Summer 2019

Functional Analysis of Cyc2 in Microbial Mat Communities of Mariana Arc and Back-arc Hydrothermal Vents

Christina A. Turner
Western Washington University, turner52@wwu.edu

Functional Analysis of Cyc2 in Microbial Mat Communities of Mariana Arc and Back-arc Hydrothermal Vents

By
Christina A. Turner

Accepted in Partial Completion
of the Requirements for the Degree
Master of Science

ADVISORY COMMITTEE

Chair, Dr. Craig Moyer
Dr. Heather Fullerton
Dr. Matt Zinkgraf
Dr. Jeff Young

GRADUATE SCHOOL

David L. Patrick, Acting Dean

MASTER'S THESIS

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Western Washington University, I grant to Western Washington University the non-exclusive royalty-free right to archive, reproduce, distribute, and display the thesis in any and all forms, including electronic format, via any digital library mechanisms maintained by WWU.

I represent and warrant this is my original work, and does not infringe or violate any rights of others. I warrant that I have obtained written permissions from the owner of any third party copyrighted material included in these files.

I acknowledge that I retain ownership rights to the copyright of this work, including but not limited to the right to use all or part of this work in future works, such as articles or books.

Library users are granted permission for individual, research and non-commercial reproduction of this work for educational purposes only. Any further digital posting of this document requires specific permission from the author.

Any copying or publication of this thesis for commercial purposes, or for financial gain, is not allowed without my written permission.

Christina Turner
August 8, 2019

Functional Analysis of Cyc2 in Microbial Mat Communities of Mariana Arc and Back-arc Hydrothermal Vents

A Thesis

Presented to

The Faculty of

Western Washington University

In Partial Fulfillment

Of the Requirements for the Degree

Master of Science

By

Christina Turner

August, 2019

ABSTRACT

The primary production by Zetaproteobacteria at iron-rich vents of the Mariana Arc and back-arc supports high microbial diversity. In microaerophilic circumneutral pH, Zetaproteobacteria are able to produce energy by oxidizing ferrous iron (Fe^{2+}) from the vent effluent with the presence of a key gene involved in their iron-oxidation, Cyc2, a fused cytochrome-porin. These genes are found in all Zetaproteobacteria known to date. This study combines molecular quantitative techniques with metagenomic analyses to enhance our confidence in gene quantification from environmental samples. The occurrence of *cyc2* was quantified from microbial mat communities at NW Eifuku, NW Rota-1, Snail and Urashima Vents using primers designed from a composite metagenome to better understand the potential impact of Zetaproteobacteria metabolism on these iron-rich microbial communities. Based on a multiple sequence alignment and phylogenetic analysis, two distinct groups of the Zetaproteobacteria *cyc2* were identified and termed Cluster 1 *cyc2* and Cluster 3 *cyc2* based on the nomenclature provided by Chan et al. (2018). Multidimensional scaling of Zetaproteobacteria SSU rRNA and *cyc2* gene abundances showed that the presence of the two *cyc2* sequence variants are strongly driven by an array of different vent fluid chemicals. We also report on the phylogenetic relationship between the two *cyc2* variants and other representative FeOB, as well as the distribution of *cyc2* gene abundances across the Mariana hydrothermal vent sites along the Arc and back-arc. Overall, this provides insight on the ecological and evolutionary implications of Zetaproteobacteria iron-oxidation represented at the community-level across a wide-range of geochemical parameters.

ACKNOWLEDGEMENTS

Much appreciation goes out to my advisor, Dr. Craig Moyer, for the insight, opportunities and support that he has provided me. I would also like to thank Dr. Heather Fullerton for her expertise, encouragement and advice, and for all the work she did for the metagenome workflow. Thank you to Dr. Matt Zinkgraf for help with coding and statistics and Dr. Jeff Young for his support and insightful discussions. Thank you to Kevin Hager, as well, for his advice and help with sample information and sequencing. Funding for this study was provided by WWU's Office of Research and Sponsored Programs and Truc and Jerry Thon through the Thon family scholarship for graduate studies in biology.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

There are many different types of iron-oxidizing bacteria (FeOB) that contribute to an abundant and diverse pool of cytochromes involved in iron-oxidation (He et al., 2017; Hedrich et al., 2011). Cytochromes confirmed to play a role in microbial iron-oxidation have been found in the genomes of acidophilic FeOB (*Acidithiobacillus ferrooxidans*) (Yarzábal et al., 2002), phototrophic FeOB (*Chlorobium phaeoferrooxidans*) (Crowe et al., 2017), marine neutrophilic FeOB (Emerson and Moyer, 2002; Emerson et al., 2007) and fresh water neutrophilic FeOB (*Ferriphaselus amnicola*) (Kato et al., 2015). The abundance of confirmed iron-oxidizing cytochromes from metabolically diverse FeOB supports the possibility that there are more cytochromes involved in iron-oxidation that have yet to be discovered.

Zetaproteobacteria are marine neutrophilic FeOB that are widespread and more metabolically diverse than previously thought, with an iron-oxidation mechanism that is still not fully understood (Chiu et al., 2017; McAllister et al., 2019). The byproduct of their Fe^{2+}oxidation, which are insoluble Fe-(oxy)hydroxide biopolymers, form external twisted-stalk, tubular-sheath, or dread-like structures which compose the flocculent orange microbial mats (Chan et al., 2011; Kato et al., 2015; Laufer et al., 2017). Zetaproteobacteria not only provide other vent organisms (macrobes and microbes) with a food source and niche, but they also promote an efficient metabolism by producing a physicochemical gradient of oxygen and Fe^{2+} (Chan et al., 2016), making them the ecosystem engineers of microbial mats at iron-rich hydrothermal vents (Hager et al., 2017). Zetaproteobacteria have been found in marine, subsurface, coastal, and terrestrial habitats, including CO_2-rich terrestrial springs (Emerson et al., 2016; Probst et al., 2016; Probst

et al., 2018), estuaries (Chiu et al., 2017; Field et al., 2016), oceanic crustal fluids (Kato et al., 2009), crustal spreading centers (Scott et al., 2015), seamounts (Fleming et al., 2013), and worm burrows (Beam et al., 2018). Their cosmopolitan distribution and iron-oxidation metabolism suggests that the genes involved in this mechanism may also allow Zetaproteobacteria taxa to adapt to a diverse range of habitats (McAllister et al., 2019).

A key gene involved in Zetaproteobacteria iron-oxidation, *Cyc2* has been identified (Kato et al., 2015; Chan et al., 2018); the abundance of this gene within a microbial community is still unknown. Zetaproteobacteria-specific *cyc2* is an outer-membrane, fused cytochrome-porin that transfers an electron from extracellular Fe^{2+} across the cellular membrane and into the electron transport chain to produce ATP (Barco et al., 2015, Singer et al., 2013). All sequenced Zetaproteobacterial genomes have at least one copy of *cyc2* (McAllister et al., 2019) and the gene was identified in the proteome of the Zetaproteobacteria strain PV-1 ($Cyc2_{PV\text{-}1}$) (Barco et al., 2015). More recently, heterologous expression of *cyc2* showed Fe oxidase activity, confirming its crucial role in the oxidation of Fe^{2+} by Zetaproteobacteria (Chan et al., 2018). The highly conserved N-terminal region encodes the cytochrome, which contains a single heme-binding motif, and the remaining variable sequence of the gene encodes for a porin (Barco et al., 2015; He et al., 2017; Chan et al., 2018). The primary goal of this study is to quantify the presence of *cyc2* from microbial communities along the Mariana Arc and back-arc vents to understand the potential functional versatility of *cyc2*.

The Mariana hydrothermal system has geochemically complex vent habitats that are distributed with varying proximity along the Arc and back-arc volcanoes (Resing et al., 2009) and

harbors diverse microbial communities that are often fueled by the primary production of Zetaproteobacteria (Davis and Moyer, 2008; Hager et al., 2017). The volcanic activity along the Arc and back-arc is due to the convergence and subduction of the Pacific plate under the Philippine plate (Stern et al., 2002). The differential magma sources that feed the vents result in vent fluid with heterogeneous chemical composition, which provides the habitat with an abundant and diverse lithotrophic energy source (Nakagawa and Takai, 2008).

Although Zetaproteobacteria have been found in the microbial communities of both Mariana and Hawaiian hydrothermal systems, the geochemistry and ecology of Lō'ihi Seamount and the Mariana vents are fundamentally different (Davis and Moyer, 2008; McAllister et al., 2011; Yoshikawa et al., 2012). The chemical composition of the vent effluent at Lō'ihi is homogeneous across the system, making the geochemical constraints on the vent communities consistent (Glazer and Rouxel, 2009). Opposingly, the composition of the vent effluent along the Arc and back-arc of the Mariana is highly variable between vent sites, which results in vent communities that are under different geochemical pressures. This difference translates into enhanced geochemical complexity between the two systems and is mainly due to the underlying geology. The microbial communities of Lō'ihi have also been shown to be less diverse than that of the Mariana communities (Moyer et al., 1995; Davis and Moyer, 2008; Hager et al., 2017). Fullerton et al. (2017) identified *cyc2*$_{PV\text{-}1}$ from metagenome-assembled genomes of Zetaproteobacteria taxa of Lō'ihi microbial communities, but no functional gene abundance estimates were conducted. By comparing *cyc2* genomic sequences from ecologically and geochemically different hydrothermal systems, we hope to gain an understanding of *cyc2* distribution to better elucidate ecological

connections among microbial communities at iron-rich vent sites at Lō'ihi Seamount (Central Pacific) and along the Mariana Arc and back-arc (Western Pacific).

This study uses quantitative polymerase chain reaction (qPCR) vetted by metagenomic data to determine the abundance and distribution of the *cyc2* gene, which is indicative of Zetaproteobacteria iron-oxidation. To get the most accurate representation of iron-oxidation by Zetaproteobacteria in the community, primers to amplify *cyc2* were designed using a composite metagenome of Mariana microbial mats from hydrothermal habitats. Sequence analysis of Mariana mat genomic *cyc2* shows two variants of *cyc2* that are evolutionarily distinct, and our analysis will examine if these variants associate to different Mariana vent sites based on local geochemistry parameters. We will also examine if certain vent-derived hydrothermal fluid chemicals potentially have a role in regulating Zetaproteobacteria iron-oxidation at Mariana iron-rich vents, as well as a potentially similar and simpler relationship between vent geochemistry and iron-oxidation occurring at Lō'ihi. Quantifying the abundance of small subunit ribosomal RNA (SSU rRNA) genes across iron-rich Mariana vents will provide further insights regarding the distribution of Zetaproteobacteria. Combined with the abundance of the two *cyc2* genetic variants, the hydrothermal vent conditions that potentially support Zetaproteobacterial growth is suggested. Overall, the ecological implications of Zetaproteobacteria iron-oxidation is characterized at the community level across Mariana Arc and back-arc microbial mats for iron-rich vents.

METHODS

Sample collection and site description

Samples were collected with a remotely operated vehicle (ROV) *Jason* II on board the R/V *Roger Revelle* during cruise 1413 from hydrothermal vents along the Mariana Island Arc and back-arc from the following locations: NW Eifuku, NW Rota-1, Snail and Urashima Vents (Figure 1). Microbial mats were sampled with either a BioMat sampler, a fine-scale syringe sampler (Breier et al., 2012), a scoop, or by using a suction sampling device (Table 1). RNA*later* (ThermoFisher Scientific, Waltham, MA) was either added to the sample remotely at depth or immediately once aboard ship. After initial contact with RNA*later* at 4°C for at least 8 hrs, samples were kept at -80°C until DNA was extracted. Details of samples, sampling sites and geochemistry are shown in Table 1. Geochemistry data at vent locations were collected and analyzed by Hager et al. (2017); however, fluid samples were not collected from Old Iron Slides and Tip Ice so fluids collected from nearby vents were used as a proxy. Two additional samples from Snail Vent, J42-1W and J42-2W, were used in this study that were previously collected (Davis and Moyer, 2008).

Metagenomic sequencing, assembly and binning

DNA was extracted from the microbial mat samples using the manufacturer's protocol for FastDNA SPIN Kit for Soil (MP Biomedicals, Santa Ana, CA) with minor modifications. In the first step, 250 µl of 0.5 M sodium citrate pH 5.8 was added to 728 µl of the supplied sodium phosphate buffer. FastPrep instrument (MP Biomedicals) was used for lysis with two rounds of bead beating for 45 sec at a speed setting of 5.5 for each. Extracts were placed on ice between runs for 30 sec. For the final step, DNA was eluted with 100 µl 1.0 mM Tris pH 8.0. A Qubit 2.0

fluorometer (ThermoFisher Scientific) was used to quantify DNA yields.

Four samples (799B156, 800B12456, 797D234, 801X126) were then purified and concentrated with an Aurora Nucleic Acid Extraction System (Boreal Genomics Inc., Vancouver, BC, Canada) and DNA libraries were prepared with a Nextera DNA Library Kit (Illumina, San Diego, CA). Library size and concentration were quantified with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) before sequencing with an Illumina MiSeq reaction kit v3 (2x300bp) (Illumina Inc., San Diego, CA). Reads were trimmed and filtered for quality control using Trimmomatic version 0.33 with MAXINFO 40:0.3 (Bolger et al., 2014). All trimmed reads were then combined and assembled using IDBA-UD (Peng et al., 2012) into a composite assembly, which was then annotated by IMG/MER (Chen et al., 2019). The annotated composite assembly is available through the IMG database (IMG Genome ID 3300013233).

Assembled contigs were then binned using MaxBin2.0 (Wu et al., 2014). CheckM was used to assign taxonomy and assess completeness of the genome bins (Parks et al., 2015). Reads were mapped back to the bins using Bowtie2 (Langmead and Salzberg, 2013) to determine the abundance of taxonomically placed reads for each of the four samples. ZetaHunter (McAllister et al., 2018) was used to assign Zetaproteobacteria operational taxonomic units (OTU) to binned contigs based on the small subunit ribosomal RNA (SSU rRNA) sequence represented by each bin.

SSU rRNA qPCR analysis

Quantitative polymerase chain reaction (qPCR) was performed for 28 Mariana microbial mat samples using previously described Bacterial and Zetaproteobacterial primers (Jesser et al., 2015) (Table 3). qPCR amplifications were performed in triplicate with a StepOnePlus Real-Time

PCR system (ThermoFisher Scientific). A reaction mixture with a total volume of 20 µl included 10 µl 2X Power SYBR Green Master mix (ThermoFisher Scientific), forward and reverse primers (0.3 µM each) and 1 ng of template gDNA. Cycling conditions were 95°C for 10 min and 40 cycles of 95°C for 15 sec and 50 to 60°C for 1 min. Thermocycling was followed by melting curve analysis from 60 to 95°C. A dilution series of known cloned SSU gene plasmid standards were used as controls after being linearized using the restriction enzyme *Not*I. Gene copy numbers were calculated using estimated C_T values relative to plasmid-control derived standard curves. Relative abundance of Zetaproteobacteria was determined by calculating the ratio between Zetaproteobacteria SSU abundance and Bacteria SSU abundance. QPCR data were used only if primers exhibited better than 90% efficiency and yielded single-peak amplicons during post-PCR melt curve analysis.

Primer design for Zetaproteobacterial *cyc2*

Cyc2 sequences from sequenced Zetaproteobacteria isolate genomes and single amplified genomes (SAG's) were obtained from IMG/MER or GenBank using the gene ID numbers (Supp. Table 1). These *cyc2* sequences were used as reference to search for *cyc2* from the Mariana composite metagenome. The blastp tool was used with each reference sequence to search the composite metagenome for putative *cyc2* genes using an E-value cutoff of 1.00e-5 or greater. Putative *cyc2* genes that did not contain the N-terminal conserved region were removed from further analysis. A total of 4824 genes were identified. Duplicate genes and sequences less than 600 bp were removed. Remaining 509 potential Zetaproteobacteria *cyc2* genes were further filtered by removing genes that did not correspond to a Zetaproteobacteria OTU-designated bin as a

conservative measure to only identify Zetaproteobacterial sequences. Twenty-one confirmed Zetaproteobacterial *cyc2* genes were identified in the Mariana composite metagenome.

Nucleotide sequences of the *cyc2* genes were aligned using Geneious Alignment (Geneious v.10.2.6, global alignment with free end gaps, cost matrix: 70% similarity, gap open penalty: 25, gap extension penalty: 15, refinement iterations: 6). When aligned, the *cyc2* sequences grouped into two distinct clusters, referred as Cluster 1 and Cluster 3 *cyc2* for this study (Chan et al., 2018). Separate primer sets were designed for Cluster 1 and Cluster 3 *cyc2* sequences for both cloning and qPCR assays. Primers for the qPCR assays are nested within the cloning primers and do not overlap (Supp. Fig.1).

Cloning for qPCR standards

Primers designed on the composite metagenome were used to amplify Cluster 1 and Cluster 3 *cyc2* from Mariana microbial mat samples using the respective cloning primers shown in Table 3. PCR amplifications were done in duplicate using a reaction mixture of 10X PCR buffer, 2.5 mM $MgCl_2$, 200 µM each deoxynucleoside triphosphate (dNTP), 1 µM each forward and reverse primer, 10 µg of bovine serum albumin (BSA), 1 µM *Taq* DNA polymerase, 5 ng of DNA template. Cycling conditions were 95°C for 10 min and 30 cycles of 95°C for 1 min, 60°C for 1.5 min., 72°C for 1 min, and a final elongation step of 72°C for 7 min. PCR products for Cluster 1 and Cluster 3 *cyc2* were visualized on a 1.5% agarose gel. Amplicon sizes are shown in Table 3. Amplicons were purified using AMPure XP beads (Beckman Coulter, Indianapolis, IN). Cleaned PCR products were transformed into pCR8 using pCR™8/GW/TOPO™ TA Cloning Kit with One Shot™ TOP10 *E. coli* (ThermoFisher Scientific). Colonies were screened by PCR amplification with 2x

BlueStar™ PCR Master mix (MCLAB, San Francisco, CA) using M13F and M13R primers. Positive clones were then grown in LB spectinomycin broth overnight at 37℃ while shaken. Plasmids were purified using QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA). Correct insert size was checked for with a restriction enzyme digest using *Eco*R1. A final step to confirm successful *cyc2* insertion was done by amplifying a nested region of the insert from the Cluster 1 and Cluster 3 *cyc2* plasmids with their respective qPCR primers shown in Table 3. Plasmid digests and PCR products were visualized on a 2% agarose gel and clones with the correct size and type of insert were sequenced using M13F and M13R primers. Two putative positive *cyc2* clones were selected that shared sequence similarity (E-value > 1.00e-05) with Zetaproteobacterial genes from the NCBI database, one from Cluster 1 and one from Cluster 3. These were then used as our Zetaproteobacteria *cyc2* standards after plasmids were linearized with *Eco*R1, for all of our *cyc2* qPCR assays.

Cyc2 qPCR analysis

qPCR amplifications were performed in triplicate with a StepOnePlus Real-Time PCR system (ThermoFisher Scientific). Primer pairs specific for Cluster 1 *cyc2* and Cluster 3 *cyc2* were used (Table 3). A reaction mixture with a total volume of 20 µl included 10 µl 2X Power SYBR Green Master mix (ThermoFisher Scientific), forward and reverse primers (0.3 µM each) and 1 ng of template DNA. Cycling conditions were 95°C for 10 min and 40 cycles of 95°C for 30 s and 55°C for 30 s, followed by melting curve analysis from 60 to 95°C. qPCR results were used only if primers exhibited >72% efficiency and yielded single-peak amplicons with consistent melting temperatures during the post-PCR melt curve analysis.

Statistical analyses

Non-parametric statistical analyses were done to assess ecological relationships based on differences in gene abundance and environmental variables across microbial communities. Kendall's Tau correlation coefficients were determined between *cyc2* and SSU gene copy numbers (Supp. Table 2). Univariate Shapiro-Wilk's normality tests were run for SSU rRNA and cyc2 gene copy numbers. Gene copy numbers were log10 transformed to better fit a normal distribution. Five one-way non-parametric multivariate analyses of variance (MANOVA) were run for *cyc2* gene copy numbers and independent variables of Zetaproteobacteria percent abundance, vent field, site location, sample location, and geochemistry (Supp. Table 3). Samples that had non-detectable qPCR gene copy numbers were considered to be below the limit of detection and removed from any statistical analyses. MANOVA was chosen based on the non-normal distribution of the gene copy numbers and to analyze two continuous response variables against different single factors. Resulting p-values from the multivariate analysis of variance was adjusted for multiple comparisons using a False Discovery Rate (FDR) of 0.05 with the Benjamini-Hochberg method. This method was chosen over the Bonferroni method due to the relatively small pool of variables tested. The abundance of Cluster 1 *cyc2*, Cluster 3 *cyc2* and relative abundance of Zetaproteobacteria were ordinated using non-parametric multidimensional scaling (NMDS) analysis with the function metaMDS using the Bray-Curtis dissimilarity matrices in the R package Vegan v2.5-4 (Bray and Curtis, 1957; Oksanen et al., 2019). Only the independent variables with a p-value of less than 0.05 from the MANOVA analyses were plotted. Geochemical variables that showed significance from the MANOVA tests were ordinated using the function env_fit in Vegan

v2.5-4 and plotted onto the gene abundance ordination. The function env_fit combines environmental variables as a vector to an ordination, which statistically tests the correlation of the geochemical variables with the sample sites that are ordinated based on gene abundance. The significance of the variables was assessed using 10,000 permutations and p-values were corrected using an FDR of 0.05 with the Benjamini-Hochberg method. NMDS plots were made using the R package ggplot2 v3.1.1. Correlation among environmental variables (temperature, pH, geochemistry) was calculated using the cor() function in R v1.0.143.

Phylogenetic analysis

An alignment of 87 amino acid *cyc*2 sequences from our Mariana composite metagenome, Lō'ihi composite metagenome (Fullerton et al., 2017), Zetaproteobacteria SAGs (Field et al., 2015), Zetaproteobacteria isolates (Singer et al., 2011; Field et al., 2015; Fullerton et al., 2015; Makita et al., 2016; Chiu et al., 2017; Mori et al., 2017), and a few representative *Proteobacteria* (Chan et al., 2018) was made with MAFFT v7.388 (algorithm used: FFT-NS-I x1000, scoring matrix: BLOSUM62, Gap open penalty: 1.53, Offset value: 0.123; Katoh and Standley, 2013) (Supp. Table 5). Sequences missing the N-terminal conserved region or with less than 100 amino acids were not used. The resulting alignment was used to create a maximum likelihood bootstrapped phylogenetic tree with RAxML 8.2.11 (Protein model used: GAMMA BLOSUM62, algorithm: rapid bootstrapping, number of starting trees or bootstrap replicates: 100, parsimony random seed: 1) (Stamatakis, 2014). A consensus tree was made from 100 bootstrapped trees with a support threshold of 70%.

RESULTS AND DISCUSSION

Sample information

NW Eifuku is farthest from the other three vent fields and is at the northern tip of the island arc approximately 750 km north of NW Rota-1. NW Rota-1 is also located on the Arc approximately 125 km north of Guam and approximately 200 km NE from Urashima and Snail. Snail is on the back-arc spreading center and Urashima is off-axis of the spreading center approximately 5 km SE of Snail. Within the NW Eifuku Vent field, Yellow Cone Marker 124 and Marker 146 are approximately 12.5 m apart. At NW Rota-1, Tip Ice and Old Iron Slides are approximately 150 m apart. Golden Horn was approximately 10 m from the adjoining Saipanda Horn and SnapSnap Vent sites within the Urashima Vent field.

Urashima Vent field has chimney structures with layers of flocculent orange microbial mat, one of the largest of these structures is Golden Horn chimney at 13 m high (Nakamura et al., 2013; Toki et al., 2015). Golden Horn sample sites were located at the top, middle and base of this chimney structure. Although there are no large chimney structures, NW Eifuku Yellow Cone sites have small chimneys composed of yellow to orange microbial mats. Tip Ice at NW Rota-1 was primarily dominated by microbial mats that are thin and white. Old Iron Slides at NW Rota-1 had yellow-orange thin tufts of microbial mat. The Snail vent field has thick orange mats with a black surface that is most likely characteristic of the presence of manganese (Hager et al., 2017).

Geochemistry of vent sites

Geochemistry was analyzed from vent fluid collected either at or very near where the microbial mat samples were collected (Table 2). The pH of the vent fluids ranged from 5.41 to

6.66 and the pH of two mid-water background samples were 7.51 and 7.55, respectively. The concentration of total hydrogen sulfide (H_2S) was abundant at Tip Ice (816 μM) and Old Iron Slides (190 μM), both located within the NW-Rota-1 Vent field. The remaining vent sites ranged from <0.2 to 4.0 μM total H_2S. Molecular hydrogen (H_2) was detected at Champagne Vent at NW Efuku and was not detected at NW Rota-1 nor Snail Vents. At Urashima, the vent fluids with the highest concentration of H_2 was found at Golden Horn base and it decreased towards the top of the chimney. The concentration of H_2 at the middle of Golden Horn (16 nM) is only 1 nM above that of the mid-water background H_2 concentration (15 nM).

NW Eifuku had the greatest concentration of iron (Fe), with 96.7 μM at Yellow Cone 146 and 221 μM at Yellow Cone 124, respectively. Snail Marker 108 had the lowest concentration of molybdenum (Mo) and all the Urashima sites had high Mo concentrations. Arsenic (As) was most concentrated at Golden Horn top. Uranium (U) concentrations were lowest at Yellow Cone Marker 124 and highest at the base of Golden Horn. Marker 108 at Snail had vent fluid with the highest temperature at 36°C, while Golden Horn base has the lowest temperature that was sampled at 10°C. Based on Pearson's correlation values (Supp. Fig. 2), temperature is negatively correlated with the concentration of Mo and positively correlated with silicon (Si). No other chemicals show a strong correlation with temperature that were measured.

Cyc2 sequences

The Zetaproteobacteria specific cytochrome-porin fused outer-membrane protein, Cyc2, that contained a conserved motif from positions 28 to 44 in the amino acid sequence were considered unique and used to distinguish among Cyc2 homologs from other outer-membrane

cytochromes (Supp. Fig. 1). Full length Cyc2 sequences range from approximately 350 to 580 amino acids (Chan et al., 2018). Zetaproteobacteria SAG and isolate Cyc2 range from 88 to 475 amino acids (Supp. Table 1) and Mariana composite metagenome Cyc2 range from 175 to 492 amino acids in length (Supp. Table 5). The putative Cyc2 identified from the Mariana composite metagenome were either "hypothetical proteins" or "cytochromes" in IMG/MER annotation.

In both the phylogenetic tree (Fig. 2) and the multiple sequence alignment (Supp. Fig. 1), our *cyc2* data grouped into two distinct clusters. Both Cluster 1 and Cluster 3 Cyc2 (using the nomenclature from Chan et al., 2018) from the Mariana composite metagenome contain a single-heme binding motif in the N-terminal region (Supp. Fig. 1). The two *cyc2* gene clusters can be distinguished based on the third residue of this N-terminal single heme-binding motif; Cluster 1 Cyc2 has a CLXCH sequence and Cluster 3 Cyc2 is either CNXCH or CSXCH in sequence (Supp. Fig. 1). Our Cluster 1 *cyc2* clone contains a 721 bp insert and the closest BLAST was to a hypothetical protein AUJ58_02665 from a Zetaproteobacteria bacterium (CG1_02_55_237) and our Cluster 3 *cyc2* clone contains a 968 bp insert with its closest similar sequence to a probable cytochrome c1 precursor protein from *Mariprofundus ferrooxydans* PV-1 (Supp. Table 6). Based on these BLAST results with an E-value cutoff of greater than 1.00e-5, our two cloned *cyc2* genes are confirmed to be of Zetaproteobacteria origin.

The highly conserved residues in the Mariana composite metagenome Cyc2 amino acid sequences (shown in Supp. Fig. 1) are also found in the Cyc2 sequences from Zetaproteobacteria isolates and SAGs (description of isolates and SAGs shown in Supp. Table 1). Previous studies have confirmed that the N-terminal conserved region (residues 28-39), the heme-binding motif

(residues 40-44), and the C-terminal variable region (from about residue 73 and onward) encodes an outer-membrane cytochrome-porin fused protein (Barco et al., 2015; Kato et al., 2015; Chan et al., 2018). Based on the shared pattern in sequence conservation and variability, there is strong evidence that the Cyc2 found in the Mariana microbial mat communities also encodes for a cytochrome-porin fused outer-membrane protein.

SSU rRNA and *cyc2* gene abundance

Zetaproteobacteria relative abundance of the Mariana Arc and back-arc microbial mats varied from 0.6 % to 72.4 % (Table 1). The microbial community sampled from Tip Ice had the lowest relative abundance of Zetaproteobacteria (0.6 %), which can be explained by the high concentration of H_2S in the vent effluent (Table 2; Supp. Table 7). This finding corresponds with a previous study of Mariana region microbial communities occasionally being dominated by sulfur-oxidizers resulting in relatively low community complexity (Hager et al., 2017). Also, the microbial community at Tip Ice does not associate with either Cluster 1 or Cluster 3 *cyc2* abundance, supported by the statistically insignificant p-value ($p\text{-value} > 0.05$) and by the lack of clustering in ordination space shown in Fig. 3 (Supp. Table 4). Combined, these findings suggest that sulfur-oxidizer abundant microbial communities are negatively associated with Zetaproteobacteria *cyc2* abundance.

QPCR assays of Zetaproteobacteria SSU, Bacteria SSU, Cluster 1 *cyc2* and Cluster 3 *cyc2* gene abundance shows a greater number of *cyc2* genes than SSU genes in all the Mariana microbial mat samples examined (Supp. Table 7). Furthermore, Cluster 1 *cyc2* abundance is consistently greater than Cluster 3 *cyc2* abundance in all mat samples analyzed. In addition, non-parametric

Kendall's tau correlation analyses on *cyc2* and SSU gene copy numbers shows that there are significant correlations (p-value < 0.05) between Cluster 1 *cyc2*, Cluster 3 *cyc2*, Zetaproteobacteria SSU, and Bacterial SSU abundances with tau correlation coefficients ranging between 0.34 and 0.54 (Supp. Table 2). This indicates a positive correlation between SSU rRNA and Cluster 1 and Cluster 3 *cyc2* gene abundances. These findings indicate that Cluster 1 *cyc2* is most abundant in the communities, and that there is potentially greater diversity of *cyc2* variants in the iron-rich communities that have not yet been identified and are being represented by the Cluster 1 and Cluster 3 *cyc2* qPCR primers.

Confirmation of Zetaproteobacteria specificity of all the *cyc2* cloning and qPCR primers were done by testing the primers *in silico* on *cyc2* sequences from Zetaproteobacteria isolates and SAGs, and on cytochromes of non-Zetaproteobacteria FeOBs. An additional confirmation step was done by testing the *cyc2* cloning and qPCR primers on the *cyc2* sequences from the Mariana composite metagenome. From the Zetaproteobacteria-specific *cyc2* primers designed, only the primer pairs that amplified at least 75% of the total *cyc2* sequences occurring in the composite metagenome were used. Linearized plasmids were used to generate a standard curve in both the *cyc2* qPCR assays. Each of these *cyc2* gene inserts were cloned and sequenced from Mariana mat samples. This allows for our estimates of Cluster 1 and Cluster 3 *cyc2* gene copy numbers to most accurately estimate Zetaproteobacteria metabolisms in Mariana microbial communities at iron-rich vents.

Phylogenetic analysis of *cyc2*

Cluster 1 and Cluster 3 *cyc2* sequences from Lō'ihi genomic bins, Mariana genomic bins,

Zetaproteobacteria SAGs and selected isolates, grouped into two distinct clades on a phylogenetic tree (Fig. 2). One clade consists of Cluster 1 *cyc2* sequences, which diverges from cytochromes of non-Zetaproteobacteria FeOB including acidophilic (*Acidithiobacillus ferrooxidans*, *Thiomonas sp.*), freshwater (*Gallionella spp.*, *Sideroxydans lithotrophicus*, *Ferriphaselus amnicola*) and phototrophic (*Chlorobium ferrooxidans*) FeOB. The other clade consists of Cluster 3 *cyc2* sequences and diverges independently of Cluster 1 *cyc2* and non-Zetaproteobacteria FeOB cytochromes. Similar findings confirm that *cyc2* homologs from marine FeOB grouped separately from freshwater FeOB (Kato et al., 2015), and that Cluster 1 *cyc2* was distinctly different from Cluster 3 *cyc2* (Chan et al., 2018). The divergence of Cluster 1 *cyc2* from a metabolically diverse group of FeOB compared to the independent Cluster 3 *cyc2* clade suggests that Cluster 1 *cyc2* could be more functionally diverse than Cluster 3 *cyc2*. Phylogenetic placement also indicates that Cluster 1 *cyc2* has diverged more recently than Cluster 3 *cyc2* (Fig. 2). The microbial samples from Golden Horn top, Golden Horn middle and Saipanda Horn/SnapSnap have a strong association with Cluster 3 *cyc2* gene abundance (Fig. 3). Considering the ancestral phylogenetic placement of Cluster 3 *cyc2*, Zetaproteobacteria iron-oxidation at the Golden Horn top, Golden Horn middle and Saipanda Horn/SnapSnap sites potentially use a more ancestral mode of iron-oxidation that involves the use of predominantly more Cluster 3 *cyc2* genes.

Phylogenetic placement of the Zetaproteobacteria OTU-designated *cyc2* sequences shows no dominant or specific OTU distribution across the two different *cyc2* clusters (Fig. 2), indicating that there is no discernible relationship among Zetaproteobacteria OTU taxonomy and the potential iron-oxidation physiology in Mariana Vent communities. In addition, *cyc2* sequences from the

Lō'ihi and Mariana composite metagenomes share the same phylogenetic relationship and show no distinct groupings (Fig. 2). This suggests that both *cyc2* Cluster 1 and Cluster 3, in iron-rich microbial communities at Lō'ihi and Mariana hydrothermal sites, originated from a shared gene pool.

Environmental drivers of *cyc2* gene abundance

Nonparametric one-way MANOVA analysis between Cluster 1 and Cluster 3 *cyc2* gene abundance with Fe concentration had a p-value greater than 0.05, indicating that Fe does not explain any differences between Cluster 1 *cyc2* and Cluster 3 *cyc2* abundance across these vent sites. Both sites at NW Eifuku sites had the greatest concentration of Fe (Table 2). MANOVA analyses also showed that the difference between Cluster 1 and Cluster 3 *cyc2* gene abundance is significantly correlated with other independent variables: Zetaproteobacteria relative abundance, vent field, sample site, H_2S, NH_4^+, H_2, Mg^{2+}, As, Mo and U (Supp. Table 3). Figure 3 shows that Zetaproteobacteria and Cluster 1 *cyc2* gene abundance is most closely associated with NH_4^+, H_2, and Mg^{2+} and that Cluster 3 *cyc2* gene abundance is most closely associated with Mo, U and As. This indicates that NH_4^+, H_2 and Mg^{2+} may play a role in driving Cluster 1 *cyc2* gene presence, and that Mo, U and As may play a role in driving Cluster 3 *cyc2* gene presence in a given microbial mat community.

Although our vent sites do not cluster into ordination space, there is a slight separation of Urashima sites (Golden Horn top, middle, and Saipanda Horn/SnapSnap) from all others. These Urashima sites are trending towards Cluster 3 *cyc2* gene abundance rather than to Cluster 1 *cyc2*, indicating a close association between these vent sites and Cluster 3 *cyc2* gene presence.

Furthermore, Cluster 3 *cyc2* genes also showed a close association with the communities at Saipanda Horn/SnapSnap sites, which are influenced by greater concentrations of Mo, U and As (Fig. 3). The vent chemistry analyzed from Golden Horn chimney shows that As concentrations are lowest at the base of the chimney and highest at the top, while H_2 concentrations are highest at the bottom and lowest at the top, i.e., inversely proportional (Table 2). Pearson's correlation analysis shows that As and H_2 are negatively associated with a correlation value of approximately -0.5 (Supp. Fig. 2). This suggests that the concentration of H_2 and As are geochemical indicators of Cluster 3 *cyc2* gene abundance and there is, potentially, a relationship between Cluster 3 *cyc2* and Cluster 1 *cyc2* gene presence that can be explained by decreased H_2 and increased As concentration.

The sites that ordinate towards Cluster 1 *cyc2* gene abundance (NW Eifuku, Snail, and NW Rota-1 Vents) do not show distinct differences among each other, indicated by their random distribution in the NMDS plot (Fig. 3). However, these vent sites are significantly associated with Zetaproteobacteria and Cluster 1 *cyc2* gene abundance and H2 concentration, represented by the sample sites trending along these vectors, as well as by p-values of less than 0.05 (Supp. Table 4). Compared to the sites that ordinate towards Cluster 1 *cyc2* gene abundance, only three of the vent sites (all from Urashima) ordinate towards Cluster 3 *cyc2* gene abundance. These findings suggest that Cluster 1 *cyc2* is found in a wider range of iron-rich vent habitats compared to Cluster 3 *cyc2*. Combined with gene abundance data, the NMDS distribution of the vent sites also suggest that Cluster 1 *cyc2* genes are potentially more prominent in the Zetaproteobacteria iron-oxidation metabolism compared to Cluster 3 *cyc2* genes.

Cluster 1 *cyc2* gene abundance is strongly associated with H_2 concentration while Cluster 3 *cyc2* genes are associated with Mo, U and As. These findings indicate that Zetaproteobacteria iron-oxidation could be using different molecular mechanisms based on the vent site and the geochemical constraints, such as H_2, As, Mo and U concentrations. Ammonium (NH_4^+) concentration is highest at Yellow Cone Marker 146 and Mg^{2+} concentration is highest at Yellow Cone Marker 124. Both Yellow Cone Marker sites ordinate towards Zetaproteobacteria and Cluster 1 *cyc2* gene presence, indicating a strong association between NW Eifuku samples with Zetaproteobacteria and Cluster 1 *cyc2* gene abundance (Table 2; Fig. 3). Out of the geochemical constraints that ordinate towards Zetaproteobacteria and Cluster 1 *cyc2* gene abundance, H_2 is most closely associated followed by NH_4^+ then Mg^{2+}. These findings suggest that H_2, NH_4^+ and Mg^{2+} may be potential drivers for iron-oxidation, but specifically for Zetaproteobacteria iron-oxidation that involves Cluster 1 *cyc2* and not Cluster 3 *cyc2* genes.

Zetaproteobacteria relative abundance and Cluster 1 *cyc2* gene abundance are closely associated to each other based on the ordination of *cyc2* and SSU rRNA copy numbers in the microbial communities we examined (Fig. 3). In addition, qPCR analysis of Cluster 1 and Cluster 3 *cyc2* genes showed that Cluster 1 *cyc2* is more abundant than Cluster 3 *cyc2* in all our microbial communities (Supp. Table 7). This indicates that Cluster 1 *cyc2* gene abundance is associated with microbial communities that are relatively abundant in Zetaproteobacteria, suggesting that Cluster 1 *cyc2* might be playing a larger role in Zetaproteobacteria metabolism compared to Cluster 3 *cyc2*.

CONCLUSION

Ecological implications of Zetaproteobacteria iron-oxidation are represented at the community-level across iron-rich vents along the Mariana Arc and back-arc. Here, we quantified the c*yc2* gene from environmental samples using qPCR primers designed from a composite metagenome to most accurately represent the Zetaproteobacteria iron-oxidation in these microbial mat communities. We found two distinct types, or sequence variants, of the *cyc2* gene detected from Zetaproteobacteria bins that were identified from our Mariana metagenomic analysis. Further, both types of Mariana genomic *cyc2* genes are likely cytochrome-porin fused outer-membrane proteins based on shared sequence similarities with findings from previous studies. We confirmed that our Zetaproteobacteria *cyc2* genes were present in Mariana microbial mat communities as well as in the genomes of cultured isolates.

Zetaproteobacteria *cyc2* gene abundance is more closely associated with microbial communities that are dominated by iron-oxidizers rather than sulfur-oxidizers. Between the two types of *cyc2* detected, Cluster 1 *cyc2* is more abundant than Cluster 3 *cyc2* across all Mariana microbial mat communities we investigated. Based on the greater abundance of *cyc2* genes compared to SSU rRNA gene copies, there is the possibility that the *cyc2* genes are as yet more diverse than our current estimates and there may be more *cyc2* sequence variants that have yet to be detected. On the other hand, based on our detailed composite metagenomic analysis, we feel that we have detected the majority of the *cyc2* genes that exist in these hydrothermal habitats. Therefore, our Cluster 1 and Cluster 3 *cyc2* specific qPCR primers can provide a robust initial approximation of *cyc2* gene abundance and thereby provide insights into the iron oxidation

potential of both iron- and sulfur-dominated microbial mat communities.

Environmental parameters play a role in the type and abundance of Zetaproteobacteria *cyc2* present in an iron-rich community. Cluster 1 and Cluster 3 *cyc2* gene abundance are driven by different geochemical constraints across the vent sites we examined with Cluster 1 *cyc2* gene abundance being driven by NH_4^+, H_2, and Mg^{+2} and Cluster 3 *cyc2* gene abundance driven by the presence of Mo, U and As. Cluster 3 *cyc2* gene abundance seems to have a relationship with decreased H_2 concentration and increased As concentration. Cluster 1 *cyc2* genes were found in a wider range of iron-rich vent habitats compared to Cluster 3 *cyc2*, therefore Cluster 1 *cyc2* is more prominent within Zetaproteobacteria iron-oxidation metabolism across all the hydrothermal habitats we studied.

The evolutionary history between Cluster 1 and Cluster 3 Cyc2 suggests a greater functional diversity of Cluster 1 *cyc2* genes and that Cluster 3 *cyc2* genes were ancestral to Cluster 1 *cyc2* genes, as well as to the representative FeOB cytochromes used in our phylogenetic analysis. Hydrothermal vent habitats that potentially have a greater impact from the more ancestral type of iron-oxidation metabolism includes Golden Horn (Top and Middle), Saipanda Horn and SnapSnap, all from the Urashima Vents area. Comparing Mariana Arc and back-arc microbial communities with those from Lō'ihi Seamount showed that both Cluster 1 and Cluster 3 *cyc2* genes are phylogenetically similar at both locations.

This study provides insight regarding the impact of Zetaproteobacterial Cyc2-based iron-oxidation metabolism across hydrothermal vents from two regions in the Pacific. We also have enhanced our understanding regarding the molecular adaptations that have enabled some

Zetaproteobacteria taxa to be widely distributed and others to be niche-specialized. In the future, we hope to expand our approach of using functional gene qPCR assays designed based on metagenomes to examine additional hydrothermal vent microbial communities in more detail with respect to their ecological and evolutionary relationships.

REFERENCES

Barco RA, Emerson D, Sylvan JB, Orcutt BN, Jacobson Meyers ME, Ramírez GA, et al. (2015) 'New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotrophy', Applied and Environmental Microbiology, 81(17), 5927–5937. doi: 10.1128/AEM.01374-15.

Beam JP, Scott JJ, Mcallister SM, Chan CS, Mcmanus J, Meysman FJR, et al. (2018) 'Biological rejuvenation of iron oxides in bioturbated marine sediments', The ISME Journal. Springer US, 12, 1389–1394. doi: 10.1038/s41396-017-0032-6.

Bolger, AM, Lohse, M and Usadel, B. (2014) 'Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data', Bioinformatics, 30(15), 2114–2120. doi: 10.1093/bioinformatics/btu170.

Bray, JR and Curtis, JT. (1957) 'An ordination of the upland forest communities of southern Wisconsin', Ecological monographs, 27(4), 325–349. doi: doi:10.2307/1942268.

Breier JA, Gomez-Ibanez D, Reddington E, Huber JA, Emerson D. (2012) 'A precision multi-sampler for deep-sea hydrothermal microbial mat studies', Deep Sea Research Part I: Oceanographic Research Papers, 70, 83–90. doi: 10.1016/j.dsr.2012.10.006.

Chan CS, Fakra SC, Emerson D, Fleming EJ, Edwards KJ. (2011) 'Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation', The ISME Journal. Nature Publishing Group, 5(4), 717–727. doi: 10.1038/ismej.2010.173.

Chan CS, McAllister SM, Leavitt AH, Glazer BT, Krepski ST, Emerson D. (2016) 'The architecture of iron microbial mats reflects the adaptation of chemolithotrophic iron oxidation in freshwater and marine environments', Frontiers in Microbiology, 7(796). doi: 10.3389/fmicb.2016.00796.

Chan CS, McAllister SM, Garber A, Hallahan BJ, Rozovsky S. (2018) 'Fe oxidation by a fused cytochrome-porin common to diverse Fe-oxidizing bacteria', bioRxiv, 1–37. doi: 10.1101/228056.

Chen I-MA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. (2019) 'IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes', Nucleic Acids Research. Oxford University Press, 47, 666–677. doi: 10.1093/nar/gky901.

Chiu BK, Kato S, McAllister SM, Field EK, Chan CS. (2017) 'Novel pelagic iron-oxidizing Zetaproteobacteria from the Chesapeake Bay oxic-anoxic transition zone', Frontiers in Microbiology, 8, 1–16. doi: 10.3389/fmicb.2017.01280.

Crowe SA, Hahn AS, Morgan-Lang C, Thompson KJ, Simister RL, Llirós M, et al. (2017) 'Draft Genome Sequence of the Pelagic Photoferrotroph Chlorobium phaeoferrooxidans', Genome Announc 5: e01584-16, 5(13), 13–14. doi: 10.1128/genomeA.01584-16.

Davis, RE and Moyer, CL. (2008) 'Extreme spatial and temporal variability of hydrothermal microbial mat communities along the Mariana Island Arc and southern Mariana back-arc system', Journal of Geophysical Research: Solid Earth, 113(8), 1–17. doi: 10.1029/2007JB005413.

Embley RW, Baker ET, Chadwick Jr. WW, Lupton JE, Resing JA, Massoth GJ, et al. (2004) 'Explorations of Mariana Arc Volcanoes Reveal New Hydrothermal Systems', EOS Trans., 85(4), 37–44.

Emerson, D and Moyer, CL. (2002) 'Neutrophilic Fe-Oxidizing Bacteria Are Abundant at the Loihi Seamount Hydrothermal Vents and Play a Major Role in Fe Oxide Deposition', Applied and Environmental Microbiology, 68(6), 3085–3093. doi: 10.1128/AEM.68.6.3085.

Emerson D, Rentz JA, Lilburn TG, Davis RE, Aldrich H, Chan C, et al. (2007) 'A novel lineage of proteobacteria involved in formation of marine Fe-oxidizing microbial mat communities', PLoS One, 2(8). e667. doi: 10.1371/journal.pone.0000667

Emerson JB, Thomas BC, Alvarez W, Banfield JF. (2016) 'Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla', Environmental Microbiology, 18(6), 1686–1703. doi: 10.1111/1462-2920.12817.

Field EK, Sczyrba A, Lyman AE, Harris CC, Woyke T, Stepanauskas R, et al. (2015) 'Genomic insights into the uncultivated marine Zetaproteobacteria at Loihi Seamount', ISME Journal, Nature Publishing Group, 1–14. doi: 10.1038/ismej.2014.183.

Field EK, Kato S, Findlay AJ, MacDonald DJ, Chiu BK, Luther III GW, et al. (2016) 'Planktonic marine iron oxidizers drive iron mineralization under low-oxygen conditions', Geobiology, 14(5), 499–508. doi: 10.1111/gbi.12189.

Fleming EJ, Davis RE, Mcallister SM, Chan CS, Moyer CL, Tebo BM, et al. (2013) 'Hidden in plain sight: discovery of sheath-forming, iron-oxidizing Zetaproteobacteria at Loihi Seamount, Hawaii, USA', FEMS Microbiol Ecol., 85, 116–27. doi: 10.1111/1574-6941.12104

Fullerton H, Hager KW, Moyer CL. (2015) Draft genome sequence of Mariprofundus ferrooxydans strain JV-1, isolated from Loihi Seamount, Hawaii. Genome Announc. 3(5), e01118-15. doi:10.1128/genomeA.01118-15.

Fullerton H, Hager KW, McAllister SM, Moyer CL. (2017) 'Hidden diversity revealed by genome-resolved metagenomics of iron-oxidizing microbial mats from Lo'ihi Seamount, Hawai'i', ISME Journal. Nature Publishing Group, 11(8), 1900–1914. doi: 10.1038/ismej.2017.40.

Hager KW, Fullerton H, Butterfield DA, Moyer CL. (2017) 'Community structure of lithotrophically-driven hydrothermal microbial mats from the Mariana Arc and back-arc', Frontiers in Microbiology, 8(1578). doi: 10.3389/fmicb.2017.01578.

He S, Barco RA, Emerson D, Roden EE. (2017) 'Comparative Genomic Analysis of Neutrophilic Iron(II) Oxidizer Genomes for Candidate Genes in Extracellular Electron Transfer', Frontiers in Microbiology, 8, 1–17. doi: 10.3389/fmicb.2017.01584.

Hedrich S, Schlömann M, Johnson DB. (2011) 'The iron-oxidizing proteobacteria', Microbiology, 157(6), 551–1564. doi: 10.1099/mic.0.045344-0.

Jesser KJ, Fullerton H, Hager KW, Moyer CL. (2015) 'Quantitative PCR analysis of functional genes in iron-rich microbial mats at an active hydrothermal vent system (Lo'ihi seamount, Hawai'i)', Applied and Environmental Microbiology, 81(9), 2976–2984. doi: 10.1128/AEM.03608-14.

Kato S, Yanagawa K, Sunamura M, Takano Y, Ishibashi JI, Kakegawa T, et al. (2009), 'Abundance of Zetaproteobacteria within crustal fluids in back-arc hydrothermal fields of the Southern Mariana Trough', Environ Microbiol., 11(12), 3210–22. doi: 10.1111/j.1462-2920.2009.02031.x

Kato S, Ohkuma M, Powell DH, Krepski ST, Oshima K, Hattori M, et al. (2015) 'Comparative genomic insights into ecophysiology of neutrophilic, microaerophilic iron oxidizing bacteria', Frontiers in Microbiology, 6(1265), 1–16. doi: 10.3389/fmicb.2015.01265.

Katoh, K and Standley, DM. (2013) ‘MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability’, Mol. Biol. Evol., 30(4), 772–780. doi: 10.1093/molbev/mst010.

Langmead, B and Salzberg, SL. (2013) ‘Fast gapped-read alignment with Bowtie 2’, Nat Methods, 9(4), 357–359. doi: 10.1038/nmeth.1923.Fast.

Laufer K, Nordhoff M, Halama M, Martinez R, Obst M, Nowak M, et al. (2017) ‘Microaerophilic Fe(II)-oxidizing Zetaproteobacteria isolated from low-Fe marine coastal sediments – physiology and characterization of their twisted stalks’, Applied and Environmental Microbiology, 83(8), 1–20. doi: 10.1128/AEM.03118-16.

Lutz, RA and Kennish, MJ. (1993) ‘Ecology of deep-sea hydrothermal vent communities: a review’, Reviews of Geophysics, 31(3), 211–242.

Makita H, Kikuchi S, Mitsunobu S, Takaki Y, Yamanaka T, Toki T, et al. (2016) ‘Comparative Analysis of Microbial Communities in Iron-Dominated Flocculent Mats in Deep Sea Hydrothermal Environments’, Applied and Environmental Microbiology, Applied and Environmental Microbiology, 82(19), 5741-5755. doi: 10.1128/AEM.01151-16.

Massoth GJ, de Ronde CEJ, Lupton JE, Feely RA, Baker ET, Lebon GT, et al. (2003) ‘Chemically Rich and Diverse Submarine Hydrothermal Plumes of the Southern Kermadec Volcanic Arc (New Zealand)’, Landscape Evolution: Denudation, Climate and Tectonics over Different Time and Space Scales, edited by K. Gallagher, S. J. Jones, and J. Wainwright, Geol. Soc. Spec. Publ., 219, 119–139. doi: 10.1144/ GSL.SP.2003.219.01.06.

McAllister SM, Davis RE, McBeth JM, Tebo BM, Emerson D, Moyer CL. (2011) 'Biodiversity and emerging biogeography of the neutrophilic iron-oxidizing Zetaproteobacteria', Appl Environ Microbiol, 77(15), 5445–5457. doi: 10.1128/AEM.00533-11.

McAllister, SM, Moore, RM and Chan, CS. (2018) 'ZetaHunter, a Reproducible Taxonomic Classification Tool for Tracking the Ecology of the Zetaproteobacteria and Other Poorly Resolved Taxa', Microbiology Resource Announcements, 7(7), 1–3. doi: 10.1128/MRA.00932-18.

McAllister SM, Moore RM, Gartman A, Luther GW, Emerson D, Chan CS. (2019) 'The Fe (II) - oxidizing Zetaproteobacteria: historical, ecological and genomic perspectives', FEMS Microbiol Ecol., 95(4), 1–18.

McCollom, TM and Shock, EL. (1997) 'Geochemical constraints on chemolithoautotrophic metabolism by microorganisms in seafloor hydrothermal systems', 61(20), 4375–4391.

Mori JF, Scott JJ, Hager KW, Moyer CL, Küsel K, Emerson D. (2017) 'Physiological and ecological implications of an iron- or hydrogen-oxidizing member of the Zetaproteobacteria, Ghiorsea bivora, gen. nov., sp. Nov.', ISME Journal, 11(11), 2624–2636. doi: 10.1038/ismej.2017.132.

Moyer, CL, Dobbs, FC and Karl, DM (1995) 'Phylogenetic Diversity of the Bacterial Community from a Microbial Mat at an Active, Hydrothermal Vent System, Loihi Seamount, Hawaii', Appl Environ Microbiol, 61(4), 1555–1562.

Nakagawa, S and Takai, K. (2008) 'Deep-sea vent chemoautotrophs: Diversity, biochemistry and ecological significance', FEMS Microbiology Ecology, 65(1), 1–14. doi: 10.1111/j.1574-6941.2008.00502.x.

Nakamura K, Toki T, Mochizuki N, Asada M, Ishibasi J, Nogi Y, et al. (2013) 'Discovery of a new hydrothermal vent based on an underwater, high-resolution geophysical survey', Deep Sea Research Part I: Oceanographic Research Papers, 74, 1–10. doi: 10.1016/j.dsr.2012.12.003.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, et al. (2019) 'vegan: Community Ecology Package', R package. Available at: https://cran.r-project.org/package=vegan.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015) 'CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes', Genome Research, 25, 1043–1055. doi: 10.1101/gr.186072.114.Freely.

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(11), 1420–1428. doi: 10.1093/bioinformatics/bts174.

Probst AJ, Castelle CJ, Singh A, Brown CT, Anantharaman K, Sharon I, et al. (2016) 'Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2 concentrations', Environmental microbiology, 00(00), 00–00. doi: 10.1111/1462-2920.13362.

Probst AJ, Ladd B, Jarett JK, Geller-McGrath DE, Sieber CMK, Emerson JB, et al. (2018) 'Differential depth distribution of microbial function and putative symbionts through sediment-hosted aquifers in the deep terrestrial subsurface', Nature Microbiology, 3, 328-336. doi: 10.1038/s41564-017-0098-y.

Resing JA, Baker ET, Lupton JE, Walker SL, Butterfield DA, Massoth GJ, et al. (2009) 'Chemistry of hydrothermal plumes above submarine volcanoes of the Mariana Arc', Geochemistry, Geophysics, Geosystems, 10(2), 1–23. doi: 10.1029/2008GC002141.

de Ronde CEJ, Baker ET, Massoth GJ, Lupton JE, Wright IC, Feely RA, et al. (2001) 'Intra-oceanic subduction-related hydrothermal venting, Kermadec volcanic arc, New Zealand', Earth and Planetary Science Letters, 193(3–4), 359–369.

de Ronde CEJ, Baker ET, Massoth GJ, Lupton JE, Wright IC, Sparks RJ, et al. (2007) 'Submarine hydrothermal activity along the mid-Kermadec Arc, New Zealand: Large-scale effects on venting', Geochemistry, Geophysics, Geosystems, 8(7), 1–27. doi: 10.1029/2006GC001495.

Schloss, P.D and Handelsman, J. (2005) 'Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness', 71(3), 1501–1506. doi: 10.1128/AEM.71.3.1501.

Scott JJ, Breier JA, Iii GWL, Emerson D. (2015) 'Microbial Iron Mats at the Mid-Atlantic Ridge and Evidence that Zetaproteobacteria May Be Restricted to Iron-Oxidizing Marine Systems', 10(3),1–19. doi: 10.1371/journal.pone.0119284.

Singer E, Emerson D, Webb EA, Barco RA, Kuenen JG, et al. (2011) 'Mariprofundus ferrooxydans PV-1 the first genome of a marine Fe(II) oxidizing Zetaproteobacterium', PLoS ONE, 6(9), e25386. doi:10.1371/journal.pone.0025386.

Singer E, Heidelberg JF, Dhillon A, Edwards KJ. (2013) 'Metagenomic insights into the dominant Fe(II) oxidizing Zetaproteobacteria from an iron mat at Lo'ihi, Hawai'i', Front Microbiol., 4(52), 1–9. doi: 10.3389/fmicb.2013.00052.

Stamatakis, A. (2014) 'RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies', Bioinformatics, 30(9), 1312-1313. doi: 10.1093/bioinformatics/btu0332.

Stern RJ. (2002), ‘Subduction zones’, Rev Geophys, 40(4). 1012. doi: 10.1029/2001RG000108.

Toki T, Ishibashi J, Noguchi T, Tawata M, Tsunogai U, Yamanaka T, et al. (2015) ‘Chemical and Isotopic Compositions of Hydrothermal Fluids at Snail, Archaean, Pika, and Urashima Sites in the Southern Mariana Trough’, Subseafloor Biosphere Linked to Hydrothermal Systems: TAIGA Concept, 587–602. doi: 10.1007/978-4-431-54865-2.

Van Dover CL, German CR, Speer KG, Parson LM, Vrijenhoek RC. (2002) ‘Evolution and Biogeography of Deep-Sea Vent and Seep Invertebrates’, Science, 295, 1253–1257. doi: 10.1126/science.1067361.

Wu YW, Tang YH, Tringe SG, Simmons BA, Singer SW. (2014) ‘MaxBin: An automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm’, Microbiome, 2(1), 1–18. doi: 10.1186/2049-2618-2-26.

Yarzábal A, Brasseur G, Ratouchniak J, Lund K, Lemesle-meunier D, Demoss JA, Bonnefoy V. (2002) ‘The High-Molecular-Weight Cytochrome c Cyc2 of Acidithiobacillus ferrooxidans Is an Outer Membrane Protein’, J. Bacteriol., 184(1), 313–317. doi: 10.1128/JB.184.1.313.

Yoshikawa, S, Okino, K and Asada, M. (2012) ‘Geomorphological variations at hydrothermal sites in the southern Mariana Trough: Relationship between hydrothermal activity and topographic characteristics’, Marine Geology, 303–306, 172–182. doi: 10.1016/j.margeo.2012.02.013.

FIGURE LEGENDS

Figure 1. Map of sample sites

Bathymetric map of the Mariana region. Vent fields are circled in black. Sites that samples were collected are from the following respective vent fields. NW Eifuku: Yellow Cone Marker 124, Yellow Cone Marker 146. Urashima: Golden Horn, Saipanda Horn, and Snap Snap. NW Rota-1: Tip ice, Iceberg, and Old Iron Slides. Snail Vent: Marker 108. Modified from Hager et al., 2017.

Figure 2. Zetaproteobacteria *cyc2* protein phylogenetic tree

Maximum likelihood phylogenetic tree of FeOB cytochromes. IMG gene ID numbers shown in parentheses. Bootstrap values out of 100 shown.

Figure 3. NMDS of gene abundance and geochemistry

Nonparametric multidimensional scale plot of Cluster 1 cyc2 (C1), Cluster 3 cyc2 (C3) and relative Zetaproteobacteria abundance (Zeta) gene abundances, represented by black vectors. Geochemical data were ordinated based on association using environmental correlation analysis (grey arrows). Only environmental variables with MANOVA P-values <0.05 were plotted (Supp. Table 6).

Supplemental Figure 1. *cyc2* multiple sequence alignment

Multiple sequence alignment of 38 *cyc2* sequences from Zetaproteobacteria isolates, SAGs and confirmed *cyc2* sequences from the composite Mariana metagenome. Arrows represent the approximate annealing locations of the primers. Colored residues share 50% similarity across all sequences. * represents all residues in that column share 100% similarity. IMG gene ID numbers are shown in parentheses.

Supplemental Figure 2. Correlation of environmental variables. Correlation analysis between environmental variables of geochemistry, pH and temp.

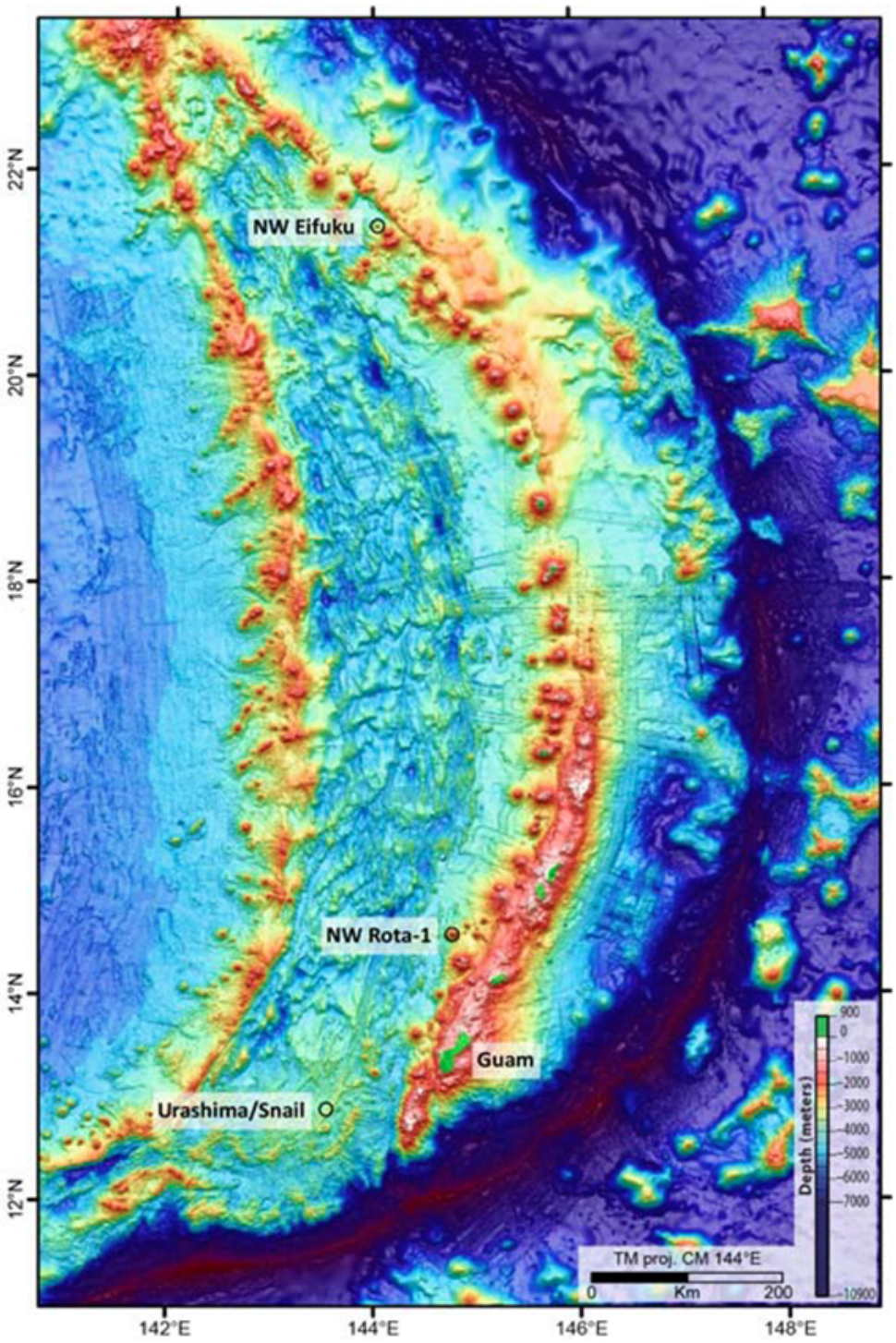

NW Eifuku
NW Rota-1
Guam
Urashima/Snail
22°N
20°N
18°N
16°N
14°N
12°N
142°E
144°E
146°E
148°E
TM proj. CM 144°E
0
Km
200
Depth (meters)
900
0
-1000
-2000
-3000
-4000
-5000
-6000
-7000
-10900

Figure 1.

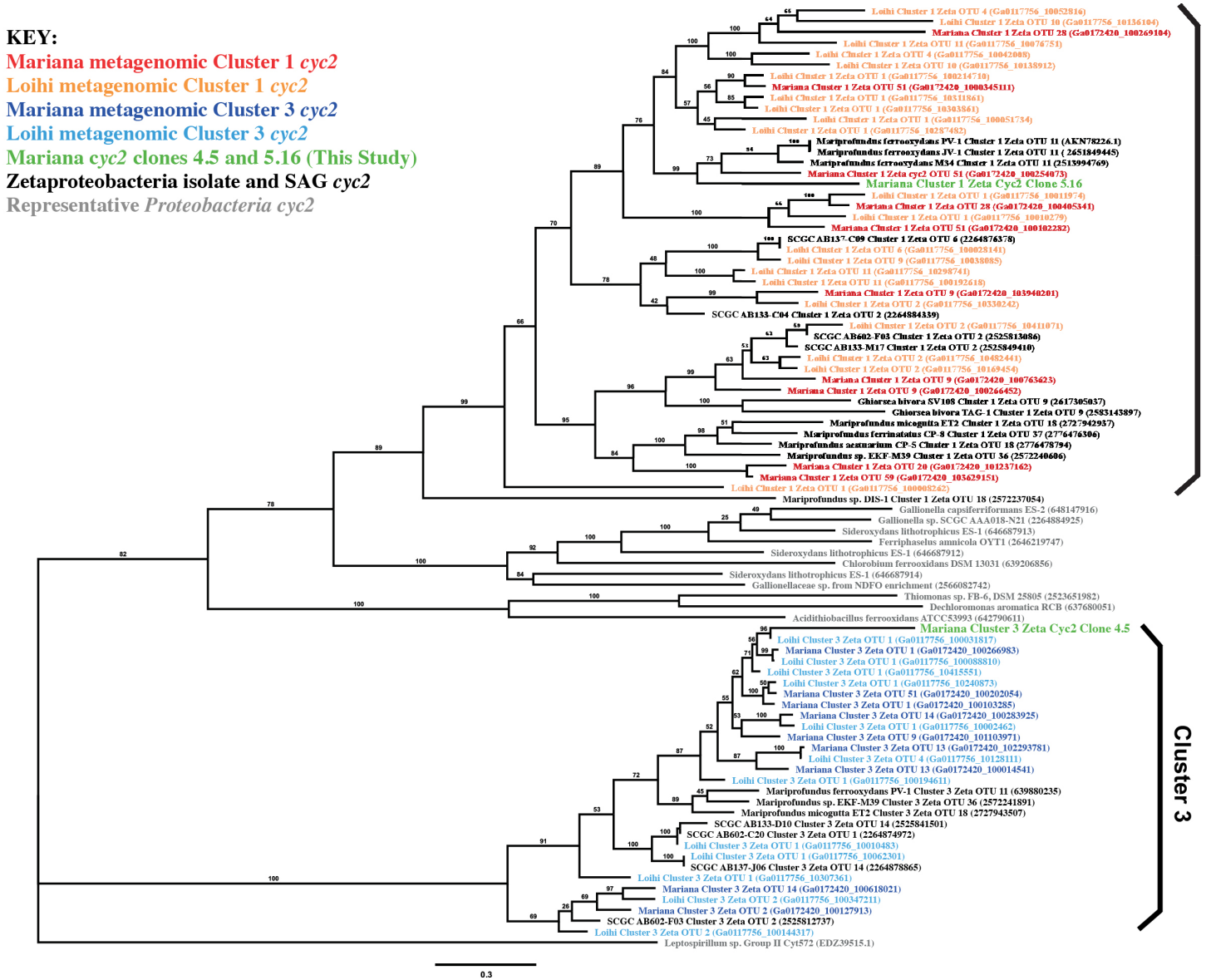

KEY:
Mariana metagenomic Cluster 1 cyc2
Loihi metagenomic Cluster 1 cyc2
Mariana metagenomic Cluster 3 cyc2
Loihi metagenomic Cluster 3 cyc2
Mariana cyc2 clones 4.5 and 5.16 (This Study)
Zetaproteobacteria isolate and SAG cyc2
Representative Proteobacteria cyc2
Cluster 1
Cluster 3
Mariana Cluster 1 Zeta Cyc2 Clone 5.16
Mariana Cluster 3 Zeta Cyc2 Clone 4.5
Leptospirillum sp. Group II Cyt572 (EDZ39515.1)
0.3

Figure 2.

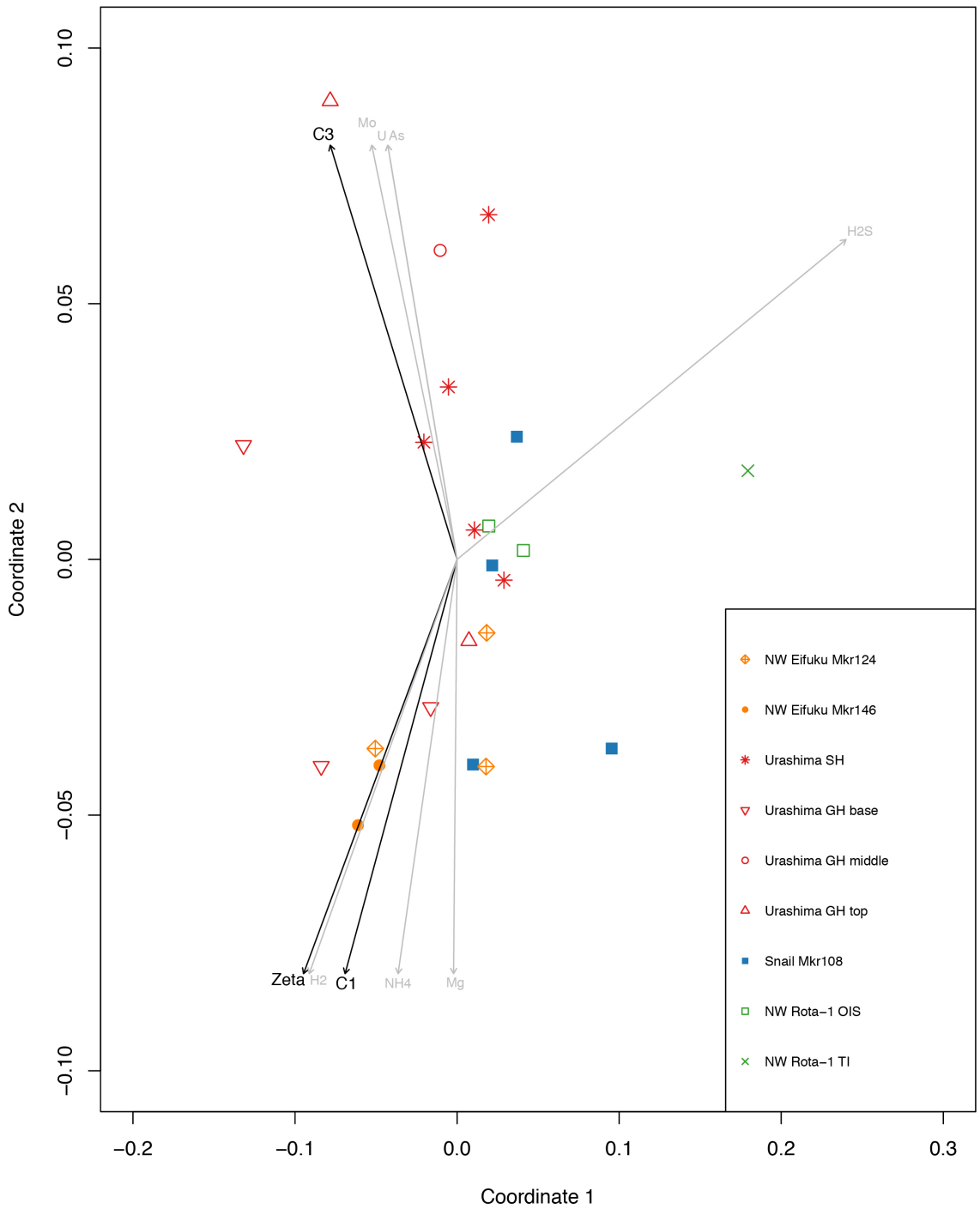
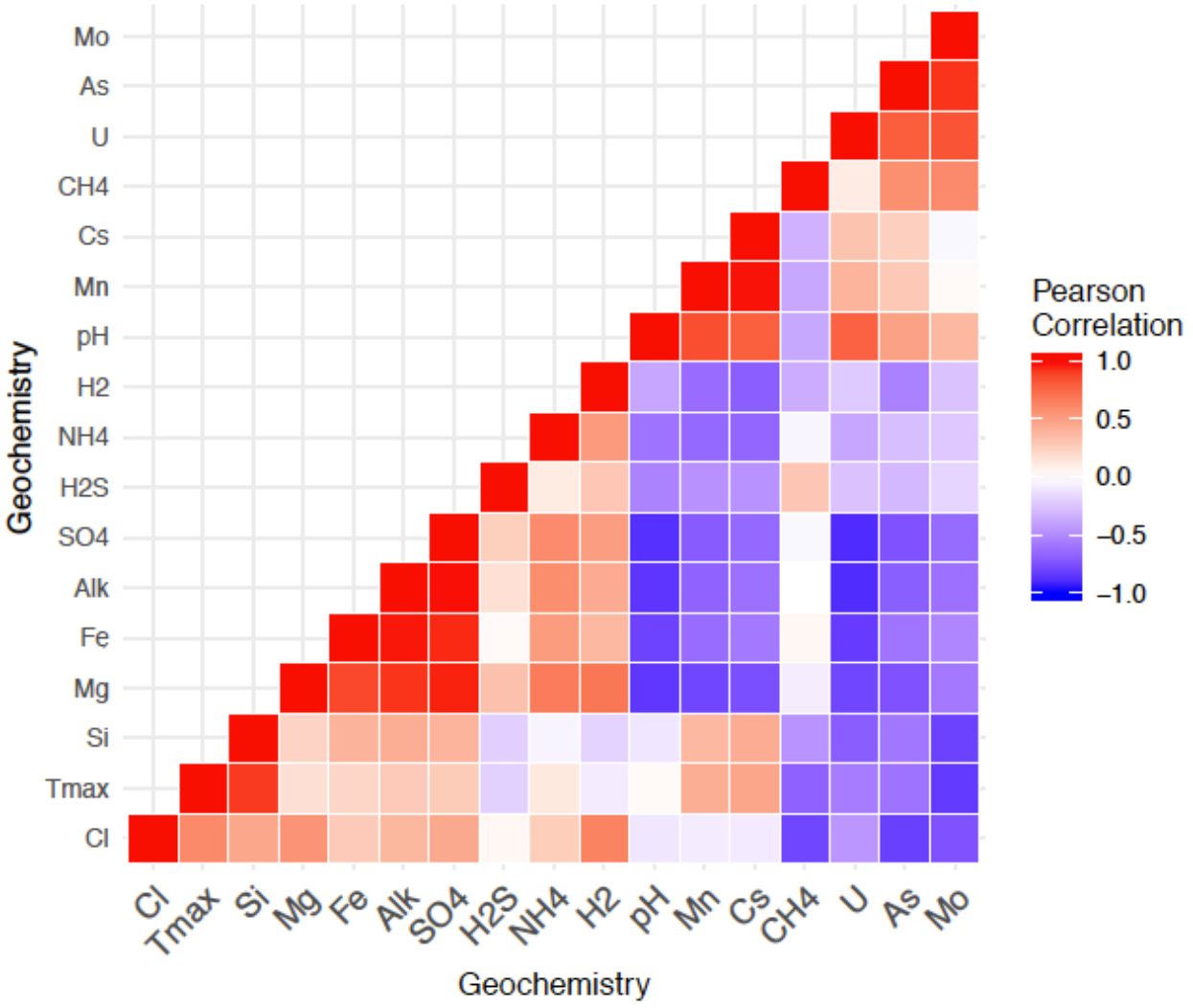

0.10
0.05
0.00
−0.05
−0.10
Coordinate 2
−0.2
−0.1
0.0
0.1
0.2
0.3
Coordinate 1
C3
Mo
U
As
H2S
Zeta
H2
C1
NH4
Mg
NW Eifuku Mkr124
NW Eifuku Mkr146
Urashima SH
Urashima GH base
Urashima GH middle
Urashima GH top
Snail Mkr108
NW Rota−1 OIS
NW Rota−1 TI

Figure 3.

Supplemental Figure 1.



Supplemental Figure 2.

Table 1. Sample information.

Vent location	Sample site	Sample name	Percent Zetaproteobacteria*	Depth (m)	Latitude	Longitude
NW Eifuku	Yellow Cone Mkr 124	798LSc3	39.6%	1584	21°29.274' N	144°2.519' E
NW Eifuku	Yellow Cone Mkr 146	798C346	49.4%	1579	21°29.265' N	144°2.519' E
NW Eifuku	Yellow Cone Mkr 124	798D12346	62.4%	1580	21°29.275' N	144°2.520' E
NW Eifuku	Yellow Cone Mkr 146	798C12	29.8%	1579	21°29.265' N	144°2.519' E
NW Eifuku	Yellow Cone Mkr 146	798LSc1	24.4%	1579	21°29.265' N	144°2.519' E
NW Eifuku	Yellow Cone Mkr 146	799B156	12.2%	1581	21°29.264' N	144°2.524' E
NW Eifuku	Yellow Cone Mkr 146	799D124	20.5%	1580	21°29.264' N	144°2.524' E
NW Eifuku	Yellow Cone Mkr 146	799D56	26.4%	1581	21°29.264' N	144°2.524' E
NW Eifuku	Yellow Cone Mkr 146	799LSc1	23.1%	1580	21°29.264' N	144°2.524' E
NW Eifuku	Yellow Cone Mkr 124	799LSc4	40.4%	1583	21°29.275' N	144°2.519' E
NW Rota-1	Old Iron Slides	800B12456	15.2%	567	14°36.056' N	144°46.656' E
NW Rota-1	Old Iron Slides	800LSc8	24.2%	567	14°36.056' N	144°46.656' E
NW Rota-1	Tip ice/Ice berg	800LSc2	0.6%	527	14°36.061' N	144°46.577' E
Snail	Mkr 108	J42-1W**	24.2%	2850	12°57.190' N	143°37.125' E
Snail	Mkr 108	J42-2W**	12.3%	2850	12°57.190' N	143°37.125' E
Snail	Mkr 108	797D156	26.9%	2850	12°57.166' N	143°37.142' E
Snail	Mkr 108	797D234	21.4%	2850	12°57.166' N	143°37.142' E
Urashima	Saipanda horn	797LSc2	39.4%	2928	12°55.333' N	143°38.950' E
Urashima	Saipanda horn	797C12	40.5%	2928	12°55.333' N	143°38.950' E
Urashima	Snap snap	797B12	11.7%	2928	12°55.333' N	143°38.950' E
Urashima	Snap snap	797B56	28.1%	2928	12°55.333' N	143°38.950' E
Urashima	Snap snap	797LSc1	33.7%	2928	12°55.333' N	143°38.950' E
Urashima	Golden horn (base)	801LSc8	60.2%	2928	12°55.340' N	143°38.957' E
Urashima	Golden horn (base)	801LSc1	72.4%	2928	12°55.340' N	143°38.957' E
Urashima	Golden horn (middle)	801X345	25.5%	2928	12°55.343' N	143°38.953' E
Urashima	Golden horn (top)	801SS	59.5%	2923	12°55.343' N	143°38.953' E
Urashima	Golden horn (top)	801LSc4	57.6%	2922	12°55.343' N	143°38.953' E
Urashima	Golden horn (base)	801X126	31.7%	2931	12°55.343' N	143°38.953' E

*Out of total bacteria

**Davis & Moyer, 2008

Mkr = marker

Table 2. Vent geochemistry. Modified from Hager et al., 2017.

Field	Site	Tmax (°C)	pH	Alk (meq/L)	Total H_2S (µM)	NH_4 (µM)	Si (µM)	H_2 (uM)	CH_4 (uM)	Cl (mmol/kg)	SO_4 (mmol/kg)	Mg (mmol/kg)	Fe (µM)	Mn (µM)	As (nM)	Mo (nM)	Cs (nM)	U (nM)
Snail	Mkr 108 site	36	6.66	2.088	<0.4	<0.6	1282	bdl	0.15	548.2 ± 1.9	26.8 ± 0.1	49.2 ± 0.40	19.3	82.6	41	47.7	33	8.2
Urashima	GoldenHorn (base)	10	6.57	2.307	<0.2	<0.9	523	0.030	0.23	541.3 ± 1.0	27.2 ± 0.0	51.0 ± 0.05	54.0	37.9	61	118	14	11.9
Urashima	GoldenHorn (middle)	10.3	6.49	2.224	<0.2	<0.9	589	0.016	0.52	538.8 ± 0.4	27.1 ± 0.1	50.4 ± 0.04	67.0	47.4	71	114	16	11.4
Urashima	GoldenHorn (top)	15.9	6.25	2.165	0.4	1.2	752	0.009	0.71	494.3 ± 0.4	24.1 ± 0.0	45.1 ± 0.11	66.1	66.1	114	131	26	10.8
Urashima	Saipanda Horn/Snap-Snap	18.7	6.52	2.23	<0.4	12.7	627	0.012	0.36	541.3 ± 0.5	27.7 ± 0.0	50.8 ± 0.50	48.5	52.9	81	116	19	10.6
Urashima	Mid-water background	2	7.51	2.399	<0.2	<0.9	159	0.015	0.01	545.1 ± 0.8	28.3 ± 0.0	52.6 ± 0.24	15.3	1.1	29	118	2	12.8
NW Eifuku	Yellow Cone (Mkr124)	22.5	5.59	9.475	0.8	0.3	1240	bdl	0.46	539.0 ± 0.3	40.2 ± 0.1	55.7 ± 0.60	221	25.3	31	61.0	14	3.86
NW Eifuku	Yellow Cone (Mkr146)	30.5	5.67	5.92	4.0	776	679	0.022	0.28	540.5 ± 0.6	33.7 ± 0.1	54.4 ± 0.17	96.7	12.6	71	83	8	8.27
NW Eifuku	Mid-water background	2	7.55	2.394	0.4	1.0	149	0.002	0.00	535.3 ± 2.1	27.7 ± 0.1	52.3 ± 0.18	0.09	<0.045	26	110	2	12.4
NW Rota-1	Fault shrimp (Old Iron Slides)	10.7	5.72	2.507	190	1.53	263	bdl	0.87	534.6 ± 1.0	28.9 ± 0.0	51.85 ± 0.09	<0.07	5.5	NA	NA	NA	NA
NW Rota-1	Menagerie (Tip Ice/Iceberg)	19.1	5.41	2.076	816	1.17	747	bdl	0.48	535.1 ± 0.4	29.6 ± 0.0	52.0 ± 0.27	<0.07	38.1	28	61.5	7	6.4

*table made form single most representative HFS sample from each site
bdl=below detection limit

Table 3. Primer sequences.

Purpose	Primer	Primer sequence (5'-3')	Target gene or region	Degeneracy	Primer length (bp)	Amplicon size (bp)	Annealing temp. (°C)	Source of primer or reference
Cluster 1 *cyc2* cloning	ZetaCyc2-F	GCWATTCCRGYVTTTGCACG	Cluster 1 *cyc2*	24	20	~800	60	this study
	Reads1020-R	TCGCCCATTTCACCCTGYATCTG	Cluster 1 *cyc2*	2	23			this study
Cluster 3 *cyc2* cloning	ZetaCyc2-F	GCWATTCCRGYVTTTGCACG	Cluster 3 *cyc2*	24	20	~500	60	this study
	Reads19-R	GACATCATRGTCATRCCRTA	Cluster 3 *cyc2*	8	20			this study
Cluster 1 *cyc2* qPCR assay	Reads1020-F	CAGACYGGCGCWGCTTGCYT	Cluster 1 *cyc2*	8	20	~100	55	this study
	ZetaCyc2-R1	TCRCCRACGTCRGTGAAK	Cluster 1 *cyc2*	16	18			this study
Cluster 3 *cyc2* qPCR assay	Reads19-F	TGKCSTGYAATTCMTGYCA	Cluster 3 *cyc2*	32	19	~90	55	this study
	ZetaCyc2-R2	ACCTGCWKCRAARCGRAA	Cluster 3 *cyc2*	32	18			this study
Zeta SSU qPCR assay	Zeta542F	GAAAGGDGCAAGCGTTGTT	Zeta SSU	3	19	~120	60	McAllister et al., 2011
	Zeta658R	TGCTACACDCGGAATTCCGC	Zeta SSU	3	20			McAllister et al., 2011
Bact SSU qPCR assay	Bact533F	GTGCCAGCAGCCGCGGTAA	Bact SSU	0	19	~150	60	McAllister et al., 2011
	Bact684R	TCTACGSATTTYACYSCTAC	Bact SSU	16	20			McAllister et al., 2011

Supplemental Table 1. Zetaproteobacteria genomes & *cyc2* reference sequences.

Zetaproteobacteria isolate or SAG	*cyc2 group*	*cyc2* IMG gene ID	Gene size (bp)	Genome completeness	Number of SSU genes in genome*	Number of *cyc2* genes in genome*	Reference
Mariprofundus ferrooxydans PV-1	Cluster 1	AKN78226.1	1320	99%	2	2	Singer et al., 2011
	Cluster 3	639880235	1452				
Ghiorsea bivora TAG-1	Cluster 1	2583143897	1164	Not reported	2	1	Mori et al., 2017
Ghiorsea bivora SV108	Cluster 1	2617305037	1173	Not reported	1	1	Mori et al., 2017
Mariprofundus sp. EKF-M39	Cluster 1	2572240606	1248	99%	1	2	Field et al., 2015
	Cluster 3	2572241891	1425				
Mariprofundus sp. DIS-1	Cluster 1	2572237054	1287	Not reported	2	1	Mori et al., 2017
Mariprofundus micogutta ET2	Cluster 1	2727942937	1242	Not reported	2	2	Makita et al., 2016
	Cluster 3	2727943507	1416				
Mariprofundus ferrooxydans M34	Cluster 1	2513994769	1275	99%	None reported	1	Mori et al., 2017
Mariprofundus ferrooxydans JV-1	Cluster 1	2651849445	1305	99%	1	1	Fullerton et al., 2015
Mariprofundus ferrinatatus CP-8	Cluster 1	2776476306	1215	Not reported	1	1	Chiu et al., 2017
Mariprofundus aestuarium CP-5	Cluster 1	2776478794	1218	Not reported	2	1	Chiu et al., 2017
SCGC AB602-C20	Cluster 3	2264874972	555	23%	1	1	Field et al., 2015
SCGC AB602-F03	Cluster 1	2525813086	1158	41%	None reported	2	Field et al., 2015
	Cluster 3	2525812737	1356				
SCGC AB137-J06	Cluster 3	2264878865	693	32%	1	1	Field et al., 2015
SCGC AB137-C09	Cluster 1	2264876378	1239	83%	1	1	Field et al., 2015
SCGC AB133-C04	Cluster 1	2264884339	264	62%	1	1	Field et al., 2015
SCGC AB133-M17	Cluster 1	2525849410	1161	57%	1	1	Field et al., 2015
SCGC AB133-D10	Cluster 3	2525841501	402	49%	1	1	Field et al., 2015

*As reported in IMG/MER

Supplemental Table 2. Kendall's tau coefficient values for cyc2 and SSU gene copy number.

	Cluster 1 *cyc2*	**Cluster 3 *cyc2***	**Zeta SSU**
Cluster 3 *cyc2*	3.43E-01		
Zeta SSU	5.50E-01	4.80E-01	
Bact SSU	4.22E-01	4.59E-01	5.45E-01

*all P-values were < 0.05

Supplemental Table 3. P-values for nonparametric one-way MANOVA for Cluster 1 and Cluster 3 *cyc2* gene abundance as the dependent variable.

Independent variable	**No. of degrees of freedom**	**P-value***
% Zeta	1	5.64E-02
Vent field	3	**3.08E-02**
Sample Site	8	**1.20E-03**
Tmax	1	1.37E-01
pH	1	2.65E-01
Alk	1	1.36E-01
H_2S	1	5.09E-02
NH_4	1	5.67E-02
Si	1	2.91E-01
H_2	1	**4.07E-02**
CH_4	1	2.91E-01
Cl	1	2.34E-01
SO_4	1	1.29E-01
Mg	1	7.49E-02
Fe	1	2.65E-01
Mn	1	2.65E-01
As	1	5.09E-02
Mo	1	**3.14E-02**
Cs	1	3.71E-01
U	1	7.49E-02

Significant p-values (< 0.05) are in bold

*Adjusted p-values

Supplemental Table 4. P-values from the correlation of environmental variables with ordinated gene abundance.

Independent variable	R^2 value	P-value*
% Zeta	4.42E-01	**1.27E-02**
Vent field	4.00E-01	**1.27E-02**
Sample site	1.00E+00	1.00E+00
Tmax	1.94E-01	1.93E-01
pH	1.06E-01	3.69E-01
H_2S	4.32E-01	**2.83E-02**
Alk	2.34E-01	1.52E-01
NH_4	2.12E-01	1.94E-01
Si	1.41E-01	3.07E-01
H_2	6.31E-01	**2.83E-02**
CH_4	1.07E-01	3.69E-01
Cl	1.31E-01	3.25E-01
SO_4	2.18E-01	1.65E-01
Mg	2.20E-01	1.52E-01
Fe	1.84E-01	1.99E-01
Mn	1.04E-01	3.69E-01
As	3.52E-01	5.31E-02
Mo	4.24E-01	**2.29E-02**
Cs	6.23E-02	5.87E-01
U	2.93E-01	1.10E-01
Cluster 1 *cyc2*	9.13E-01	**1.10E-03**
Cluster 3 *cyc2*	9.91E-01	**1.10E-03**

Significant p-values (< 0.05) are in bold

*Adjusted p-values

Supplemental Table 5. Metagenomic *cyc2* gene information.

Source	Cluster	Gene size (amino acids)	IMG gene ID	Zetaproteo-bacteria OTU
Mariana	1	422	Ga0172420_1000345111	51
Mariana	1	466	Ga0172420_100269104	28
Mariana	1	441	Ga0172420_100254073	51
Mariana	1	470	Ga0172420_100405341	28
Mariana	1	459	Ga0172420_100102282	51
Mariana	1	242	Ga0172420_103940201	9
Mariana	1	208	Ga0172420_100763623	9
Mariana	1	387	Ga0172420_100266452	9
Mariana	1	175	Ga0172420_103629151	59
Mariana	1	217	Ga0172420_101237162	20
Mariana	3	492	Ga0172420_100266983	1
Mariana	3	468	Ga0172420_100202054	51
Mariana	3	462	Ga0172420_100103285	1
Mariana	3	479	Ga0172420_100283925	14
Mariana	3	471	Ga0172420_101103971	9
Mariana	3	425	Ga0172420_102293781	13
Mariana	3	465	Ga0172420_100014541	13
Mariana	3	336	Ga0172420_100618021	14
Mariana	3	448	Ga0172420_100127913	2
Loihi	1	429	Ga0117756_10303861	1
Loihi	1	473	Ga0117756_10311861	1
Loihi	1	428	Ga0117756_100214710	1
Loihi	1	230	Ga0117756_10287482	1
Loihi	1	285	Ga0117756_100051734	1
Loihi	1	429	Ga0117756_10136104	10
Loihi	1	430	Ga0117756_10052816	4
Loihi	1	384	Ga0117756_10076751	11
Loihi	1	421	Ga0117756_10138912	10
Loihi	1	432	Ga0117756_10042008	4
Loihi	1	236	Ga0117756_10011974	1
Loihi	1	444	Ga0117756_10010279	1
Loihi	1	413	Ga0117756_100028141	6
Loihi	1	418	Ga0117756_10038085	9
Loihi	1	398	Ga0117756_100192618	11
Loihi	1	271	Ga0117756_10298741	11
Loihi	1	181	Ga0117756_10330242	2
Loihi	1	205	Ga0117756_10411071	2
Loihi	1	192	Ga0117756_10169454	2
Loihi	1	406	Ga0117756_10482441	2
Loihi	1	433	Ga0117756_100008262	1
Loihi	3	494	Ga0117756_100088810	1
Loihi	3	464	Ga0117756_100031817	1
Loihi	3	463	Ga0117756_10415551	1
Loihi	3	164	Ga0117756_10240873	1
Loihi	3	470	Ga0117756_10002462	1
Loihi	3	471	Ga0117756_10128111	4
Loihi	3	519	Ga0117756_100194611	1
Loihi	3	477	Ga0117756_10010483	1
Loihi	3	466	Ga0117756_10062301	1
Loihi	3	476	Ga0117756_10307361	1
Loihi	3	455	Ga0117756_100347211	2
Loihi	3	468	Ga0117756_100144317	2

Supplemental Table 6. *cyc2* clone sequence information.

Clone	*cyc2* group	Gene length (bp)	NCBI accession number (TBD)	Top 4 BLAST hits	E-value	Query coverage (%)	Percent Identity (%)
5.16	Cluster 1	721	X	hypothetical protein AUJ58_02665 [Zetaproteobacteria bacterium CG1_02_55_237]	9.0E-103	85	75.25
				hypothetical protein [Mariprofundus ferrooxydans]	1.0E-91	84	70.33
				outer membrane cytochrome c [Mariprofundus ferrooxydans PV-1]	6.0E-88	84	67.94
				hypothetical protein [Mariprofundus ferrooxydans]	2.0E-87	84	67.46
4.5	Cluster 3	968	X	probable cytochrome c1 precursor protein [Mariprofundus ferrooxydans PV-1]	2.0E-95	57	75.66
				hypothetical protein MMIC_P1597 [Mariprofundus micogutta]	3.0E-93	57	75.96
				hypothetical protein [Mariprofundus ferrooxydans]	3.0E-12	12	85.37
				hypothetical protein [Mariprofundus micogutta]	1.0E-10	12	80.49

Supplemental Table 7. Gene abundance of *cyc2* and SSU ribosomal genes.

Vent location	Sample site	Sample name	Zetaproteobacteria SSU Ave. Gene Copies/ng DNA	Zetaproteobacteria SSU standard deviation	Bacteria SSU Ave. Gene Copies/ng DNA	Bacteria SSU standard deviation	Cluster 1 *cyc2* Ave. Gene Copies/ng DNA	Cluster 1 *cyc2* standard deviation	Cluster 3 *cyc2* Ave. Gene Copies/ng DNA	Cluster 3 *cyc2* standard deviation
NW Eifuku	Yellow Cone Mkr 124	798LSc3	6.75E+04	1.18E+04	1.71E+05	5.87E+04	4.08E+06	1.11E+06	8.86E+04	2.96E+04
NW Eifuku	Yellow Cone Mkr 146	798C346	3.25E+05	7.33E+04	6.57E+05	1.56E+05	1.89E+07	2.09E+06	1.34E+05	1.07E+04
NW Eifuku	Yellow Cone Mkr 124	798D1234	1.09E+05	2.39E+04	1.74E+05	2.45E+04	1.76E+07	1.34E+07	1.51E+05	5.09E+04
NW Eifuku	Yellow Cone Mkr 146	798C12	1.68E+05	3.21E+04	5.63E+05	8.68E+04	2.52E+07	5.12E+06	1.56E+05	6.10E+04
NW Eifuku	Yellow Cone Mkr 146	798LSc1	2.17E+04	2.28E+03	8.91E+04	1.60E+04	n.d.	n.d.	n.d.	n.d.
NW Eifuku	Yellow Cone Mkr 146	799B156	1.62E+04	2.99E+03	1.33E+05	1.85E+04	2.83E+06	6.18E+05	n.d.	n.d.
NW Eifuku	Yellow Cone Mkr 146	799D124	1.00E+04	3.57E+03	4.88E+04	1.25E+04	n.d.	n.d.	7.77E+03	2.70E+03
NW Eifuku	Yellow Cone Mkr 146	799D56	1.81E+04	5.68E+03	6.86E+04	1.35E+04	2.45E+06	3.87E+05	n.d.	n.d.
NW Eifuku	Yellow Cone Mkr 146	799LSc1	1.14E+04	2.18E+03	4.95E+04	7.76E+03	n.d.	n.d.	n.d.	n.d.
NW Eifuku	Yellow Cone Mkr 124	799LSc4	4.33E+04	3.94E+03	1.07E+05	2.06E+04	4.85E+06	1.24E+06	7.74E+04	2.05E+04
NW Rota-1	Old Iron Slides	800B12456	2.02E+04	8.06E+03	1.33E+05	3.66E+04	4.01E+06	9.21E+05	1.27E+05	5.92E+03
NW Rota-1	Old Iron Slides	800LSc8	2.30E+04	3.94E+03	9.51E+04	7.96E+03	2.81E+06	6.83E+05	7.30E+04	3.99E+04
NW Rota-1	Tip ice/Ice berg	800LSc2	1.22E+03	1.47E+02	2.20E+05	5.14E+04	8.33E+05	1.18E+05	1.13E+04	1.98E+03
Snail	Mkr 108	J42-1W	2.00E+04	9.01E+02	8.27E+04	1.08E+04	1.78E+06	2.50E+05	1.11E+05	2.44E+04
Snail	Mkr 108	J42-2W	9.45E+03	2.63E+03	7.67E+04	6.50E+03	3.95E+06	1.71E+06	2.06E+04	4.31E+03
Snail	Mkr 108	797D156	2.77E+04	4.34E+03	1.03E+05	1.97E+04	7.38E+06	4.37E+05	7.51E+04	1.30E+04
Snail	Mkr 108	797D234	1.96E+04	1.29E+03	9.14E+04	1.49E+04	3.90E+06	1.63E+06	1.02E+05	2.38E+04
Urashima	Saipanda horn	797LSc2	4.28E+04	1.43E+04	1.09E+05	2.35E+04	2.48E+06	6.73E+05	9.28E+04	1.86E+04
Urashima	Saipanda horn	797C12	2.80E+04	4.24E+03	6.92E+04	1.12E+04	2.76E+06	1.11E+06	1.24E+05	2.17E+04
Urashima	Snap snap	797B12	1.39E+04	1.16E+03	1.19E+05	1.93E+04	1.78E+06	2.52E+05	2.89E+05	4.94E+04
Urashima	Snap snap	797B56	6.43E+04	2.35E+04	2.29E+05	6.03E+04	3.30E+06	8.89E+05	2.92E+05	1.77E+04
Urashima	Snap snap	797LSc1	3.57E+04	8.26E+03	1.06E+05	2.96E+04	4.61E+06	1.40E+06	2.86E+05	1.38E+04
Urashima	Golden horn (base)	801LSc8	8.83E+04	1.82E+04	1.47E+05	1.83E+04	6.39E+06	1.67E+06	1.35E+05	3.52E+04
Urashima	Golden horn (base)	801LSc1	1.21E+05	2.45E+04	1.67E+05	3.10E+04	1.79E+07	1.69E+06	2.63E+05	4.79E+04
Urashima	Golden horn (middle)	801X345	7.07E+04	8.29E+03	2.78E+05	7.29E+04	2.69E+06	7.20E+05	4.63E+05	7.68E+04
Urashima	Golden horn (top)	801SS	6.17E+04	6.02E+03	1.04E+05	2.18E+04	3.76E+06	6.64E+05	1.04E+05	9.11E+03
Urashima	Golden horn (top)	801LSc4	8.49E+04	1.39E+04	1.47E+05	3.35E+04	3.40E+06	4.20E+05	2.37E+06	3.02E+05
Urashima	Golden horn (base)	801X126	7.73E+04	1.66E+04	2.44E+05	6.22E+04	2.35E+07	6.70E+06	1.77E+06	8.58E+05

n.d. = not detected