J* **4:** 700-710.

Teske A, Hinrichs K-U, Edgcomb V, de Vera Gomez A, Kysela D, Sylva SP *et al* (2002). Microbial Diversity of Hydrothermal Sediments in the Guaymas Basin: Evidence for Anaerobic Methanotrophic Communities. *Applied and Environmental Microbiology* **68:** 1994-2007.

Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics* **14:** 178-192.

Thrash JC, Ben T, Brandon KS, Zachary CL, Tanja W, Edward FD *et al* (2014). Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *The ISME Journal*.

Toner BM, Fakra SC, Manganini SJ, Santelli CM, Marcus MA, Moffett JW *et al* (2009). Preservation of iron(II) by carbon-rich matrices in a hydrothermal plume. *Nature Geosci* **2:** 197-201.

Vignais PM, Billoud B (2007). Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chemical Reviews* **107:** 4206-4272.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* **326:** 578-582.

Winn CD, Karl DM, Massoth GJ (1986). Microorganisms in deep-sea hydrothermal plumes. *Nature* **320:** 744-746.

Wright JJ, Konwar KM, Hallam SJ (2012). Microbial ecology of expanding oxygen minimum zones. *Nat Rev Micro* **10:** 381-394.

Yamamoto M, Takai K (2011). Sulfur metabolisms in epsilon- and gamma-Proteobacteria in deep-sea hydrothermal fields. *Frontiers in Microbiology* **2**.

Zellmer KE, Taylor B (2001). A three-plate kinematic model for Lau Basin opening. *Geochemistry, Geophysics, Geosystems* **2:** 1020.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:** 821-829.

CHAPTER IV

SULFUR OXIDATION GENES IN DIVERSE DEEP-SEA VIRUSES

Karthik Anantharaman[a], Melissa B. Duhaime[b], John A. Breier[c], Kathleen Wendt[d], Brandy M. Toner[d], and Gregory J. Dick[a,b,e]

a. Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA
b. Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA
c. Applied Ocean Physics and Engineering, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
d. Department of Soil, Water and Climate, University of Minnesota – Twin Cities, St. Paul, MN, USA
e. Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA

4.1 Abstract

Viruses are the most abundant biological entities in the oceans and a pervasive cause of mortality of microorganisms that drive biogeochemical cycles. Although the ecological and evolutionary impacts of viruses on marine phototrophs are now well-recognized, little is known about the parallel impacts on ubiquitous marine lithotrophs, which drive substantial primary production in the deep oceans. Here we report the discovery of 18 genome sequences of double-stranded DNA viruses that putatively infect abundant and widespread sulfur-oxidizing SUP05 *Gammaproteobacteria*. 15 of these viral genomes contain host-derived auxiliary metabolic genes (AMGs) encoding the alpha and gamma subunits of the reverse-acting dissimilatory sulfite

reductase *(rdsr)* that oxidizes elemental sulfur, which is abundant in the hydrothermal plumes studied here. Our findings implicate viruses as a key agent in the biogeochemical cycle of sulfur and suggest that viruses are a reservoir of genetic diversity for bacterial lithotrophic machinery that underpins chemosynthesis in the deep oceans.

4.2 Introduction

Chemolithoautotrophic bacteria are ubiquitous in the dark oceans (Swan et al 2011), where they serve as a sink for CO_2 (Aristegui et al 2009) through primary production that contributes up to 53% of the particulate organic carbon exported from the photic zone (Reinthaler et al 2010). Bacteria of the uncultured SUP05 clade of Gammaproteobacterial sulfur oxidizers (GSOs) are among the most abundant and widespread marine chemolithoautotrophs, fixing carbon and oxidizing reduced sulfur species and hydrogen in diverse marine environments such as hydrothermal vent plumes (Anantharaman et al 2013) , hydrothermal vent-associated animals (Newton et al 2007, Petersen et al 2011), and oxygen minimum zones (Walsh et al 2009), where they underpin cryptic links between the sulfur and nitrogen cycles (Canfield et al 2010). Although viruses are abundant in these deep-sea ecosystems (Hara et al 1996), little is known about viruses that infect lithotrophic primary producers.

4.3 Materials and Methods

Sample Collection. Samples were collected on two cruises aboard the *R/V Thomas G. Thompson* in June/July 2009. The details of samples and the sampling locations are provided in

Supplementary table 1. A total of 12 hydrothermal plume and background deep-sea samples were collected. Of these, 9 samples were collected from the rising plume using a Suspended Particle Rosette Sampler (Breier et al 2009) (SUPR) mounted on remotely operated vehicle *ROV Jason II* while 3 samples were collected using a Conductivity Temperature, and Depth Rosette (CTD). The hydrothermal plumes were recognized visually and using optical scatter by turbidity anomalies. Three types of backgrounds were collected, sample TN236-J2449-2 (depth=2155m) was a near bottom background collected with SUPR, sample TN236-J2445-20.2 (depth~800m) was an above plume background collected with SUPR while sample TN236-CTD-Mar-BP1 (depth=1785m) was a below neutrally buoyant plume background collected with CTD-Rosette. Water samples collected with SUPR (10-60*l*) were filtered *in situ* on to 0.8-µm 37-mm polycarbonate SUPOR membranes and preserved shipboard in RNAlater (Ambion, Austin, TX, USA). Water samples collected by CTD-Rosette (20*l*) were pressure filtered with N_2 gas shipboard on to 0.2-µm 47-mm polycarbonate membranes and preserved in RNAlater (Ambion, Austin, TX, USA).

Extraction of nucleic acids and multiple displacement amplification of DNA. DNA was extracted from ¼ filters as described previously (Dick and Tebo 2010). Multiple displacement amplification (MDA) of genomic DNA was performed using the illustra Ready-To-Go GenomiPhi V3 DNA Amplification Kit (GE Healthcare, Piscataway, NJ, USA).

DNA sequencing and pre-assembly data processing. Genomic DNA was purified using standard protocols (Illumina, Inc., San Diego, CA, USA). Shotgun sequencing of DNA was performed with Illumina HiSeq2000 at the University of Michigan DNA Sequencing Core. The raw shotgun sequencing reads were deprelicated with the following thresholds (100% identity

over 100% lengths) followed by trimming of dereplicated reads using the adaptive read trimmer, Sickle.

De novo genomic assembly. Samples from the six vent sites (Kilo Moana, Abe, Mariner, Tahi Moana, Tui Malila, Guaymas) were each assembled *de novo* to obtain six separate assemblies. Whole genome *de novo assemblies* were performed using IDBA-UD (Peng et al 2012) with the following parameters (--mink 50, --maxk 92, --step 4 or 6, --min_contig 500). Assemblies were then repeated in an iterative manner by removing the reads used in formation of contigs at the higher kmer with Velvet (Zerbino and Birney 2008) (kmer 91 to 51, steps of 4) using the following parameters (-exp_cov auto –ins_length * -ins_length_sd 20 –read_trkg yes –min_contig_lgth 2500), followed by MetaVelvet (Namiki et al 2012) (kmer 91 to 51, steps of 4 or 6) using the following parameters (-scaffolding yes, -min_contig_lgth 2500). rRNA reads were identified using RiboPicker (Schmieder et al 2012) with a custom database (5s+16s+23s rRNA) and assembled separately using IDBA-UD with the following parameters (--mink 50, --maxk 92, --step 4). Repetition of *de novo* whole-genome assemblies using two different assemblers was done to check the veracity of generated contigs. All major trends were similar across both sets of assemblies. All data presented in this paper is from the assemblies generated by IDBA-UD. * - Parameter varies by sample in the range 214-223.

Annotations. Assembled contigs were first annotated through the Integrated Microbial Genomes (IMG) automated online pipeline (Markowitz et al 2008). Prediction of open reading frames was done using Prodigal (Hyatt et al 2010) with default parameters. Manual curation of the four *rdsrA* containing viral genomes (Lau77, Lau85, Lau87 and Lau218) was performed using a combination of blastp (Altschul et al 1990) against NCBI-nr and InterProScan 4.x(Quevillon et al 2005) against the InterPro data v43.1 (July 2013).

Binning. Binning of assembled viral genomes was performed using a combination of tetranucleotide frequencies, contig coverage and %GC content in emergent self-organizing maps (Dick et al 2009). 39 viral genomes representing diverse viruses were downloaded from the CAMERA database 'Moore Marine Phage/Virus Genomes' and used in conjunction with 3 bacterial genomes SUP05 GB-1, GB-2 (Anantharaman et al 2013) and *Bathymodiolus* endosymbionts (Petersen et al 2011) as reference genomes. The bacterial genomes were used on the ESOM to demonstrate the difference in signature of the viruses from SUP05 bacteria.

Comparative genomics. Comparative genomics of the viruses identified in this study and was performed against the complete phage genome in GenBank to which each Lau virus shared the greatest number of significantly similar homologs. Synteny with known phages (Lau218 to Pelagibacter phage HTVC019P (Zhao et al 2013) (NC_020483); Lau85 to Synechococcus phage S-SSM7 (Sullivan et al 2010) (NC_015287); Lau87 to Enterobacteria phage T5 (NC_005859); Lau77 to Enterobacteria phage VB_KleM-RaK2 (Šimoliūnas et al 2013) (NC_019526) was determined based on reciprocal best blastP hits between known the known phage isolates and assembled Lau viruses. Fig. 4.1 and Supplementary fig. 3 were generated in Circos (Krzywinski et al 2009).

Sequence alignment and phylogeny. Alignment of *rdsrA, rdsrC and terL* amino acid sequences was performed by MUSCLE (Edgar 2004) using default parameters followed by manual refinement. Representative bacterial and viral *rdsrA* and *rdsrC* gene amino acid sequences were aligned and compared with reference sequences, *Desulfovibrio vulgaris* str.Hildenborough (Karkhoff-Schweizer et al 1995) (P45574) (sulfate reducing bacterium, *dsrA*) and *Allochromatium vinosum* DSM180 (Weissgerber et al 2011) (AAC35394) (sulfur oxidizing

bacterium, *rdsrA*) to identify conserved residues across the *sulfite reductase* domain identified previously (Dahl et al 1993, Dhillon et al 2005, Oliveira et al 2008).

Phylogenetic analysis of *rdsrA* genes was inferred by Maximum Likelihood implemented in RaxML (Stamatakis 2006) using the PROTGAMMAGTR algorithms and bootstrapped 1000 times with the following parameters: -f a –m PROTGAMMAGTR –N 1000 –x 777 –p 333

All *terL* sequences identified previously (Duhaime et al 2011) were supplemented with additional sequences having the best blastp hits to *terL* sequences from ELSC. Alignment of *terL* amino acid sequences was performed by MUSCLE using default parameters followed by manual refinement. Phylogenetic analysis of *terL* genes was inferred by Maximum Likelihood implemented in RaxML (Stamatakis 2006) using the PROTGAMMAJTT algorithms and bootstrapped 1000 times with the following parameters: -f a –m PROTGAMMAJTT –N 1000 –x 777 –p 333. All observed patterns were similar between the best tree generated and consensus trees. Fig. 4.2 and Supplementary fig. 2 were generated using the best tree with bootstrap support and branch lengths.

Thermodynamic modeling. Equilibrium thermodynamic reaction path modeling was used to predict Fe mineral precipitation, chemical concentrations, and activity coefficients resulting from the mixing of seawater with end member fluid from A1 vent in the ABE hydrothermal field (Mottl et al 2011). Our approach follows those of previous studies (Bowers et al 1985, Janecky and Seyfried 1984, McCollom 2000a). Our specific plume model implementation has been previously described (Anantharaman et al 2013, Breier et al 2012)). The following is a brief description of the aspects of this model pertinent to this study.

The ABE-A1 plume reaction path is modeled through a mixing process that ends at a vent fluid to seawater dilution of 1 part in 10,000, representing the dilution achieved at the non-buoyant plume heights sampled in this study. Vent fluid composition (Supplementary table 4) is based on measurements made from samples collected in 2005 (Mottl et al 2011, Seewald et al 2005), and assumptions for N species, Cu, Zn, and Ba are described previously (Anantharaman et al 2013, Mottl et al 2011). Measurements made on samples collected in 2009, at the time of this study, using previously described methods (Mottl et al 2011) showed ~15% greater H_2S and ~57% lower Fe than in 2009 (Flores et al 2012). Thus, at the time of this study, S_0 plume concentrations would have been greater than these predictions. In situ pH was calculated from measurements of pH at 25° C using an equilibrium reaction path model that increased the temperature of the measured fluid to the original vent fluid temperature. Background seawater dissolved O_2 concentration was based on WOCE measurements from section P06 (Talley 2007) . (Note, the available data predates this study; actual vent chemistry during this study may have differed.)

Reaction path modeling was performed with REACT, part of the Geochemist's Workbench package (Bethke 2007). Conductive cooling was neglected and mixture temperatures were a strict function of conservative end-member mixing. Precipitated minerals were allowed to dissolve and their constituents to re-precipitate based on thermodynamic equilibrium constraints. Thermodynamic data was predicted by SUPCRT95 (Johnson et al 1992) for the temperature range of 1-425^{o}C (specifically 1, 25, 60, 100, 225, 290, 350, and 425^{o}C) and a pressure of 500 bar, a pressure and temperature range that encompasses all known deep sea vents. SUPCRT95 uses previously published thermodynamic data for minerals, gases, and aqueous species (Helgeson et al 1978, McCollom and Shock 1997, Saccocia and Seyfried Jr 1994, Shock and

Helgeson 1988, Shock et al 1989, Shock et al 1997, Sverjensky et al 1997). Thermodynamic data for pyrolusite, bixbyite, hausmannite, marcasite, and $Fe(OH)_3$ were added for our study (Robie et al 1979, Wagman et al 1982). The B-dot activity model was used (Helgeson 1969, Helgeson and Kirkham 1974) . Temperature dependent activity coefficients were used for aqueous CO_2 and water in an NaCl solution (Bethke 2007, Cleverley and Bastrakov 2005, Drummond 1981). A general limitation of REACT is that it does not predict the thermodynamic behaviour of solid solutions. Thus minerals such as sphalerite, pyrrhotite, chalcopyrite, and isocubanite are treated as separate phases with ideal stoichiometry. This may influence the predicted plume mineral assemblage.

In a previous study (Anantharaman et al 2013), we suppressed all aqueous phase redox couples in order to estimate upper limit constraints on potential chemosynthetic metabolic energy. In this case, we use these same assumptions but because of our interest in chemical speciation in this study we have added additional assumptions related to mineral formation following previous studies (Breier et al 2012). The precipitation of hematite was suppressed to allow Fe hydroxide to precipitate on the basis that the latter is a closer approximation than the former to the more common amorphous Fe oxyhydroxides, which precipitate preferentially due to kinetic effects. The precipitation of Mg bearing minerals, and silicates, with the exception of amorphous silica, were also suppressed for simplicity. Some in this group have been found as minor plume constituents, others such as quartz appear kinetically inhibited; but in any case, the suppression of this group does not influence the precipitation of the minerals of interest in this study. Precipitated minerals were allowed to dissolve and their constituents to re-precipitate based on thermodynamic equilibrium constraints.

Micro-probe X-ray Diffraction (μXRD). Plume particle mineralogy was examined using the X-ray microprobe beamline 10.3.2, Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA (Marcus et al 2004). X-ray fluorescence (XRF) mapping at multiple incident energies was used to determine the spatial distribution of particles on the filter, as well as the elemental composition of the particles. Fluorescence was measured using a Canberra 7-element Ge detector for: (1) a map below the PbL_3 absorption energy for As, Ni, Zn, Cu, and Fe; (2) a map below the absorption energy of FeK for Mn; (3) maps above and below the absorption energy of VK were used to separate V from Ti and map lighter elements such as Ca, S, and Cl. These maps were deadtime corrected, registered, and combined using custom beamline software. The composite XRF map was used to identify particles for micro-probe X-ray diffraction measurements.

X-ray diffraction patterns were collected at an incident energy of 17 keV (λ=0.729 angstrom) with 240-second exposure, and a beam spot size on the sample of 6 × 11 μm. The XRD patterns were radially integrated to obtain profiles of intensity versus 2θ using the freeware *Fit2D* after calibration of sample-to-detector distance an alumina standard (Al_2O_3) (Hammersley et al 1996). The intensity versus 2θ data was processed in the software package *JADE* v9.1. A background XRD pattern for the filter substrate (polycarbonate) was subtracted from each XRD pattern collected from sample particles. Phase identification through peak matching in *JADE* was guided by the elemental composition of the particle as measured by point X-ray fluorescence (XRF) spectra. The level of confidence in the phase identification was ranked using *JADE*'s "figure of merit" parameter.

4.4 Results and Discussion

We conducted shotgun metagenomic sequencing on samples from five different hydrothermal vent plumes and associated deep ocean waters at the Eastern Lau Spreading Center (ELSC) in the Western Pacific Ocean and one plume at Guaymas Basin (GB) in the Gulf of California (Dick and Tebo 2010, Lesniewski et al 2012) (Supplementary table 1). *De novo* assembly of sequence reads and binning by emergent self-organizing maps (ESOM) of tetranucleotide signatures (Dick et al 2009) (Supplementary fig. 1) revealed five genomic 'bins' (henceforth Lau77, Lau85, Lau87, Lau218 and Lau220) of putative SUP05 viruses that contain 18 double stranded DNA (dsDNA) viral genome sequences. Phylogeny of the viral large terminase gene (*terL*) (Supplementary fig. 2) (which reflects phage DNA packaging mechanisms (Casjens et al 2005)), synteny with well-characterized phage of known taxonomy (Supplementary fig. 3) and results of protein sequence similarity searches against public sequence databases (Supplementary fig. 4) indicate that the five viruses belong to three marine viral families of the orders *Caudovirales* (dsDNA viruses, no RNA stage), *Podoviridae*, *Siphoviridae* and *Myoviridae* (Supplementary table 2).

15 of the 18 viral genomes (from four of the five SUP05 viral genomic bins) contain genes encoding the alpha (*rdsrA*) and gamma (*rdsrC*) subunits of the reverse-acting dissimilatory sulfite reductase (*rdsr*) complex for elemental sulfur oxidation (Fig. 4.1).

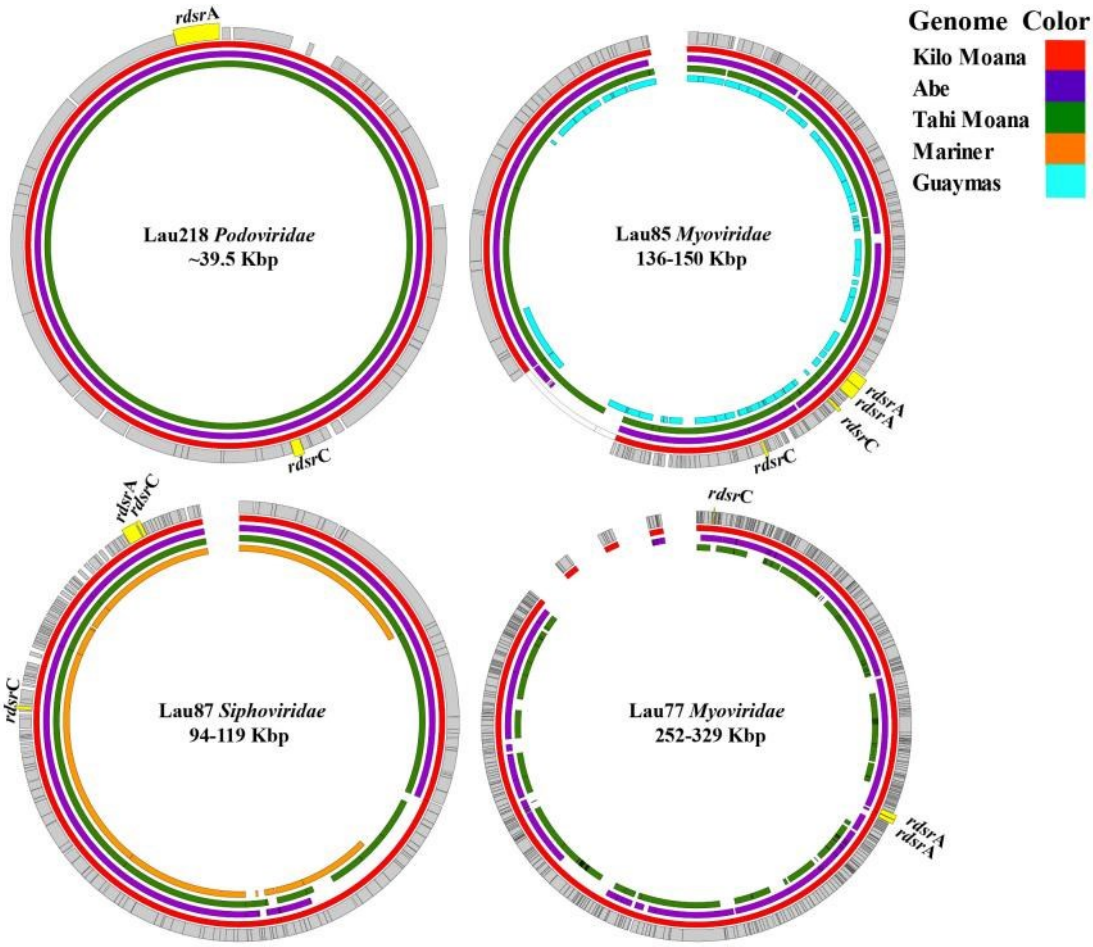


Fig. 4.1 Gene content of 15 phage genomes from 3 viral families retrieved from Lau and Guaymas basins. Colored nested circles represent syntenous viral genomes/contiguous genomic fragments from locations indicated in the legend. Grey – Identified genes/ORFs on Kilo Moana strain of each of the 4 viruses. *rdsrA* and *rdsrC* genes are highlighted in yellow.

No other *rdsr* genes or other sulfur oxidation genes were present on the viral genomes. Analysis of bacterial genome sequences recovered from ELSC and GB metagenomes revealed the presence of 5 different bacterial genomic bins possessing the *rdsr* complex: Lau10 (SUP05 *Gammaproteobacteria*), Lau51 (SUP05 *Gammaproteobacteria*), Lau60 (unclassified *Gammaproteobacteria*), Lau62 (uncultured EC-01-9C-26 *Gammaproteobacteria*) and Lau20 (Sar324 *Deltaproteobacteria*). Co-localized *rdsr* genes in the order *rdsrABEFHCMKLJOPN*

were found on Lau10, Lau60 and Lau62, while Lau51 and Lau20 Sar324 (Sheik et al 2013) possessed only *rdsrABC* and *rdsrAB*, respectively. Regions flanking the bacterial *rdsr* gene clusters showed no similarity to the viral genome sequences, suggesting that viral *rdsr* genes were derived from selective retention of *rdsrA* and *rdsrC* genes rather than recent homologous recombination with bacterial genomic DNA.

Phylogenetic analyses indicated that all viral *rdsrA* genes recovered here affiliate with SUP05 *Gammaproteobacteria* (74-96% amino acid identity, Supplementary fig. 5) and are distinct from other bacterial *rdsrA* genes from ELSC, GB and other marine environments (Fig. 4.2). We identified two distinct groups of *rdsrA* sequences that each include both viral and bacterial sequences. All viral *rdsrA* genes fall into Group 1 except for Lau85, which contains two copies of *rdsrA* with one representative in each group. Bacterial representatives of Group 1 include the SUP05 GB-1 and GB-2 from GB plumes as well as *Bathymodiolus* mussel symbionts (Petersen et al 2011), while Group 2 is populated by SUP05 from oxygen minimum zones (Walsh et al 2009) and symbionts of deep-sea clams (Newton et al 2007). The tight phylogenetic clustering of *rdsrA* genes sequences of three distinct phage families with SUP05 bacteria in two separate lineages suggests that the phage *rdsrA* genes originated from SUP05 and were transferred to viruses. These observations are analogous to those of core photosynthesis genes in cyanobacterial phages and other microbe-derived AMGs (Breitbart 2012, Breitbart 2007) (e.g. *psbA*, *psbD, mazG*) that are similar but not identical to known hosts, forming sub-clusters distinct from host proteins (Ignacio-Espinoza and Sullivan 2012, Lindell et al 2004).

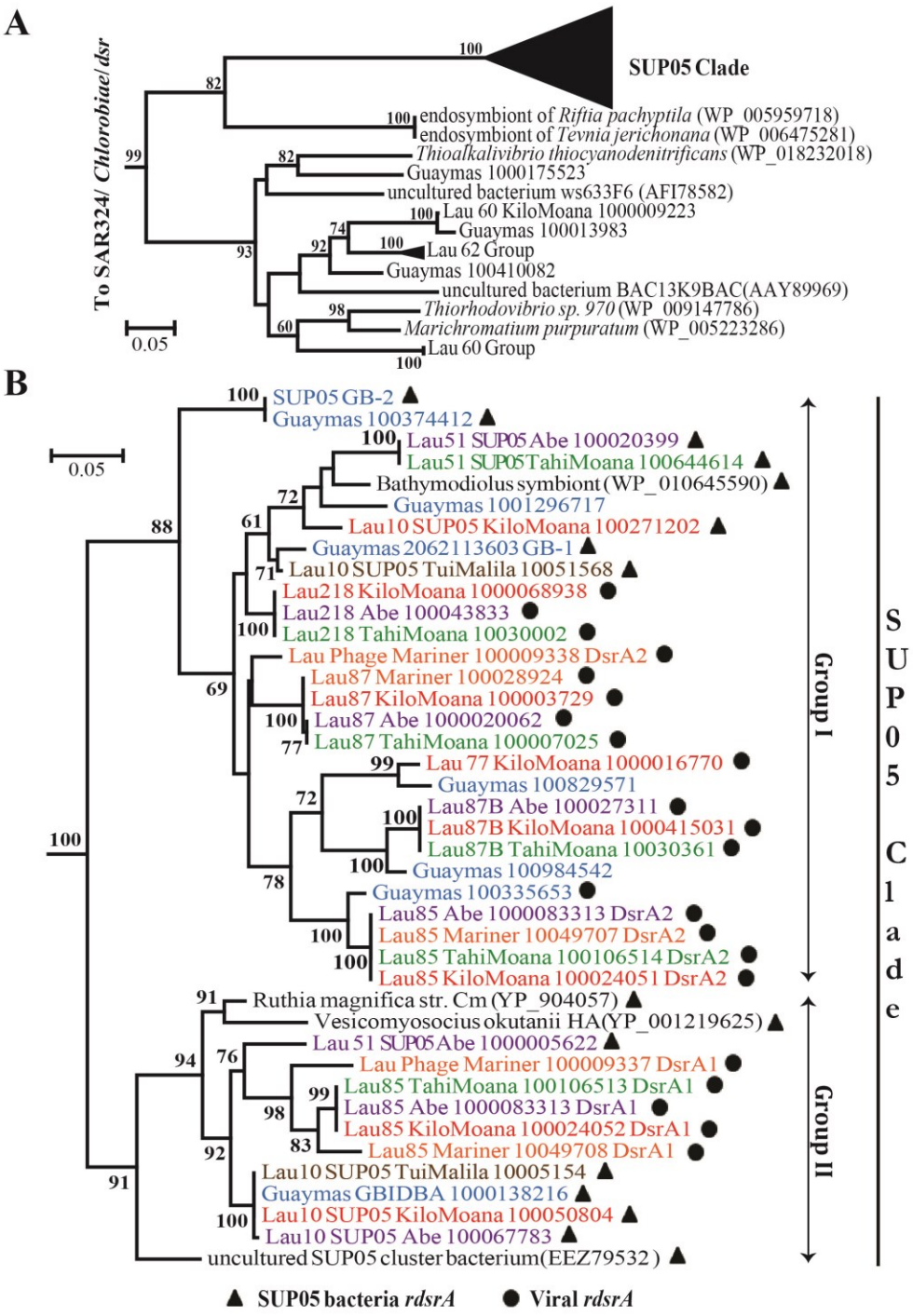


Fig. 4.2 A. Phylogenetic tree of rdsrA genes inferred by Maximum Likelihood. B. Detailed view of the SUP05 rdsrA clade. Group 1 and Group 2 sub-clades are shown on the right. Sequences are colored by geographical origin; Blue – Guaymas Basin; Red – Kilo Moana (Lau Basin); Green – Tahi Moana (Lau Basin); Purple – Abe (Lau Basin); Brown – Tui Malila (Lau Basin); Orange – Mariner (Lau Basin).

The amino acid sequences deduced from the viral *rdsrA* and *rdsrC* genes indicate the capacity to serve as functional sulfur-oxidizing enzymes. Phage RdsrA contain all conserved sulfite reductase residues and secondary structure elements for α-helix and β-sheets (Supplementary fig. 6). Similarly, a multiple alignment of *rdsrC* amino acid sequences indicated the presence of highly conserved residues across two distinct groups (Supplementary fig. 7). We also identified additional AMGs with high amino acid sequence identity to SUP05 in viral bins Lau77, Lau85 and Lau87, including multiple iron-sulfur cluster proteins (for cluster assembly, binding, biogenesis, delivery and insertion), 4Fe-4S ferredoxin, cytochrome c and 2-thiouridine synthesis/sulfur relay (TusA) proteins (Supplementary table 3). The existence of these additional SUP05-like genes on viral genomes supports their specificity to SUP05 bacteria and suggests a role for viral genes in supplementing host metabolism.

AMGs alleviate biochemical 'bottlenecks' of bacterial hosts during viral infection by encoding proteins of rate-limiting steps in metabolic pathways (such as the rapidly turned over D1 protein central to photosynthesis) (Breitbart 2012). We infer that the consistent occurrence of *rdsr* genes in viral genomes (universal presence of *rdsrA* and *rdsrC* and absence of necessary *rdsr* genes such as *rdsrB*) reflects biochemical properties of the rdsr system and a role for virally-encoded rdsr proteins in alleviating sulfur oxidation bottlenecks. RdsrA is part of the catalytic subunit of the *rdsr* complex (RdsrAB), and its expression is regulated in most sulfur-oxidizing microorganisms. In previous laboratory and environmental studies, *rdsrA* shows increased transcription relative to *rdsrB* (Anantharaman et al 2013, Weissgerber et al 2013), suggesting that the RdsrA protein has lower translational efficiency or higher protein turnover rate than RdsrB, consistent with its role as a bottleneck-relieving viral AMG. In contrast, the *rdsrC* gene is thought to be involved in sulfur-substrate delivery (Cort et al 2008) and regulation of *rdsr* gene

expression (Grimm et al 2010) and is constitutively expressed at high levels (Grimm et al 2010, Weissgerber et al 2013). Since the phage *rdsrA* and *rdsrC* genes were not localized in a cluster in the viral genomes (Fig. 4.1), we hypothesize that they are transcribed independently, as observed previously in both sulfur oxidizing (Grimm et al 2010) and sulfate reducing microorganisms (Karkhoff-Schweizer et al 1993). Overall, this scenario is analogous to the cyanobacteria/cyanophage model, whereby cyanophage carry genes (Mann et al 2003, Sullivan et al 2006) and express proteins (Lindell et al 2005) of a small subset of the photosystem II subunits – one of which turns over rapidly during photosynthesis and declines in abundance during infection (Lindell et al 2005, Sullivan et al 2006). The presence of *rdsrA* and *rdsrC* on viral genomes may offer selective advantages to the viruses by supplementing host pathways during infection. First, enhanced expression of *rdsrA* could replenish proteins involved in a rate limiting reaction in the host, as previously demonstrated with cyanobacterial phage D1 proteins involved in photosynthesis (Lindell et al 2005). Second, phage *rdsrC* could maintain or increase high transcription levels to ensure efficient delivery of sulfur-substrate to the *rdsrAB* complex during infection. Thus, phage AMGs that can supplement or sustain sulfur oxidation metabolism in their hosts may ensure continued viral infection and replication.

Oxidation of elemental sulfur is amongst the most energy yielding lithotrophic reactions in hydrothermal vent environments (Anantharaman et al 2013, McCollom 2000b, Petersen et al 2011), and it is the rate-limiting step in microbial oxidation of reduced sulfur species to sulfate (Grimm et al 2010). Recent studies indicate that globally abundant SUP05 bacteria form intracellular sulfur globules that serve as a store of electron donor in the energy-starved dark oceans (Anantharaman et al 2013, Newton et al 2007, Walsh et al 2009). To estimate the importance of elemental sulfur as an energy source in the ELSC deep-sea hydrothermal

ecosystem, we utilized a coupled bioenergetic-thermodynamic model of the hydrothermal plume at ABE in ELSC. Our model indicates that elemental sulfur oxidation accounts for greater than 92% of the total lithotrophic energy available in the plume (Fig. 4.3A). In order to identify the form of sulfur associated with particles in the hydrothermal plumes, we utilized X-ray fluorescence (XRF) and microprobe X-ray diffraction (μXRD). XRF maps show that sulfur is abundant in the plumes (Fig. 4.3B), while μXRD identified elemental sulfur to be widely present in the hydrothermal plumes at Abe (Fig. 4.3C) and Mariner (Supplementary fig. 8). Although our methods could not conclusively identify intracellular elemental sulfur, our results demonstrate that elemental sulfur presents an abundant source of energy for SUP05 bacteria in hydrothermal plumes and deep ocean waters of GB (Anantharaman et al 2013) and ELSC.

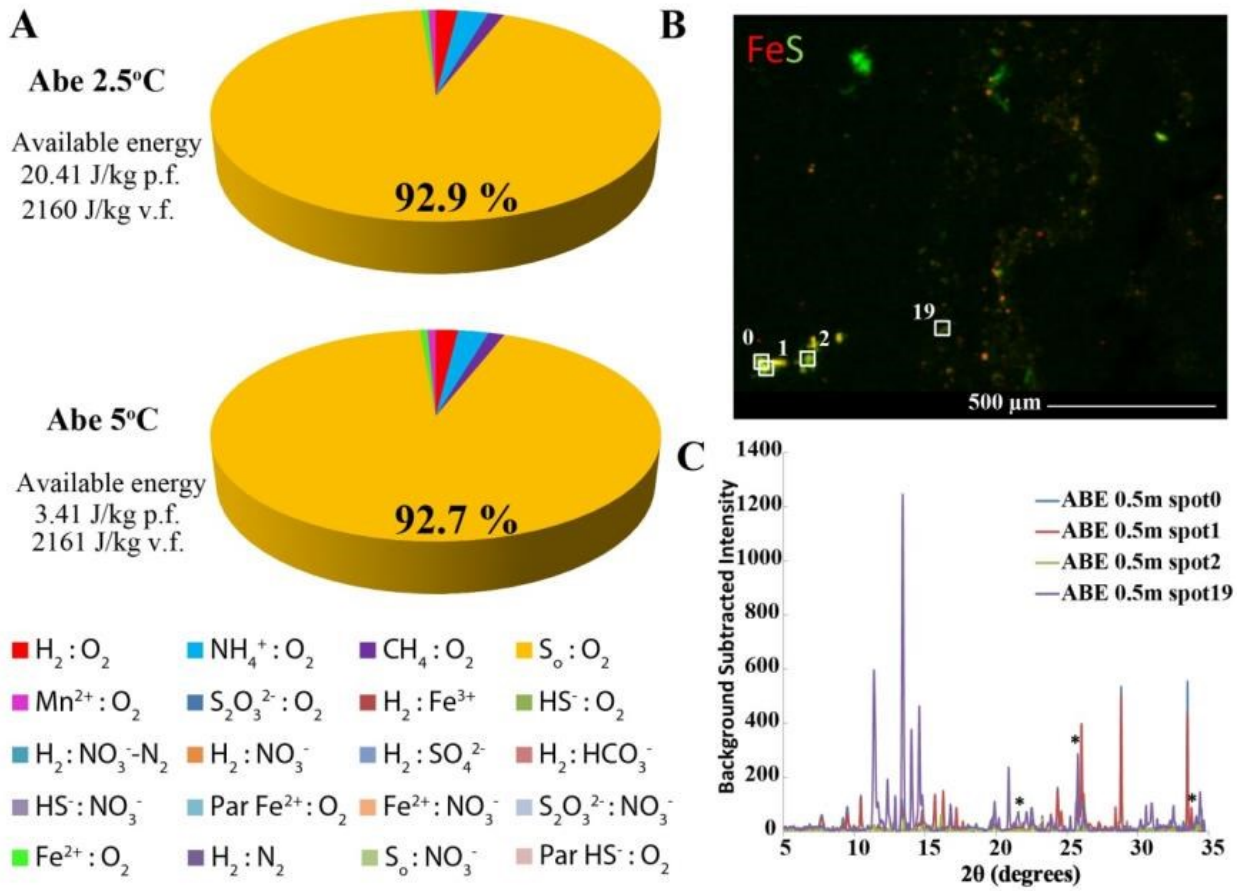


Fig. 4.3 A. Modeled free energies of catabolic reactions as a percentage of total available free energy in the Abe hydrothermal plume at 2.5 and 5oC. Total available free energy in the plume is normalized per kilogram plume fluid (p.f.) and per kilogram vent fluid (v.f.). **B. Distribution of iron (displayed in red) and sulfur (displayed in green) in particles collected at 0.5 m above the ABE vent.** Locations where elemental sulfur was detected by micro-probe X-ray diffraction measurements are indicated as spots 0, 1, 2, and 19. **C. Radially integrated diffractograms with elemental sulfur peaks annotated (*) at 22.0, 25.7, and 34.1 degrees 2-theta.** Elemental sulfur was detected in particle aggregates with other crystalline phases, such as pyrite, as indicated by additional non-elemental sulfur peaks.

4.5 Conclusions

The abundance and diversity of viruses infecting SUP05 bacteria in hydrothermal plumes suggests that chemolithoautotrophs in the deep sea face viral predation pressures similar to their photosynthetic counterparts in the surface waters (Avrani et al 2011). The remarkable synteny

and conservation of the four viruses studied here (95-99% genome nucleotide identity) across hydrothermal vent environments (GB and ELSC), ocean basins (Eastern and Western Pacific Ocean) and time (2004-2009) suggests that these viruses are persistent in marine environments dominated by SUP05 bacteria. Analyses of the Pacific Ocean Virome (POV) dataset (Hurwitz and Sullivan 2013) (Supplementary fig. 9), which notably contains viral communities from oxygen minimum zones dominated by SUP05 (Walsh et al 2009), revealed the presence of GSO-like *rdsrA* and *rdsrC* genes (Supplementary table 5), consistent with the prevalence of phage-encoded sulfur oxidation beyond hydrothermal plumes and in the wider pelagic oceans. To date, SUP05 has evaded growth in laboratory cultures, thus direct host-phage manipulations and validation of the underlying mechanisms of phage-influenced sulfur oxidation are impossible. Yet, this study demonstrates the sequence-based elucidation of microbial community dynamics through the discovery of phages that infect a widespread deep-sea bacterium.

Our results provide evidence for phage AMGs associated with chemolithotrophy. Phage-encoded sulfur oxidation is an unprecedented ecological strategy for viruses to access vast inorganic metabolic energy sources in the form of abundant environmental and intracellular elemental sulfur. These findings support the developing paradigm that viral AMGs serve to relieve metabolic bottlenecks during infection by prolonging host fitness long enough to ensure viral propagation (Breitbart 2012). The existence of *rdsr* genes in viral genomes across the Pacific basin portends this feature to be widespread, reveals a mechanism for horizontal transfer of genes associated with sulfur cycling (Klein et al 2001) and implicates viruses in the evolutionary dynamics of a central step in the planetary cycling of sulfur.

4.6 Appendix C

CHAPTER IV Supplementary Information

Contents

9. **Supplementary Text**
10. **Supplementary Figure 1**
11. **Supplementary Figure 2**
12. **Supplementary Figure 3**
13. **Supplementary Figure 4**
14. **Supplementary Figure 5**
15. **Supplementary Figure 6**
16. **Supplementary Figure 7**
17. **Supplementary Figure 8**
18. **Supplementary Figure 9**
19. **Supplementary Figure 10**
20. **Supplementary Table 1**
21. **Supplementary Table 2**
22. **Supplementary Table 3**
23. **Supplementary Table 4**
24. **Supplementary Table 5**
25. **Supplementary Table 6**
26. **Supplementary Table 7**

Supplementary Text

CRISPR-Cas loci

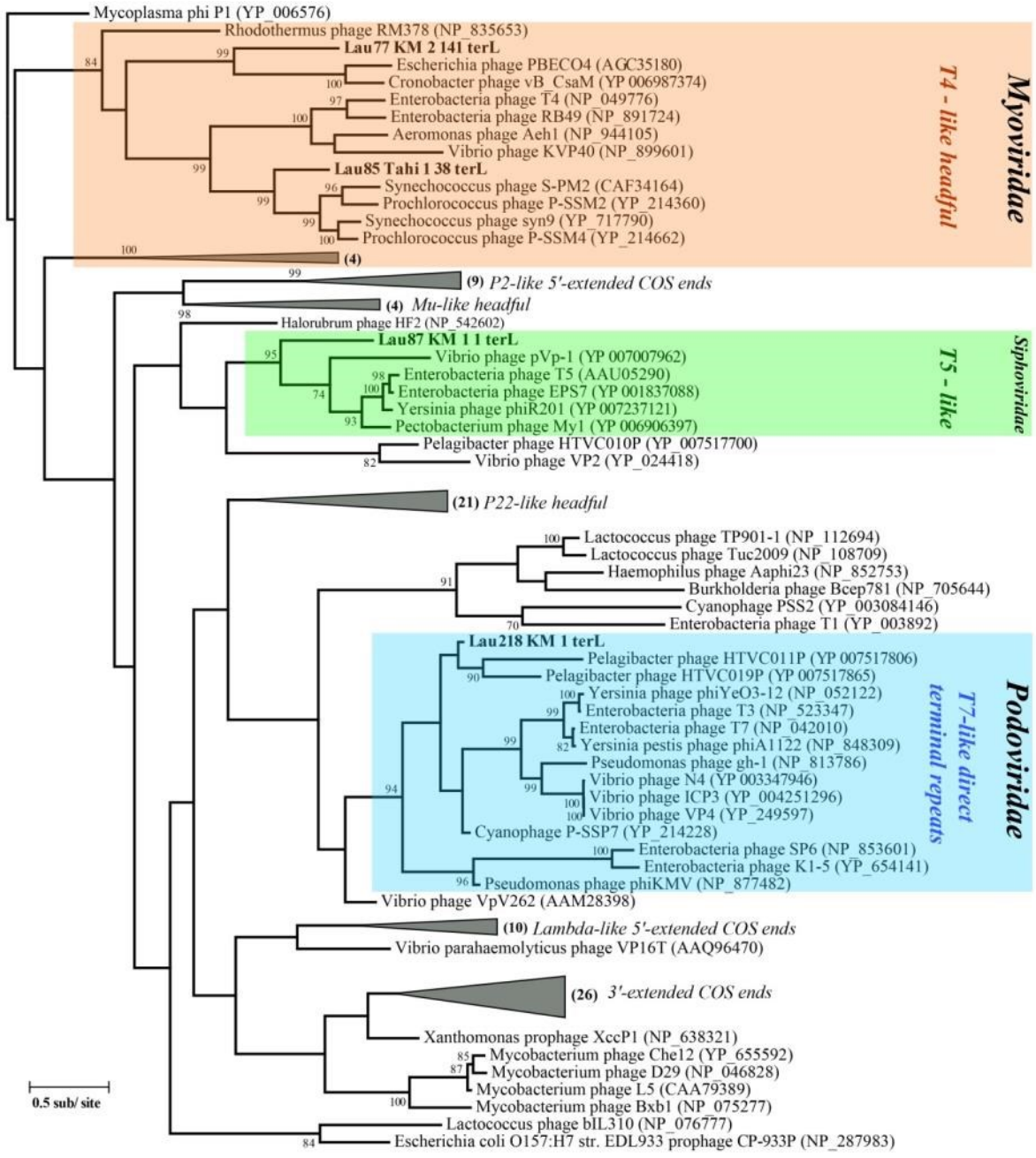
Analysis of bacterial genomes from the Eastern Lau Spreading Center (ELSC) and Guaymas Basin (GB) revealed the presence of a novel Type II-C CRISPR Cas system (Sangal et al 2013) in Lau10 SUP05 bacteria with its spacers targeting only Lau220 (Supplementary fig. 10). No other CRISPR-Cas loci were identified on the bacterial genomes at Lau and Guaymas.

Identification of Lau viruses in the Pacific Ocean Virome dataset

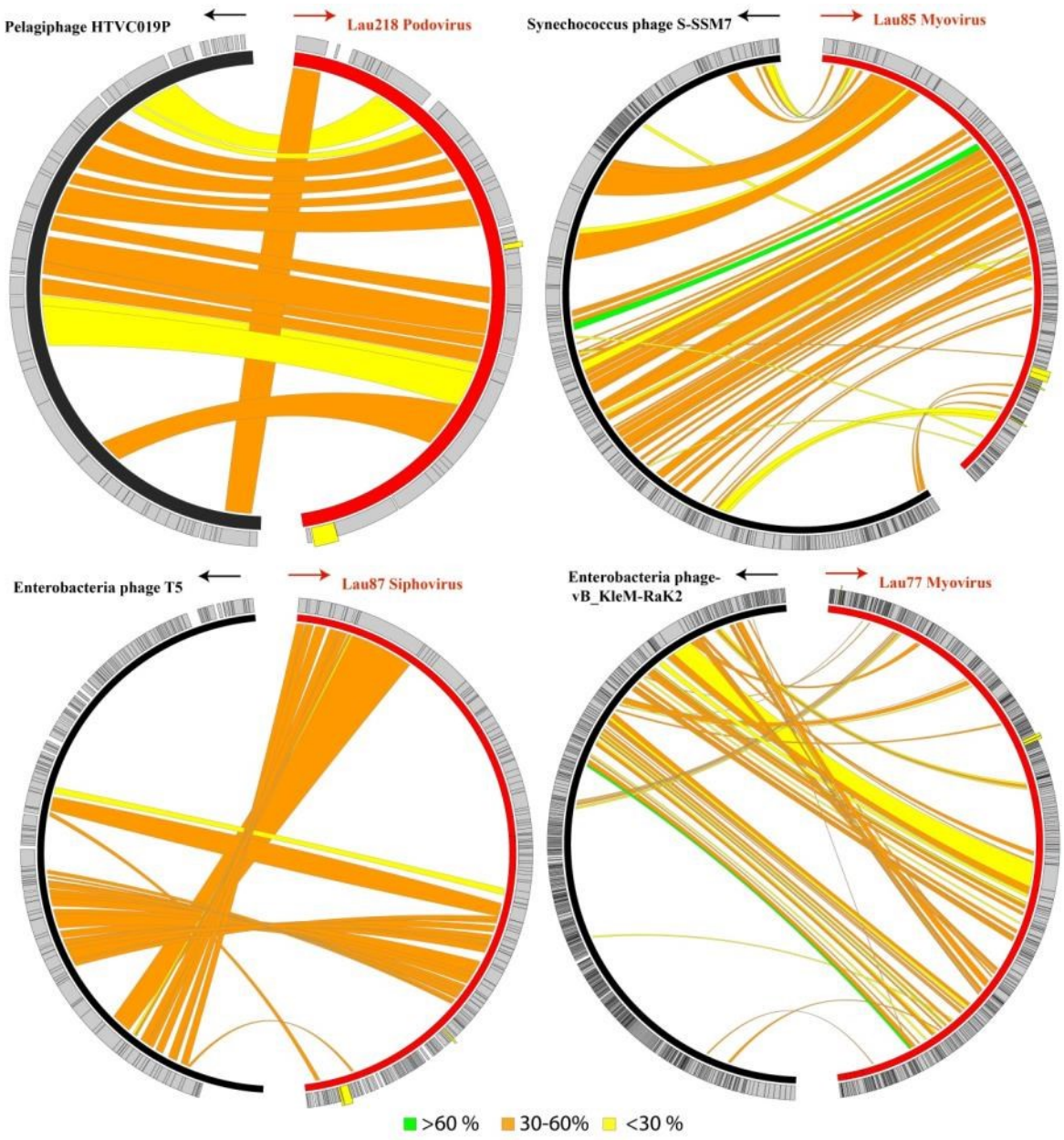
Results of Protein blasts (Blastp) of Lau77, Lau85, Lau87 and Lau218 amino acid sequences against the Pacific Ocean Virome (POV; collection of 32 viral metagenomes from the Pacific Ocean) protein clusters (Hurwitz and Sullivan 2013) indicate an amino acid identity in the range of 30-70 % suggesting that SUP05 viruses are divergent from the viruses identified in the Pacific Ocean Virome dataset (Fig. S9). However, we identified *rdsrA* and *rdsrC* amino acid sequences amongst the POV protein clusters in the "ultraclean" POV dataset (Hurwitz et al 2013) (curated to remove any POV metagenomes with either 16S rRNA genes from trace microbial contamination or gene transfer agents) using blastp (Supplementary Table S5) against custom *rdsrA* and *rdsrC* databases indicating that *rdsrA* and *rdsrC* genes may be widespread even in divergent phages in the pelagic oceans.

Bin	Annotation/Organism	Bin	Annotation/Organism
A	SUP05 GB-1	O	Prochlorococcus phage P-RSP2
	SUP05 GB-2		Prochlorococcus phage P-SSP3
B	Lau77		Cyanophage 9515-10a
C	Lau87		Cyanophage NATL1A-7
D	Bathymodiolus endosymbiont		Pseudoalteromonas phage pYD6-A
E	Lau218		Cyanophage NATL2A-133
F	Campylobacter phage CP21	P	Thermus phage phiYS40
G	Cellulophaga phage phiST		Cellulophaga phage phi47:1
H	Colwellia phage 9A		Cellulophaga phage phiSM
I	Vibrio phage helene 12B3	Q	Halorubrum phage CGphi46
	Vibrio phage PWH3a-P1		Halorubrum phage GNf2
J	Micromonas pusilla virus 12T		Deep-sea thermophilic phage D6E
	Micromonas pusilla virus PL1		Aeromonas phage pIS4-A
	Ostreococcus lucimarinus virus OlV3		Loktanella phage pCB2051-A
K	Vibrio phage henriette 12B8		Paenibacillus phage PG1
L	Emiliania huxleyi virus 201		Roseobacter phage
M	Synechococcus phage MbCM1		Salicola phage CGphi29
	Synechococcus phage S-MbCM6		Siphoviridae CAM ASM 000043
	Synechococcus phage S-IOM18		Sulfitobacter phage pCB2047-A
	Synechococcus phage S-SKS1		Sulfitobacter phage pCB2047-C
	Synechococcus phage S-SM2		Vibrio phage VD1
N	Lau85 Guaymas Basin	R	Lau220
	Lau85	S	Nitratiruptor phage NrS-1
U	Pseudoalteromonas phage PM2	T	Psychrobacter phage pOW20-A

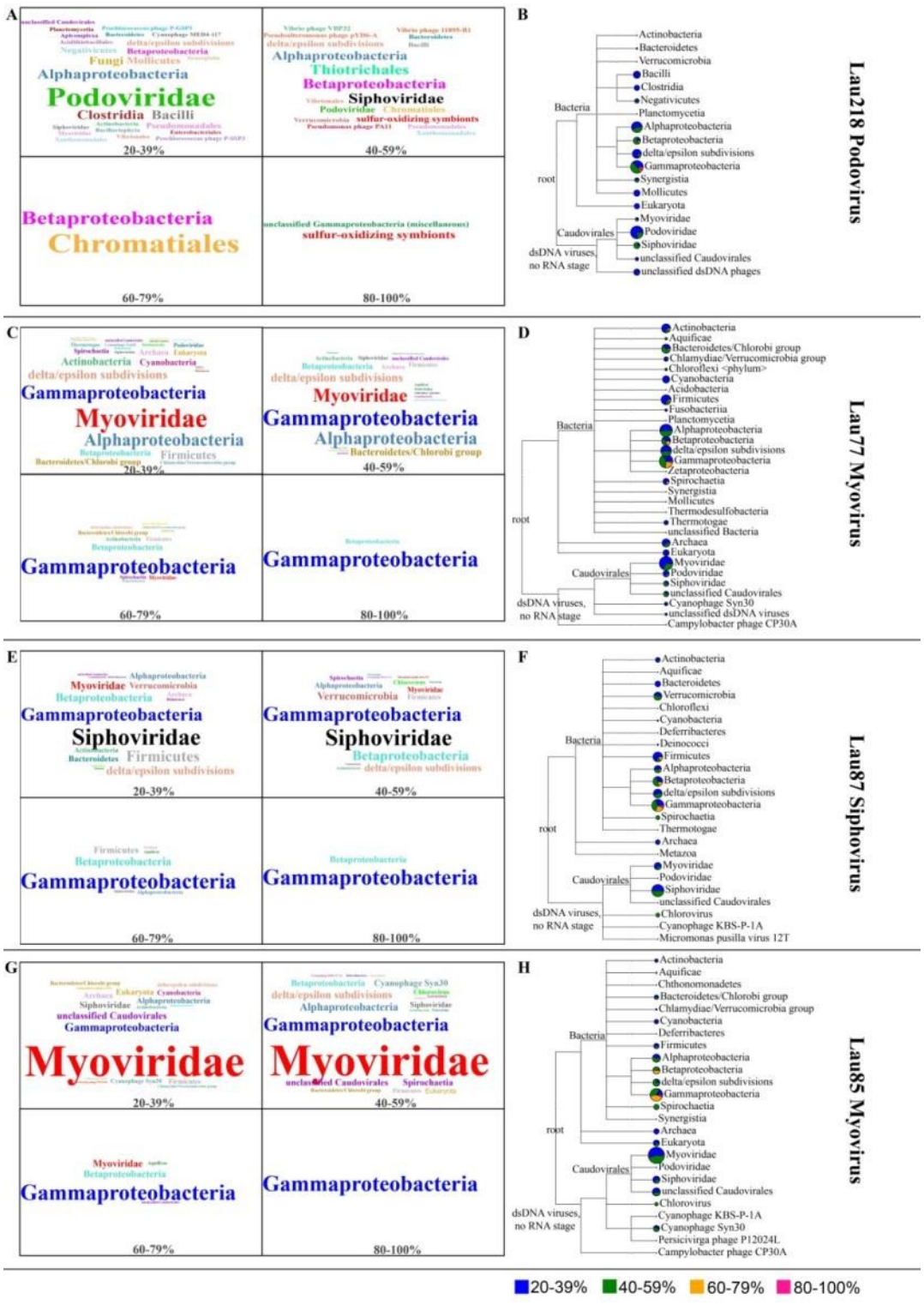
Supplementary Figure 1. Assignment of assembled Lau and Guaymas contigs to specific viral populations using ESOM implemented with tetranucleotide frequencies. Each point on the map represents a contig color coded by source as indicated by the legend. Background topography color represents euclidean distance of tetranucleotide frequency between data points, with blue indicating highest similarity, followed by green, and brown ridges representing the largest differences. Black lines indicate delineation of genomic bins.



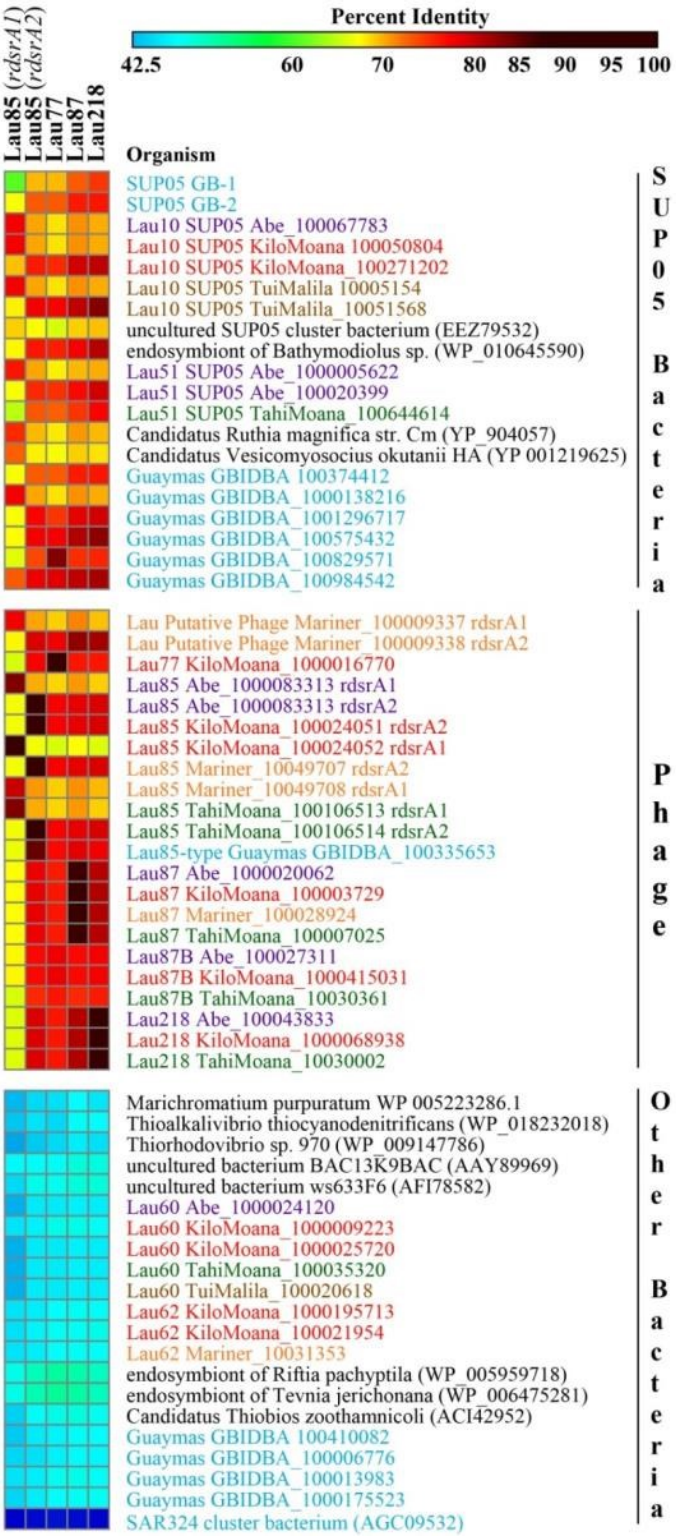
Supplementary Figure 2. Phylogenetic tree of viral large terminase (*terL)* genes inferred by Maximum Likelihood. Viral *terL* genes from Lau basin are in bold. Colored boxes indicate phylogenetic clusters containing Lau viruses.



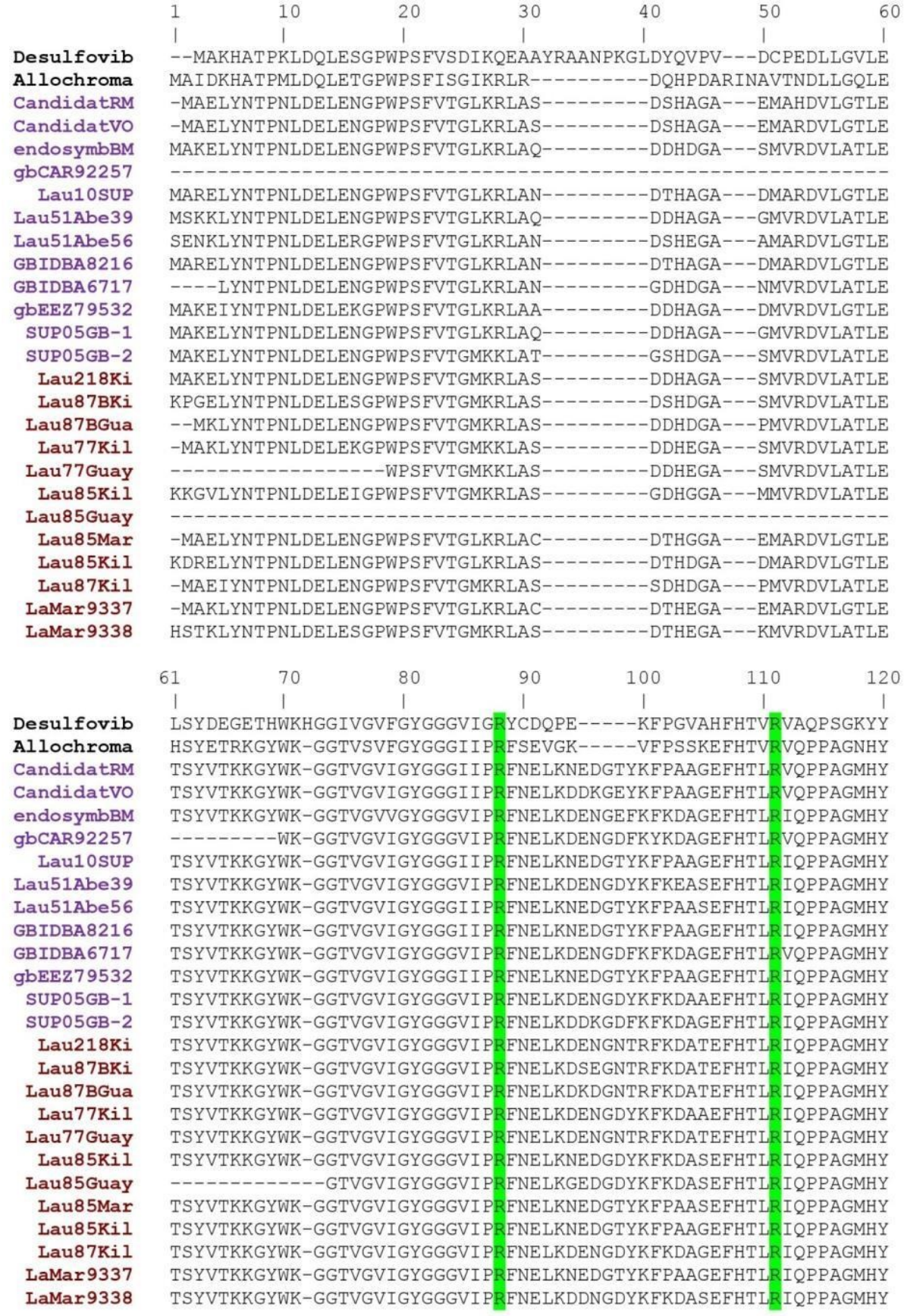
Supplementary Figure 3. Whole-genome synteny of four ELSC viruses with reference phage genomes. Colors indicate the following: Red - ELSC viral genomes; Black – Genomes of known reference phages; Grey (outermost nested circle) – Identified genes/ORFs on the genomes; Yellow- *rdsrA* and *rdsrC* genes. Ribbons indicate syntenous genes between the viruses and are colored according to sequence similarity as indicated in the legend. Arrows indicate the start position of the genomes.

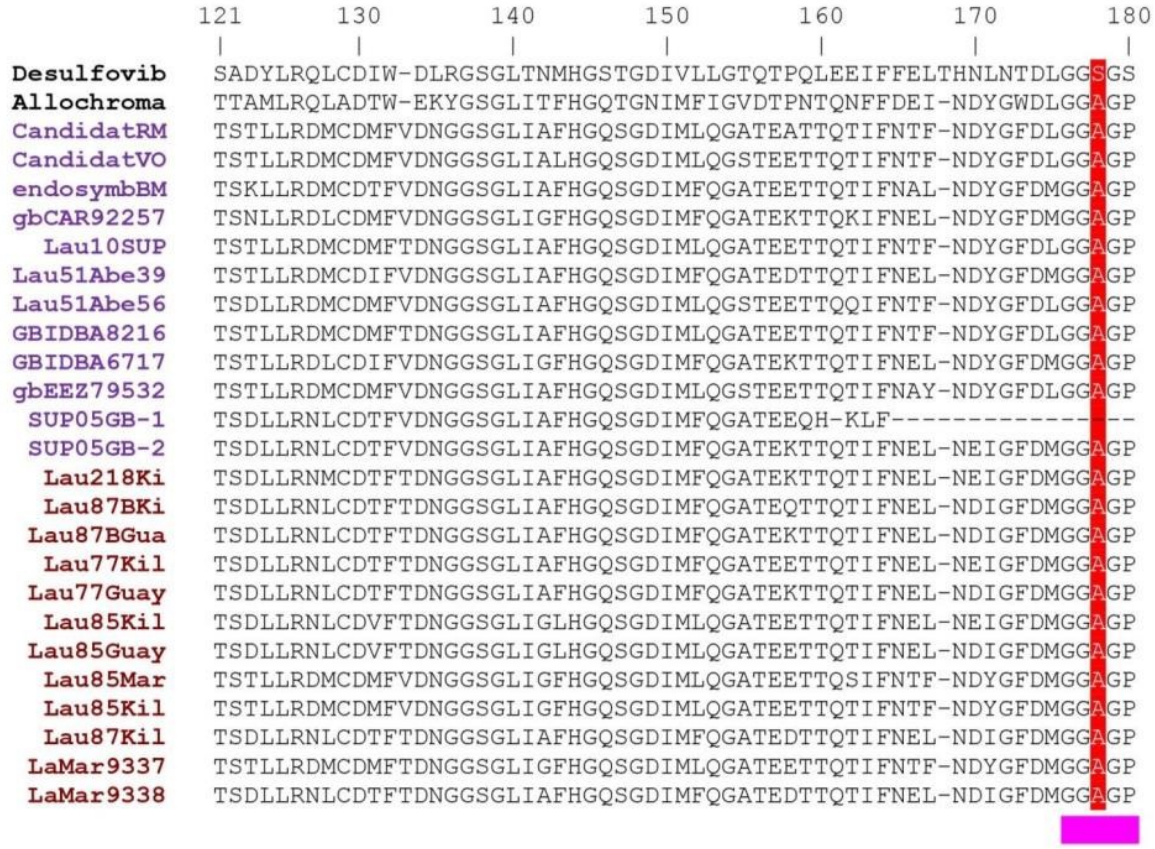


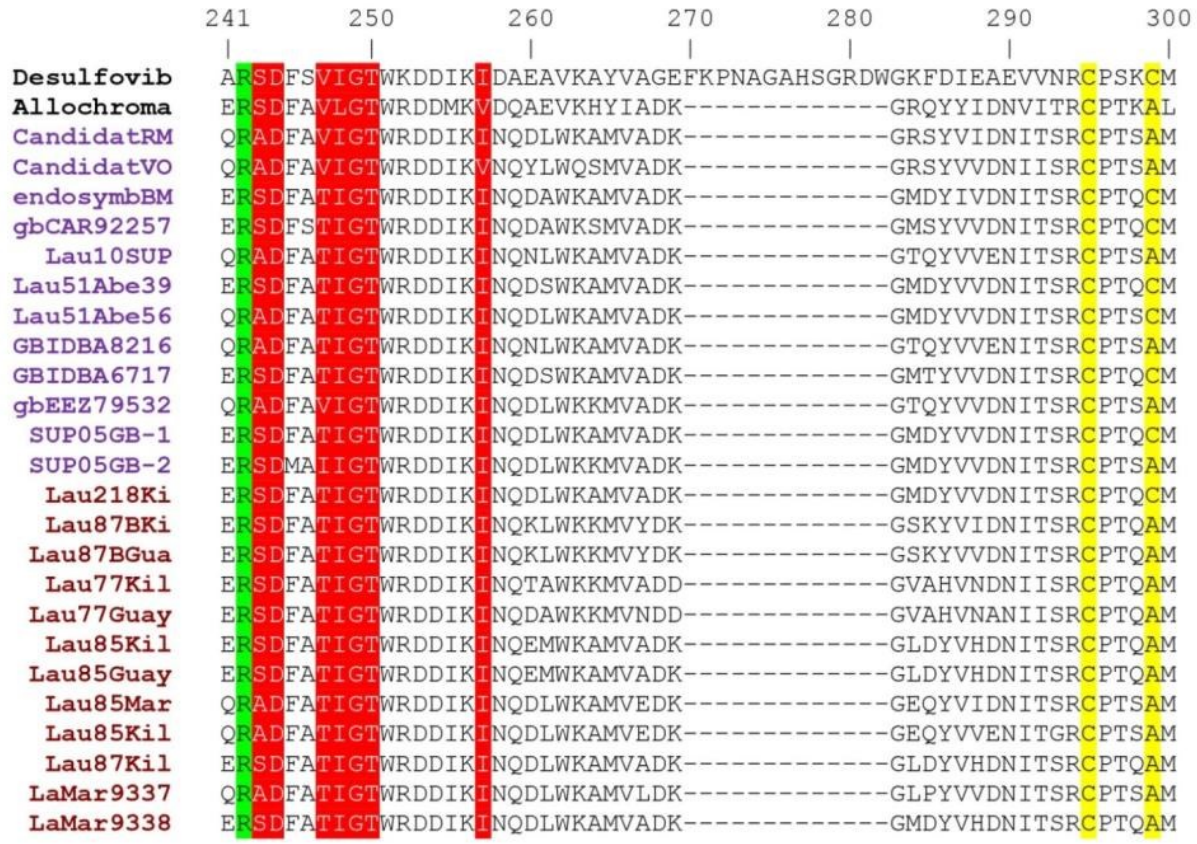
Supplementary Figure 4. Family-level weighted word cloud (A,C,E,G) and phylogenetic classification (B,D,F,H) of ELSC viral ORFs identified by blastp against NCBI-nr (cutoff: evalue >1e-5, top ten hits). Bubbles on phylogenetic tree indicate proportion of hits with colors indicating percent identity as indicated in legend.

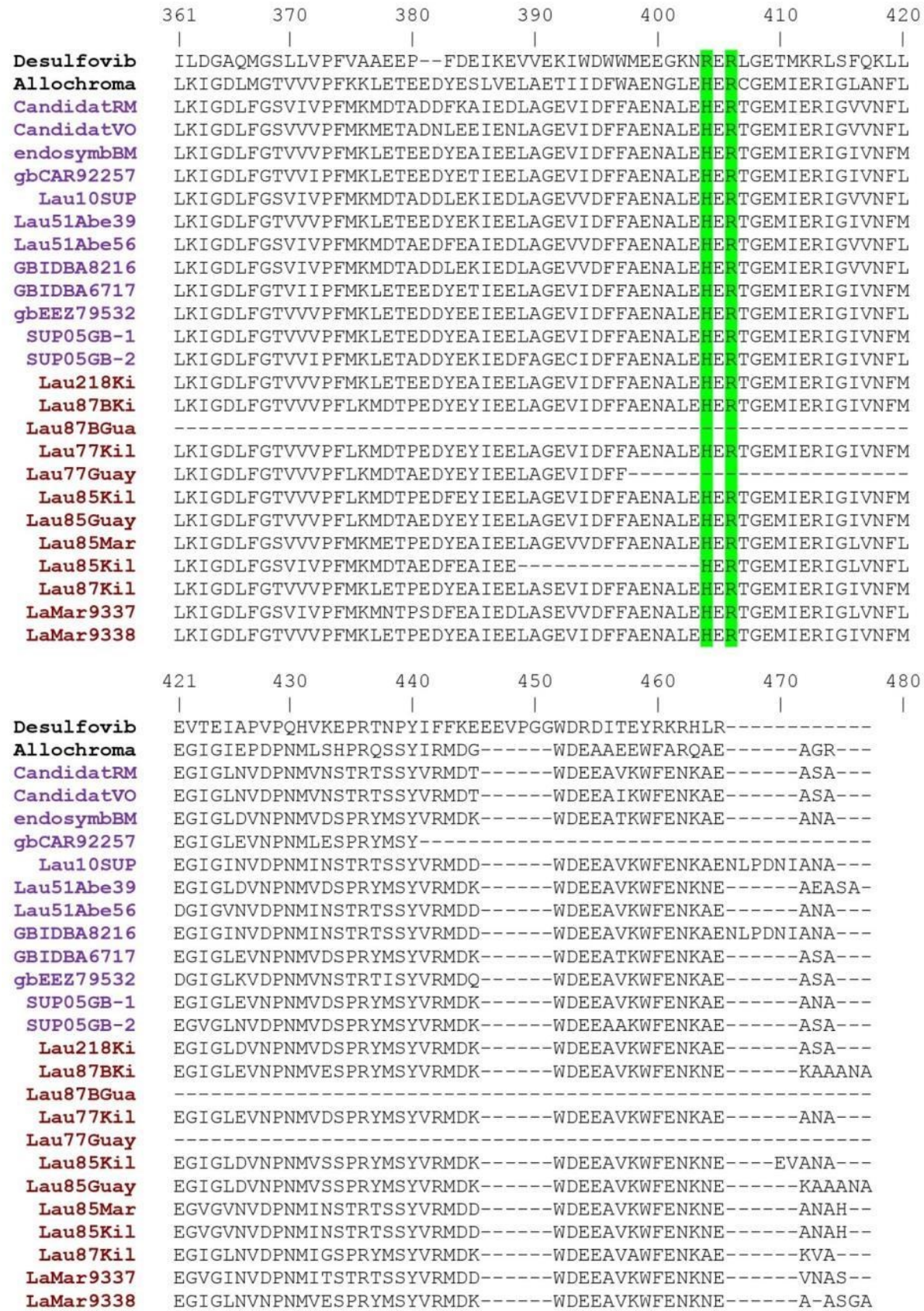


Supplementary Figure 5. **Heat map indicating comparison of amino acid identities of phage *rdsrA* to bacterial *rdsrA* sequences.** All sequences are color-coded as indicated in Fig. 1.


1 10 20 30 40 50 60
Desulfovib --MAKHATPKLDQLESGPWPSFVSDIKQEAAYRAANPKGLDYQVPV---DCPEDLLGVLE
Allochroma MAIDKHATPMLDQLETGPWPSFISGIKRLR----------DQHPDARINAVTNDLLGQLE
CandidatRM -MAELYNTPNLDELENGPWPSFVTGLKRLAS---------DSHAGA---EMAHDVLGTLE
CandidatVO -MAELYNTPNLDELENGPWPSFVTGLKRLAS---------DSHAGA---EMARDVLGTLE
endosymbBM MAKELYNTPNLDELENGPWPSFVTGLKRLAQ---------DDHDGA---SMVRDVLATLE
gbCAR92257 ----------
Lau10SUP MARELYNTPNLDELENGPWPSFVTGLKRLAN---------DTHAGA---DMARDVLGTLE
Lau51Abe39 MSKKLYNTPNLDELENGPWPSFVTGLKRLAQ---------DDHAGA---GMVRDVLATLE
Lau51Abe56 SENKLYNTPNLDELERGPWPSFVTGLKRLAN---------DSHEGA---AMARDVLGTLE
GBIDBA8216 MARELYNTPNLDELENGPWPSFVTGLKRLAN---------DTHAGA---DMARDVLGTLE
GBIDBA6717 ----LYNTPNLDELENGPWPSFVTGLKRLAN---------GDHDGA---NMVRDVLATLE
gbEEZ79532 MAKEIYNTPNLDELEKGPWPSFVTGLKRLAA---------DDHAGA---DMVRDVLGTLE
SUP05GB-1 MAKELYNTPNLDELENGPWPSFVTGLKRLAQ---------DDHAGA---GMVRDVLATLE
SUP05GB-2 MAKELYNTPNLDELENGPWPSFVTGMKKLAT---------GSHDGA---SMVRDVLATLE
Lau218Ki MAKELYNTPNLDELENGPWPSFVTGMKRLAS---------DDHAGA---SMVRDVLATLE
Lau87BKi KPGELYNTPNLDELESGPWPSFVTGMKRLAS---------DSHDGA---SMVRDVLATLE
Lau87BGua --MKLYNTPNLDELENGPWPSFVTGMKRLAS---------DDHDGA---PMVRDVLATLE
Lau77Kil -MAKLYNTPNLDELEKGPWPSFVTGMKKLAS---------DDHEGA---SMVRDVLATLE
Lau77Guay ----------WPSFVTGMKKLAS---------DDHEGA---SMVRDVLATLE
Lau85Kil KKGVLYNTPNLDELEIGPWPSFVTGMKRLAS---------GDHGGA---MMVRDVLATLE
Lau85Guay ----------
Lau85Mar -MAELYNTPNLDELENGPWPSFVTGLKRLAC---------DTHGGA---EMARDVLGTLE
Lau85Kil KDRELYNTPNLDELENGPWPSFVTGLKRLAS---------DTHDGA---DMARDVLGTLE
Lau87Kil -MAEIYNTPNLDELENGPWPSFVTGMKRLAS---------SDHDGA---PMVRDVLATLE
LaMar9337 -MAKLYNTPNLDELENGPWPSFVTGLKRLAC---------DTHEGA---EMARDVLGTLE
LaMar9338 HSTKLYNTPNLDELESGPWPSFVTGMKRLAS---------DTHEGA---KMVRDVLATLE
61 70 80 90 100 110 120
Desulfovib LSYDEGETHWKHGGIVGVFGYGGGVIGRYCDQPE-----KFPGVAHFHTVRVAQPSGKYY
Allochroma HSYETRKGYWK-GGTVSVFGYGGGIIPRFSEVGK-----VFPSSKEFHTVRVQPPAGNHY
CandidatRM TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKNEDGTYKFPAAGEFHTLRVQPPAGMHY
CandidatVO TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKDDKGEYKFPAAGEFHTLRVQPPAGMHY
endosymbBM TSYVTKKGYWK-GGTVGVVGYGGGVIPRFNELKDENGEFKFKDAGEFHTLRIQPPAGMHY
gbCAR92257 ---------WK-GGTVGVIGYGGGVIPRFNELKDENGDFKYKDAGEFHTLRVQPPAGMHY
Lau10SUP TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKNEDGTYKFPAAGEFHTLRIQPPAGMHY
Lau51Abe39 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGDYKFKEASEFHTLRIQPPAGMHY
Lau51Abe56 TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKNEDGTYKFPAASEFHTLRIQPPAGMHY
GBIDBA8216 TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKNEDGTYKFPAAGEFHTLRIQPPAGMHY
GBIDBA6717 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGDFKFKDAGEFHTLRVQPPAGMHY
gbEEZ79532 TSYVTKKGYWK-GGTVGVIGYGGGIIPRFNELKNEDGTYKFPAAGEFHTLRIQPPAGMHY
SUP05GB-1 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGDYKFKDAAEFHTLRIQPPAGMHY
SUP05GB-2 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDDKGDFKFKDAGEFHTLRIQPPAGMHY
Lau218Ki TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGNTRFKDATEFHTLRIQPPAGMHY
Lau87BKi TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDSEGNTRFKDATEFHTLRIQPPAGMHY
Lau87BGua TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDKDGNTRFKDATEFHTLRIQPPAGMHY
Lau77Kil TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGDYKFKDAAEFHTLRIQPPAGMHY
Lau77Guay TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGNTRFKDATEFHTLRIQPPAGMHY
Lau85Kil TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKNEDGDYKFKDASEFHTLRIQPPAGMHY
Lau85Guay ----------GTVGVIGYGGGVIPRFNELKGEDGDYKFKDASEFHTLRIQPPAGMHY
Lau85Mar TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKNEDGTYKFPAASEFHTLRIQPPAGMHY
Lau85Kil TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKNEDGTYKFPAAGEFHTLRIQPPAGMHY
Lau87Kil TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDENGDYKFKDAGEFHTLRIQPPAGMHY
LaMar9337 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKNEDGTYKFPAAGEFHTLRIQPPAGMHY
LaMar9338 TSYVTKKGYWK-GGTVGVIGYGGGVIPRFNELKDDNGDYKFKDASEFHTLRIQPPAGMHY

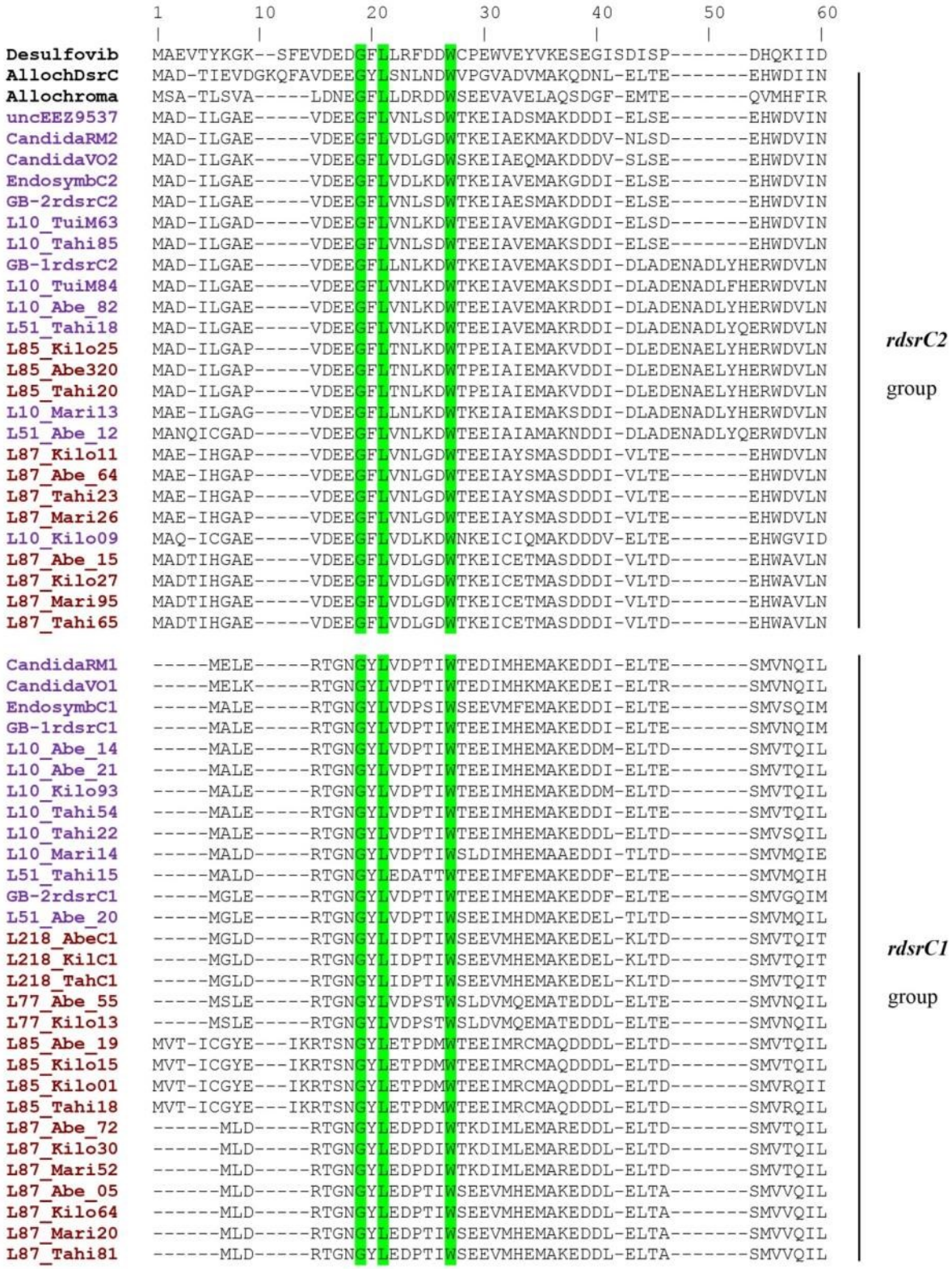

121 130 140 150 160 170 180
Desulfovib SADYLRQLCDIW-DLRGSGLTNMHGSTGDIVLLGTQTPQLEEIFFELTHNLNTDLGGSGS
Allochroma TTAMLRQLADTW-EKYGSGLITFHGQTGNIMFIGVDTPNTQNFFDEI-NDYGWDLGGAGP
CandidatRM TSTLLRDMCDMFVDNGGSGLIAFHGQSGDIMLQGATEATTQTIFNTF-NDYGFDLGGAGP
CandidatVO TSTLLRDMCDMFVDNGGSGLIALHGQSGDIMLQGSTEETTQTIFNTF-NDYGFDLGGAGP
endosymbBM TSKLLRDMCDTFVDNGGSGLIAFHGQSGDIMFQGATEETTQTIFNAL-NDYGFDMGGAGP
gbCAR92257 TSNLLRDLCDMFVDNGGSGLIGFHGQSGDIMFQGATEKTTQKIFNEL-NDYGFDMGGAGP
Lau10SUP TSTLLRDMCDMFTDNGGSGLIAFHGQSGDIMLQGATEETTQTIFNTF-NDYGFDLGGAGP
Lau51Abe39 TSTLLRDMCDIFVDNGGSGLIAFHGQSGDIMFQGATEDTTQTIFNEL-NDYGFDMGGAGP
Lau51Abe56 TSDLLRDMCDMFVDNGGSGLIAFHGQSGDIMLQGSTEETTQQIFNTF-NDYGFDLGGAGP
GBIDBA8216 TSTLLRDMCDMFTDNGGSGLIAFHGQSGDIMLQGATEETTQTIFNTF-NDYGFDLGGAGP
GBIDBA6717 TSTLLRDLCDIFVDNGGSGLIGFHGQSGDIMFQGATEKTTQTIFNEL-NDYGFDMGGAGP
gbEEZ79532 TSTLLRDMCDMFVDNGGSGLIAFHGQSGDIMLQGSTEETTQTIFNAY-NDYGFDLGGAGP
SUP05GB-1 TSDLLRNLCDTFVDNGGSGLIAFHGQSGDIMFQGATEEQH-KLF----------
SUP05GB-2 TSDLLRNLCDTFVDNGGSGLIAFHGQSGDIMFQGATEKTTQTIFNEL-NEIGFDMGGAGP
Lau218Ki TSDLLRNMCDTFTDNGGSGLIAFHGQSGDIMFQGATEKTTQTIFNEL-NEIGFDMGGAGP
Lau87BKi TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEQTTQTIFNEL-NDIGFDMGGAGP
Lau87BGua TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEKTTQTIFNEL-NDIGFDMGGAGP
Lau77Kil TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEETTQTIFNEL-NEIGFDMGGAGP
Lau77Guay TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEKTTQTIFNEL-NDIGFDMGGAGP
Lau85Kil TSDLLRNLCDVFTDNGGSGLIGLHGQSGDIMLQGATEETTQTIFNEL-NEIGFDMGGAGP
Lau85Guay TSDLLRNLCDVFTDNGGSGLIGLHGQSGDIMLQGATEETTQTIFNEL-NDIGFDMGGAGP
Lau85Mar TSTLLRDMCDMFVDNGGSGLIGFHGQSGDIMLQGATEETTQSIFNTF-NDYGFDMGGAGP
Lau85Kil TSTLLRDMCDMFVDNGGSGLIGFHGQSGDIMLQGATEETTQTIFNTF-NDYGFDMGGAGP
Lau87Kil TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEDTTQTIFNEL-NDIGFDMGGAGP
LaMar9337 TSTLLRDMCDMFTDNGGSGLIGFHGQSGDIMLQGATEETTQTIFNTF-NDYGFDMGGAGP
LaMar9338 TSDLLRNLCDTFTDNGGSGLIAFHGQSGDIMFQGATEDTTQTIFNEL-NDIGFDMGGAGP
181 190 200 210 220 230 240
Desulfovib NLRTPESC GKSRCEFACYDSQAACYELTMEYQDELHRPAFPYKFKFKFDACPNGCVASI
Allochroma CVRTAMSCVGSARCEMSCTNELKAHRLLVNNFTDDVHRPALPYKFKFKVSGCPNDCQNAI
CandidatRM AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNTV
CandidatVO AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCSNDCMNTV
endosymbBM AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
gbCAR92257 AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau10SUP AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNTV
Lau51Abe39 AVRTGMSCVGAARCEMSNTNESAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau51Abe56 AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNTV
GBIDBA8216 AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNTV
GBIDBA6717 AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
gbEEZ79532 ----------MSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSV
SUP05GB-1 ----------MSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
SUP05GB-2 AVRTGMACVGASRCEMSNTNETAALRTLVNAFLDDMHRPALPYKFKFKVSGCANDCMNSI
Lau218Ki AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau87BKi AVRTGMSCVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau87BGua AVRTGMACVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau77Kil AVRTGMSCVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau77Guay AVRTGMSCVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau85Kil AVRTGMSCVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau85Guay AVRTGMACVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
Lau85Mar AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKLKFKVSGCSNDCMNSV
Lau85Kil AVRTGMSCVGASRCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCSNDCMNSV
Lau87Kil AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI
LaMar9337 AVRTGMSCVGAARCEMSNVNEQAVLRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSV
LaMar9338 AVRTGMSCVGAARCEMSNTNEQAALRTLVNAFLDDMHRPALPYKMKFKVSGCANDCMNSI

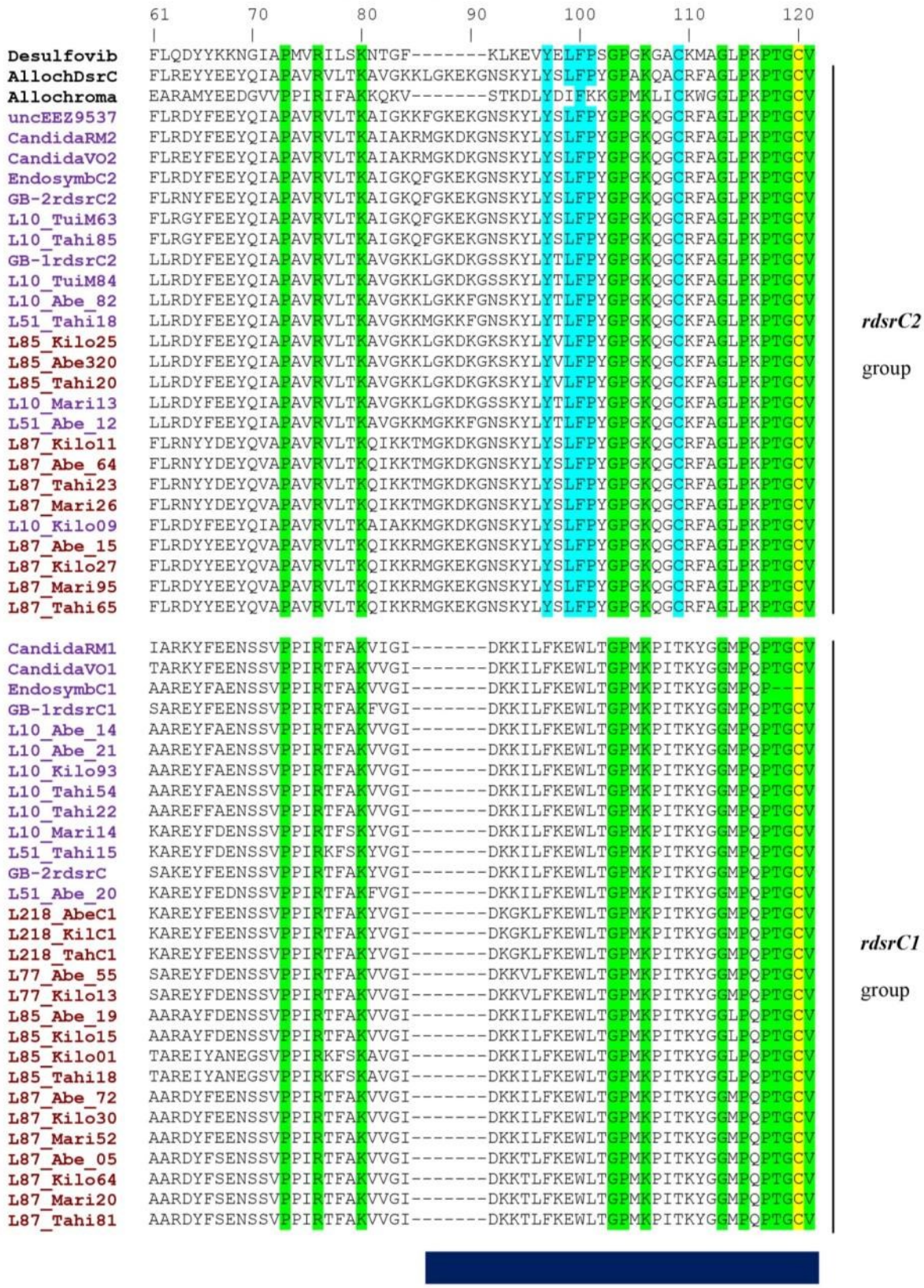

241 250 260 270 280 290 300
Desulfovib ARSDFSVIGTWKDDIKIDAEAVKAYVAGEFKPNAGAHSGRDWGKFDIEAEVVNRCPSKCM
Allochroma ERSDFAVLGTWRDDMKVDQAEVKHYIADK----------GRQYYIDNVITRCPTKAL
CandidatRM QRADFAVIGTWRDDIKINQDLWKAMVADK----------GRSYVIDNITSRCPTSAM
CandidatVO QRADFAVIGTWRDDIKVNQYLWQSMVADK----------GRSYVVDNIISRCPTSAM
endosymbBM ERSDFATIGTWRDDIKINQDAWKAMVADK----------GMDYIVDNITSRCPTQCM
gbCAR92257 ERSDFSTIGTWRDDIKINQDAWKSMVADK----------GMSYVVDNITSRCPTQCM
Lau10SUP QRADFATIGTWRDDIKINQNLWKAMVADK----------GTQYVVENITSRCPTSAM
Lau51Abe39 ERSDFATIGTWRDDIKINQDSWKAMVADK----------GMDYVVDNITSRCPTQCM
Lau51Abe56 QRADFATIGTWRDDIKINQDLWKAMVADK----------GMDYVVDNITSRCPTSCM
GBIDBA8216 QRADFATIGTWRDDIKINQNLWKAMVADK----------GTQYVVENITSRCPTSAM
GBIDBA6717 ERSDFATIGTWRDDIKINQDSWKAMVADK----------GMTYVVDNITSRCPTQCM
gbEEZ79532 QRADFAVIGTWRDDIKINQDLWKKMVADK----------GTQYVVDNITSRCPTSAM
SUP05GB-1 ERSDFATIGTWRDDIKINQDLWKAMVADK----------GMDYVVDNITSRCPTQCM
SUP05GB-2 ERSDMAIIGTWRDDIKINQDLWKKMVADK----------GMDYVVDNITSRCPTSAM
Lau218Ki ERSDFATIGTWRDDIKINQDLWKAMVADK----------GMDYVVDNITSRCPTQCM
Lau87BKi ERSDFATIGTWRDDIKINQKLWKKMVYDK----------GSKYVIDNITSRCPTQAM
Lau87BGua ERSDFATIGTWRDDIKINQKLWKKMVYDK----------GSKYVVDNITSRCPTQAM
Lau77Kil ERSDFATIGTWRDDIKINQTAWKKMVADD----------GVAHVNDNIISRCPTQAM
Lau77Guay ERSDFATIGTWRDDIKINQDAWKKMVNDD----------GVAHVNANIISRCPTQAM
Lau85Kil ERSDFATIGTWRDDIKINQEMWKAMVADK----------GLDYVHDNITSRCPTQAM
Lau85Guay ERSDFATIGTWRDDIKINQEMWKAMVADK----------GLDYVHDNITSRCPTQAM
Lau85Mar QRADFATIGTWRDDIKINQDLWKAMVEDK----------GEQYVIDNITSRCPTSAM
Lau85Kil QRADFATIGTWRDDIKINQDLWKAMVEDK----------GEQYVVENITGRCPTSAM
Lau87Kil ERSDFATIGTWRDDIKINQDLWKAMVADK----------GLDYVHDNITSRCPTQAM
LaMar9337 QRADFATIGTWRDDIKINQDLWKAMVLDK----------GLPYVVDNITSRCPTSAM
LaMar9338 ERSDFATIGTWRDDIKINQDLWKAMVADK----------GMDYVHDNITSRCPTQAM
301 310 320 330 340 350 360
Desulfovib KW-DGSKLSIDNKECVRCMHCINTMPRALH----------IGDERGASILCGAKAP
Allochroma SLNDDDTLDVNNRDCVRCMHCLNVMPKALH----------PGDDKGVTILIGGKRT
CandidatRM TLNEDNSITIDNKNCVKCMHCLNVTSPLTHKYIV-KGDVEPILATGDDKGVMIIMGGKRT
CandidatVO TLNEDDSVTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDIEPILATGDDKGVMIIMGGKRT
endosymbBM KVEADTSLTIDNKNCVKCMHCLNATSPLTHNYIK-DAAEGAILATGDDKGVTICMGGKRT
gbCAR92257 KVNDDTSLSIDNENCVKCMHCLNATSPLTHKYID-DASKGAILATGDDKGVTICMGGKRT
Lau10SUP TINEDTSLTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDVEPILATGDDKGVMIIMGGKRT
Lau51Abe39 KVEADNSLTIDNRNCVKCMHCLNATSPLNHKYTD-KPSDGAMLATGDDKGVTICMGGKRT
Lau51Abe56 KVEEDTSLTIDNKNCVRCMHCLNVTSPLTHRYIA-EGDVEPILATGDDKGVMIIMGGKRT
GBIDBA8216 TINEDTSLTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDVEPILATGDDKGVMIIMGGKRT
GBIDBA6717 KVNEDTSLTIDNANCVKCMHCLNATSPLTHKYIQKDAPAEAILATGDDKGVTICMGGKRT
gbEEZ79532 TLNDDSSLTIDNKNCVKCMHCLNATSPLTHNYIA-KGDIEPILATGDDKGVSIIMGGKRT
SUP05GB-1 KVESDTSLTIDNKNCVKCMHCLNVTSPLTHKYITKDMPTEAILATGDDKGVTICMGGKRT
SUP05GB-2 TVNADTSLTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDVEPILATGDDKGVTVCMGGKRT
Lau218Ki TVNADTSLTIDNRNCVKCMHCLNATSPLTHNYITKDMPTEAILATGDDKGVTICMGGKRT
Lau87BKi HLNKDKSLTIDNANCVKCMHCLNVTSPLTHKYIS-EGDVEPILAQGDDRGVTICMGGKRT
Lau87BGua ----------
Lau77Kil KMLPDNSLEIDNKNCVKCMHCLNVTSPKTHKYIS-DASEGAILEQGDDKGVTICMGGKRT
Lau77Guay KLLPDNSLEIDNKNCVKCMHCLNVTSPKTHKYIS-DAAEGAILEQGDDKGVTICMGGKRT
Lau85Kil HVNADTSLTIDNRNCVKCMHCLNATSPLTHNYIQKDAPAEAILATGDDKGVTICMGGKRT
Lau85Guay HVNADTSLTIDNRNCVKCMHCLNVTSPLTHKYIQKDAPAEAILATGDDKGVTICMGGKRT
Lau85Mar QMNSDKSLTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDVEPILAQGDDKGVMIIMGGKRT
Lau85Kil KMESDKSLTIDNKNCVKCMHCLNVTSPLTHKYIA-KGDVEPILAQGDDKGVMIIMGGKRT
Lau87Kil TLNEDTSLTIDNRNCVKCMHCLNVTSPLTHKYITKDMPTEAILATGDDKGVTICMGGKRT
LaMar9337 TIESDLTLNIDNKNCVKCMHCLNVTSPLTHKYIL-DGDVEPILAQGDDKGVMIIMGGKRT
LaMar9338 KVEADTSLTIDNKNCVKCMHCLNVTSPLTHKYITKDMPTEAILATGDDKGVTICMGGKRT


361 370 380 390 400 410 420
Desulfovib ILDGAQMGSLLVPFVAAEEP--FDEIKEVVEKIWDWWMEEGKNRERLGETMKRLSFQKLL
Allochroma LKIGDLMGTVVVPFKKLETEEDYESLVELAETIIDFWAENGLEHERCGEMIERIGLANFL
CandidatRM LKIGDLFGSVIVPFMKMDTADDFKAIEDLAGEVIDFFAENALEHERTGEMIERIGVVNFL
CandidatVO LKIGDLFGSVVVPFMKMETADNLEEIENLAGEVIDFFAENALEHERTGEMIERIGVVNFL
endosymbBM LKIGDLFGTVVVPFMKLETEEDYEAIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
gbCAR92257 LKIGDLFGTVVIPFMKLETEEDYETIEELAGEVIDFFAENALEHERTGEMIERIGIVNFL
Lau10SUP LKIGDLFGSVIVPFMKMDTADDLEKIEDLAGEVVDFFAENALEHERTGEMIERIGVVNFL
Lau51Abe39 LKIGDLFGTVVVPFMKLETEEDYEKIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau51Abe56 LKIGDLFGSVIVPFMKMDTAEDFEAIEDLAGEVVDFFAENALEHERTGEMIERIGVVNFL
GBIDBA8216 LKIGDLFGSVIVPFMKMDTADDLEKIEDLAGEVVDFFAENALEHERTGEMIERIGVVNFL
GBIDBA6717 LKIGDLFGTVIIPFMKLETEEDYETIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
gbEEZ79532 LKIGDLFGTVVVPFMKLETEDDYEEIEELAGEVIDFFAENALEHERTGEMIERIGIVNFL
SUP05GB-1 LKIGDLFGTVVVPFMKLETEDDYEAIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
SUP05GB-2 LKIGDLFGTVVIPFMKLETADDYEKIEDFAGECIDFFAENALEHERTGEMIERIGIVNFL
Lau218Ki LKIGDLFGTVVVPFMKLETEEDYEAIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau87BKi LKIGDLFGTVVVPFLKMDTPEDYEYIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau87BGua ----------
Lau77Kil LKIGDLFGTVVVPFLKMDTPEDYEYIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau77Guay LKIGDLFGTVVVPFLKMDTAEDYEYIEELAGEVIDFF----------
Lau85Kil LKIGDLFGTVVVPFLKMDTPEDFEYIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau85Guay LKIGDLFGTVVVPFLKMDTAEDYEYIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
Lau85Mar LKIGDLFGSVVVPFMKMETPEDYEAIEELAGEVVDFFAENALEHERTGEMIERIGLVNFL
Lau85Kil LKIGDLFGSVIVPFMKMDTAEDFEAIEE----------HERTGEMIERIGLVNFL
Lau87Kil LKIGDLFGTVVVPFMKLETPEDYEAIEELASEVIDFFAENALEHERTGEMIERIGIVNFM
LaMar9337 LKIGDLFGSVIVPFMKMNTPSDFEAIEDLASEVVDFFAENALEHERTGEMIERIGLVNFL
LaMar9338 LKIGDLFGTVVVPFMKLETPEDYEAIEELAGEVIDFFAENALEHERTGEMIERIGIVNFM
421 430 440 450 460 470 480
Desulfovib EVTEIAPVPQHVKEPRTNPYIFFKEEEVPGGWDRDITEYRKRHLR----------
Allochroma EGIGIEPDPNMLSHPRQSSYIRMDG------WDEAAEEWFARQAE------AGR---
CandidatRM EGIGLNVDPNMVNSTRTSSYVRMDT------WDEEAVKWFENKAE------ASA---
CandidatVO EGIGLNVDPNMVNSTRTSSYVRMDT------WDEEAIKWFENKAE------ASA---
endosymbBM EGIGLDVNPNMVDSPRYMSYVRMDK------WDEEATKWFENKAE------ANA---
gbCAR92257 EGIGLEVNPNMLESPRYMSY----------
Lau10SUP EGIGINVDPNMINSTRTSSYVRMDD------WDEEAVKWFENKAENLPDNIANA---
Lau51Abe39 EGIGLDVNPNMVDSPRYMSYVRMDK------WDEEAVKWFENKNE------AEASA-
Lau51Abe56 DGIGVNVDPNMINSTRTSSYVRMDD------WDEEAVKWFENKAE------ANA---
GBIDBA8216 EGIGINVDPNMINSTRTSSYVRMDD------WDEEAVKWFENKAENLPDNIANA---
GBIDBA6717 EGIGLEVNPNMVDSPRYMSYVRMDK------WDEEATKWFENKAE------ASA---
gbEEZ79532 DGIGLKVDPNMVNSTRTISYVRMDQ------WDEEAVKWFENKAE------ASA---
SUP05GB-1 EGIGLEVNPNMVDSPRYMSYVRMDK------WDEEAVKWFENKAE------ANA---
SUP05GB-2 EGVGLNVDPNMVDSPRYMSYVRMDK------WDEEAAKWFENKAE------ASA---
Lau218Ki EGIGLDVNPNMVDSPRYMSYVRMDK------WDEEAVKWFENKAE------ASA---
Lau87BKi EGIGLEVNPNMVESPRYMSYVRMDK------WDEEAVKWFENKNE------KAAANA
Lau87BGua ----------
Lau77Kil EGIGLEVNPNMVDSPRYMSYVRMDK------WDEEAVKWFENKAE------ANA---
Lau77Guay ----------
Lau85Kil EGIGLDVNPNMVSSPRYMSYVRMDK------WDEEAVKWFENKNE----EVANA---
Lau85Guay EGIGLDVNPNMVSSPRYMSYVRMDK------WDEEAVKWFENKNE------KAAANA
Lau85Mar EGVGVNVDPNMINSTRTSSYVRMDD------WDEEAVKWFENKNE------ANAH--
Lau85Kil EGVGVNVDPNMINSTRTSSYVRMDD------WDEEAVKWFENKNE------ANAH--
Lau87Kil EGIGLNVDPNMIGSPRYMSYVRMDK------WDEEAVAWFENKAE------KVA---
LaMar9337 EGVGINVDPNMITSTRTSSYVRMDD------WDEEAVKWFENKNE------VNAS--
LaMar9338 EGIGLNVNPNMVESPRYMSYVRMDK------WDEEAVKWFENKNE------A-ASGA

Supplementary Figure 6. Sequence alignment of SUP05 and phage *rdsrA* with *Allochromatium vinosum* DSM180 *rdsrA* and *Desulfovibro vulgaris* 'Hildenborough' *dsrA*. Genes are color-coded with SUP05-derived sequences in purple and phage-derived sequences in brown. Strictly conserved residues are highlighted in colors as follows: (i) Blue – $CX_5CX_nCX_3C$ motif identified for binding of the siroheme-[4Fe4S] cofactor in dissimilatory sulfite reductase A1 domain in Oliveira et al (2008) and Dahl et al(1993) (ii) Green – Substrate binding sites in *dsrA* identified in Oliveira et al (2008) (iii) Yellow – Conserved residues for binding one [4Fe4S]cluster in dissimilatory sulfite reductase A2 domain identified in Oliveira et al (2008) (iv) Red - Conserved residues across the entire sulfite reductase domain (assimilatory, dissimilatory and reverse-acting dissimilatory) identified in Dhillon et al 2005). The secondary structure elements are indicated below the alignments with orange boxes depicting α-helices and pink boxes indicating β-sheets. Annotations with accession numbers are as follows:

Desulfovib: *Desulfovibrio vulgaris* str.Hildenborough (P45574)
Allochroma: *Allochromatium vinosum* DSM180 (AAC35394)
CandidatRM: Candidatus *Ruthia magnifica* str.Cm (YP_904057
CandidatVO: Candidatus *Vesicomyosocius okutanii* str.HA (YP_001219625)
endosymbBM: endosymbiont of *Bathymodiolus* sp. (WP_010645590)
gbCAR92257: uncultured african shelf marine bacterium (CAR92257)
Lau10SUP: KiloMoana_100050804
Lau51Abe39: Abe_100020399
Lau51Abe56: Lau51_Abe_1000005622
GBIDBA8216: GBIDBA_1000138216
GBIDBA6717: GBIDBA_1001296717
gbEEZ79532: uncultured SUP05 bacterium (EEZ79532)
SUP05GB-1: SUP05_GB-1(2062113603)
SUP05GB-2: SUP05_GB-2(2062112241)
Lau218Ki: Lau_218_KiloMoana_1000068938
Lau87BKi: Lau_87B_KiloMoana_1000415031
Lau87BGua: Lau87_Guaymas_GBIDBA_100984542
Lau77Kil: Lau77_KiloMoana_1000016770
Lau77Guay: Lau77_Guaymas_GBIDBA_100829571
Lau85Kil: Lau85_KiloMoana_100024051_DsrA2
Lau85Guay: Lau85_DsrA2_Guaymas_GBIDBA_100335653
Lau85Mar: Lau85_Mariner_10049708_DsrA1
Lau85Kil: Lau_85_KiloMoana_100024052_DsrA1
Lau87Kil: Lau_87_KiloMoana_100003729_DsrA
LaMar9337: Lau_Putative_Phage_Mariner_100009337_DsrA1
LaMar9338: Lau_Putative_Phage_Mariner_100009338_DsrA2

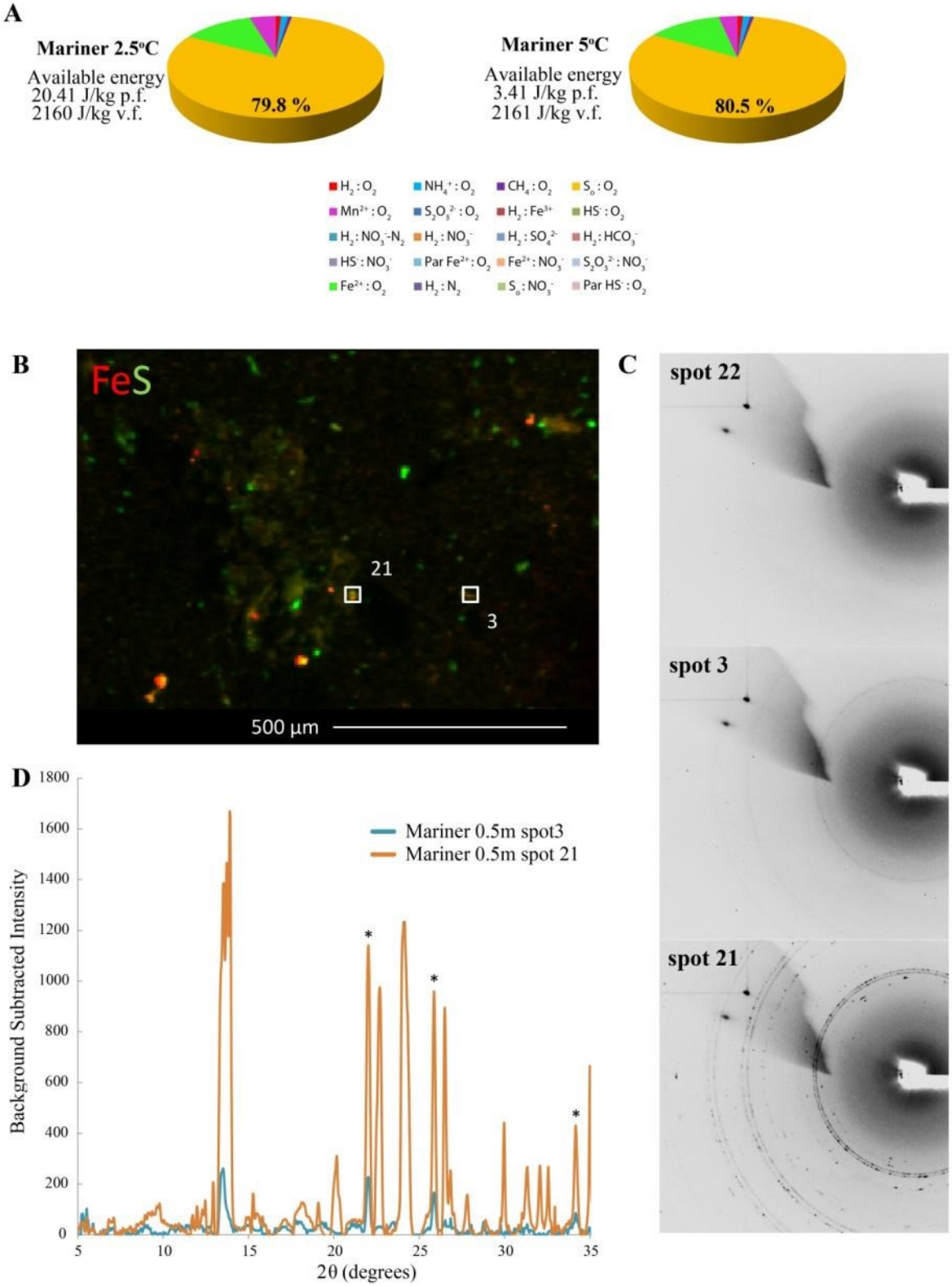

1 10 20 30 40 50 60
Desulfovib MAEVTYKGK--SFEVDEDGFLLRFDDWCPEWVEYVKESEGISDISP-------DHQKIID
AllochDsrC MAD-TIEVDGKQFAVDEEGYLSNLNDWVPGVADVMAKQDNL-ELTE-------EHWDIIN
Allochroma MSA-TLSVA-----LDNEGFLLDRDDWSEEVAVELAQSDGF-EMTE-------QVMHFIR
uncEEZ9537 MAD-ILGAE-----VDEEGFLVNLSDWTKEIADSMAKDDDI-ELSE-------EHWDVIN
CandidaRM2 MAD-ILGAE-----VDEEGFLVDLGDWTKEIAEKMAKDDDV-NLSD-------EHWDVIN
CandidaVO2 MAD-ILGAK-----VDEEGFLVDLGDWSKEIAEQMAKDDDV-SLSE-------EHWDVIN
EndosymbC2 MAD-ILGAE-----VDEEGFLVDLKDWTKEIAVEMAKGDDI-ELSE-------EHWDVIN
GB-2rdsrC2 MAD-ILGAE-----VDEEGFLVNLSDWTKEIAESMAKDDDI-ELSE-------EHWDVIN
L10_TuiM63 MAD-ILGAD-----VDEEGFLVNLKDWTEEIAVEMAKGDDI-ELSD-------EHWDVIN
L10_Tahi85 MAD-ILGAE-----VDEEGFLVNLSDWTEEIAVEMAKSDDI-ELSE-------EHWDVLN
GB-1rdsrC2 MAD-ILGAE-----VDEEGFLLNLKDWTKEIAVEMAKSDDI-DLADENADLYHERWDVLN
L10_TuiM84 MAD-ILGAE-----VDEEGFLVNLKDWTKEIAVEMAKSDDI-DLADENADLFHERWDVLN
L10_Abe_82 MAD-ILGAE-----VDEEGFLVNLKDWTEEIAVEMAKRDDI-DLADENADLYHERWDVLN
L51_Tahi18 MAD-ILGAE-----VDEEGFLVNLKDWTEEIAVEMAKRDDI-DLADENADLYQERWDVLN
L85_Kilo25 MAD-ILGAP-----VDEEGFLTNLKDWTPEIAIEMAKVDDI-DLEDENAELYHERWDVLN
L85_Abe320 MAD-ILGAP-----VDEEGFLTNLKDWTPEIAIEMAKVDDI-DLEDENAELYHERWDVLN
L85_Tahi20 MAD-ILGAP-----VDEEGFLTNLKDWTPEIAIEMAKVDDI-DLEDENAELYHERWDVLN
L10_Mari13 MAE-ILGAG-----VDEEGFLLNLKDWTKEIAIEMAKSDDI-DLADENADLYHERWDVLN
L51_Abe_12 MANQICGAD-----VDEEGFLVNLKDWTEEIAIAMAKNDDI-DLADENADLYQERWDVLN
L87_Kilo11 MAE-IHGAP-----VDEEGFLVNLGDWTEEIAYSMASDDDI-VLTE-------EHWDVLN
L87_Abe_64 MAE-IHGAP-----VDEEGFLVNLGDWTEEIAYSMASDDDI-VLTE-------EHWDVLN
L87_Tahi23 MAE-IHGAP-----VDEEGFLVNLGDWTEEIAYSMASDDDI-VLTE-------EHWDVLN
L87_Mari26 MAE-IHGAP-----VDEEGFLVNLGDWTEEIAYSMASDDDI-VLTE-------EHWDVLN
L10_Kilo09 MAQ-ICGAE-----VDEEGFLVDLKDWNKEICIQMAKDDDV-ELTE-------EHWGVID
L87_Abe_15 MADTIHGAE-----VDEEGFLVDLGDWTKEICETMASDDDI-VLTD-------EHWAVLN
L87_Kilo27 MADTIHGAE-----VDEEGFLVDLGDWTKEICETMASDDDI-VLTD-------EHWAVLN
L87_Mari95 MADTIHGAE-----VDEEGFLVDLGDWTKEICETMASDDDI-VLTD-------EHWAVLN
L87_Tahi65 MADTIHGAE-----VDEEGFLVDLGDWTKEICETMASDDDI-VLTD-------EHWAVLN
rdsrC2
group
CandidaRM1 -----MELE-----RTGNGYLVDPTIWTEDIMHEMAKEDDI-ELTE-------SMVNQIL
CandidaVO1 -----MELK-----RTGNGYLVDPTIWTEDIMHKMAKEDEI-ELTR-------SMVNQIL
EndosymbC1 -----MALE-----RTGNGYLVDPSIWSEEVMFEMAKEDDI-ELTE-------SMVSQIM
GB-1rdsrC1 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDI-ELTE-------SMVNQIM
L10_Abe_14 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDM-ELTD-------SMVTQIL
L10_Abe_21 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDI-ELTE-------SMVTQIL
L10_Kilo93 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDM-ELTD-------SMVTQIL
L10_Tahi54 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDI-ELTE-------SMVTQIL
L10_Tahi22 -----MALE-----RTGNGYLVDPTIWTEEIMHEMAKEDDL-ELTD-------SMVSQIL
L10_Mari14 -----MALD-----RTGNGYLVDPTIWSLDIMHEMAAEDDI-TLTD-------SMVMQIE
L51_Tahi15 -----MALD-----RTGNGYLEDATTWTEEIMFEMAKEDDF-ELTE-------SMVMQIH
GB-2rdsrC1 -----MGLE-----RTGNGYLVDPTIWTEEIMHEMAKEDDF-ELTE-------SMVGQIM
L51_Abe_20 -----MGLE-----RTGNGYLVDPTIWSEEIMHDMAKEDEL-TLTD-------SMVMQIL
L218_AbeC1 -----MGLD-----RTGNGYLIDPTIWSEEVMHEMAKEDEL-KLTD-------SMVTQIT
L218_KilC1 -----MGLD-----RTGNGYLIDPTIWSEEVMHEMAKEDEL-KLTD-------SMVTQIT
L218_TahC1 -----MGLD-----RTGNGYLIDPTIWSEEVMHEMAKEDEL-KLTD-------SMVTQIT
L77_Abe_55 -----MSLE-----RTGNGYLVDPSTWSLDVMQEMATEDDL-ELTE-------SMVNQIL
L77_Kilo13 -----MSLE-----RTGNGYLVDPSTWSLDVMQEMATEDDL-ELTE-------SMVNQIL
L85_Abe_19 MVT-ICGYE---IKRTSNGYLETPDMWTEEIMRCMAQDDDL-ELTD-------SMVTQIL
L85_Kilo15 MVT-ICGYE---IKRTSNGYLETPDMWTEEIMRCMAQDDDL-ELTD-------SMVTQIL
L85_Kilo01 MVT-ICGYE---IKRTSNGYLETPDMWTEEIMRCMAQDDDL-ELTD-------SMVRQII
L85_Tahi18 MVT-ICGYE---IKRTSNGYLETPDMWTEEIMRCMAQDDDL-ELTD-------SMVRQIL
L87_Abe_72 ------MLD-----RTGNGYLEDPDIWTKDIMLEMAREDDL-ELTD-------SMVTQIL
L87_Kilo30 ------MLD-----RTGNGYLEDPDIWTKDIMLEMAREDDL-ELTD-------SMVTQIL
L87_Mari52 ------MLD-----RTGNGYLEDPDIWTKDIMLEMAREDDL-ELTD-------SMVTQIL
L87_Abe_05 ------MLD-----RTGNGYLEDPTIWSEEVMHEMAKEDDL-ELTA-------SMVVQIL
L87_Kilo64 ------MLD-----RTGNGYLEDPTIWSEEVMHEMAKEDDL-ELTA-------SMVVQIL
L87_Mari20 ------MLD-----RTGNGYLEDPTIWSEEVMHEMAKEDDL-ELTA-------SMVVQIL
L87_Tahi81 ------MLD-----RTGNGYLEDPTIWSEEVMHEMAKEDDL-ELTA-------SMVVQIL
rdsrC1
group


61
70
80
90
100
110
120
Desulfovib FLQDYYKKNGIAPMVRILSKNTGF-------KLKEVYELFPSGPGKGACKMAGLPKPTGCV
AllochDsrC FLREYYEEYQIAPAVRVLTKAVGKKLGKEKGNSKYLYSLFPYGPAKQACRFAGLPKPTGCV
Allochroma EARAMYEEDGVVPPIRIFAKKQKV-------STKDLYDIFKKGPMKLICKWGGLPKPTGCV
uncEEZ9537 FLRDYFEEYQIAPAVRVLTKAIGKKFGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
CandidaRM2 FLRDYFEEYQIAPAVRVLTKAIAKRMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
CandidaVO2 FLREYFEEYQIAPAVRVLTKAIAKRMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
EndosymbC2 FLRDYFEEYQIAPAVRVLTKAIGKQFGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
GB-2rdsrC2 FLRNYFEEYQIAPAVRVLTKAIGKQFGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L10_TuiM63 FLRGYFEEYQIAPAVRVLTKAIGKQFGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L10_Tahi85 FLRGYFEEYQIAPAVRVLTKAIGKQFGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
GB-1rdsrC2 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGSSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L10_TuiM84 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGSSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L10_Abe_82 LLRDYFEEYQIAPAVRVLTKAVGKKLGKKFGNSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L51_Tahi18 LLRDYFEEYQIAPAVRVLTKAVGKKMGKKFGNSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L85_Kilo25 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGKSKYLYVLFPYGPGKQGCKFAGLPKPTGCV
L85_Abe320 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGKSKYLYVLFPYGPGKQGCKFAGLPKPTGCV
L85_Tahi20 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGKSKYLYVLFPYGPGKQGCKFAGLPKPTGCV
L10_Mari13 LLRDYFEEYQIAPAVRVLTKAVGKKLGKDKGSSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L51_Abe_12 LLRDYFEEYQIAPAVRVLTKAVGKKMGKKFGNSKYLYTLFPYGPGKQGCKFAGLPKPTGCV
L87_Kilo11 FLRNYYDEYQVAPAVRVLTKQIKKTMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Abe_64 FLRNYYDEYQVAPAVRVLTKQIKKTMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Tahi23 FLRNYYDEYQVAPAVRVLTKQIKKTMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Mari26 FLRNYYDEYQVAPAVRVLTKQIKKTMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L10_Kilo09 FLRDYFEEYQIAPAVRVLTKAIAKKMGKDKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Abe_15 FLRDYYEEYQVAPAVRVLTKQIKKRMGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Kilo27 FLRDYYEEYQVAPAVRVLTKQIKKRMGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Mari95 FLRDYYEEYQVAPAVRVLTKQIKKRMGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
L87_Tahi65 FLRDYYEEYQVAPAVRVLTKQIKKRMGKEKGNSKYLYSLFPYGPGKQGCRFAGLPKPTGCV
rdsrC2
group
CandidaRM1 IARKYFEENSSVPPIRTFAKVIGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
CandidaVO1 TARKYFEENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
EndosymbC1 AAREYFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQP-----
GB-1rdsrC1 SAREYFEENSSVPPIRTFAKFVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Abe_14 AAREYFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Abe_21 AAREYFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Kilo93 AAREYFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Tahi54 AAREYFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Tahi22 AAREFFAENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L10_Mari14 KAREYFDENSSVPPIRTFSKYVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L51_Tahi15 KAREYFDENSSVPPIRKFSKYVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
GB-2rdsrC SAKEYFEENSSVPPIRTFAKYVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L51_Abe_20 KAREYFEDNSSVPPIRTFAKFVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L218_AbeC1 KAREYFEENSSVPPIRTFAKYVGI-------DKGKLFKEWLTGPMKPITKYGGMPQPTGCV
L218_KilC1 KAREYFEENSSVPPIRTFAKYVGI-------DKGKLFKEWLTGPMKPITKYGGMPQPTGCV
L218_TahC1 KAREYFEENSSVPPIRTFAKYVGI-------DKGKLFKEWLTGPMKPITKYGGMPQPTGCV
L77_Abe_55 SAREYFDENSSVPPIRTFAKVVGI-------DKKVLFKEWLTGPMKPITKYGGMPQPTGCV
L77_Kilo13 SAREYFDENSSVPPIRTFAKVVGI-------DKKVLFKEWLTGPMKPITKYGGMPQPTGCV
L85_Abe_19 AARAYFDENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGLPQPTGCV
L85_Kilo15 AARAYFDENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGLPQPTGCV
L85_Kilo01 TAREIYANEGSVPPIRKFSKAVGI-------DKKILFKEWLTGPMKPITKYGGLPQPTGCV
L85_Tahi18 TAREIYANEGSVPPIRKFSKAVGI-------DKKILFKEWLTGPMKPITKYGGLPQPTGCV
L87_Abe_72 AARDYFEENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L87_Kilo30 AARDYFEENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L87_Mari52 AARDYFEENSSVPPIRTFAKVVGI-------DKKILFKEWLTGPMKPITKYGGMPQPTGCV
L87_Abe_05 AARDYFSENSSVPPIRTFAKVVGI-------DKKTLFKEWLTGPMKPITKYGGMPQPTGCV
L87_Kilo64 AARDYFSENSSVPPIRTFAKVVGI-------DKKTLFKEWLTGPMKPITKYGGMPQPTGCV
L87_Mari20 AARDYFSENSSVPPIRTFAKVVGI-------DKKTLFKEWLTGPMKPITKYGGMPQPTGCV
L87_Tahi81 AARDYFSENSSVPPIRTFAKVVGI-------DKKTLFKEWLTGPMKPITKYGGMPQPTGCV
rdsrC1
group

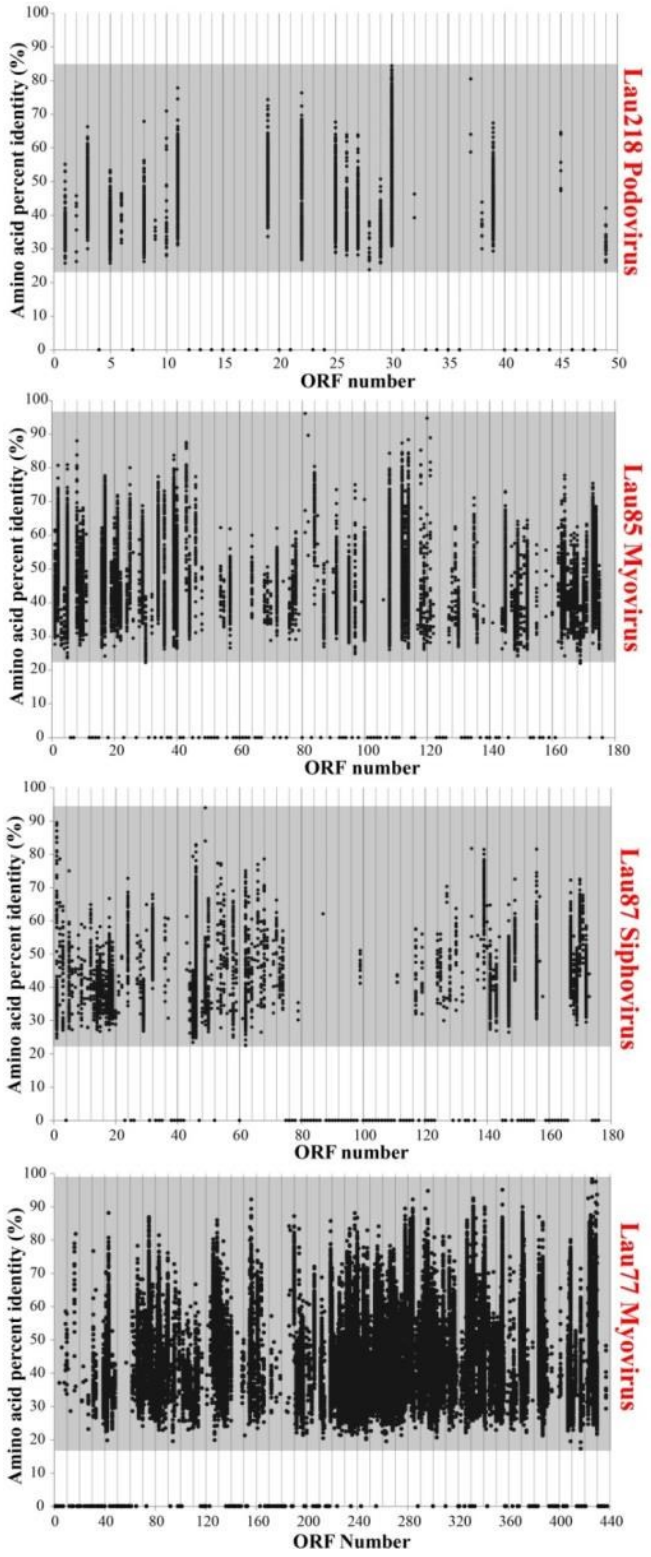
Supplementary Figure 7. Sequence alignment of SUP05 and phage *rdsrC* with *Allochromatium vinosum* DSM180 *rdsrC* and *Desulfovibro vulgaris* 'Hildenborough' *dsrC*. Genes are color-coded with SUP05-derived sequences in purple and phage-derived sequences in brown. Strictly conserved residues are highlighted in colors as follows: (i) Yellow – Highly conserved cysteine in all *dsrC* proteins. (ii) Green – Residues with 100% identity across the alignment. (iii) Cyan – Conserved residues in C-terminal of only *rdsrC2* group. Dark blue box indicates the location of the C-terminal arm.

Desulfovib: *Desulfovibrio vulgaris* str. Hildenborough (YP_011988)
AllochDsrC: *Allochromatium vinosum* DSM180 (ADC62195)
Allochroma: *Allochromatium vinosum* DSM 180 (ADC61307)
uncEEZ9537: uncultured SUP05 cluster bacterium (EEZ79537)
CandidaRM2: Candidatus *Ruthia magnifica* (YP_904052)
CandidaVO2: Candidatus *Vesicomyosocius okutanii* HA(YP_001219620)
EndosymbC2: endosymbiont of *Bathymodiolus* sp. (WP_010645598)
GB-2rdsrC2: SUP05 GB-2 (2062112236) rdsrC2
L10_TuiM63: Lau10_TuiMalila_10051563
L10_Tahi85: Lau10_TahiMoana_10042585
GB-1rdsrC2: SUP05_GB-1 (2062113608) rdsrC2
L10_TuiM84: Lau10_TuiMalila_10051184
L10_Abe_82: Lau10_Abe_100133582
L51_Tahi18: Lau51_TahiMoana_100090918
L85_Kilo25: Lau85_KiloMoana_1001303025
L85_Abe320: Lau85_Abe_1000083320
L85_Tahi20: Lau85_TahiMoana_100106520
L10_Mari13: Lau10_Mariner_10194713
L51_Abe_12: Lau51_Abe_1000037012
L87_Kilo11: Lau87_KiloMoana_1000037211
L87_Abe_64: Lau87_Abe_1000020064
L87_Tahi23: Lau87_TahiMoana_100007023
L87_Mari26: Lau87_Mariner_100028926
L10_Kilo09: Lau10_KiloMoana_100050809
L87_Abe_15: Lau87_Abe_100027315
L87_Kilo27: Lau87_KiloMoana_1000415027
L87_Mari95: Lau87_Mariner_10010695
L87_Tahi65: Lau87_TahiMoana_10030365
CandidaRM1: Candidatus *Ruthia magnifica* (YP_904067)
CandidaVO1: Candidatus *Vesicomyosocius okutanii* HA (YP_001219635)
EndosymbC1: endosymbiont of *Bathymodiolus* sp. (WP_010645576)
GB-1rdsrC1: SUP05 GB-1 rdsrC1 (2062399953)
L10_Abe_14: Lau10_Abe_100098014
L10_Abe_21: Lau10_Abe_100146021
L10_Kilo93: Lau10_KiloMoana_100089893
L10_Tahi54: Lau10_TahiMoana_10050554
L10_Tahi22: Lau10_TahiMoana_10094822
L10_Mari14: Lau10_Mariner_100363914
L51_Tahi15: Lau51_TahiMoana_100055515

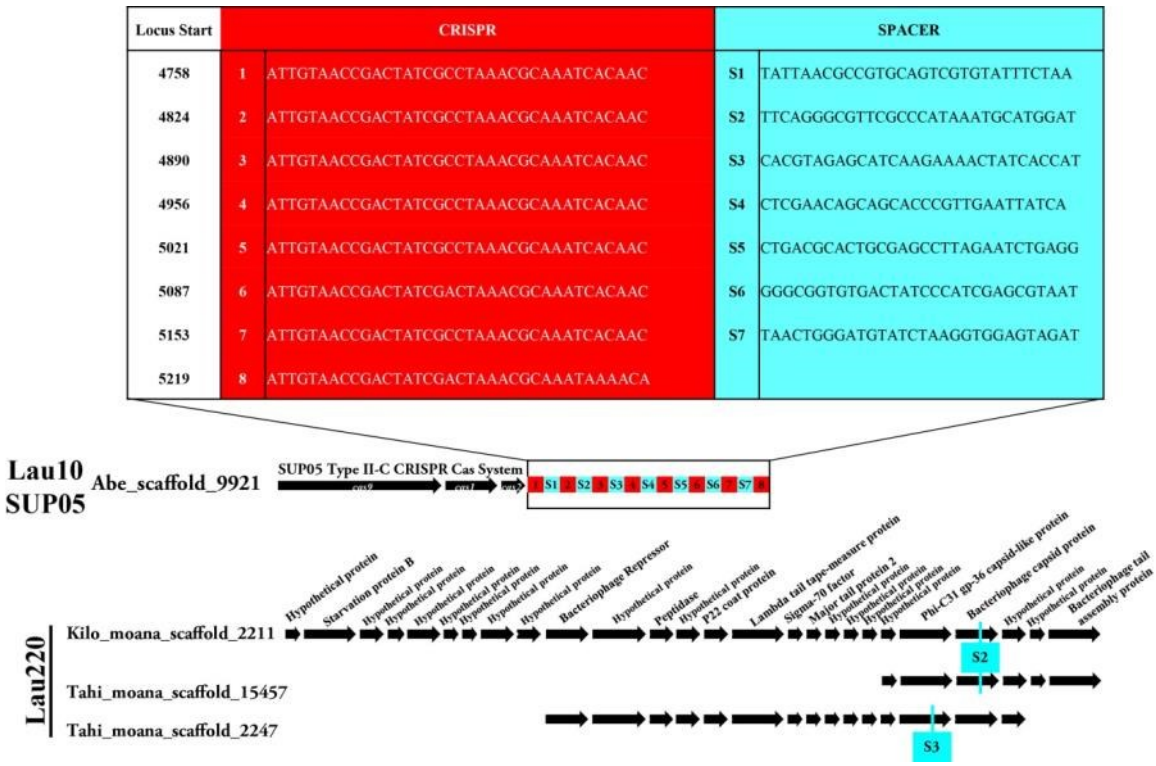
GB-2rdsrC1: SUP05 GB-2 rdsrC1 (2062169932)
L51_Abe_20: Lau51_Abe_1000005620
L218_AbeC1: Lau218_Abe
L218_KilC1: Lau218_KiloMoana
L218_TahC1: Lau218_TahiMoana
L77_Abe_55: Lau77_Abe_1000028055
L77_Kilo13: Lau77_KiloMoana_1000023813
L85_Abe_19: Lau85_Abe_100171319
L85_Kilo15: Lau85_KiloMoana_1000577115
L85_Kilo01: Lau85_KiloMoana_100130301
L85_Tahi18: Lau85_TahiMoana_100411318
L87_Abe_72: Lau87_Abe_100049072
L87_Kilo30: Lau87_KiloMoana_1000539530
L87_Mari52: Lau87_Mariner_10008452
L87_Abe_05: Lau87_Abe_100002005
L87_Kilo64: Lau87_KiloMoana_100050664
L87_Mari20: Lau87_Mariner_100018220
L87_Tahi81: Lau87_TahiMoana_100007081


A
Mariner 2.5°C
Available energy
20.41 J/kg p.f.
2160 J/kg v.f.
79.8 %
Mariner 5°C
Available energy
3.41 J/kg p.f.
2161 J/kg v.f.
80.5 %
$H_2 : O_2$
$NH_4^+ : O_2$
$CH_4 : O_2$
$S_o : O_2$
$Mn^{2+} : O_2$
$S_2O_3^{2-} : O_2$
$H_2 : Fe^{3+}$
$HS^- : O_2$
$H_2 : NO_3^- - N_2$
$H_2 : NO_3^-$
$H_2 : SO_4^{2-}$
$H_2 : HCO_3^-$
$HS^- : NO_3^-$
Par $Fe^{2+} : O_2$
$Fe^{2+} : NO_3^-$
$S_2O_3^{2-} : NO_3^-$
$Fe^{2+} : O_2$
$H_2 : N_2$
$S_o : NO_3^-$
Par $HS^- : O_2$
B
FeS
21
3
500 µm
C
spot 22
spot 3
spot 21
D
Background Subtracted Intensity
2θ (degrees)
Mariner 0.5m spot3
Mariner 0.5m spot 21
0
200
400
600
800
1000
1200
1400
1600
1800
5
10
15
20
25
30
35

Supplementary Figure 8. **A**. **Modeled free energies of catabolic reactions as a percentage of total available free energy in the Mariner hydrothermal plume at 2.5 and 5°C.** Total available free energy in the plume is normalized per kilogram plume fluid (p.f.) and per kilogram vent fluid (v.f.).**B. Distribution of iron (displayed in red) and sulfur (displayed in green) in particles collected at 0.5 m above the Mariner vent.** Locations where elemental sulfur was detected by micro-probe X-ray diffraction measurements are indicated as spots 3 and 21. **C. Radially integrated diffractograms with elemental sulfur peaks annotated (*) at 22.0, 25.7, and 34.1 degrees 2-theta.** Elemental sulfur was detected in particle aggregates with other crystalline phases, such as pyrite, as indicated by additional non-elemental sulfur peaks. **D. X-ray diffraction patterns for spots 3 and 21 of the Mariner 0.5 m sample are displayed with a background pattern (spot 22).** Elemental sulfur was detected in 2 of 27 diffractograms for this sample. The background pattern (spot 22) is subtracted from each sample pattern after radial integration and conversion of the data to intensity versus degrees 2-theta.



Supplementary Figure 9. Protein Blast (blastp) of ELSC viral genomes against the Pacific Ocean Virome (POV) protein clusters. Circles indicate best hits for a particular ORF. Grey boxes indicate range of protein identities.



Supplementary Figure 10. CRISPR-Cas and spacer loci identified on ELSC Lau10 SUP05 and Lau220 viral genomes.

Supplementary Table 1: Sampling details

Site/Vent	Sample ID	Sample type	Date (DD/MM/YYYY)	Latitude/Longitude	Depth (m)	Filter size(µm)	No. of sequence reads
Abe/A1	TN235-J2426-20	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	1960	0.8	168325584
Abe	TN236-J2449-2	Near bottom background	04/07/2009	S 22 45.677706 W 176 11.369574	2155	0.8	169488288
Abe/A1	TN235-J2426-7	Rising Plume	25/05/2009	S 20 45.672883 W 176 11.434418	2159	0.8	181094744
Kilo Moana/KM1	TN236-J2436-16	Rising Plume	15/06/2009	S 20 3.229502 W 176 8.015363	2605	0.8	174530426
Kilo Moana/KM4	TN235-J2424-20	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2440	0.8	157276514
Kilo Moana/KM4	TN235-J2424-8	Rising Plume	22/05/2009	S 20 3.234200 W 176 8.008000	2639	0.8	118751402
Kilo Moana/KM1	TN236-CTD-KM-IP2	Neutrally buoyant plume	13/06/2009	S 20 3.246489 W 176 8.011308	2315	0.2	188964668
Mariner/MA3	TN236-J2440-18	Rising Plume	20/06/2009	S 22 10.818293 W 176 36.086423	1890	0.8	185135248
Mariner	TN236-CTD-Mar-BP1	Below plume background	19/06/2009	S 20 10.8035165 W 176 36.074496	1785	0.2	154673072
Tahi Moana	TN236-J2445-20.2	Above plume background	29/06/2009	S 20 40.927843 W 176 11.001806	~1300	0.2	181482188
Tahi Moana 1a/SP2	TN236-J2450-9	Rising Plume	05/07/2009	S 20 40.894100 W 176 10.940463	2229	0.8	186087990
Tui Malila/TM1	TN236-J2447-10	Rising Plume	01/07/2009	S 21 59.401181 W 176 34.124651	1919	0.8	187067650
Guaymas Basin	GOC11-1#2	Neutrally buoyant plume	11/7/2004	N 27°30.95 W 111 25.5	1993	0.2	171620910

Supplementary Table 2: Details of Viral Bins

Bin	Taxonomy	Strain	No. of Contigs	Total size(bp)	ORFs	No.of *rdsrA*	No. of *rdsrC*	Avg. nucleotide identity to strain-Kilo Moana
Lau218	*Viruses;dsDNA viruses;Caudovirales;Podoviridae*	Kilo Moana	1	39,311	48	1	1	x
		Abe	1	39,425	50	1	1	99.63
		Tahi Moana	1	38,756	51	1	1	99.1
Lau85	*Viruses;dsDNA viruses;Caudovirales;Myoviridae*	Kilo Moana	2	136,837	176	2	2	x
		Abe	5	137,686	177	2	2	99.38
		Tahi Moana	2	148,281	187	2	2	99.57
		Mariner	9	62,682	-	2	1	98.3
		Guaymas	18	117,847	184	1	1	94.72
Lau87	*Viruses;dsDNA viruses;Caudovirales;Siphoviridae*	Kilo Moana	3	119,296	175	1	1	x
		Abe	3	106,741	159	1	1	98.96
		Mariner	2	94,717	144	1	1	96.51
		Tahi Moana	2	118,938	175	1	1	97.63
Lau77	*Viruses;dsDNA viruses;Caudovirales;Myoviridae*	Kilo Moana	6	333,285	438	1	1	x
		Abe	14	294,153	401	1	1	98.86
		Tahi Moana	24	252,462	340	1	1	99.43
Lau220	*Viruses;dsDNA viruses**	Kilo Moana	10	98,420	140	0	0	x
		Abe	14	96,484	152	0	0	85.56
		Tahi Moana	11	87,061	115	0	0	83.08

*Lau220 was not classified due to lack of *terL* genes and due to lack of syntenous genomes.

Supplementary Table 3. Phage Annotation Tables**.**

A. Genome: uncultured Podovirus Lau218 str. KiloMoana

Id	Start	Stop	Strand	Annotation	Blast organism	Blastp id	e-value
Lau218_KiloMoana_1	1	1719	+	maturase, NTPase containing	Azorhizobium caulinodans ORS 571	47.01	7E-166
Lau218_KiloMoana_2	2237	2377	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_3	3105	3332	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_4	3381	3953	+	N6 adenine-specific DNA methyltransferase, N12 class	Sulfurimonas gotlandica	47.22	6E-42
Lau218_KiloMoana_5	3998	4195	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_6	4233	4388	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_7	4403	4768	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_8	4797	4991	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_9	5079	5249	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_10	5242	5442	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_11	5439	7934	-	DNA-directed RNA polymerase, phage-type	Pseudomonas sp. TJI-51	26.74	5E-51
Lau218_KiloMoana_12	8487	9146	+	phage ssDNA-binding protein	Stenotrophomonas maltophilia	30.56	2E-14
Lau218_KiloMoana_13	9170	10840	+	DNA primase/helicase, Bacteriophage T7-like	Azorhizobium caulinodans ORS 571	42.7	3E-152
Lau218_KiloMoana_14	10882	11259	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_15	11252	12193	+	DNA-directed DNA polymerase	Stenotrophomonas maltophilia	35.61	1E-38
Lau218_KiloMoana_16	12193	12585	+	endodeoxyribonuclease	Enterobacter phage EcP1	42.06	1E-15
Lau218_KiloMoana_17	12592	12786	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_18	12797	13543	+	exonuclease	Pseudomonas putida KT2440	41.09	6E-42
Lau218_KiloMoana_19	13513	13740	+	IQ motif-containing protein	Bizionia argentinensis	47.76	9E-11
Lau218_KiloMoana_20	13730	14158	+	DNA N-6-adenine-methyltransferase	Clostridium botulinum	38.19	1E-22
Lau218_KiloMoana_21	14155	15888	+	ribonucleotide reductase, large subunit	Pelagibacter phage HTVC019P	53.2	0
Lau218_KiloMoana_22	15984	16241	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_23	16412	16681	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_24	16686	17003	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_25	17014	17112	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_26	17072	17284	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_27	17297	17587	+	rdsrC sulfur transfer protein	Candidatus Ruthia magnifica str. Cm	81.25	7E-53
Lau218_KiloMoana_28	17603	17878	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau218_KiloMoana_29	17875	18498	+	Thymidylate synthase ThyX	Moraxella catarrhalis	55.39	3E-75
Lau218_KiloMoana_30	18498	18677	+	FmdB family transcriptional regulator	#N/A	#N/A	#N/A
Lau218_KiloMoana_31	18674	19624	+	N6 adenine-specific DNA methyltransferase, N12 class	Eremococcus coleocola	28.52	2E-17
Lau218_KiloMoana_32	19712	20689	+	Ribonucleotide reductase, small subunit	Pelagibacter phage HTVC019P	54.78	5E-128
Lau218_KiloMoana_33	20711	20875	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_34	20879	21082	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_35	21093	22580	+	phage head-to-tail joining protein, podovirus-type	Azorhizobium caulinodans ORS 571	46.45	4E-132
Lau218_KiloMoana_36	22650	23432	+	phage capsid assembly protein, T7-like	Pelagibacter phage HTVC019P	37.34	1E-33
Lau218_KiloMoana_37	23544	24422	+	phage capsid protein	Azorhizobium caulinodans ORS 571	38.77	3E-53
Lau218_KiloMoana_38	24496	25068	+	phage tail protein	Pseudomonas sp. TJI-51	30.07	2E-12
Lau218_KiloMoana_39	25079	27280	+	phage tail protein	Pseudomonas putida KT2440	29.68	3E-59
Lau218_KiloMoana_40	27277	30195	+	phage fiber protein	Marine cyanobacterial siphovirus PSS2	48.18	3E-15
Lau218_KiloMoana_41	30195	30503	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_42	30500	30853	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_43	30850	31110	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_44	31113	31655	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_45	31627	33930	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_46	34026	37550	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau218_KiloMoana_47	37563	38870	+	rdsrA Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	93.79	0
Lau218_KiloMoana_48	38961	39188	+	phage protein	Synechococcus phage S-RIP2;Cyanophage KBS-P-1A	39.66	0.000003

B. Genome: uncultured Siphovirus Lau87 str. KiloMoana

Id	Start	Stop	Strand	Annotation	Blast organism	Blastp id	e-value
Lau87_KiloMoana_Scaffold_1_1	1	1302	+	GTA-like phage terminase, large subunit	Pectobacterium phage My1	49.19	2E-141
Lau87_KiloMoana_Scaffold_1_2	1430	1801	+	hypothetical phage protein	Vibrio phage pVp-1	44.44	1E-29
Lau87_KiloMoana_Scaffold_1_3	1798	2991	+	GTA-like phage portal protein	Enterobacteria phage EPS7	44.58	3E-104
Lau87_KiloMoana_Scaffold_1_4	2994	3320	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_5	3345	5288	+	GTA-like phage protease and capsid protein	Verrucomicrobia bacterium SCGC AAA164-A08	40.58	5E-82
Lau87_KiloMoana_Scaffold_1_6	5362	5949	+	Phage gp6-like head-tail connector protein	Verrucomicrobia bacterium SCGC AAA164-A08	35.68	2E-31

Lau87_KiloMoana_Scaffold_1_7	5949	7433	+	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	37.96	7E-13
Lau87_KiloMoana_Scaffold_1_8	7430	7870	+	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	43.88	2E-28
Lau87_KiloMoana_Scaffold_1_9	7944	9101	+	phage major tail protein	Vibrio phage pVp-1	31.45	6E-51
Lau87_KiloMoana_Scaffold_1_10	9304	9747	+	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	54.23	1E-47
Lau87_KiloMoana_Scaffold_1_11	9798	10109	+	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	36.14	3E-08
Lau87_KiloMoana_Scaffold_1_12	10125	21491	+	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	42.5	2E-48
Lau87_KiloMoana_Scaffold_1_13	21494	21844	+	lambda-like minor tail protein, GpM	Thiomonas sp. 3As	37.5	1E-15
Lau87_KiloMoana_Scaffold_1_14	21841	22665	+	lambda-like minor tail protein, GpL	Burkholderia graminis	38.91	7E-56
Lau87_KiloMoana_Scaffold_1_15	22667	23380	+	JAB and Endopeptidase multi-domain protein	Candidatus Glomeribacter gigasporarum	43.59	6E-71
Lau87_KiloMoana_Scaffold_1_16	23374	23979	+	hypothetical phage protein	Rahnella aquatilis HX2	32.06	1E-21
Lau87_KiloMoana_Scaffold_1_17	23976	24569	+	phage tail assembly protein, lambda-like	Xanthomonas campestris pv. campestris str. B100	39.6	3E-37
Lau87_KiloMoana_Scaffold_1_18	24569	30307	+	fibronectin cell adhesion tail protein, GTA-like	Cupriavidus sp. HPC(L)	45.7	1E-132
Lau87_KiloMoana_Scaffold_1_19	30325	31443	+	chaperone of phage endosialidase	Micromonas pusilla virus 12T	46.39	2E-12
Lau87_KiloMoana_Scaffold_1_20	31443	31787	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_21	31784	32164	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_22	32273	33112	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_23	33128	33337	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_24	33393	35378	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_25	35648	35890	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_26	35894	37180	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_1_27	37209	38531	+	hypothetical phage protein	#N/A	#N/A	#N/A
NEXT CONTIG							
Lau87_KiloMoana_Scaffold_3_1	2	646	+	hypothetical phage protein	Pelagibacter phage HTVC011P	32.72	6E-17
Lau87_KiloMoana_Scaffold_3_2	656	976	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_3	1004	1201	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_4	1203	1733	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_5	1743	1952	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_6	1984	2181	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_7	2183	2959	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_8	2972	3283	+	hypothetical phage protein	Arcobacter nitrofigilis DSM 7299	56.52	3E-08
Lau87_KiloMoana_Scaffold_3_9	3346	4212	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_10	4216	4581	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_11	4591	5322	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_12	5304	5843	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau87_KiloMoana_Scaffold_3_13	5830	6426	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_14	6413	7324	+	Elongator protein 3/MiaB/NifB domain-containing protein	Ignicoccus hospitalis KIN4/I	28.96	1E-16
Lau87_KiloMoana_Scaffold_3_15	7321	7653	+	hypothetical phage protein	gamma proteobacterium HTCC2207	51.38	7E-32
Lau87_KiloMoana_Scaffold_3_16	7650	9227	+	Tryptophan halogenase	Streptomyces rimosus	33.13	4E-71
Lau87_KiloMoana_Scaffold_3_17	9194	10348	-	Ribonucleotide reductase, small subunit	Providencia rettgeri	36.04	8E-63
Lau87_KiloMoana_Scaffold_3_18	10348	11154	-	thymidylate synthase, thyX	Gluconacetobacter europaeus	56.41	8E-66
Lau87_KiloMoana_Scaffold_3_19	11135	11290	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_20	11287	11868	-	conserved domain CHP02466, unknown function	Prochlorococcus phage P-SSM2	25.63	2E-11
Lau87_KiloMoana_Scaffold_3_21	11883	14174	-	Ribonucleotide reductase, class I, alpha subunit	Haemophilus parainfluenzae	40.13	0
Lau87_KiloMoana_Scaffold_3_22	14216	15049	-	Exonuclease	Verrucomicrobia bacterium SCGC AAA164-A08	53.85	4E-89
Lau87_KiloMoana_Scaffold_3_23	15046	15489	-	tRNA endonuclease	Pectobacterium phage My1	53.85	1E-36
Lau87_KiloMoana_Scaffold_3_24	15491	15943	-	thioredoxin domain-containing protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_25	15943	17118	-	NTP hydrolase (possible frame shift split)	Verrucomicrobia bacterium SCGC AAA164-A08	40.41	3E-74
Lau87_KiloMoana_Scaffold_3_26	17237	17692	-	NTP hydrolase	Verrucomicrobia bacterium SCGC AAA164-A08	68.67	4E-61
Lau87_KiloMoana_Scaffold_3_27	17661	18614	-	phosphoesterase domain	Verrucomicrobia bacterium SCGC AAA164-A08	53.31	2E-115
Lau87_KiloMoana_Scaffold_3_28	18684	19337	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	45.41	1E-47
Lau87_KiloMoana_Scaffold_3_29	19328	19702	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	34.62	2E-13
Lau87_KiloMoana_Scaffold_3_30	19752	21041	-	Helicase	Verrucomicrobia bacterium SCGC AAA164-A08	58.96	0
Lau87_KiloMoana_Scaffold_3_31	21059	22030	-	ABC transporter	alpha proteobacterium HIMB114	42.86	5E-80
Lau87_KiloMoana_Scaffold_3_32	22192	22368	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_33	22320	22853	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	29.81	0.000001
Lau87_KiloMoana_Scaffold_3_34	22804	25065	-	DNA polymerase	Verrucomicrobia bacterium SCGC AAA164-A08	53.78	0
Lau87_KiloMoana_Scaffold_3_35	25133	25981	-	DNA primase	Verrucomicrobia bacterium SCGC AAA164-A08	39.49	2E-73
Lau87_KiloMoana_Scaffold_3_36	25988	27376	-	DNA helicase	Vibrio phage pVp-1	38.32	2E-95
Lau87_KiloMoana_Scaffold_3_37	27376	27675	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_38	27744	28484	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	58.8	2E-92
Lau87_KiloMoana_Scaffold_3_39	28511	29272	-	NAD-dependent DNA ligase, subunit B	Vibrio phage pVp-1	44.22	3E-58
Lau87_KiloMoana_Scaffold_3_40	29361	30293	-	NAD-dependent DNA ligase, subunit A	Verrucomicrobia bacterium SCGC AAA164-A08	58.2	4E-112
Lau87_KiloMoana_Scaffold_3_41	30286	30582	-	ssDNA-binding transcriptional regulator	Verrucomicrobia bacterium SCGC AAA164-A08	48.31	1E-19
Lau87_KiloMoana_Scaffold_3_42	30590	30829	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	44	4E-17
Lau87_KiloMoana_Scaffold_3_43	30849	31508	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	37.17	3E-32
Lau87_KiloMoana_Scaffold_3_44	31512	31868	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	52.99	1E-30
Lau87_KiloMoana_Scaffold_3_45	31996	32463	-	hypothetical phage protein	Leptonema illini	40.28	0.0000004

Lau87_KiloMoana_Scaffold_3_46	32637	32858	-	hypothetical phage protein	alpha proteobacterium IMCC14465	45.21	3E-13
Lau87_KiloMoana_Scaffold_3_47	33044	33229	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_48	33226	33387	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_49	33412	33522	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_50	33577	33990	-	hypothetical phage protein	Candidatus Vesicomyosocius okutanii HA	46.2	2E-29
Lau87_KiloMoana_Scaffold_3_51	34167	35042	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_52	35151	35420	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_53	35676	35864	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_54	35864	36151	-	rdsrC sulfur transfer protein	Candidatus Ruthia magnifica str. Cm (Calyptogena magnifica)	86.17	9E-55
Lau87_KiloMoana_Scaffold_3_55	36155	36460	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_56	36460	36747	-	hypothetical phage protein	Candidatus Vesicomyosocius okutanii HA	43.08	0.000006
Lau87_KiloMoana_Scaffold_3_57	36845	37651	-	unknown function, DUF 3050	Magnetococcus marinus MC-1	42.26	9E-65
Lau87_KiloMoana_Scaffold_3_58	37998	38228	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_59	38371	38631	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_60	38615	38908	-	DUF 955; possible HGT regulation protein	gamma proteobacterium SCGC AAA001-B15	42.22	4E-08
Lau87_KiloMoana_Scaffold_3_61	38944	39117	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_62	39213	39569	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_63	39547	39741	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_64	39814	40146	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_65	40796	41131	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_66	41556	41912	-	hypothetical phage protein	Psychromonas ingrahamii 37	36.96	5E-13
Lau87_KiloMoana_Scaffold_3_67	41875	42093	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_68	42090	42245	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_69	42215	42421	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_70	42418	42603	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_71	42578	42799	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_72	42790	43026	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_73	43073	43222	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_74	43219	43344	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_75	43346	43510	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_76	43507	43638	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_77	43601	43747	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_78	43737	43964	-	hypothetical phage protein	#N/A	#N/A	#N/A

Lau87_KiloMoana_Scaffold_3_79	43952	44215	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_80	44229	44462	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_81	44610	44810	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_82	44897	45271	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_83	45271	45468	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_84	45468	45638	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_85	45638	45862	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_86	45872	46108	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_87	46110	46265	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_88	46362	46649	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_89	46925	47257	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	45.63	8E-19
Lau87_KiloMoana_Scaffold_3_90	47478	47621	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_91	47618	47959	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	52.38	4E-10
Lau87_KiloMoana_Scaffold_3_92	48129	48296	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_93	48296	48619	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_94	48637	48756	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_95	48756	48935	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_96	49393	49713	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	36.96	0.0000002
Lau87_KiloMoana_Scaffold_3_97	50179	50496	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	42.39	8E-13
Lau87_KiloMoana_Scaffold_3_98	50679	51353	-	Ribonuclease H	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_99	51420	51671	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_100	51947	52270	-	hypothetical phage protein	Verrucomicrobia bacterium SCGC AAA164-A08	40.22	2E-12
Lau87_KiloMoana_Scaffold_3_101	52658	52921	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_102	52995	53453	+	hypothetical phage protein	Enterobacteria phage T5	40.82	1E-21
Lau87_KiloMoana_Scaffold_3_103	53450	53650	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_104	53634	54083	+	hypothetical phage protein	Verrucomicrobiae bacterium DG1235	27.03	0.0000004
Lau87_KiloMoana_Scaffold_3_105	54094	54333	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_106	54273	54647	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_107	54723	56027	+	rdsrA Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	90.28	0
Lau87_KiloMoana_Scaffold_3_108	56024	56161	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_109	56161	56484	+	rdsrC sulfur transfer protein	Candidatus Vesicomyosocius okutanii HA	81.31	2E-60
Lau87_KiloMoana_Scaffold_3_110	56542	56763	+	TusA-like protein	Candidatus Vesicomyosocius okutanii HA	82.86	7E-33
Lau87_KiloMoana_Scaffold_3_111	56765	57151	+	NifU-like FeS cluster assembly protein	Neisseria mucosa	79.51	5E-68

Lau87_KiloMoana_Scaffold_3_112	57136	57363	+	4Fe-4S ferredoxin-type, iron-sulpur binding protein	endosymbiont of Bathymodiolus sp.	84.42	1E-35
Lau87_KiloMoana_Scaffold_3_113	57360	57704	+	Fe-S biogenesis and delivery protein	Candidatus Ruthia magnifica str. Cm (Calyptogena magnifica)	79.28	1E-57
Lau87_KiloMoana_Scaffold_3_114	57704	58048	+	Glutaredoxin	Candidatus Ruthia magnifica str. Cm (Calyptogena magnifica)	77.06	1E-57
Lau87_KiloMoana_Scaffold_3_115	58050	58370	+	Fe-S biogenesis protein	Vibrio campbellii ATCC BAA-1116	64.76	8E-42
Lau87_KiloMoana_Scaffold_3_116	58377	58502	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_117	58570	58797	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_118	58881	59120	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_119	59151	59909	+	PhoH-like protein	Simiduia agarivorans SA1 = DSM 21679	56.07	2E-76
Lau87_KiloMoana_Scaffold_3_120	59899	60069	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_121	60072	60545	+	unknown function, DUF 1353	Pelagibacter phage HTVC008M	50.94	4E-49
Lau87_KiloMoana_Scaffold_3_122	60834	61364	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_123	61366	61707	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_3_124	61743	61934	+	hypothetical phage protein	#N/A	#N/A	#N/A
NEXT CONTIG							
Lau87_KiloMoana_Scaffold_2_1	2	754	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_2	766	1683	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_3	1683	2147	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_4	2197	3084	+	hypothetical phage protein	Bdellovibrio bacteriovorus str. Tiberius	34.96	0.0000002
Lau87_KiloMoana_Scaffold_2_5	3095	4072	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_6	4086	4574	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_7	4624	5175	+	triple helix repeat domain-containing protein	Lactobacillus vaginalis	32.06	5E-10
Lau87_KiloMoana_Scaffold_2_8	5185	5385	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_9	5492	6319	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_10	6312	6512	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_11	6526	8052	+	4Fe4S cluster-containing radical SAM	Clostridium sp. BNL1100	26.44	1E-15
Lau87_KiloMoana_Scaffold_2_12	8053	9090	+	Aldolase-type TIM barrel containing protein	Clostridium sp. BNL1100	23.55	0.000002
Lau87_KiloMoana_Scaffold_2_13	9077	10147	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_14	10129	11259	+	4Fe4S cluster-containing radical SAM	Magnetospirillum sp. SO-1	28.69	7E-29
Lau87_KiloMoana_Scaffold_2_15	11260	11730	+	hypothetical phage protein	Commensalibacter intestini	41.23	4E-15
Lau87_KiloMoana_Scaffold_2_16	11853	14393	+	cell surface adhesion-related domains	Geobacter sp. M21	34.36	3E-08
Lau87_KiloMoana_Scaffold_2_17	14390	14794	+	hypothetical phage protein	pseudomallei group	29.23	3E-09
Lau87_KiloMoana_Scaffold_2_18	14794	15381	+	triple helix repeat domain-containing protein	Bacillus cereus	50	1E-08
Lau87_KiloMoana_Scaffold_2_19	15383	16348	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau87_KiloMoana_Scaffold_2_20	16357	16761	+	hypothetical phage protein	Prochlorococcus phage P-RSM4	33.61	9E-13
Lau87_KiloMoana_Scaffold_2_21	16780	17187	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_22	17174	17788	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_23	17788	18504	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau87_KiloMoana_Scaffold_2_24	18597	18776	+	hypothetical phage protein	#N/A	#N/A	#N/A

C. Genome: uncultured Myovirus Lau85 str. TahiMoana

Id	Start	Stop	Strand	Annotation	Blast organism	Blastp id	e-value
Lau85_TahiMoana_Scaffold_1_1	2	2461	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_2	2660	3391	-	phage base plate protein, T4-like	Synechococcus phage S-CAM1	39.92	3E-47
Lau85_TahiMoana_Scaffold_1_3	3441	6860	-	tail tube protein	Synechococcus phage S-SM2	27.75	5E-10
Lau85_TahiMoana_Scaffold_1_4	6863	7372	-	hypothetical phage protein	Salmonella phage S16	45.18	1E-35
Lau85_TahiMoana_Scaffold_1_5	7374	7814	-	phage packaging protein, endonuclease domain	Prochlorococcus phage P-SSM2	51.39	1E-47
Lau85_TahiMoana_Scaffold_1_6	7869	8801	+	phage tail-tube assembly protein, T4-like	Synechococcus phage S-PM2	28.99	1E-16
Lau85_TahiMoana_Scaffold_1_7	8804	9304	+	phage baseplate wedge protein, T4-like	Aeromonas phage CC2	30.69	8E-15
Lau85_TahiMoana_Scaffold_1_8	9304	10659	+	hypothetical phage protein	Aeromonas phage CC2	25.43	1E-14
Lau85_TahiMoana_Scaffold_1_9	10669	11712	+	phage baseplate protein, T4-like	Vibrio phage KVP40	44.86	5E-72
Lau85_TahiMoana_Scaffold_1_10	11723	13684	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_11	13684	13974	+	PAAR domain-containing phage protein	Acidithiobacillus caldus	51.61	3E-28
Lau85_TahiMoana_Scaffold_1_12	14001	14408	+	phage outter wedge baseplate T4-like protein, copper amine oxidase domain	Aeromonas phage phiAS5	42.31	2E-16
Lau85_TahiMoana_Scaffold_1_13	14418	16178	+	possible baseplate wedge	Pelagibacter phage HTVC008M	34.7	3E-98
Lau85_TahiMoana_Scaffold_1_14	16191	24545	+	Soluble quinoprotein glucose dehydrogenase	Leptospira interrogans	46.62	5E-32
Lau85_TahiMoana_Scaffold_1_15	24625	25638	+	phage baseplate structural protein, T4-like	Aeromonas phage Aeh1	29.89	3E-37
Lau85_TahiMoana_Scaffold_1_16	25713	28718	+	hypothetical phage protein	Salmonella phage SKML-39	27.71	6E-86
Lau85_TahiMoana_Scaffold_1_17	28936	31113	+	hypothetical phage protein	Prochlorococcus phage P-RSM4	40.91	1E-49
Lau85_TahiMoana_Scaffold_1_18	31124	31327	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_19	31340	31543	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_20	31563	39530	+	hypothetical phage protein	Polaribacter irgensii	32.33	2E-09
Lau85_TahiMoana_Scaffold_1_21	39631	43200	+	Cupredoxin	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_22	43247	44227	+	hypothetical phage protein	Synechococcus phage S-SSM7	41.98	3E-10
Lau85_TahiMoana_Scaffold_1_23	44246	45046	+	hypothetical phage protein	Synechococcus phage S-SSM7	31.91	3E-13
Lau85_TahiMoana_Scaffold_1_24	45072	46046	+	hypothetical phage protein	Synechococcus phage S-SSM7	41.98	3E-10

Lau85_TahiMoana_Scaffold_1_25	46060	46788	+	hypothetical phage protein	Synechococcus phage S-SSM7	28.21	6E-09
Lau85_TahiMoana_Scaffold_1_26	46926	47984	+	hypothetical phage protein	Synechococcus phage S-SSM7	54.9	7E-132
Lau85_TahiMoana_Scaffold_1_27	47974	49797	+	Carbamoyltransferase	Synechococcus phage S-SSM7	65.06	0
Lau85_TahiMoana_Scaffold_1_28	49794	51584	+	Carbamoyltransferase	Synechococcus phage S-SSM7	65.06	0
Lau85_TahiMoana_Scaffold_1_29	51581	52465	+	ADP-L-glycero-D-manno-heptose-6-epimerase	Escherichia hermannii	44.41	9E-73
Lau85_TahiMoana_Scaffold_1_30	52453	52806	-	RmlC-like cupin domain-containing protein	Synechococcus phage S-CAM1	40	1E-15
Lau85_TahiMoana_Scaffold_1_31	52796	54040	-	carbohydrate kinase PfkB with cytidyltransferase-like domain	Alcanivorax dieselolei B5	39.96	5E-102
Lau85_TahiMoana_Scaffold_1_32	54100	54975	+	phage neck protein	Pelagibacter phage HTVC008M	28.66	4E-39
Lau85_TahiMoana_Scaffold_1_33	54980	55591	+	phage neck protein, T4-like domain	Synechococcus phage S-SSM5	33.33	2E-22
Lau85_TahiMoana_Scaffold_1_34	55596	56411	+	phage tail sheath protein	Prochlorococcus phage P-SSM2	38.14	3E-45
Lau85_TahiMoana_Scaffold_1_35	56427	56768	+	terminase small subunit	Pelagibacter phage HTVC008M	66	2E-16
Lau85_TahiMoana_Scaffold_1_36	56765	56869	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_37	56882	57301	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_1_38	57320	58912	+	terminase, large subunit	Cyanophage P-RSM6	53.53	0
Lau85_TahiMoana_Scaffold_1_39	58993	60600	+	phage tail sheath	Pelagibacter phage HTVC008M	54.66	5E-106
Lau85_TahiMoana_Scaffold_1_40	60612	61109	+	phage tail tube, T4-like domain	Enterobacteria phage RB69	36.75	9E-27
Lau85_TahiMoana_Scaffold_1_41	61159	61302	+	hypothetical phage protein	#N/A	#N/A	#N/A
NEXT CONTIG							
Lau85_TahiMoana_Scaffold_2_1	162	1205	+	phage prohead core	Pelagibacter phage HTVC008M	40.24	8E-48
Lau85_TahiMoana_Scaffold_2_2	1234	2427	+	phage major capsid protein	Pelagibacter phage HTVC008M	51.64	2E-140
Lau85_TahiMoana_Scaffold_2_3	2528	3025	+	phage tail-tube protein, T4-like	Cyanophage Syn1	29.22	3E-13
Lau85_TahiMoana_Scaffold_2_4	3040	3450	+	UvsY, ssDNA binding protein	Pelagibacter phage HTVC008M	37.32	4E-20
Lau85_TahiMoana_Scaffold_2_5	3457	4911	+	Helicase	Synechococcus phage S-SKS1	52.02	3E-168
Lau85_TahiMoana_Scaffold_2_6	4908	5210	+	hypothetical phage protein	Synechococcus phage S-MbCM6	29.7	0.000005
Lau85_TahiMoana_Scaffold_2_7	5731	6342	+	sigma factor for late transcription	Pelagibacter phage HTVC008M	57.45	2E-50
Lau85_TahiMoana_Scaffold_2_8	6343	7383	+	phosphoesterase domain, possible endonuclease activity	Prochlorococcus phage P-HM2	40.74	5E-79
Lau85_TahiMoana_Scaffold_2_9	7380	9053	+	recombination endonuclease subunit	Synechococcus phage S-SM1	44.79	4E-156
Lau85_TahiMoana_Scaffold_2_10	9056	10021	+	ABC transporter type 1, transmembrane domain containing protein	alpha proteobacterium HIMB114	42.22	4E-78
Lau85_TahiMoana_Scaffold_2_11	10137	10337	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_12	10455	10676	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_13	10688	10888	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_14	10977	13214	+	hypothetical phage protein	Synechococcus phage S-CAM1	36.63	2E-137
Lau85_TahiMoana_Scaffold_2_15	13426	14643	+	porphyrin biosynthetic protein, AAA ATPase-containing	Synechococcus phage S-PM2	51.99	5E-129
Lau85_TahiMoana_S	14705	15184	+	hypothetical phage protein	#N/A	#N/A	#N/A

caffold_2_16							
Lau85_TahiMoana_Scaffold_2_17	15184	15495	+	conserved phage protein of unknown function, T7-like Gp1.7 domain	Synechococcus phage S-SSM7	46.15	3E-15
Lau85_TahiMoana_Scaffold_2_18	15564	16196	+	Gp45 sliding clamp domain-containing DNA polymerase accessory	Pelagibacter phage HTVC008M	45.5	4E-55
Lau85_TahiMoana_Scaffold_2_19	16189	17112	+	sliding clamp loader, AAA ATPase-containing	Pelagibacter phage HTVC008M	53.5	3E-114
Lau85_TahiMoana_Scaffold_2_20	17112	17411	+	Phosphoribosyl-ATP pyrophosphohydrolase	Staphylococcus phage JD007	48.35	5E-19
Lau85_TahiMoana_Scaffold_2_21	17411	17728	+	Thioredoxin-like fold containing protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_22	17712	18131	+	clamp loader small subunit	Synechococcus phage S-SKS1	51.49	5E-46
Lau85_TahiMoana_Scaffold_2_23	18103	18525	+	Translation repressor RegA	Synechococcus phage S-SKS1	51.49	5E-46
Lau85_TahiMoana_Scaffold_2_24	18534	19823	+	hypothetical phage protein	Cyanophage Syn19	31.49	6E-21
Lau85_TahiMoana_Scaffold_2_25	19855	20247	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_26	20257	20970	+	EF-Hand 1, calcium-binding domain containing protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_27	20974	21417	+	endonuclease (SNase-like)	Cyanophage P-RSM1	44.78	2E-28
Lau85_TahiMoana_Scaffold_2_28	21472	22068	+	exonuclease, ribonuclease H-like domain containing	Enterobacteria phage RB32	29.35	3E-21
Lau85_TahiMoana_Scaffold_2_29	22068	22379	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_30	22381	22995	+	Oxoglutarate/iron-dependent dioxygenase	Synechococcus phage S-CRM01	30.85	3E-21
Lau85_TahiMoana_Scaffold_2_31	23036	23278	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_32	23294	25774	+	DNA polymerase, family B	Synechococcus phage S-SM2	50.59	0
Lau85_TahiMoana_Scaffold_2_33	25784	26230	+	rdsrC sulfur transfer protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_34	26302	27348	+	RecA DNA recombination and repair protein	Enterobacteria phage Phi1	60.77	5E-140
Lau85_TahiMoana_Scaffold_2_35	27335	27508	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_36	27498	27701	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_37	27698	28807	+	DNA primase/helicase, P-loop NTPase-containing	Prochlorococcus phage P-SSM2	57.79	4E-147
Lau85_TahiMoana_Scaffold_2_38	28849	29061	+	DNA primase/helicase	Cyanophage MED4-213	66.67	5E-12
Lau85_TahiMoana_Scaffold_2_39	29104	30318	+	hypothetical phage protein	Pelagibacter phage HTVC008M	37.98	9E-82
Lau85_TahiMoana_Scaffold_2_40	30320	31672	+	hypothetical phage protein	Cyanophage Syn1	29.44	3E-10
Lau85_TahiMoana_Scaffold_2_41	31735	31965	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_42	31970	32278	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_43	32275	32853	+	nicotinate-nucleotide adenylyltransferase	Cyanophage S-SSM4	41.88	2E-38
Lau85_TahiMoana_Scaffold_2_44	32850	33128	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_45	33128	33829	+	hypothetical phage protein	Simiduia agarivorans SA1	30.7	0.0000001
Lau85_TahiMoana_Scaffold_2_46	33847	34815	+	hypothetical phage protein	Aeromonas phage CC2	31.92	4E-31
Lau85_TahiMoana_Scaffold_2_47	34882	35214	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_48	35216	35476	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau85_TahiMoana_Scaffold_2_49	35473	36549	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_50	36563	36721	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_51	36714	36872	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_52	36943	37335	+	hypothetical phage protein	Pelagibacter phage HTVC008M	40.87	2E-21
Lau85_TahiMoana_Scaffold_2_53	37332	37751	+	hypothetical phage protein	Pelagibacter phage HTVC008M	37.17	6E-18
Lau85_TahiMoana_Scaffold_2_54	37748	38035	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_55	38076	39098	+	DNA primase	Synechococcus phage S-RIM8 A.HR1	43.53	2E-93
Lau85_TahiMoana_Scaffold_2_56	39095	39286	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_57	39297	39644	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_58	39666	40001	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_59	40012	40230	+	putative regulatory protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_60	40231	40647	+	Rhodanese-like domain-containing protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_61	40670	40840	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_62	40837	41415	+	hypothetical phage protein	Deftia phage phiW-14	30.32	7E-13
Lau85_TahiMoana_Scaffold_2_63	41429	43753	+	PA14 domain-containing protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_64	43810	44061	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_65	44042	44206	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_66	44283	44558	+	phage domain of unknown function	Enterobacteria phage vB_EcoM_ACG-C40	35.9	0.000003
Lau85_TahiMoana_Scaffold_2_67	44551	46368	+	DNA topoisomerase, type IIA	Vibrio phage KVP40	40.82	2E-142
Lau85_TahiMoana_Scaffold_2_68	46365	47660	+	DNA topoisomerase, type IIA-like domain	Aeromonas phage 31	38.2	7E-92
Lau85_TahiMoana_Scaffold_2_69	47661	47861	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_70	47858	48706	+	5'-3' exonuclease, Rnase H activity	Synechococcus phage S-CAM1	49.64	2E-92
Lau85_TahiMoana_Scaffold_2_71	48706	48849	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_72	48846	49025	+	putative regulatory protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_73	49030	49221	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_74	49224	50561	+	DNA ligase	Cronobacter phage vB_CsaM_GAP32	34.14	4E-68
Lau85_TahiMoana_Scaffold_2_75	50558	51523	+	Phosphoribosylformylglycinamidine cyclo-ligase	Sorghum bicolor	45.08	2E-71
Lau85_TahiMoana_Scaffold_2_76	51523	51849	+	phosphoheptose isomerase, HAD-like domain	Synechococcus phage S-SM2	53.12	8E-23
Lau85_TahiMoana_Scaffold_2_77	52011	52877	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_78	53032	53442	+	hypothetical phage protein	Candidatus Ruthia magnifica str. Cm	36.04	1E-08
Lau85_TahiMoana_Scaffold_2_79	53686	54996	+	rdsrA Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	88.68	0
Lau85_TahiMoana_Scaffold_2_80	55019	56377	+	rdsrA Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	88.68	0
Lau85_TahiMoana_Scaffold_2_81	56355	56513	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau85_TahiMoana_Scaffold_2_82	56513	56893	+	FeS cluster assembly protein	Thioalkalivibrio sp. K90mix	79.03	9E-65
Lau85_TahiMoana_Scaffold_2_83	56881	57114	+	4Fe-4S ferredoxin-type, iron-sulpur binding domain	endosymbiont of Bathymodiolus sp.	87.01	4E-41
Lau85_TahiMoana_Scaffold_2_84	57111	57437	+	FeS cluster insertion protein	Candidatus Vesicomyosocius okutanii HA	76.85	5E-53
Lau85_TahiMoana_Scaffold_2_85	57434	57754	+	FeS cluster insertion protein	Leptothrix cholodnii;Leptothrix cholodnii SP-6	56.19	8E-38
Lau85_TahiMoana_Scaffold_2_86	57774	57998	+	TusA-like protein	endosymbiont of Bathymodiolus sp.	86.3	1E-38
Lau85_TahiMoana_Scaffold_2_87	58007	58099	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_88	58109	58453	+	rdsrC sulfur transfer protein	endosymbiont of Bathymodiolus sp.	77.19	1E-56
Lau85_TahiMoana_Scaffold_2_89	58566	58820	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_90	58825	59178	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_91	59178	59303	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_92	59367	59594	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_93	59675	59947	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_94	59989	60621	+	PhoH-like protein	Prochlorococcus phage P-RSM4	45.24	3E-55
Lau85_TahiMoana_Scaffold_2_95	60667	60879	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_96	60911	62269	+	DNA methylase N-4/N-6	Monosiga brevicollis MX1	28.9	2E-10
Lau85_TahiMoana_Scaffold_2_97	62359	62511	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_98	62548	63174	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_99	63180	63851	+	Restriction endonuclease type II-like	Synechococcus phage S-ShM2	41.46	7E-45
Lau85_TahiMoana_Scaffold_2_100	63841	64230	+	hypothetical phage protein	Ralstonia pickettii;Ralstonia pickettii 12J	45.76	6E-25
Lau85_TahiMoana_Scaffold_2_101	64214	64405	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_102	64590	64754	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_103	65455	65553	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_104	65550	65765	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_105	65781	65969	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_106	66237	66524	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_107	66521	66814	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_108	67008	67745	+	Thymidylate synthase, ThyX	Gluconacetobacter europaeus	54.95	3E-68
Lau85_TahiMoana_Scaffold_2_109	67781	67972	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_110	67982	68125	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_111	68122	68217	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_112	68217	68525	+	rdsrC sulfur transfer protein	Candidatus Vesicomyosocius okutanii HA	72.63	3E-44
Lau85_TahiMoana_Scaffold_2_113	68678	70429	+	ribonucleotide reductase large subunit	Francisella philomiragia	71.89	0
Lau85_TahiMoana_Scaffold_2_114	70444	71649	+	ribonucleotide reductase large subunit	Francisella sp. TX077308	70.59	0

Lau85_TahiMoana_Scaffold_2_115	71723	71908	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_116	71905	72084	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_117	72102	72482	+	GroES (chaperonin 10)-like protein	Ignavibacterium album Mat9-16	41.58	4E-14
Lau85_TahiMoana_Scaffold_2_118	72490	73923	+	nucleic acid-modifying 2-oxoglutarate (2OG)-Fe(II)-dependent dioxygenase	Persicivirga phage P12024L	34.2	2E-22
Lau85_TahiMoana_Scaffold_2_119	73923	75071	+	hypothetical phage protein	Serratia phage phiMAM1	24.86	8E-19
Lau85_TahiMoana_Scaffold_2_120	75046	76083	+	hypothetical phage protein	Serratia phage phiMAM1	32.02	4E-49
Lau85_TahiMoana_Scaffold_2_121	76083	76631	+	P-loop NTPase	Mycobacterium phage PG1	32.12	1E-13
Lau85_TahiMoana_Scaffold_2_122	76675	77253	+	Oxoglutarate/iron-dependent dioxygenase	Cyanophage Syn30	30.98	8E-21
Lau85_TahiMoana_Scaffold_2_123	77272	77694	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_124	77691	77864	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_125	77992	78186	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_126	78199	78378	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_127	78380	78667	+	Thioredoxin	Hyperthermus butylicus DSM 5456	35.16	1E-10
Lau85_TahiMoana_Scaffold_2_128	78738	79298	+	(p)ppGpp synthetase	Microcoleus vaginatus	38.33	1E-34
Lau85_TahiMoana_Scaffold_2_129	79409	79630	+	hypothetical phage protein	alpha proteobacterium IMCC14465	49.32	1E-14
Lau85_TahiMoana_Scaffold_2_130	80126	80575	+	HSP20-like chaperone	Nitratireductor pacificus	37.23	3E-28
Lau85_TahiMoana_Scaffold_2_131	80597	80845	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_132	80808	80933	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_133	80933	81049	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_134	81075	81485	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_135	81855	81998	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_136	82039	82428	-	conserved phage protein of unknown function	Pelagibacter phage HTVC008M	56.49	2E-41
Lau85_TahiMoana_Scaffold_2_137				[false prediction, not a protein]	0	0	0
Lau85_TahiMoana_Scaffold_2_138	82568	83008	-	hypothetical phage protein	Verrucomicrobiae bacterium DG1235	29.45	0.000002
Lau85_TahiMoana_Scaffold_2_139	83008	83874	-	hypothetical phage protein	Campylobacter phage CP21	27.3	1E-20
Lau85_TahiMoana_Scaffold_2_140	83984	84256	+	T4-like transcriptional regulator, dsDNA-binding domain	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_141	84337	84552	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_142	84554	84751	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_143	84751	85044	+	hypothetical phage protein	Halobacterium sp. DL1	36.84	9E-08
Lau85_TahiMoana_Scaffold_2_144	85094	85720	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_145	85717	85914	+	conserved protein of unknown function, DUF1289	#N/A	#N/A	#N/A
Lau85_TahiMoana_Scaffold_2_146	86160	86774	+	phage helicase assembly protein, T4-like	Pelagibacter phage HTVC008M	27.5	2E-24
Lau85_TahiMoana_Scaffold_2_147	86818	86925	+	hypothetical phage protein	#N/A	#N/A	#N/A

D. Uncultured Myovirus Lau77 str. KiloMoana

Id	Start	Stop	Strand	Annotation	Blast organism	Blastp id	e-value
Lau77_KiloMoana_1_1	20	172	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_2	157	369	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_3	369	611	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_4	611	973	+	protein binding domain-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_5	970	1098	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_6	1095	1349	+	hypothetical phage protein	Vibrio phage SIO-2	41.54	0.000004
Lau77_KiloMoana_1_7	1395	1589	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_8	1628	1978	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_9	2118	3215	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_10	3294	4265	+	SbmA/BacA-like domain, ABC-like transporter	alpha proteobacterium HIMB114	43.9	6E-81
Lau77_KiloMoana_1_11	4265	4468	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_12	4471	4713	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_13	4716	5006	+	rdsrC/TusE Sulfur transfer protein	Candidatus Ruthia magnifica str. Cm	85.42	2E-56
Lau77_KiloMoana_1_14	5027	5242	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_15	5262	5357	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_16	5351	5602	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_17	5665	6090	+	hypothetical phage protein	Acinetobacter haemolyticus	31.82	0.0000003
Lau77_KiloMoana_1_18	6113	6250	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_19	6340	6516	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_20	6529	6630	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_21	6623	7015	+	hypothetical phage protein	endosymbiont of Bathymodiolus sp.	51.97	4E-38
Lau77_KiloMoana_1_22	7061	8461	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_23	8458	9585	-	Dihydroorotate dehydrogenase, class 2	Vibrio nigripulchritudo	40.48	1E-81
Lau77_KiloMoana_1_24	9650	10015	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_25	10126	10698	+	Cytokinin riboside 5'-monophosphate phosphoribohydrolase LOG-domain containing protein	Joostella marina	51.09	1E-51
Lau77_KiloMoana_1_26	11138	11227	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_27	11172	11501	+	HTH ArsR-type transcriptional regulator	Candidatus Ruthia magnifica str. Cm	68.42	2E-38
Lau77_KiloMoana_1_28	11618	11959	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_29	11956	12222	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_30	12222	12398	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_31	12481	12888	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_1_32	12890	13267	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_33	13268	14119	+	Formyltetrahydrofolate deformylase	endosymbiont of Bathymodiolus sp.	70.32	1E-149
Lau77_KiloMoana_1_34	14121	14396	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_35	14478	14900	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_36	14909	15094	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_37	15203	15514	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_38	15719	15865	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_39	15946	16344	+	hypothetical phage protein	gamma proteobacterium SCGC AAA001-B15	38.71	9E-12
Lau77_KiloMoana_1_40	16341	16487	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_41	16566	17330	+	UDP-glucose/GDP-mannose dehydrogenase	Exiguobacterium antarcticum B7	31.67	1E-24
Lau77_KiloMoana_1_42	17393	17680	+	GroES (chaperonin 10)-like protein	Nitrosococcus halophilus Nc 4	62.77	2E-32
Lau77_KiloMoana_1_43	17690	19333	+	Chaperonin Cpn60/TCP-1	Alteromonas sp. SN2	75.05	#N/A
Lau77_KiloMoana_1_44	19333	19485	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_45	19588	19869	+	hypothetical phage protein	gamma proteobacterium SCGC AAA001-B15	39.53	5E-12
Lau77_KiloMoana_1_46	19943	20902	+	Lipase, class 3	Teredinibacter turnerae T7901	32.48	2E-42
Lau77_KiloMoana_1_47	20899	20994	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_48	20999	21499	-	Recombination endonuclease VII	Pseudomonas phage phikF77	37.4	3E-10
Lau77_KiloMoana_1_49	21559	21720	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_50	21727	21948	+	hypothetical phage protein	Acidovorax sp. JS42	38.81	6E-09
Lau77_KiloMoana_1_51	22540	22944	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_52	23120	23374	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_53	23568	23891	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_54	23930	24202	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_55	24427	25362	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_56	25462	25755	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_57	25764	26018	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_58	26112	26501	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_59	26673	26867	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_60	26864	27145	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_61	27142	27366	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_62	27359	27679	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_63	27769	27951	+	hypothetical phage protein	Sphaerochaeta pleomorpha str. Grapes	67.27	1E-15
Lau77_KiloMoana_1_64	27951	28169	+	Translation initiation factor IF-1	Thioalkalivibrio sulfidophilus HL-EbGr7	51.39	2E-20
Lau77_KiloMoana_1_65	28169	29155	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_1_66	29254	29862	+	phage baseplate structural protein Gp9/Gp10, T4-like	Cyanophage Syn30	31.78	6E-09
Lau77_KiloMoana_1_67	29882	32563	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_68	32574	33314	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_69	33350	33592	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_70	33589	34071	+	DNA polymerase III, alpha subunit	Cronobacter phage vB_CsaM_GAP32	37.5	1E-17
Lau77_KiloMoana_1_71	34102	34542	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_72	34542	34943	+	Chaperonin Cpn10	Enterobacteria phage vB_KleM-RaK2	43.84	6E-15
Lau77_KiloMoana_1_73	34940	35170	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_74	35170	35418	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_75	35453	36367	+	sliding clamp loader subunit, AAA-type ATPase domain	Cronobacter phage vB_CsaM_GAP32	38.44	3E-67
Lau77_KiloMoana_1_76	36364	36693	+	RmlC-like cupin	gamma proteobacterium SCGC AAA007-O20	48.42	2E-23
Lau77_KiloMoana_1_77	36718	37290	+	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	28.34	4E-10
Lau77_KiloMoana_1_78	37277	37813	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_79	37820	38500	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_80	38509	39165	+	SGNH hydrolase-type esterase	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_81	39149	39721	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_82	39726	40406	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_83	40501	42093	+	Protein of unknown function DUF112, transmembrane	Marinomonas mediterranea MMB-1	35.43	4E-69
Lau77_KiloMoana_1_84	42105	43265	+	hypothetical phage protein	Polymorphum gilvum SL003B-26A1	26.93	1E-17
Lau77_KiloMoana_1_85	43274	43594	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_86	43587	43928	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_87	43957	44823	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_88	44879	45649	+	SGNH hydrolase-type esterase domain-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_89	45639	46055	+	P-loop containing NTPase	Candidatus Pelagibacter ubique	52.03	3E-37
Lau77_KiloMoana_1_90	46154	48448	+	ATP-dependent Clp protease ATP-binding protein, subunit ClpA	Brevundimonas diminuta	40.99	#N/A
Lau77_KiloMoana_1_91	48459	48623	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_92	48628	48723	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_93	48726	49025	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_94	49034	49486	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_95	49495	50205	+	hypothetical phage protein	Synechococcus phage S-CRM01	33.49	8E-28
Lau77_KiloMoana_1_96	50210	50413	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_97	50424	51497	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_98	51528	51956	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_1_99	51969	52583	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_100	52580	53713	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_101	53710	54780	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_102	54777	56189	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_103	56186	56422	+	KTSC domain-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_104	56419	59760	+	PcfJ-like protein	Azotobacter vinelandii DJ	28.82	5E-13
Lau77_KiloMoana_1_105	59757	60107	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_106	60097	61488	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_107	61535	62464	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_108	62624	63259	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_109	63407	64324	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_110	64342	64869	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_111	64939	65466	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_112	65470	66780	+	Protein of unknown function DUF112, transmembrane	Oligotropha carboxidovorans OM5	27.05	1E-39
Lau77_KiloMoana_1_113	66767	67834	+	Adenine nucleotide alpha hydrolases-like domain-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_114	67860	69113	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_115	69273	72767	+	EF-Hand 1, calcium-binding site-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_116	72769	73101	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_117	73131	73433	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_118	73423	74010	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_119	74024	74242	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_120	74251	74559	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_121	74584	75321	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_122	75336	75491	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_123	75488	75865	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_124	75964	76227	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_125	76240	77226	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_126	77266	79968	+	DNA-directed DNA polymerase, family B	Cronobacter phage vB_CsaM_GAP32	40.93	1E-126
Lau77_KiloMoana_1_127	79965	80915	+	S-adenosyl-L-methionine-dependent methyltransferase	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_128	80981	81715	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	28.05	1E-11
Lau77_KiloMoana_1_129	81719	83170	+	hypothetical phage protein	Azospirillum sp. B510	27.88	1E-47
Lau77_KiloMoana_1_130	83170	83709	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_131	83709	84149	+	hypothetical phage protein	Methylobacterium extorquens PA1	45.16	2E-24

Lau77_KiloMoana_1_132	84146	85231	+	hypothetical phage protein	Methylobacterium extorquens PA1	42.51	1E-93
Lau77_KiloMoana_1_133	85228	85857	+	Pyridoxal phosphate-dependent transferase	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_134	85902	86576	+	GTP cyclohydrolase I	SAR86 cluster bacterium SAR86E	72.52	4E-120
Lau77_KiloMoana_1_135	86576	87118	+	Isopenicillin N synthase	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_136	87223	87432	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_137	87533	88012	+	Phosphoribosyltransferase	Novosphingobium sp. Rr 2-17	31.58	1E-18
Lau77_KiloMoana_1_138	88087	88488	-	hypothetical phage protein	Candidatus Ruthia magnifica str. Cm	48.65	4E-12
Lau77_KiloMoana_1_139	88686	89150	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_140	89286	93863	+	Subtilase	Candidatus Puniceispirillum marinum IMCC1322	41.08	6E-88
Lau77_KiloMoana_1_141	93893	94726	+	Ion transport domain-containing protein	Colwellia psychrerythraea 34H	60.61	4E-108
Lau77_KiloMoana_1_142	94841	95404	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_143	95492	95716	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_144	95709	95933	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_145	96028	96231	+	Hypothetical glycosyl hydrolase 12	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_146	96263	96421	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_147	96490	96867	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_148	96880	97119	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_149	97142	97813	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_150	97815	98462	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_151	98464	98781	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_152	98774	98947	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_153	98988	99329	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_154	99343	100017	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_155	100014	100868	+	7-carboxy-7-deazaguanine synthase	Vibrio phage KVP40	34.56	2E-37
Lau77_KiloMoana_1_156	100869	101963	+	S-adenosylmethionine synthetase	Magnetospirillum magneticum AMB-1	56.02	4E-125
Lau77_KiloMoana_1_157	101956	102252	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_158	102260	102925	+	hypothetical phage protein	Collimonas fungivorans Ter331	25.91	2E-09
Lau77_KiloMoana_1_159	102922	103614	+	hypothetical phage protein	Collimonas fungivorans Ter331	26.38	0.000001
Lau77_KiloMoana_1_160	103611	104360	+	SGNH hydrolase-type esterase	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_161	104385	105389	+	5'-3' exonuclease, Rnase H ribonuclease activity	Cronobacter phage vB_CsaM_GAP32	49.84	4E-110
Lau77_KiloMoana_1_162	105382	105825	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_163	105825	106028	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_164	106066	106359	+	hypothetical phage protein	Azospirillum brasilense Sp245	50	1E-15
Lau77_KiloMoana_1_165	106334	106717	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_1_166	106680	107675	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_167	107749	108279	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_168	108279	108692	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_169	108683	109078	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_170	109072	109296	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_171	109410	109664	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_172	109661	110164	+	Glycine zipper 2TM domain-containing protein	gamma proteobacterium SCGC AAA001-B15	31.85	1E-13
Lau77_KiloMoana_1_173	110177	110497	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_174	110451	110738	+	TusA-like domain-containing protein	Candidatus Vesicomyosocius okutanii HA	57.53	7E-25
Lau77_KiloMoana_1_175	110823	111344	+	DsrEFH-like protein	Candidatus Vesicomyosocius okutanii HA	70.76	2E-85
Lau77_KiloMoana_1_176	111416	111796	+	Cytochrome c	gamma proteobacterium SCGC AAA007-O20	39.64	5E-13
Lau77_KiloMoana_1_177	111793	112176	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_178	112216	112491	+	Heavy metal-associated domain-containing protein	Candidatus Vesicomyosocius okutanii HA	63.64	4E-30
Lau77_KiloMoana_1_179	112626	112952	+	Conserved hypothetical protein CHP02001	Asticcacaulis biprosthecium	46.75	6E-09
Lau77_KiloMoana_1_180	113088	113486	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_181	113746	114126	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_182	114212	114484	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_183	114584	114877	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_184	114885	115514	+	hypothetical phage protein	Nitratireductor indicus	29.9	4E-16
Lau77_KiloMoana_1_185	115614	116531	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_1_186	116769	118070	+	rdsrA, Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	88.94	#N/A
NEXT CONTIG							
Lau77_KiloMoana_2_1	3	1124	+	rdsrA, Sulfite reductase, dissimilatory-type alpha subunit	endosymbiont of Bathymodiolus sp.	88.46	#N/A
Lau77_KiloMoana_2_2	1102	1260	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_3	1570	1797	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_4	1896	2273	+	NIF system FeS cluster assembly, NifU	Candidatus Ruthia magnifica str. Cm	82.79	9E-65
Lau77_KiloMoana_2_5	2341	2673	+	Monothiol glutaredoxin	Candidatus Ruthia magnifica str. Cm	76.15	3E-57
Lau77_KiloMoana_2_6	2670	2990	+	FeS cluster biogenesis protein	Ralstonia eutropha JMP134	59.43	2E-41
Lau77_KiloMoana_2_7	3064	3468	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_8	3490	3735	+	4Fe-4S ferredoxin-type, iron-sulpur binding domain-containing protein	Candidatus Vesicomyosocius okutanii HA	89.47	1E-40
Lau77_KiloMoana_2_9	3732	4076	+	Iron-sulfur cluster insertion protein ErpA	Candidatus Vesicomyosocius okutanii HA	78.9	3E-56
Lau77_KiloMoana_2_10	4171	5133	+	Aldolase-type TIM barrel-containing protein	Turneriella parva DSM 21527	31.02	1E-45
Lau77_KiloMoana_2	5144	6415	+	Aldolase-type TIM barrel-	Bdellovibrio exovorus JSS	32.79	1E-63

_11				containing protein			
Lau77_KiloMoana_2_12	6467	6634	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_13	6724	7278	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_14	7360	7743	+	Mechanosensitive ion channel MscS	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_15	7760	8674	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_16	8671	9183	+	Adenine phosphoribosyl transferase	Desulfovibrio hydrothermalis AM13	49.06	4E-49
Lau77_KiloMoana_2_17	9176	9811	+	Nucleoside phosphorylase	Ruminococcus gnavus	27.88	1E-19
Lau77_KiloMoana_2_18	9948	10304	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_19	10301	10690	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_20	10653	11072	+	Staphylococcal nuclease (SNase-like)	Synechococcus phage S-MbCM6	47.15	5E-26
Lau77_KiloMoana_2_21	11121	11945	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_22	11949	12947	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_23	12947	14044	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_24	14044	15120	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_25	15190	15765	+	6-pyruvoyl tetrahydropterin synthase/QueD family protein	Lentisphaera araneosa	39.46	8E-26
Lau77_KiloMoana_2_26	15766	15966	-	Cold shock protein, CspA	Haloplasma contractile	73.44	7E-26
Lau77_KiloMoana_2_27	16007	16912	+	hypothetical phage protein	Teredinibacter turnerae T7901	24.23	2E-14
Lau77_KiloMoana_2_28	16899	17255	-	S-adenosylmethionine decarboxylase family protein	Anaerolinea thermophila UNI-1	43.48	6E-24
Lau77_KiloMoana_2_29	17374	17532	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_30	17544	17726	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_31	17818	18012	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_32	18106	18246	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_33	18294	19376	+	AAA+ ATPase domain-containing protein	Arcobacter nitrofigilis DSM 7299	42.66	1E-82
Lau77_KiloMoana_2_34	19376	20656	+	metalopeptidase/metal-binding domain-related proteins	Enterobacteria phage vB_KleM-RaK2	35.84	1E-82
Lau77_KiloMoana_2_35	20735	21073	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_36	21076	21711	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_37	21727	22068	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_38	22082	22255	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_39	22255	23205	+	NAD-dependent epimerase/dehydratase	Eubacteriaceae bacterium ACC19a	37.26	8E-56
Lau77_KiloMoana_2_40	23214	24386	+	hypothetical phage protein	Xenorhabdus nematophila	23.73	2E-19
Lau77_KiloMoana_2_41	24387	25475	+	DegT/DnrJ/EryC1/StrS aminotransferase	Caldicellulosiruptor saccharolyticus DSM 8903	33.42	3E-43
Lau77_KiloMoana_2_42	25444	27051	-	hypothetical phage protein	Photorhabdus luminescens subsp. laumondii TTO1	29.33	1E-18
Lau77_KiloMoana_2_43	27054	28133	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_44	28127	29317	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_2_45	29331	30362	+	hypothetical phage protein	Streptomyces sp. C	22.86	2E-13
Lau77_KiloMoana_2_46	30328	31851	-	Carbamoyltransferase	Collimonas fungivorans Ter331	44.87	1E-134
Lau77_KiloMoana_2_47	31949	33022	-	Pleckstrin homology domain-containing protein	Teredinibacter turnerae T7901	24.54	5E-12
Lau77_KiloMoana_2_48	33001	34311	-	Aldolase-type TIM barrel-containing protein	Bdellovibrio exovorus JSS	44.67	3E-142
Lau77_KiloMoana_2_49	34304	35005	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_50	34998	36200	-	4Fe4S-binding SPASM domain-containing protein	Bdellovibrio exovorus JSS	31.87	4E-43
Lau77_KiloMoana_2_51	36197	36916	-	SGNH hydrolase-type esterase domain-containing protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_52	36940	38226	-	4Fe4S-binding SPASM domain-containing protein	planctomycete KSU-1	38.76	4E-95
Lau77_KiloMoana_2_53	38223	39089	-	4Fe4S-binding SPASM domain-containing protein	Mesorhizobium sp. STM 4661	32.34	2E-33
Lau77_KiloMoana_2_54	39086	40426	-	Aldolase-type TIM barrel-containing protein	Bdellovibrio exovorus JSS	53.02	1E-170
Lau77_KiloMoana_2_55	40423	41292	-	NAD-dependent epimerase/dehydratase	Methanocorpusculum labreanum Z	42.28	8E-71
Lau77_KiloMoana_2_56	41289	42344	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_57	42368	42676	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_58	42689	43876	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_59	44259	45665	-	Aldolase-type TIM barrel-containing protein	Bdellovibrio exovorus JSS	40.46	6E-112
Lau77_KiloMoana_2_60	45973	46950	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_61	46932	47837	-	NAD-dependent epimerase/dehydratase	Methanocaldococcus infernus ME	29.21	1E-28
Lau77_KiloMoana_2_62	47839	48762	-	hypothetical phage protein	Rhodobacter sphaeroides KD131	34	2E-45
Lau77_KiloMoana_2_63	48958	49305	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_64	49310	50128	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_65	50121	50939	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_66	50936	51241	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_67	51318	52289	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_68	52365	54569	-	hypothetical phage protein	Commensalibacter intestini	32.63	3E-13
Lau77_KiloMoana_2_69	54601	55530	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_70	55533	57416	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_71	57428	58033	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	42	3E-38
Lau77_KiloMoana_2_72	58037	58162	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_73	58190	70072	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	22.93	2E-50
Lau77_KiloMoana_2_74	70076	73492	-	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	32.08	2E-122
Lau77_KiloMoana_2_75	73546	73923	-	phage baseplate outer wedge, T4-like	Cronobacter phage vB_CsaM_GAP32	29.82	0.0000002
Lau77_KiloMoana_2_76	73927	75345	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	26.02	3E-24
Lau77_KiloMoana_2_77	75345	77846	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_78	77848	78144	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	36.62	0.000002

Lau77_KiloMoana_2_79	78144	79205	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	28.31	4E-18
Lau77_KiloMoana_2_80	79365	80303	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	39	2E-50
Lau77_KiloMoana_2_81	80300	81310	+	Phosphoesterase domain-containing protein	Enterobacteria phage vB_KleM-RaK2	41.35	8E-95
Lau77_KiloMoana_2_82	81307	82872	+	4Fe4S-binding SPASM domain-containing protein	Bdellovibrio exovorus JSS	28.5	4E-36
Lau77_KiloMoana_2_83	82881	84617	+	RecF/RecN/SMC N-terminal domain-containing protein	Cronobacter phage vB_CsaM_GAP32	32.72	3E-112
Lau77_KiloMoana_2_84	85244	85573	-	hypothetical phage protein	Ahrensia sp. R2A130	31.19	5E-11
Lau77_KiloMoana_2_85	85575	86768	-	hypothetical phage protein	Eggerthella sp. YY7918	24.37	1E-11
Lau77_KiloMoana_2_86	86807	87499	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_87	87563	88075	-	hypothetical phage protein	Oscillatoria nigro-viridis PCC 7112	24.68	0.000006
Lau77_KiloMoana_2_88	88075	88356	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_89	88299	89126	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	29.86	5E-28
Lau77_KiloMoana_2_90	89180	91816	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_91	91872	92252	+	hypothetical phage protein	Helicobacter bilis	28.7	6E-09
Lau77_KiloMoana_2_92	92266	93999	+	phage head protein, T4-like Gp20	Cronobacter phage vB_CsaM_GAP32	45.98	2E-166
Lau77_KiloMoana_2_93	94011	94238	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_94	94353	95093	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_95	95138	95467	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_96	95471	96097	+	phage prohead protease, T4-like	Cronobacter phage vB_CsaM_GAP32	53.11	1E-55
Lau77_KiloMoana_2_97	96107	97225	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	31.32	5E-50
Lau77_KiloMoana_2_98	97235	98434	+	phage major capsid protein, T4-like Gp23	Cronobacter phage vB_CsaM_GAP32	64	#N/A
Lau77_KiloMoana_2_99	98510	99268	+	SGNH hydrolase-type esterase domain	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_100	99268	99915	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_101	100003	100644	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_102	100785	101060	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_103	101116	101868	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_104	101879	102193	+	hypothetical phage protein	Geobacillus sp. Y412MC61	35.63	0.000003
Lau77_KiloMoana_2_105	102195	102623	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_106	102623	103981	+	hypothetical phage protein	Commensalibacter intestini	41.86	3E-14
Lau77_KiloMoana_2_107	104056	105342	+	hypothetical phage protein	Ostreococcus lucimarinus CCE9901	40.23	2E-12
Lau77_KiloMoana_2_108	105564	106802	+	hypothetical phage protein	Cyanophage P-RSM6	38.61	3E-25
Lau77_KiloMoana_2_109	106771	110307	+	hypothetical phage protein	Cyanophage P-RSM6	31.31	3E-10
Lau77_KiloMoana_2_110	110425	112800	+	hypothetical phage protein	Cyanophage PSS2	45.7	9E-28
Lau77_KiloMoana_2_111	112800	113474	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_112	113459	113890	-	NUDIX hydrolase domain	Cronobacter phage vB_CsaM_GAP32	38.71	1E-22

Lau77_KiloMoana_2_113	113887	115281	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	37.29	6E-41
Lau77_KiloMoana_2_114	115283	115696	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_115	115696	116982	-	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	40.2	1E-32
Lau77_KiloMoana_2_116	117097	118215	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_117	118255	119016	+	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	38.46	7E-48
Lau77_KiloMoana_2_118	119018	119584	+	Manganese/iron superoxide dismutase	Alloprevotella tannerae	34.9	5E-31
Lau77_KiloMoana_2_119	119581	120162	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_120	120192	120716	+	Ribonuclease H-like domain-containing protein	Hafnia alvei	40.97	3E-24
Lau77_KiloMoana_2_121	120713	121042	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_122	121174	124539	+	phage tail sheath protein	Enterobacteria phage vB_KleM-RaK2	32.18	3E-73
Lau77_KiloMoana_2_123	124583	125131	-	HNH endonuclease	Aurantimonas manganoxydans	44.51	5E-42
Lau77_KiloMoana_2_124	125323	125439	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_125	125503	125724	-	hypothetical phage protein	alpha proteobacterium IMCC14465	47.95	7E-14
Lau77_KiloMoana_2_126	125845	126270	-	HNH endonuclease	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_127	126375	126989	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	33.33	3E-25
Lau77_KiloMoana_2_128	126996	128462	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	28.08	1E-14
Lau77_KiloMoana_2_129	128462	128815	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_130	128815	129237	+	phage head completion protein, endonuclease I activity, T7-like	Cronobacter phage vB_CsaM_GAP32	45.26	2E-29
Lau77_KiloMoana_2_131	129239	129718	+	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	32.84	0.0000003
Lau77_KiloMoana_2_132	129766	130362	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_133	130359	130682	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_134	130692	130982	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_135	130991	131266	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_136	131271	131870	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_137	131867	133363	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_138	133447	133788	+	conserved phage protein Mycobacteriophage D29, Gp61	Capnocytophaga granulosa	50.43	2E-30
Lau77_KiloMoana_2_139	133803	134531	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_140	134632	135639	+	Aldolase-type TIM barrel-containing protein	Turneriella parva DSM 21527	30.28	1E-43
Lau77_KiloMoana_2_141	135648	137216	+	Terminase, large subunit	Cronobacter phage vB_CsaM_GAP32	45.79	5E-146
Lau77_KiloMoana_2_142	137273	138562	+	hypothetical phage protein	Methylobacterium extorquens	45.21	2E-08
Lau77_KiloMoana_2_143	138552	138968	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_144	139127	140089	+	single-stranded DNA-binding, T4-like Gp32	Enterobacteria phage vB_KleM-RaK2	42.81	1E-74
Lau77_KiloMoana_2_145	140138	140998	+	hypothetical phage protein	#N/A	#N/A	#N/A

Lau77_KiloMoana_2_146	140998	142044	+	DNA recombination/repair protein RecA-like, ATP-binding domain	Cronobacter phage vB_CsaM_GAP32	56.62	7E-134
Lau77_KiloMoana_2_147	142044	142277	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_148	142277	142729	+	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	44.9	3E-36
Lau77_KiloMoana_2_149	142738	143760	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_150	143757	144506	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_151	144535	145314	+	Cytidyltransferase-like and Mannose-6-phosphate isomerase domain-containing protein	Synechococcus phage S-SSM7	50	2E-33
Lau77_KiloMoana_2_152	145316	145999	+	hypothetical phage protein	Hylemonella gracilis	56.25	5E-90
Lau77_KiloMoana_2_153	146008	146997	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_154	147054	148055	+	P-loop containing nucleoside triphosphate hydrolase	Acinetobacter phage Acj61	36.82	4E-51
Lau77_KiloMoana_2_155	148187	149644	+	Helicase/UvrB domain-containing protein	Cronobacter phage vB_CsaM_GAP32	49.27	8E-168
Lau77_KiloMoana_2_156	149635	150663	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_157	150672	151019	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_158	151089	151589	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	31.13	2E-19
Lau77_KiloMoana_2_159	151590	152036	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_160	152011	152706	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	27.56	2E-22
Lau77_KiloMoana_2_161	152707	153762	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_162	153863	154447	+	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	38.54	1E-33
Lau77_KiloMoana_2_163	154416	154796	-	hypothetical phage protein	Enterobacteria phage vB_KleM-RaK2	32.08	3E-16
Lau77_KiloMoana_2_164	154809	155627	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_165	155634	156344	+	PhoH-like protein	Synechococcus phage syn9	46.06	8E-62
Lau77_KiloMoana_2_166	156344	156619	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_167	156586	157314	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_168	157705	158673	+	DNA primase	Methylobacterium sp. GXF4	32.82	3E-45
Lau77_KiloMoana_2_169	158696	160078	+	DNA helicase, DnaB-like	Enterobacteria phage vB_KleM-RaK2	36.06	2E-88
Lau77_KiloMoana_2_170	160104	160361	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_171	160339	161091	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_172	161093	161245	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_173	161238	161804	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_174	161862	162140	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_175	162222	162452	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_176	162440	162904	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_177	162909	163427	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2	163378	164337	+	hypothetical phage protein	#N/A	#N/A	#N/A

_178							
Lau77_KiloMoana_2_179	164337	164666	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_180	164663	164998	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_181	165000	166706	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_182	166721	167347	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_183	167369	168088	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	47.86	5E-60
Lau77_KiloMoana_2_184	168120	169757	+	putative phage tail spike protein, pectin lyase-domain containing	Frankia sp. EUN1f	40	0.0000005
Lau77_KiloMoana_2_185	169836	172106	+	Ribonucleotide reductase, class I, alpha subunit	Proteus penneri	59.11	#N/A
Lau77_KiloMoana_2_186	172247	173374	+	Ribonucleotide reductase small subunit	Cronobacter phage vB_CsaM_GAP32	62.07	1E-172
Lau77_KiloMoana_2_187	173371	173628	+	Glutaredoxin	gamma proteobacterium SCGC AAA001-B15	42.17	9E-17
Lau77_KiloMoana_2_188	173630	173932	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_189	173941	174240	+	PAAR motif-containing protein	Vibrio cholerae	43.88	9E-16
Lau77_KiloMoana_2_190	174304	174558	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_191	174549	174812	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_192	174809	175555	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_193	175525	175746	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_194	175743	176051	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_195	176052	176390	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_196	176422	176691	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_197	176693	176974	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_198	177005	178759	+	hypothetical phage protein	Bradyrhizobium oligotrophicum S58	29.32	4E-16
Lau77_KiloMoana_2_199	178996	179310	+	ATP-dependent Clp protease adaptor protein ClpS	Magnetococcus marinus MC-1	39.77	2E-16
Lau77_KiloMoana_2_200	179310	179663	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_201	179809	181017	+	UDP-glucose/GDP-mannose dehydrogenase	alpha proteobacterium HIMB59	49.51	1E-123
Lau77_KiloMoana_2_202	181014	182030	+	hypothetical phage protein	Streptomyces bingchenggensis BCW-1	44.64	1E-92
Lau77_KiloMoana_2_203	182030	182875	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_204	182872	184128	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_205	184101	184682	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_206	184759	184953	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_207	184970	185251	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_208	185220	185402	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_209	185568	185987	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_210	186170	186643	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2	186967	187134	+	hypothetical phage protein	#N/A	#N/A	#N/A

_211							
Lau77_KiloMoana_2_212	187131	187373	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_213	187517	187867	+	Metallopeptidase	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_214	187949	188353	+	Cytochrome c domain	gamma proteobacterium SCGC AAA007-O20	60.98	1E-27
Lau77_KiloMoana_2_215	188435	188674	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_216	188674	189261	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_217	189267	189593	+	TM2 domain-containing protein	Rhodobacter capsulatus SB 1003	47.62	6E-19
Lau77_KiloMoana_2_218	189611	189916	+	hypothetical phage protein	Candidatus Vesicomyosocius okutanii HA	62.38	3E-38
Lau77_KiloMoana_2_219	190035	190253	+	hypothetical phage protein	gamma proteobacterium SCGC AAA007-O20	46.55	6E-12
Lau77_KiloMoana_2_220	190405	190800	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_221	190825	192435	+	4Fe4S-binding SPASM domain-containing protein	planctomycete KSU-1	33.08	3E-71
Lau77_KiloMoana_2_222	192482	192589	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_223	192580	193029	+	Clp protease proteolytic subunit/Translocation-enhancing protein TepA	Mitsuokella multacida	64.05	2E-64
Lau77_KiloMoana_2_224	193152	193430	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_225	193455	194648	-	Aldolase-type TIM barrel-containing protein	Bdellovibrio exovorus JSS	27.84	8E-17
Lau77_KiloMoana_2_226	194716	194898	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_227	194859	195164	+	phage protein, T7-like Gp1.7	Prochlorococcus phage P-SSM2	46.99	4E-14
Lau77_KiloMoana_2_228	195164	195358	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_229	195420	195851	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_230	195864	196259	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_231	196472	196843	+	Rossmann-like alpha/beta/alpha sandwich fold-containing protein	Flavobacteria bacterium BAL38	63.16	2E-48
Lau77_KiloMoana_2_232	196821	196967	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_233	197303	197707	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_234	197741	197989	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_KiloMoana_2_235	198004	198495	+	hypothetical phage protein	#N/A	#N/A	#N/A
NEXT CONTIG							
Lau77_Kilo_Moana_scaffold_13447_1	2	244	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_13447_2	391	1074	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_13447_3	1110	3506	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_13447_4	3512	4168	+	hypothetical phage protein	Cronobacter phage vB_CsaM_GAP32	40.31	6E-34
NEXT CONTIG							
Lau77_Kilo_Moana_scaffold_20335_1	2	1108	-	hypothetical phage protein	Pelagibacter phage HTVC010P	54.66	9E-132
Lau77_Kilo_Moana_scaffold_20335_2	1108	1689	-	hypothetical phage protein	Pelagibacter phage HTVC010P	42.37	9E-40

Lau77_Kilo_Moana_scaffold_20335_3	1763	2296	-	hypothetical phage protein	Pelagibacter phage HTVC010P	38.2	1E-27
Lau77_Kilo_Moana_scaffold_20335_4	2309	3178	-	hypothetical phage protein	Pelagibacter phage HTVC010P	57.65	8E-103
Lau77_Kilo_Moana_scaffold_20335_5	3340	4227	-	hypothetical phage protein	Pelagibacter phage HTVC010P	29.67	3E-32
NEXT CONTIG							
Lau77_Kilo_Moana_scaffold_33604_1	3	566	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_2	568	801	+	hypothetical phage proteins	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_3	755	1627	-	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_4	1684	2541	-	hypothetical phage proteins	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_5	2604	2759	+	hypothetical phage protein	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_6	2769	2867	+	hypothetical phage proteins	#N/A	#N/A	#N/A
Lau77_Kilo_Moana_scaffold_33604_7	3000	3341	+	Mycobacteriophage D29, Gp61 domain	Capnocytophaga granulosa	55.14	8E-31
Lau77_Kilo_Moana_scaffold_33604_8	3380	4048	+	hypothetical phage protein	#N/A	#N/A	#N/A

Gene annotation of reverse-acting dissimilatory sulfite reductase (*rdsr*)-containing putative viral genomes. *rdsr* genes are highlighted in yellow and other SUP05-like genes in cyan.

Supplementary Table 4. Modeling end-member fluid properties.

	A1 vent[a]	Seawater
T (°C)	309	2
pH[b]	4.3	8
O_2, aqueous	0	0.15[c]
NH_4^+ [d]	13.6	0
N_2, aqueous	0.48[e]	0.58[e]
NO_3^-	0	0.035
NO_2^-	0	0.001
H_2, aqueous [f]	0.2	0.0000004
SO_4^{2-}	0	28.0
H_2S, aqueous	3.6	0
$\sum CO_2$, aqueous	8	1.8
CH_4, aqueous [f]	0.04	0
Cl^-	534	540
Na^+	430	464
Ca^{2+}	39	10.2
Mg^{2+}	0	52.2
K^+	25	10.1
SiO_2, aqueous	17	0.17
Fe	0.27	0
Mn^{2+}	0.47	0
Cu^+ [g]	0.03	0
Zn^{2+} [g]	0.07	0
Ba^{2+} [g]	0.05	0

All concentrations in mmol/kg vent fluid.

(a) Vent chemistry data (Mottl et al 2011).

(b) *In situ* pH based on 25 °C measurement.

(c) WOCE section P06 background dissolved O_2, NO_3^-, and NO_2^- (Talley 2007).

(d) Predicted to exist as NH_3 in vent fluid.

(e) Seawater dissolved N_2 (Weiss and Craig 1973); vent fluid dissolved N_2 assumed to be 83% of seawater concentration (Brandes et al 1998).

(f) Dissolved gases(Seewald et al 2005).

(g) Based on EPR 21° N (Von Damm et al 1985) for lack of more relevant data.

Supplementary Table 5. *rdsrA* and *rdsrC* gene clusters in Pacific Ocean Virome "Ultraclean" Protein Clusters(Hurwitz et al 2013) dataset as identified by protein blasts against ncbi-nr.

POV Cluster	Best hit Annotation/Organism	Best hit % identity	Bit Score
M5OD-32bf9769178a97562f0f4bfb24afa79d	*rdsrA*,uncultured Thiohalocapsa sp. PB-PSB1 (WP_023412096)	90	188
LJ4S-30a39c11d1d45b1825b75037ce18355f	*rdsrA*, Candidatus Vesicomyosocius okutanii (YP_001219625)	94	158
LJ4S-e3d58f574ed0cd6d7a07c7474e6e81f8	*rdsrA*, uncultured bacterium BAC13K9BAC (AAY89969)	80	129
LF26S-f425969cfe3d696e0666e41cad24d6c2	*rdsrC*, uncultured bacterium BAC13K9BAC (AAY89972)	71	125
LJ4D-fb6c9460c2cd9ac34e873a9ea7beb931	*rdsrC*, marine gamma proteobacterium HTCC2148 (WP_007227926)	63	108
LJ4D-1bf8ca159adda6441f4410c01638ac9f	*rdsrC*, marine gamma proteobacterium HTCC2148 (WP_007227926)	70	101
M5OD-4a496e172b80e7f9a83a0044da3c1c00	*rdsrC*, gamma proteobacterium NOR5-3 (WP_009022409)	89	81
LJ4D-a4ceee95cf5026c64b2c581029b896c1	*rdsrC*, marine gamma proteobacterium HTCC2148 (WP_007227926)	69	60

4.7 References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Anantharaman K, Breier JA, Sheik CS, Dick GJ (2013). Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proceedings of the National Academy of Sciences* **110:** 330-335.

Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54:** 1501-1529.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011). Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474:** 604-608.

Bethke CM (2007). *Geochemical and biogeochemical reaction modeling*, Second edn. Cambridge University Press: Cambridge.

Bowers TS, Von Damm KL, Edmond JM (1985). Chemical evolution of mid-ocean ridge hot springs. *Geochimica Et Cosmochimica Acta* **49:** 2239-2252.

Brandes JA, Boctor NZ, Cody GD, Cooper BA, Hazen RM, Yoder HS (1998). Abiotic nitrogen reduction on the early Earth. *Nature* **395:** 365-367.

Breier JA, Rauch CG, McCartney K, Toner BM, Fakra SC, White SN *et al* (2009). A suspended-particle rosette multi-sampler for discrete biogeochemical sampling in low-particle-density waters. *Deep Sea Research Part I: Oceanographic Research Papers* **56:** 1579-1589.

Breier JA, Toner BM, Fakra SC, Marcus MA, White SN, Thurnherr AM *et al* (2012). Sulfur, sulfides, oxides and organic matter aggregated in submarine hydrothermal plumes at 9°50′N East Pacific Rise. *Geochimica Et Cosmochimica Acta* **88:** 216-236.

Breitbart M (2012). Marine Viruses: Truth or Dare. *Annual Review of Marine Science* **4:** 425-448.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330:** 1375-1378.

Casjens SR, Gilcrease EB, Winn-Stapley DA, Schicklmaier P, Schmieger H, Pedulla ML *et al* (2005). The generalized transducing Salmonella bacteriophage ES18: complete genome sequence and DNA packaging strategy. *J Bacteriol* **187:** 1091-1104.

Cleverley JS, Bastrakov EN (2005). K2GWB: Utility for generating thermodynamic data files for The Geochemist's WorkbenchÆ at 0-1000°C and 1-5000 bar from UT2K and the UNITHERM database. *Computers & Geosciences* **31:** 756-767.

Cort JR, Selan U, Schulte A, Grimm F, Kennedy MA, Dahl C (2008). Allochromatium vinosum DsrC: Solution-State NMR Structure, Redox Properties, and Interaction with DsrEFH, a Protein Essential for Purple Sulfur Bacterial Sulfur Oxidation. *Journal of Molecular Biology* **382:** 692-707.

Dahl C, Kredich NM, Deutzmann R, Trlfper HG (1993). Dissimilatory sulphite reductase from Archaeoglobus fulgidus: physico-chemical properties of the enzyme and cloning, sequencing and analysis of the reductase genes. *Journal of General Microbiology* **139:** 1817-1828.

Dhillon A, Goswami S, Riley M, Teske A, Sogin M (2005). Domain evolution and functional diversification of sulfite reductases. *Astrobiology* **5:** 18-29.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.

Dick GJ, Tebo BM (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12:** 1334-1347.

Drummond SE (1981). Boiling and mixing of hydrothermal fluids: chemical effects on mineral precipitation., Pennsylvania State University.

Duhaime MB, Wichels A, Waldmann J, Teeling H, Glockner FO (2011). Ecogenomics and genome landscapes of marine Pseudoalteromonas phage H105/1. *ISME J* **5:** 107-121.

Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32:** 1792-1797.

Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Geoff Wheat C *et al* (2012). Inter-field variability in the microbial communities of hydrothermal vent deposits from a back-arc basin. *Geobiology* **10:** 333-346.

Grimm F, Dobler N, Dahl C (2010). Regulation of dsr genes encoding proteins responsible for the oxidation of stored sulfur in Allochromatium vinosum. *Microbiology* **156:** 764-773.

Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Hausermann D (1996). Two-dimensional detector software: From real detector to idealised image or two-theta scan. *High Pressure Research* **14:** 235-248.

Hara S, Koike I, Terauchi K, Kamiya H, Tanoue E (1996). Abundance of viruses in deep oceanic waters. *Marine Ecology Progress Series* **145:** 269-277.

Helgeson HC (1969). Thermodynamics of hydrothermal systems at elevated temperatures and pressures. *American Journal of Science* **267:** 729-804.

Helgeson HC, Kirkham DH (1974). Theoretical prediction of the thermodynamic behavior of aqueous electrolytes at high pressures and temperatures; II, Debye-Huckel parameters for activity coefficients and relative partial molal properties. *American Journal of Science* **274:** 1199-1261.

Helgeson HC, Delaney JM, Nesbitt HW, Bird DK (1978). Summary and critique of the thermodynamic properties of rock-forming minerals. *American Journal of Science* **278-A:** 1-229.

Hurwitz BL, Hallam SJ, Sullivan MB (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14:** R123.

Hurwitz BL, Sullivan MB (2013). The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *PLoS One* **8:** e57355.

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

Janecky DR, Seyfried WE (1984). Formation of massive sulfide deposits on oceanic ridge crests - incremental reaction models for mixing between hydrothermal solutions and seawater. *Geochimica Et Cosmochimica Acta* **48:** 2723-2738.

Johnson JW, Oelkers EH, Helgeson HC (1992). SUPCRT92: A software package for calculating the standard molal thermodynamic properties of minerals, gases, aqueous species, and reactions from 1 to 5000 bar and 0 to 1000°C. *Computers & Geosciences* **18:** 899-947.

Karkhoff-Schweizer RR, Bruschi M, Voordouw G (1993). Expression of the γ-subunit gene of desulfoviridin-type dissimilatory sulfite reductase and of the α- and β-subunit genes is not coordinately regulated. *European Journal of Biochemistry* **211:** 501-507.

Karkhoff-Schweizer RR, Huber DP, Voordouw G (1995). Conservation of the genes for dissimilatory sulfite reductase from Desulfovibrio vulgaris and Archaeoglobus fulgidus allows their detection by PCR. *Appl Environ Microbiol* **61:** 290-296.

Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, Abicht H *et al* (2001). Multiple Lateral Transfers of Dissimilatory Sulfite Reductase Genes between Major Lineages of Sulfate-Reducing Prokaryotes. *Journal of Bacteriology* **183:** 6028-6035.

Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D *et al* (2009). Circos: An information aesthetic for comparative genomics. *Genome Research*.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J*.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86-89.

Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003). Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

Marcus MA, MacDowell AA, Celestre R, Manceau A, Miller T, Padmore HA *et al* (2004). Beamline 10.3.2 at ALS: a hard X-ray microprobe for environmental and materials sciences. *Journal of Synchrotron Radiation* **11:** 239-247.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D *et al* (2008). IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Research* **36:** D534-D538.

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

McCollom TM, Shock EL (1997). Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems. *Geochimica Et Cosmochimica Acta* **61:** 4375-4391.

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

Mottl MJ, Seewald JS, Wheat CG, Tivey MK, Michael PJ, Proskurowski G *et al* (2011). Chemistry of hot springs along the Eastern Lau Spreading Center. *Geochimica et Cosmochimica Acta* **75:** 1013-1038.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40:** e155.

Newton IL, Woyke T, Auchtung TA, Dilly GF, Dutton RJ, Fisher MC *et al* (2007). The Calyptogena magnifica chemoautotrophic symbiont genome. *Science* **315:** 998-1000.

Oliveira TF, Vonrhein C, Matias PM, Venceslau SS, Pereira IAC, Archer M (2008). The Crystal Structure of Desulfovibrio vulgaris Dissimilatory Sulfite Reductase Bound to DsrC Provides Novel Insights into the Mechanism of Sulfate Respiration. *Journal of Biological Chemistry* **283:** 34141-34149.

Peng Y, Leung HCM, Yiu SM, Chin FYL (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28:** 1420-1428.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476:** 176-180.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R *et al* (2005). InterProScan: protein domains identifier. *Nucleic Acids Research* **33:** W116-W120.

Reinthaler T, van Aken HM, Herndl GJ (2010). Major contribution of autotrophy to microbial carbon cycling in the deep North Atlantic's interior. *Deep Sea Research Part II: Topical Studies in Oceanography* **57:** 1572-1580.

Robie RA, Hemingway BS, Fisher JR (1979). *Thermodynamic properties of minerals and related substances at 298.15 K and 1 Bar (10 Pascals) pressure and at higher temperatures. Bulletin 1452*. U.S. Geological Survey: Reston, VA.

Saccocia PJ, Seyfried Jr WE (1994). The solubility of chlorite solid solutions in 3.2 wt% NaCl fluids from 300-400°C, 500 bars. *Geochimica Et Cosmochimica Acta* **58:** 567-585.

Sangal V, Fineran PC, Hoskisson PA (2013). Novel configurations of type I and II CRISPR–Cas systems in Corynebacterium diphtheriae. *Microbiology* **159:** 2118-2126.

Schmieder R, Lim YW, Edwards R (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28:** 433-435.

Seewald J, McCollom T, Proskurowski G, Reeves E, Mottl M, Sharkey J *et al* (2005). Aqueous Volatiles in Lau Basin Hydrothermal Fluids. *EOS Transactions, AGU* **86:** 52.

Sheik CS, Jain S, Dick GJ (2013). Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology***:** n/a-n/a.

Shock EL, Helgeson HC (1988). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Correlation algorithms for ionic species and equation of state predictions to 5 kb and 1000°C. *Geochimica Et Cosmochimica Acta* **52:** 2009-2036.

Shock EL, Helgeson HC, Sverjensky DA (1989). Calculation of the thermodynamic and transport properties of aqueous species at high pressures and temperatures: Standard partial molal properties of inorganic neutral species. *Geochimica Et Cosmochimica Acta* **53:** 2157-2183.

Shock EL, Sassani DC, Willis M, Sverjensky DA (1997). Inorganic species in geologic fluids: Correlations among standard molal thermodynamic properties of aqueous ions and hydroxide complexes. *Geochimica Et Cosmochimica Acta* **61:** 907-950.

Šimoliūnas E, Kaliniene L, Truncaitė L, Zajančkauskaitė A, Staniulis J, Kaupinis A *et al* (2013). Klebsiella Phage vB_KleM-RaK2 — A Giant Singleton Virus of the Family Myoviridae. *PLoS One* **8:** e60717.

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

Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW (2006). Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts. *PLoS Biol* **4:** e234.

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR *et al* (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12:** 3035-3056.

Sverjensky DA, Shock EL, Helgeson HC (1997). Prediction of the thermodynamic properties of aqueous metal complexes to 1000°C and 5 kb. *Geochimica Et Cosmochimica Acta* **61:** 1359-1412.

Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for Chemolithoautotrophy Among Ubiquitous Bacteria Lineages in the Dark Ocean. *Science* **333:** 1296-1300.

Talley LD (ed) (2007) *Hydrographic atlas of the World Ocean Circulation Experiment (WOCE)*. . International WOCE Project Office: Southampton, UK

Von Damm KL, Edmond JM, Measures CI, Grant B (1985). Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochimica Et Cosmochimica Acta* **49:** 2221-2237.

Wagman DD, Evans WH, Parker VB, Schumm RH, Halow I, Bailey SM *et al* (1982). *The NBS tables of chemical thermodynamic properties : selected values for inorganic and C1 and C2 organic substances in SI units*, vol. 11, supplement no. 2. American Chemical Society and the American Institute of Physics for the National Bureau of Standards: Washington, D.C.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al* (2009). Metagenome of a Versatile Chemolithoautotroph from Expanding Oceanic Dead Zones. *Science* **326:** 578-582.

Weiss RF, Craig H (1973). Precise shipboard determination of dissolved nitrogen, oxygen, argon, and total inorganic carbon by gas chromatography. *Deep Sea Research and Oceanographic Abstracts* **20:** 291-303.

Weissgerber T, Zigann R, Bruce D, Chang YJ, Detter JC, Han C *et al* (2011). Complete genome sequence of Allochromatium vinosum DSM 180(T). *Stand Genomic Sci* **5:** 311-330.

Weissgerber T, Dobler N, Polen T, Latus J, Stockdreher Y, Dahl C (2013). Genome-Wide Transcriptional Profiling of the Purple Sulfur Bacterium Allochromatium vinosum DSM 180T during Growth on Different Reduced Sulfur Compounds. *Journal of Bacteriology* **195:** 4231-4245.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:** 821-829.

Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC *et al* (2013). Abundant SAR11 viruses in the ocean. *Nature* **494:** 357-360.

Acknowledgements: This project is funded in part by the Gordon and Betty Moore Foundation Grant GBMF2609 and National Science Foundation Grant OCE1038006. We also thank the University of Michigan Rackham Graduate School Faculty Research Fellowship Program for their support and Meng Li and Sunit Jain for their assistance. DNA sequencing was conducted at the University of Michigan DNA Sequencing Core. We thank the captain and crew of R/V *Thomas G. Thompson* for logistical support. For assistance with synchrotron measurements at BL10.3.2, we thank K.J. Edwards, J.V. Sorensen, M.A. Marcus, and S.C. Fakra. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The nucleotide sequences are available from DOE JGI-IMG/MER - Taxon Object IDs (Kilo Moana: 3300001680, Abe: 3300001681, Mariner: 3300001678, Tahi Moana: 3300001679, Tui Malila: 3300001676 and Guaymas: 3300001683) and NCBI (GenBank- Lau218 : KJ183191-3, Lau87: XXXX, Lau85: XXXX, Lau77: XXXX and sequence read archive XXXX). The authors declare no competing financial interests.

CHAPTER V

CONCLUSIONS AND FUTURE DIRECTIONS

5.1 Introduction

In this Chapter, I draw some wide-ranging inferences, describe the potential impact of my research on the field of microbial ecology and look towards the future to what I believe to be interesting directions and ultimate goals for this field. My dissertation has applied a wide variety of tools and approaches to answer questions regarding the microbial ecology of hydrothermal plumes and the deep-sea: Which microbes inhabit hydrothermal plumes? What geochemical transformations are they capable of? What is the role of geochemistry in shaping the diversity of microbes? Can we identify viruses infecting ubiquitous chemolithotrophs?

To address these questions, **Chapter II** described the energy metabolism and *in situ* transcriptional responses of two SUP05 bacteria to the disparate geochemical environments of the hydrothermal plumes of Guaymas Basin and surrounding background deep waters (in the Guaymas and Carmen Basins). This was the first genome and transcriptome of a hydrogen oxidizing microorganism in the deep ocean water column. This chapter also served as a proof of concept to validate my approach and set the stage for studying the microbial community *en masse* in the hydrothermal plumes of the Eastern Lau Spreading Center. **Chapter III** elucidated the diverse microbial community composition and the energy metabolisms along the

geochemical gradient at the Eastern Lau Spreading Center in the Western Pacific Ocean. Analyses of genome sequences highlighted the dominance of sulfur-based energy metabolism in plumes across the ELSC, revealed extensive functional redundancy associated with reduced sulfur species and discovered several novel microorganisms that had not previously been observed in the deep oceans. **Chapter IV** described the identification of five distinct viruses that infect sulfur-oxidizing SUP05 bacteria. This chapter provides the first evidence of viral auxiliary metabolic genes involved in lithotrophy and implicated viruses as an important cog in the global biogeochemical cycle of sulfur.

5.2 Hydrogen oxidation in the deep ocean

Amongst the most important conclusions that can be drawn from this dissertation is the importance of hydrogen as an electron donor in the deep-sea (Chapter II). The deep ocean (>200m) is significant sink for carbon and is the largest repository of dissolved inorganic carbon (DIC) on our planet (Aristegui et al 2009). Chemosynthesis in deep-sea hydrothermal plumes transforms this DIC into organic carbon and can potentially account for up to 25% of the total organic carbon in the deep oceans (Maruyama et al 1998). Numerous studies have demonstrated the need to identify electron donors for chemosynthetic metabolism, yet our knowledge on this subject had been previously limited to reduced sulfur species (Lesniewski et al 2012), ammonia (Lam et al 2004) and methane (de Angelis et al 1993). Hydrogen offers significant advantages to microbes that possess the capability to use it as an electron donor. Firstly, hydrogen oxidation has the potential to produce the greatest amount of energy per mole amongst all catabolic reactions involving dissolved components (McCollom 2000). Secondly, as a dissolved gas

hydrogen can diffuse easily into the cell and microbes do not require separate mechanisms for uptake as they do for solid species like iron or metal sulfides. In Chapter II, I identified two distinct types of hydrogenases in SUP05 bacteria: the '*hup*' type hydrogenases that are active in environments with high hydrogen concentration (>50nM), and '*hyd*' type hydrogenases that are transcriptionally active in the deep oceans with hydrogen concentrations as little as 0.4 nM. The presence of such distinct mechanisms for use of the same electron donor, hydrogen, potentially allows SUP05 bacteria to thrive across a variety of hydrogen concentrations. Taken together with SUP05's ability to use other electron donors such as reduced sulfur species, multiple electron acceptors (nitrate and oxygen), and multiple carbon sources (organic and inorganic), this reveals an enormous potential for metabolic flexibility in SUP05 bacteria. These results also raise several questions that could serve as avenues of future research. Firstly, the molecular evidence for hydrogen oxidation does not yield any quantitative information on the rate at which hydrogen is oxidized or the amount of carbon fixation that can be derived from it. Rather, these findings provide the impetus for *in situ* geochemical measurements of hydrogen oxidation rates, which could help reconcile gaps in our knowledge of the deep ocean carbon cycle (Aristegui et al 2009, Burd et al 2010). Secondly, transcriptomic evidence for utilization of hydrogen and reduced sulfur species fails to quantitatively identify the relative importance of electron donors for carbon fixation and thereby leaves open the possibility of niche differentiation within the diverse SUP05 group for utilizing the different electron donors. Further studies at the cellular level and rate measurements of sulfur and hydrogen oxidation in conjunction with carbon fixation are necessary to definitively resolve this question.

5.3 Complexity of microbial communities inhabiting hydrothermal plumes

An important conclusion that can be drawn from my dissertation is that deep-sea hydrothermal plumes host a complex microbial community of bacteria, archaea, eukarya and viruses. In the case of bacteria and archaea, community complexity is due in part to the simultaneous availability of multiple energy sources for chemosynthesis. In Chapter III, another conclusion was made based on DNA sequence-based analyses that show that differences in the geochemistry of hydrothermal vents do not manifest in microbial community diversity, which displays only minor variance across the ELSC geochemical gradient. Sulfur oxidation dominates microbial metabolism and supports a diversity of microorganisms that provide intriguing insights into metabolic versatility and functional redundancy in the microbial community. For example, the occurrence of sulfur oxidation genes in hydrogen and methane oxidizing organisms hints at metabolic plasticity and opportunism in chemolithotrophic microorganisms in the deep-sea that would enable them to respond to fluctuating redox environments. In addition, metabolic versatility associated with sulfur oxidation has the secondary effect of imparting functional redundancy associated with sulfur oxidation in the microbial community (Allison and Martiny 2008). In combination, this potentially allows the microbial metabolism of sulfur oxidation to be resilient and dominate hydrothermal plume environments. Hydrothermal plumes represent a dynamic setting that is heavily influenced by the geography of water masses, ocean currents and changes in hydrothermal activity. The abundance, diversity, and metabolic flexibility of sulfur oxidizing microbes suggest that the function of sulfur oxidation will be robust to any changes in community composition due to environmental changes such as the above perturbations. Lastly, the recent elucidation of a cryptic sulfur cycle in oxygen minimum zones attributed to microbial communities similar to those described here (Canfield et al 2010) suggests that metabolic

versatility and functional redundancy of sulfur oxidation may also play important roles in other marine ecosystems.

Another important insight from Chapter III is that although the microbial communities inhabiting hydrothermal plumes are dominated by microorganisms from the pelagic oceans as observed previously (Dick and Tebo 2010, Lesniewski et al 2012), they also contain novel organisms that were previously unknown in marine environments or detected only in seafloor systems. There are two likely explanations for this discrepancy. Firstly, improved sequence based analyses allow us to capture and explore a greater proportion of the microbial community than previously possible. Secondly, although there is now a growing body of evidence suggesting hydrothermal plume communities are not seeded entirely by seafloor communities as suggested previously (Winn et al 1986), my dissertation suggests that hydrothermal plume communities, while relatively stable, are seeded by a complex yet unresolved interplay of pelagic ocean and seafloor communities. Further studies involving the complementary approaches of fluid dynamics models and microbiological observations could shed light on mechanism of entrainment of microorganisms in hydrothermal plumes and definitively resolve this question.

Clearly, considering the impact of hydrothermal plumes on the global oceans and their elemental cycles (Kadko 1993), hydrothermal plume microbial communities need to be studied further. Although the work presented in this dissertation (Chapter II and Chapter III) was cultivation independent, it provides insights into potential strategies for cultivation. Pure cultures of microorganisms open up new avenues of research and could fill important gaps in our knowledge involving microbial rates of geochemical transformations and carbon fixation that cultivation-independent approaches cannot currently fill. While the work in Chapter III represents the most highly resolved spatial sampling of rising hydrothermal plumes to date, it

still represents relatively coarse scales and leaves room for further advancements. Future improvements should involve a more concerted sampling effort of rising hydrothermal plumes and potential sources of plume microbes and advanced sampling tools involving *in situ* fixation that would enable further DNA and RNA-based studies of microbial metabolism and activities. Such progress would be invaluable in elucidating fine-scale dynamics that underpin the diversity of microbial communities in hydrothermal plumes and the deep oceans. Lastly, although my dissertation advances our understanding of chemolithotrophy in hydrothermal plumes, heterotrophy remains a critically understudied but important component of the microbial community.

5.4 Viruses of chemolithotrophic microorganisms.

This dissertation provides the first evidence for existence of viruses of chemolithotrophic bacteria in the deep oceans (Chapter IV). The sequence-based discovery of phages that infect a widespread deep-sea bacterium, SUP05, amongst a complex microbial community of bacteria, archaea, eukarya and viruses can be used as a model to study community dynamics and advance the field of viral ecology that has so far been dependent on culturing as its primary tool. In addition, the discovery of viral AMGs associated with sulfur-based chemolithotrophy provides an unparalleled insight into the role of phage-encoded sulfur oxidation as an ecological strategy for viruses to access abundant elemental sulfur in the environment. However, pure cultures of SUP05 and their viruses are necessary to study the physiology of SUP05 host-phage interactions and validate the underlying mechanisms of phage-influenced sulfur oxidation. Co-cultures of SUP05 and their viruses would open up several new avenues of research in the field of

biogeochemistry and the role of viruses in chemosynthetic microbial communities. Firstly, downstream approaches should include transcriptomic and protein expression studies to identify all putative viral genes associated with sulfur oxidation. Specifically, measurement of expression of DsrA and DsrC proteins from SUP05 viruses during infection and their interaction with intracellular sulfur globules would provide definitive evidence to validate our hypotheses. Secondly, measurements of DsrA and DsrC protein turn-over in SUP05 bacteria during oxidation of elemental sulfur would be an obvious approach to help answer the question of why SUP05 viruses carry AMGs for sulfur oxidation. These studies would also aid construction of models to study SUP05-phage fitness and thereby advance our understanding of this globally relevant microbe. The results from Chapter IV also shed light on a long standing unanswered question in the global sulfur cycle pertaining to widely observed horizontal gene transfer associated with genes for sulfur-cycling by providing a mechanism for such transformations (Klein et al 2001). Our results suggest that this avenue of research should be revisited to consider viruses as the agent of horizontal gene transfer. Further sampling of the *dsr* genes in the environment are necessary to understand the diversity of the viral gene reservoir and develop a more holistic understanding of transfer of *dsr* genes to and from viruses.

In coherence with the vast number of viral AMGs identified already, and consistent observations that viral AMGs relieve metabolic bottlenecks in both microbes and elemental geochemical cycles (Breitbart 2012), it is conceivable that future investigations could identify both viruses and AMGs associated with other important chemolithotrophic microbial metabolisms in hydrothermal plumes such as ammonia, methane, nitrite, iron, and hydrogen oxidation. Finally, this study follows in the footsteps of the first reports of AMGs in viruses in photosynthetic ecosystems in the surface oceans (Lindell et al 2004, Lindell et al 2005, Mann et

al 2003), which have since inspired new avenues of research in the burgeoning, yet understudied field of microbial ecology in the dark oceans.

5.5 References

Allison SD, Martiny JBH (2008). Resistance, resilience, and redundancy in microbial communities. *Proceedings of the National Academy of Sciences*.

Aristegui J, Gasol JM, Duarte CM, Herndl GJ (2009). Microbial oceanography of the dark ocean's pelagic realm. *Limnol Oceanogr* **54:** 1501-1529.

Breitbart M (2012). Marine Viruses: Truth or Dare. *Annual Review of Marine Science* **4:** 425-448.

Burd AB, Hansell DA, Steinberg DK, Anderson TR, Arístegui J, Baltar F *et al* (2010). Assessing the apparent imbalance between geochemical and biochemical indicators of meso- and bathypelagic biological activity: What the @$♯! is wrong with present calculations of carbon budgets? *Deep Sea Research Part II: Topical Studies in Oceanography* **57:** 1557-1571.

Canfield DE, Stewart FJ, Thamdrup B, De Brabandere L, Dalsgaard T, Delong EF *et al* (2010). A cryptic sulfur cycle in oxygen-minimum-zone waters off the Chilean coast. *Science* **330:** 1375-1378.

de Angelis MA, Lilley MD, Baross JA (1993). Methane oxidation in deep-sea hydrothermal plumes of the endeavour segment of the Juan de Fuca Ridge. *Deep Sea Research Part I: Oceanographic Research Papers* **40:** 1169-1186.

Dick GJ, Tebo BM (2010). Microbial diversity and biogeochemistry of the Guaymas Basin deep-sea hydrothermal plume. *Environ Microbiol* **12:** 1334-1347.

Kadko D (1993). An assessment of the effect of chemical scavenging within submarine hydrothermal plumes upon ocean geochemistry. *Earth and Planetary Science Letters* **120:** 361-374.

Klein M, Friedrich M, Roger AJ, Hugenholtz P, Fishbain S, Abicht H *et al* (2001). Multiple Lateral Transfers of Dissimilatory Sulfite Reductase Genes between Major Lineages of Sulfate-Reducing Prokaryotes. *Journal of Bacteriology* **183:** 6028-6035.

Lam P, Cowen JP, Jones RD (2004). Autotrophic ammonia oxidation in a deep-sea hydrothermal plume. *FEMS Microbiology Ecology* **47:** 191-206.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *ISME J* **6:** 2257–2268.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 11013-11018.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86-89.

Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003). Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

Maruyama A, Urabe T, Ishibashi J, Feely R, Baker ET (1998). Global hydrothermal primary production rate estimated from the southern East Pacific Rise. *Cahiers de Biologie Marine* **39:** 249-252.

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

Winn CD, Karl DM, Massoth GJ (1986). Microorganisms in deep-sea hydrothermal plumes. *Nature* **320:** 744-746.